

POPULATION VS SAMPLE

Dr Bibhunandini Das

POPULATION AND SAMPLE

Population

- consists of all subjects (human or otherwise) that are being studied.

Sample

- is a group of subjects selected from a population
- If the subjects of a sample are properly selected, most of the time they should possess the same or similar characteristics as the subjects in the population

MERITS

- Data are obtained from each and every unit of the population
- The results obtained are likely to be more representative, accurate and reliable
- Data of complete enumeration census can be widely used as a basis for various surveys.

DEMERITS

- The effort, money and time required for carrying out complete enumeration is generally be very large.
- Cost will also be high
- Resources constitute big constraints

RANDOM VARIABLE

Random variables- Variables whose values are determined by chance

- (e.g) 3 out of every 100 automobiles the company insured were involved in accidents during a 1-year period



SAMPLING

SAMPLING

A sample is a subset of population units

It is simply the process of learning about the population on the basis of a sample drawn from it

Process of sampling involves *three* elements:

- Selecting the sample (**Sample Selection**)
- Collecting the information (**Data Collection**)
- Making an inference about the population (**Estimation**)

THEORY OF SAMPLING

Two Important Laws:

1. Law of Statistical Regularity
2. Law of Inertial of Large Numbers

Law of statistical regularity  the mathematical theory of probability

Random Sampling (Likely to possess the same characteristics as that of the population)

LAW OF INERTIA OF LARGE NUMBERS

Other things being equal, larger the size of the sample, more accurate the results are likely to be

Large Numbers are more stable as compared to small ones.

METHODS OF SAMPLING

Two broad heads:

- 1. *Probability Sampling (Random Sampling)***
- 2. *Non Probability Sampling (Non-random Sampling)***

PROBABILITY SAMPLING

Every item in the universe has a known chance (probability of being chosen for the sample)

Probability sampling does not depend upon the existence of detailed information about the universe for its effectiveness

Estimates are essentially unbiased

LIMITATIONS OF PROBABILITY SAMPLING

It requires a very high level of skill and experience for its use

Requires lot of time to plan and execute a probability sample

In comparison to non-probability sampling, the costs is high

TYPES OF PROBABILITY SAMPLING

1. Simple or unrestricted random sampling

2. Restricted random sampling:

- Stratified Sampling

- Systematic Sampling

- Cluster Sampling

SIMPLE/UNRESTRICTED RANDOM SAMPLING

Each and every unit of the population has an equal opportunity of being selected in the sample.

Items get selected is just a matter of chance

No personal bias of the investigator

All n items of the sample are selected independently of one another

CONT.....

Independence of selection- Selection of a particular item in one draw has no influence on the probabilities of selection in any other draw

All remaining items in the population have the same chance to be drawn

Sampling with replacement- each item has probability of $1/N$ of being drawn at each selection

CONT.....

Sampling without replacement-

- ❑ at the first draw, the probability of selection of each item is $1/N$
- ❑ At the second draw, it is $1/(N-1)$
- ❑ At the third draw, it is $1/(N-2)$ and so on

CONT.....

Two methods to ensure randomness:

Lottery method

Table of Random Numbers

RESTRICTED RANDOM SAMPLING

Stratified Sampling

- ❑ The universe to be sampled is subdivided (or stratified) into groups which are mutually exclusive and include in all items in the universe
- ❑ A simple random sample is then chosen independently from each group

RESTRICTED RANDOM SAMPLING

- ❑ Base of Stratification
- ❑ Number of Strata
- ❑ Sample Size within Strata
 - ❑ Proportional
 - ❑ Disproportional
- ❑ Example- 10 % of 5000 = 500
- ❑ 5 groups- (600, 500, 900, 1000, 2000)

CONT.....

Systematic Sampling

- ❑ Select one unit at random and then selecting additional units at evenly spaced intervals until the sample is formed
- ❑ Complete list of population is available

$$k=N/n$$

Where k = Sampling interval, N = Universe Size and n = Sample size

CONT.....

Multi-stage or Cluster Sampling

□ The random selection is made of primary, intermediate and final (or the ultimate) units from a given population or stratum

□ Example: Sample of 5000 households from the State of Odisha.

□ State → Districts → Villages → Households

NON-PROBABILITY SAMPLING

It does not provide every item in the universe with a known chance of being included in the sample

Sample selection without the use of randomization

Non-random sample is selected on the basis of convenience, judgments etc.

TYPES OF NON-PROBABILITY SAMPLING

1. Judgment Sampling
2. Convenience Sampling
3. Quota Sampling

CONT.....

Judgment Sampling

- The choice of sample items depends exclusively on the judgement of investigator.

Convenience Sampling

- A convenience sample is obtained by selecting convenient population units

CONT.....

Quota Sampling

- ❑ It is a type of judgement sampling and most commonly used sampling technique under non-probability sampling
- ❑ Quotas are set up according to some specific characteristics
- ❑ Example- based on income groups, age, gender etc.

SIZE OF SAMPLE

The size of the universe

The resources available

The ***degree of accuracy or precision desired***

Homogeneity and Heterogeneity of the Universe

Nature of Study (intensive and continuous study-a small sample may be suitable)

Method of Sampling adopted

Nature of respondents

PARAMETERS AND STATISTICS

- To understand the population characteristics, usually a random sample from that population is chosen to explain the population characteristics
- The population characteristics of interest is known as parameter and the corresponding sample characteristics is the sample statistics or parameter estimate
- The statistic is a summary of information about a parameter obtained from the sample, the value of a statistic depends on the particular sample that was drawn from the population.

PARAMETERS AND STATISTICS

- Its values change randomly from one random sample to the next one, therefore a statistic is a random quantity (variable).
- The probability distribution of this random variable is called sampling distribution.
- The sampling distribution of a (sample) statistic is important because it enables us to draw conclusions about the corresponding population parameter based on a random sample.

PARAMETERS AND STATISTICS

- For example, when a random sample is drawn from a normally distributed population, the sample mean is a statistic.
- The value of the sample mean based on the sample at hand is an estimate of the population mean.
- This estimated value will randomly change if a different sample is taken from that same normal population.

PARAMETERS AND STATISTICS

- The probability distribution that describes those changes is the sampling distribution of the sample mean.
- The sampling distribution of a statistic specifies all the possible values of a statistic and how often some range of values of the statistic occurs.

PARAMETERS AND STATISTICS

- Parameters are descriptive measures of an entire population that may be used as the inputs for a probability distribution function (PDF) to generate distribution curves.
- Parameters are usually signified by Greek letters to distinguish them from sample statistics.
- For example, the population mean is represented by the Greek letter mu (μ) and the population standard deviation by the Greek letter sigma (σ).

PARAMETERS AND STATISTICS

- Parameter values are usually unknown because it is infeasible to measure an entire population.
- Because of this a random sample from the population to obtain parameter estimates.
- One goal of statistical analyses is to obtain estimates of the population parameters along with the ***amount of error*** associated with these estimates.

PARAMETERS AND STATISTICS

- Different types of Parameter estimates
 - Point estimates are the single, most likely value of a parameter. For example, the point estimate of population mean (the parameter) is the sample mean (the parameter estimate).
 - Confidence intervals are a range of values likely to contain the population parameter.