Sampling

TMSL

Prof. Sweta

Population

- The population refers to the entire group of people, events, or things of interest for investigation.
- It is the group of people, events, or things of interest for which the researcher wants to make inferences

Element

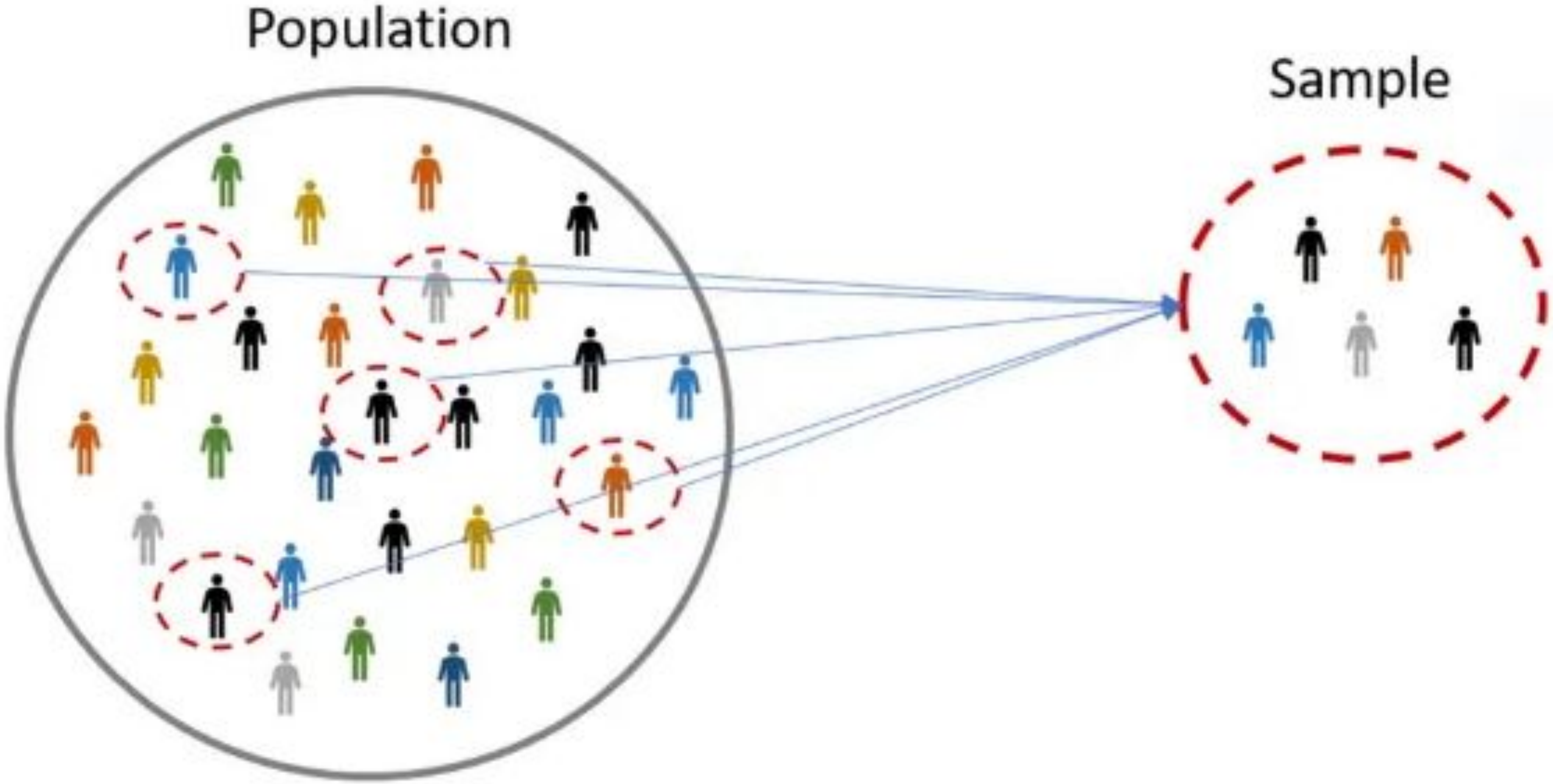
- An element is a single member of the population.

