



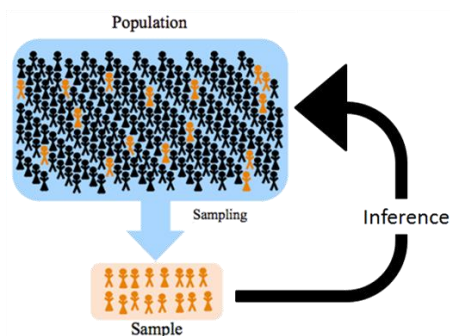
Populations and Samples

STATISTICS IN RESEARCH



- ▶ When using statistics to analyse research data, there are two important types of data sets - **populations** and **samples**.
- ▶ A **population** is all members of a specified group.
- ▶ A **sample** is a part of a population used to describe the whole group.
- ▶ A **parameter** is a measurable characteristic of a population, e.g. a *mean* or *standard deviation*.
- ▶ A **statistic** is a measurable characteristic of a sample.

Difference Between a Population and a Sample:



Examples:



Population	Sample
1.) Advertisements for IT jobs in Kenya.	The top 50 search results for advertisements for IT jobs in Kenya on May 1, 2020.
2.) Undergraduate students in Kenya	300 undergraduate students from three Kenyan universities who volunteer for your research study.
3.) All countries of the world	Countries with published data available on birth rates and GDP since 2000.

Difference Between a Statistic and a Parameter?



- ▶ The difference between a statistic and a parameter is that statistics describe a **sample**. A parameter describes an **entire population**.
- ▶ **Example 1:** You randomly poll voters in an election. You find that 55% of the population plans to vote for candidate A. That is a **statistic**. Why? *You only asked a sample of the population who they are voting for. You calculated what the population was likely to do based on the sample.*
- ▶ **Example 2:** You ask the BTC Year 3 class who likes vanilla ice cream. 90% raise their hands. You have a parameter: 90% of this class likes vanilla ice cream. You know this because you asked **everyone in the class**.

Example 1



- ▶ For the following study, describe the *population*, *sample*, *population parameters*, and *sample statistics*.

In an Infotrack poll of 1027 Kenyans surveyed by cell phones and land lines, 62% of those who responded said there should be an investigation of the torture tactics used during the Moi administration.



- The **population** is the set of all Kenyans,.
- The **sample** is the set of 1027 Kenyans surveyed.
- The **population parameter** is the percent of Kenyans who believe there should be an investigation of the torture tactics used during the Moi administration,
- The **sample** statistic is the 62% of those surveyed who believe there should be such an investigation.



- The **population** is "*all the species that live in the lake.*"
- The **sample** is "*the species that are in the bucket.*"
- The **parameter** is the number of species in the lake,
- The **statistic** is the number of species found in the bucket.



- The **population** is *all the students at the school.*
- The **sample** is *the 70 students polled.*
- The **parameter** is *the lunch preferences of the whole school.*
- The **statistic** is *the responses to the poll.*



Example 2:

► A researcher takes a big bucket of water from a lake and counts how many species of bacteria, bugs, and other creepy crawlies he finds in the bucket. Identify the:

- population,
- the sample,
- the parameter, and
- the statistic in this situation.



Example 3:

► A school takes a poll to find out what students want to eat at lunch. 70 students are randomly chosen to answer the poll questions. What are the:

- population,
- the sample,
- the parameter, and
- the statistic of this study?



- **NB:** Populations and samples *do not have to be people*....they can be a deck of cards, animals, cars etc that you have collected data about.
- There are many different ways you can get a sample from your population.

Sampling Methods



- ▶ If a sample is to be used, by whatever method it is chosen, **it is important that the individuals selected are representative of the whole population.**
- ▶ Sampling techniques can be subdivided into two groups: **probability sampling and non-probability sampling.**
- ▶ In probability sampling, you start with a complete sampling frame of all eligible individuals from which you select your sample.
- ▶ In this way, all eligible individuals have a chance of being chosen for the sample, and you will be more able to generalise the results from your study. .
- ▶ Probability sampling methods tend to be more time-consuming and expensive than non-probability sampling.

