

Sampling Fundamentals

Introduction

- The idea of gathering data from a population is one that has been used successfully over the years and is called a census.
- However, collecting data from an entire population is almost impossible because of the amount of people, places, or things within the population and also involves much time and money.
- To collect data on a smaller scale, researchers gather data from a portion or ***sample of the population***.

Definitions

- A sample is a “subgroup of a population” (Frey et al. 2000)
- A representative “taste” of a group (Berinstein 2003)
- A sample is “a smaller (but hopefully representative) collection of units from a population used to determine truths about that population” (Field, 2005)
- Cochran (1953) posits that using correct sampling methods allows researchers;
 - the ability to reduce research costs,
 - conduct research more efficiently (speed),
 - have greater flexibility, and
 - provides for greater accuracy.

Definitions

- **Population** is a complete set of elements (persons or objects) that possess some common characteristic defined by the sampling criteria established by the researcher.
- **Composed of two groups -**
 1. **Target population (universe)** is the entire group of people or objects to which the researcher wishes to generalize the study findings.
 2. **Accessible population** is the portion of the population to which the researcher has reasonable access OR a subset of the target population.
- **Sampling Frame** is a list of all the elements in the population from which the sample is drawn.

Example

Target Population

All people with AIDS

Study Population

All people with AIDS
in Mumbai area

Sample

A list of all people
with AIDS in Mumbai
area who are
diagnosed in the year
2015.

Definitions

Sampling design:

- A sample design is a definite plan for obtaining a sample from the sampling frame. Sampling design is determined before any data are collected.

Statistic(s) and parameter(s):

- A statistic is a characteristic of a sample, whereas a parameter is a characteristic of a population.
- The population mean (μ) is a parameter, whereas the sample mean (\bar{X}) is a statistic.
- **For example**, say you want to know the mean income of the subscribers to a particular magazine—a parameter of a population. You draw a random sample of 100 subscribers and determine that their mean income is **\$27,500** (a statistic). You conclude that the population mean income μ is likely to be close to **\$27,500** (a parameter) as well. This example is one of **statistical inference**.

Definitions

Sample Errors

- Sampling error is the error that arises in a data collection process as a result of taking a sample from a population rather than using the whole population.
- Statistical errors are sample error
- We have no control over it.
- It is affected by a number of factors including:
 - sample size.
 - the variability within the population.
 - sample design.

Definitions

Non-sampling error

- “Non-sampling error is the error that arises in a data collection process as a result of factors other than taking a sample.
- Not controlled by sample size
- Two types;
 1. Non Response Error
 2. Response Error

Non-Sampling Error

- ***Non Response Error***
 - occurs when units selected as part of the sampling procedure do not respond in whole or in part
- ***Response Error***
 - A response or data error is any systematic bias that occurs during data collection, analysis or interpretation.
 - Respondent error (e.g., lying, forgetting, etc.)
 - Interviewer bias
 - Recording errors
 - Poorly designed questionnaires
 - Measurement error

Good Sample Design

- The characteristics of a good sample design are:
 1. Sample design must be a true representative sample.
 2. Sample design must be such which results in a small sampling error.
 3. Sample design must be viable in the context of funds available for the research study.
 4. Sample design must be such so that systematic bias can be controlled in a better way.
 5. Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