Sample

- A sample is a subset of the population.
- It comprises some members selected from it.
- In other words, some, but not all, elements of the population forms the sample.

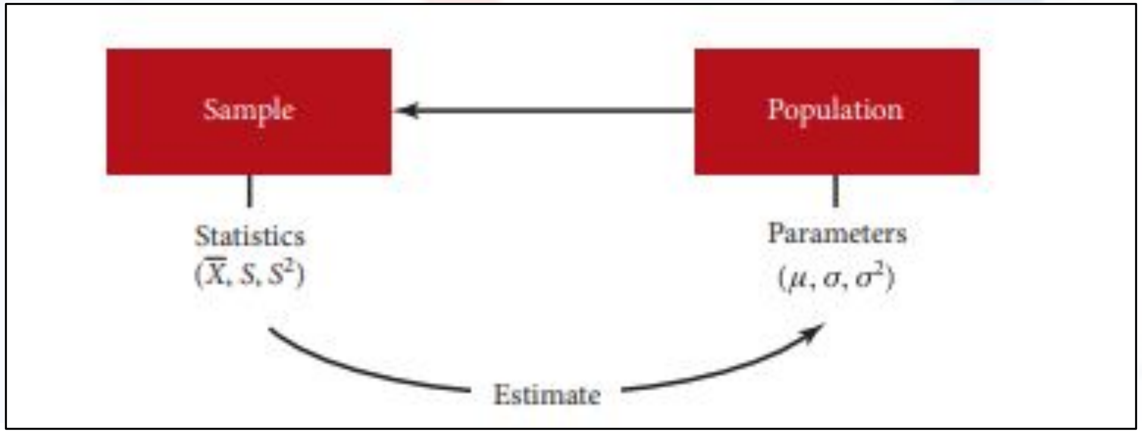
Sampling Unit

- The sampling unit is the element or set of elements available for selection at some stage of the sampling process.
- Examples of sampling units are city blocks, households, and individuals within the households


Population
Sample

Population and Sample

- When we sample, the sampling units (employees, consumers, etc.) provide responses
- For instance, a consumer responding to a survey question may give a response of “3”
- When we examine the responses that we get for our entire sample, we make use of statistics
- There is a wide variety of statistics we can use, such as the mean, the median, or the mode
- The reason we sample, however, is that we are interested in the characteristics of the population we sample from
- If we study the entire population and calculate the mean or the standard deviation, then we don’t refer to this as a statistic. Instead, we call it a parameter of the population.



Population and Sample

- The characteristics of the population such as μ (the population mean), σ (the population standard deviation), and σ2 (the population variance) are referred to as its parameters
- The central tendencies, the dispersions, and other statistics in the sample of interest to the research are treated as approximations of the central tendencies, dispersions, and other parameters of the population
- As such, all conclusions about the sample under study are generalized to the population
- In other words, the sample statistics – X (the sample mean), S (the standard deviation), and S2 (the variation in the sample) – are used as estimates of the population parameters

Statistics Symbols		
Statistics	**Population**	**Sample**
Mean =	μ	X
Variance	σ^2	S^2
Standard Deviation	σ	S

Population Mean **Sample Mean**

$$\mu = \frac{\sum \mathrm{x}}{\mathrm{N}}$$

$$\bar{X} = \frac{\sum \mathrm{x}}{\mathrm{n}}$$

Population variance

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

Sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Population standard deviation