PROBABILITY SAMPLING METHODS



1. **Simple random sample:** A sample of the population is selected, and **each member of the sample is chosen entirely by chance.** This means each member has an equally likely chance of being selected.
2. **Stratified random sample:** A population is **divided into smaller groups, or strata.** Strata are created based on members' shared characteristics or attributes (e.g. location, age, income...). Once the population has been stratified, **a random sample from each stratum** is selected. The results are combined to obtain a random sample.
3. **Cluster sampling:** The population is **already divided into groups, or clusters** e.g. like voting districts or blocks of a city. A random sample of these clusters is then selected.
4. **Systematic sample:** In a systematic sample, **every nth item is included** in the sample, where n is a natural number. For example, choosing every 6th name on a list of students is a systematic sample.

4. **Judgement (or Purposive) Sampling:** Also known as selective, or subjective, sampling, this technique **relies on the judgement of the researcher when choosing who to ask to participate.** Researchers may implicitly thus choose a "representative" sample to suit their needs, or specifically approach individuals with certain characteristics. This approach is often used by the media when canvassing the public for opinions and in qualitative research.
5. **Snowball sampling** is a **special non-probability method** used when the desired **sample characteristic is rare.** It may be extremely difficult or cost prohibitive to locate respondents in these situations. Snowball sampling relies on referrals from initial subjects to generate additional subjects. While this technique can dramatically lower search costs, it comes at the expense of introducing bias because the technique itself reduces the likelihood that the sample will represent a good cross section from the population.

Each Sampling Method has its own Advantages and Disadvantages-see link for a summary document.-

<http://www.allresearchjournal.com/archives/2017/vol3issue7/PartK/3-7-69-542.pdf>



- ▶ In non-probability (non-random) sampling, you do not start with a complete sampling frame, so some individuals have no chance of being selected.
- ▶ Consequently, you cannot estimate the effect of sampling error and there is a significant risk of ending up with a non-representative sample which produces non-generalisable results.
- ▶ However, non-probability sampling methods tend to be cheaper and more convenient, and they are useful for exploratory research and hypothesis generation.

NON-PROBABILITY SAMPLING METHODS



1. **Convenience sample:** A sampling of a population is selected based on ease or convenience, such as shoppers in a supermarket. A convenience sample **does not represent the entire population; therefore, it is a biased sample.**
2. **Voluntary response sample:** This sample consists of **people who are self-selected** and respond to a general appeal.
3. **Quota sampling** is the **non-probability** equivalent of stratified sampling. Like stratified sampling, the researcher **first identifies the strata** and their proportions as they are represented in the population. **Then convenience** sampling is used to select the required number of subjects from each stratum. This differs from stratified sampling, where the strata are filled by random sampling.

Identify the sampling technique used (random, cluster, stratified, convenience, systematic):

1. Every fifth person boarding a plane is searched thoroughly- **SYSTEMATIC SAMPLING. (every 5th person)**
2. At a local primary school, five math classes are randomly selected out of 20 and all of the students from each class are interviewed- **CLUSTER SAMPLING. (the 20 classes are already decided.)**
3. A researcher randomly selects and interviews fifty male and fifty female teachers.-**STRATIFIED SAMPLING (Male/Female strata decided by the researcher).**
4. A researcher for an airline interviews all of the passengers on five randomly selected flights. **CLUSTER SAMPLING-(The 5 random flights have already been selected)**
5. Based on 12,500 responses from 42,000 surveys sent to its alumni, a major university estimated that the annual salary of its alumni was 92,500. **SIMPLE RANDOM SAMPLE.**
6. A Strathmore student interviews everyone in the Research Methods class to determine the percentage of students that own a car. **CONVENIENCE SAMPLE-(Convenient for a SU student)**
7. A market researcher randomly selects 200 drivers under 35 years of age and 100 drivers over 35 years of age.-**STRATIFIED SAMPLE (drivers under 35/over 35.. strata)**
8. All of the teachers from 85 randomly selected nation's middle schools were interviewed.-**CLUSTER SAMPLE-(The 85 schools already selected)**
9. To avoid working late, the quality control manager inspects the last 10 items produced that day- **CONVENIENCE SAMPLE.**
10. The names of 70 contestants are written on 70 cards. The cards are placed in a bag, and three names are picked from the bag.-**SIMPLE RANDOM SAMPLE**