Sampling Design Process

Define Population



Determine Sampling Frame



Determine Sampling Procedure

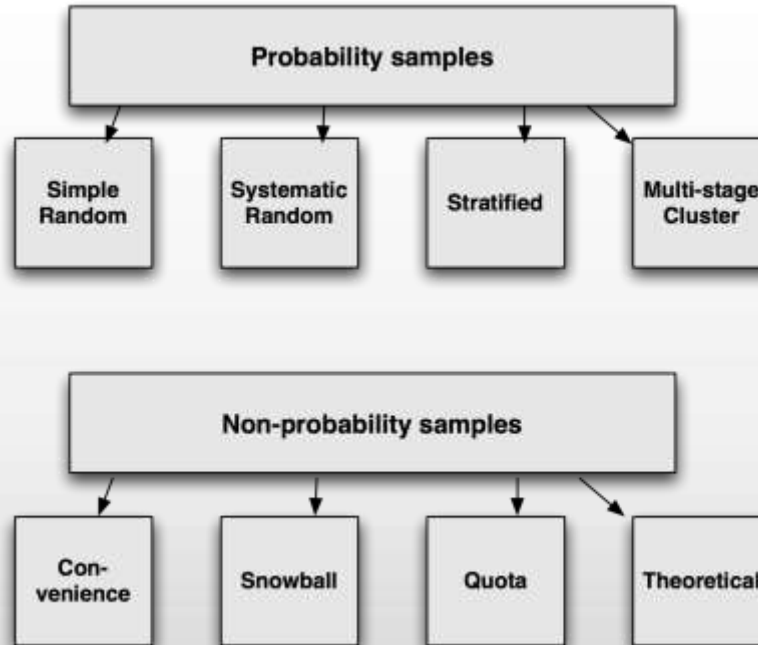


Determine Appropriate Sample Size



Execute Sampling Design

Types of samples



Probability Sampling

- Defined as having the “distinguishing characteristic that each unit in the population has a known, nonzero probability of being included in the sample” (Henry 1990).
- It is described more clearly as “every subject or unit has an equal chance of being selected” from the population (Fink 1995).
- Four types
 1. *Simple Random Sampling*
 2. *Stratified Random Sampling*
 3. *Cluster Sampling*
 4. *Systematic Random Sampling*

Non-probability sampling

- Non-probability sampling/non-parametric sampling is a sampling technique where the samples are gathered in a process that does not give all the individuals in the population equal chances of being selected.
- In this type of sampling, items for the sample are selected deliberately by the researcher
- The sample may or may not be representative of the population, and this can influence the external validity of the study.

Probability vs Non-probability sampling

Probability

- Refer from the sample as well as the population.
- Every individual of the population has equal probability to be taken into the sample.
- May be representative of the population
- The observations (data) of the probability sample are used for the inferential purpose.
- Inferential or parametric statistics are used for probability sample.

Non-probability

- There is no idea of population in non-probability sampling.
- There is no probability of selecting any individual.
- It has free distribution.
- The observations of non-probability sample are not used for generalization purpose.
- Non-parametric or non-inferential statistics are used in non probability sample.

Probability Sampling

Simple Random Sampling

- A simple random sample is one in which each element of the population has an equal and independent chance of being included in the sample
- Randomization is a method done by using a number of techniques as :
 - *Tossing a coin.*
 - Throwing a dice.
 - Lottery method.
 - *Blind folded method.*

Simple Random Sampling

- Population (N)
- In order to select a sample (n)
- assign a consecutive number from 1 to N
- Finally select sample randomly from 1 to N



Systematic Random Sampling

- **Systematic random sampling** is the random sampling method that requires selecting samples based on a system of intervals in a numbered population.
- In systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals.
- This method requires the complete information about the population.
- Let sample size = n and population size = N
- We select each N/n^{th} individual from the list.
- Thus for this technique of sampling population should be arranged in any systematic way.

Systematic Random Sampling

- Population (N)
- In order for a sample (n)
- select an element from the list at random and
- then every k^{th} element in the frame is selected,
- where k , the sampling interval is calculated as: N/n



Stratified Random Sampling

- Stratified random sampling is “one in which the population is divided into subgroups or ‘strata,’ and a random sample is then selected from each subgroup” (Fink).
- If a population from which a sample is to be drawn does not constitute a homogeneous group, stratified sampling technique is generally applied in order to obtain a representative sample.
- Under stratified sampling the population is divided into several sub-populations OR **Strata** that are individually more homogeneous than the total population and
- then select items from each stratum to constitute a sample.

Stratified Random Sampling

Population



Strata



Randomly selects subjects
proportionally from strata.