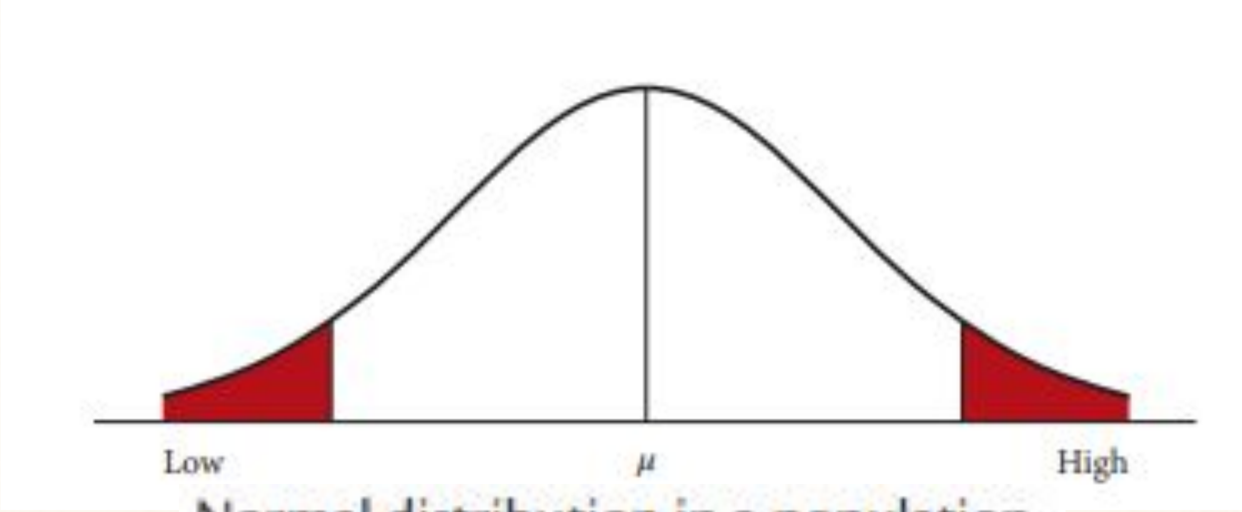
$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_1 - \mu)^2}$$

Sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Normality of Distributions

- Attributes or characteristics of the population are generally normally distributed.
- If it is needed to estimate the population characteristics from those represented in a sample with reasonable accuracy, the sample has to be chosen in such a way so that the distribution of the characteristics of interest follows the same pattern of normal distribution in the sample as it does in the population
- As the sample size n increases, the means of the random samples taken from practically any population approach a normal distribution



Normal distribution in a population

Sampling Process

- Sampling is the process of selecting a sufficient number of the right elements from the population

- It helps in understanding the properties or characteristics of the sample and thus makes it possible to generalize such properties or characteristics to the population

- The major steps in sampling include:
    1. Define the population
    2. Determine the sample frame
    3. Determine the sampling design
    4. Determine the appropriate sample size
    5. Execute the sampling process

1) Defining the population

- Sampling begins with precisely defining the target population.
- The target population must be defined in terms of elements, geographical boundaries, and time.

2) Determining the sample frame

- The sampling frame is a (physical) representation of all the elements in the population from which the sample is drawn.
- Although the sampling frame is useful in listing each element in the population, it may not always be a current, up-to-date document.
- For instance, the names of members who have recently left the organization may appear in the employee list.

3) Determining the sampling design

- There are two major types of sampling design: probability and nonprobability sampling
- Depending on the extent of generalizability desired, the demands of time and other resources, and the purpose of the study, different types of probability and nonprobability sampling designs are chosen.
- The following points should be considered while determining the choice of sampling design:
  - › What is the relevant target population of focus to the study?
  - › What exactly are the parameters we are interested in investigating?
  - › What kind of a sampling frame is available?
  - › What costs are attached to the sampling design?
  - › How much time is available to collect the data from the sample?

4) Determining the sample size

- The factors affecting decisions on sample size are as:
    1. The research objective
    2. The extent of precision desired (the confidence interval)
    3. The acceptable risk in predicting that level of precision (confidence level)
    4. The amount of variability in the population itself
    5. The cost and time constraints
    6. In some cases, the size of the population itself

5) Executing the sampling process

- A failure to obtain information from a number of subjects included in the sample (nonresponse) may lead to nonresponse error.
- Nonresponse error exists to the extent that those who responded to the survey differ from those who did not on (one of the) characteristics of interest in the study.