Technique	Advantages	Disadvantages
Simple random	Highly representative if all subjects participate; the ideal	Not possible without complete list of population members; potentially uneconomical to achieve; can be disruptive to isolate members from a group; time-scale may be too long, data/sample could change
Stratified random	Can ensure that specific groups are represented, even proportionally, in the sample(s) (e.g., by gender), by selecting individuals from strata list	More complex, requires greater effort than simple random; strata must be carefully defined
Cluster Sampling	Possible to select randomly when no single list of population members exists, but local lists do; data collected on groups may avoid introduction of confounding by isolating members	Clusters in a level must be equivalent and some natural ones are not for essential characteristics (e.g., geographic: numbers equal, but unemployment rates differ)

Technique	Advantages	Disadvantages
Convenience /Volunteer Sampling	Inexpensive way of ensuring sufficient numbers of a study	Can be highly unrepresentative
Quota Sampling	Ensures selection of adequate numbers of subjects with appropriate characteristics	Not possible to prove that the sample is representative of designated population
Snowball sampling	Possible to include members of groups where no lists or identifiable clusters even exist (e.g., drug abusers, criminals)	No way of knowing whether the sample is representative of the population
Judgement (or Purposive) Sampling	Is time-and cost-effective to perform whilst resulting in a range of responses (particularly useful in qualitative research).	it is prone to errors of judgement by the researcher and the findings, and while it is potentially broad, will not necessarily be representative.

Specific terms are used to describe problems with surveys and samples.



- Sampling frame:** A list of all the members of the target population.
- Census:** Getting desired information from everyone in the target population.
- Random sample:** Each member of the population has an equal chance of being selected for the sample.
- Bias:** Systematic unfairness in sample selection or data collection.
- Convenience sample:** Chosen solely for convenience, not based on randomness.
- Volunteer/self-selected sample:** Sample where people determine on their own to be involved.
- Non-response bias:** Occurs when someone in the sample doesn't return or doesn't finish the survey.
- Response bias:** When the respondent takes the survey but doesn't give correct information.
- Under coverage:** Sampling frame doesn't include adequate representation from certain groups within the target population.

Example Question.....



On April 20, 1999, Eric Harris and Dylan Klebold entered Columbine High School in the US, and began shooting teachers and students. Thirteen individuals died, and the Research community was again asked to explain such violent behavior. A researcher decided to interview Columbine students to obtain their perspectives on the factors that motivated the two young men to commit such a horrendous act. A group of ten students has already decided to meet and discuss the events of that day.

The researcher asks if it would be all right to attend the meeting and ask them some questions. The students agree, and the researcher records their thoughts. Based on this information, the researcher concludes that a primary reason for the violent behavior was the peer dynamics in the school that created groups of outcasts.

Answers:



- Who appears to constitute the **population** of interest?
Columbine High School Students.
- Which type of sampling procedure best describes that used by the researcher? **Cluster/Voluntary/Convenience sampling. (Explain why...)**
- What are the **limitations** of this sampling method, and in what specific ways could the sampling method have affected the findings?
Limitations:
 - Cluster sampling is prone to biases. If the clusters that represent the entire population were formed under a biased opinion, the interpretations about the entire population would be biased as well.*
 - In this particular case, there was bias in that the researcher only interviewed the group of ten students that had already decided to meet and discuss the events of that day. The rest of the student population was not interviewed, thus the interpretations the researcher made about the Columbine High School students may not be accurate.*



- What specific steps would you have taken to obtain a representative sample?

Pick a proper Cluster Sample: e.g. since the students are already grouped according to classes, the researcher should go to each class and pick a random sample of students. This will ensure that a representative sample is obtained.