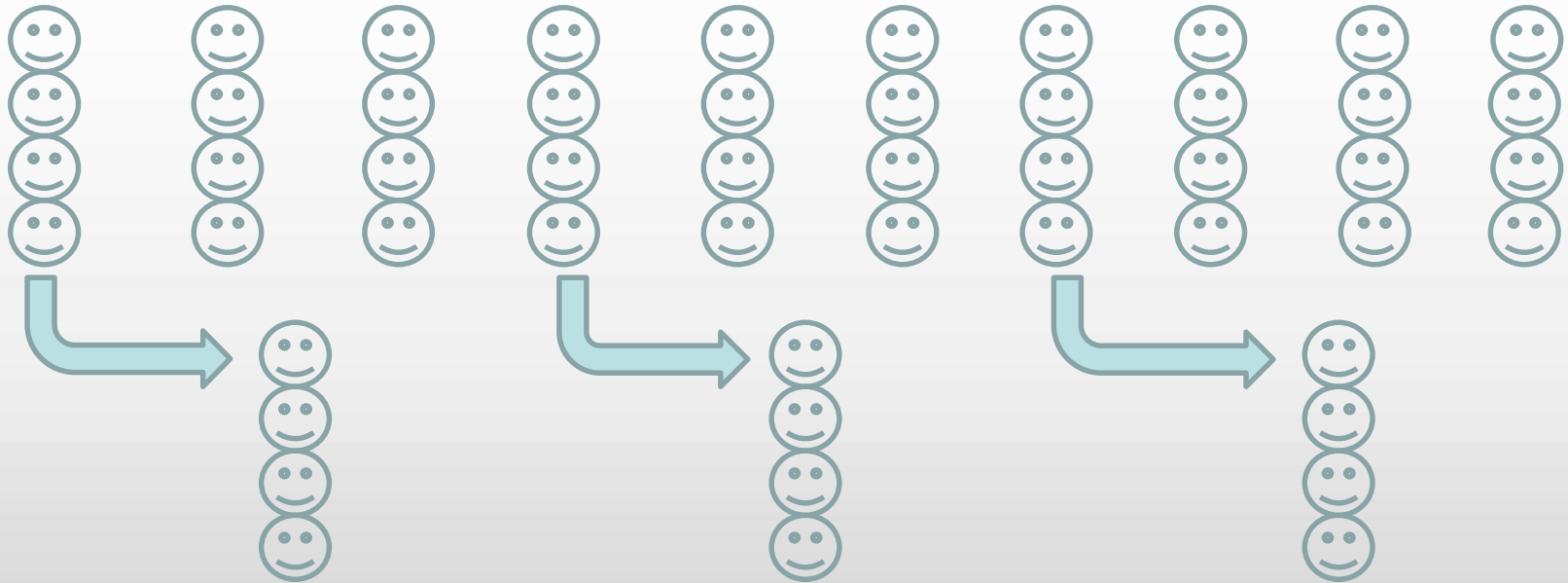
Strata



Cluster Sampling

- In Cluster sampling the sample units contain groups of elements (**clusters**) instead of individual members or items in the population.
- This is considered if the total area of interest happens to be a big one.
- Divide the area into a number of smaller non-overlapping areas and then to randomly select a number of these smaller areas, usually called **clusters**
- A major difference between cluster and stratified sampling relates to the fact that in cluster sampling a cluster is perceived as a sampling unit, whereas in stratified only specific elements of strata are accepted as sampling unit.