Issues Of Precision And Confidence In Determining Sample Size

- **Precision** refers to how close our estimate is to the true population characteristic.

- Precision is a function of the range of variability in the sampling distribution of the sample mean.

- If we take a number of different samples from a population and take the mean of each of these, we will usually find that they are all different, are normally distributed, and have a dispersion associated with them.

- The **standard error** is calculated by: $$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

where S is the standard deviation of the sample, n is the sample size

- To reduce the standard error, we need to increase the sample size.

- The smaller the variation in the population, the smaller the standard error.

- **Confidence** denotes how certain we are that our estimates will hold true for the population

- Confidence reflects the level of certainty with which we can state that our estimates of the population parameters, based on our sample statistics, will hold true.

- The level of confidence can range from 0 to 100%

- A 95% confidence is the conventionally accepted level for most business research, most commonly expressed by denoting the significance level as $p \leq 0.05$.

- In other words, we say that at least 95 times out of 100, our estimate will reflect the true population characteristic.

$$\mu = X \pm KS$$

- For a 90% confidence level, the K value is 1.645.
- For a 95% confidence level, the K value is 1.96.
- For a 99% confidence level, the K value is 2.576.

Trade-Off Between Confidence and Precision

- We have noted that if we want more precision, or more confidence, or both, the sample size needs to be increased – unless, of course, there is very little variability in the population itself.

- However, if the sample size (n) cannot be increased, for whatever reason – say, we cannot afford the costs of increased sampling – then, with the same n, the only way to maintain the same level of precision is to forsake the confidence with which we can predict our estimates.

- That is, we reduce the confidence level or the certainty of our estimate.

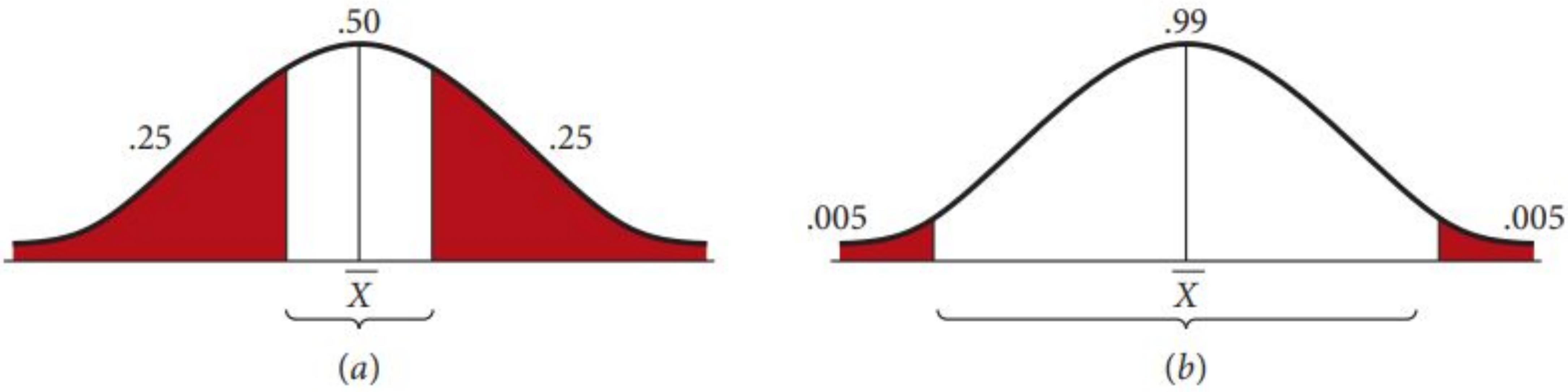


Illustration of the trade-off between precision and confidence. (a) More precision but less confidence; (b) more confidence but less precision.