Cluster Sampling



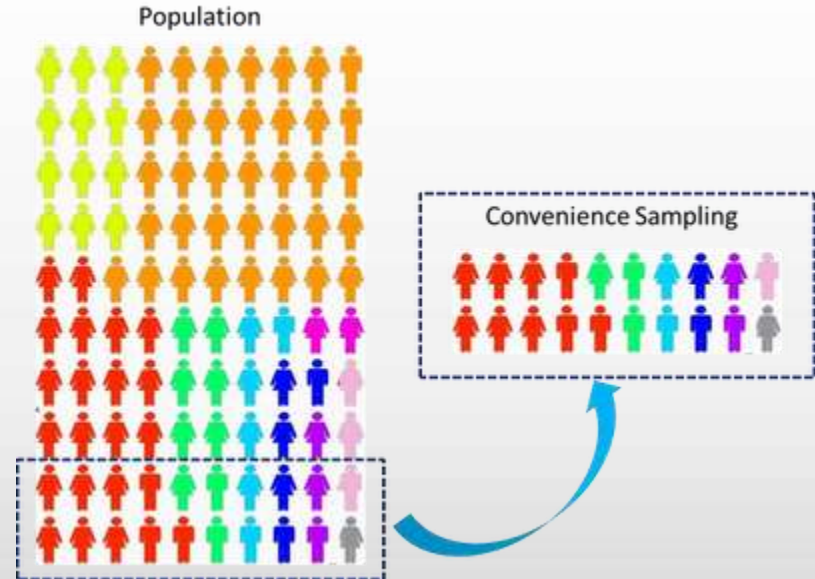
Non-probability sampling

Quota sampling

- **Quota sampling** is a method for selecting survey participants that is a non-probabilistic version of stratified sampling.
- A population is first segmented into sub-groups.
- Then judgment is used to select the subjects or units from each segment based on a specified proportion.
- In quota sampling, there is non-random sampling which makes this a non-probability sampling technique.

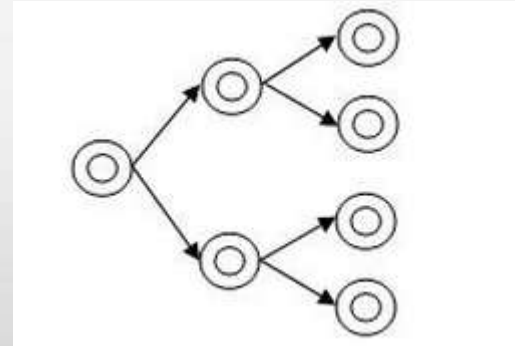
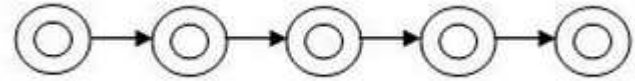
Convenience sampling

- A method that relies on data collection from population members who are conveniently available to participate in study



Snowball sampling

- This sampling method involves primary data sources nominating another potential primary data sources to be used in the research.



Purposive sampling

- Purposive sampling, also referred to as judgment, selective or subjective sampling is a non-probability sampling method that is characterised by a deliberate effort to gain representative samples by including groups or typical areas in a sample.
- Here researcher has sufficient knowledge of topic to select sample and subjects are chosen in this sampling method according to the type of the topic.

References

1. Latham B. Sampling: What is it? Quantitative Research Methods. 2007: 1-13. Downloaded from [http://webpages.acs.ttu.edu/rlatham/Coursework/5377\(Quant\)\)/Sampling_Methodology_Paper.pdf](http://webpages.acs.ttu.edu/rlatham/Coursework/5377(Quant))/Sampling_Methodology_Paper.pdf)
2. Cochran WG. Sampling Techniques. New York: John Wiley & Sons, Inc., 1953.
3. <http://www.umsi.edu/~lindquists/sample.html>
4. <https://www.quora.com/What-is-the-difference-between-a-parameter-and-a-statistic/answer/Akshay-Pratap-Singh-1?srid=5sSf>
5. Kothari CR. Research Methodology, Methods and Techniques. 2nd Edition. 2004, New Age International (P) Ltd. New Delhi, Chapter 8, Sampling Fundamentals. 153-183
6. Dr Nic. Sampling error and non-sampling error. Statistics Learning center. 4 September, 2014. Available from <https://learnandteachstatistics.wordpress.com/2014/09/04/sampling-and-non-sampling-error/>
7. Sharada. Sampling: A Scientific Method of Data Collection. Powerpoint presentation. Available from <http://www.slideshare.net/rambhu21/sampling-and-sampling-errors-19870549>
8. <http://www.copernicusconsulting.net>
9. [Explorable.com](http://www.explorable.com) (May 17, 2009). Non-Probability Sampling. Retrieved Apr 20, 2016 from Explorable.com: <https://explorable.com/non-probability-sampling>
10. Singh YK. Fundamental of Research Methodology and Statistics. New Age International (P). 2006. Chapter 5, Research Planning and Sampling: 77-98
11. Jackson C. Systematic Random Samples: Definition, Formula & Advantages. <http://study.com/>. Available from: <http://study.com/academy/lesson/systematic-random-samples-definition-formula-advantages.html>
12. <http://research-methodology.net/sampling/>
13. https://en.wikipedia.org/wiki/Quota_sampling