Population & Sample

Researchers are often interested in answering questions about the population like:

- Which political party will win the election?
- What will be the average cutoff mark in the competitive examination?
- What percentage of citizens in a certain city support a certain law?
- What is the effect of certain medicine on the population of disease patients?

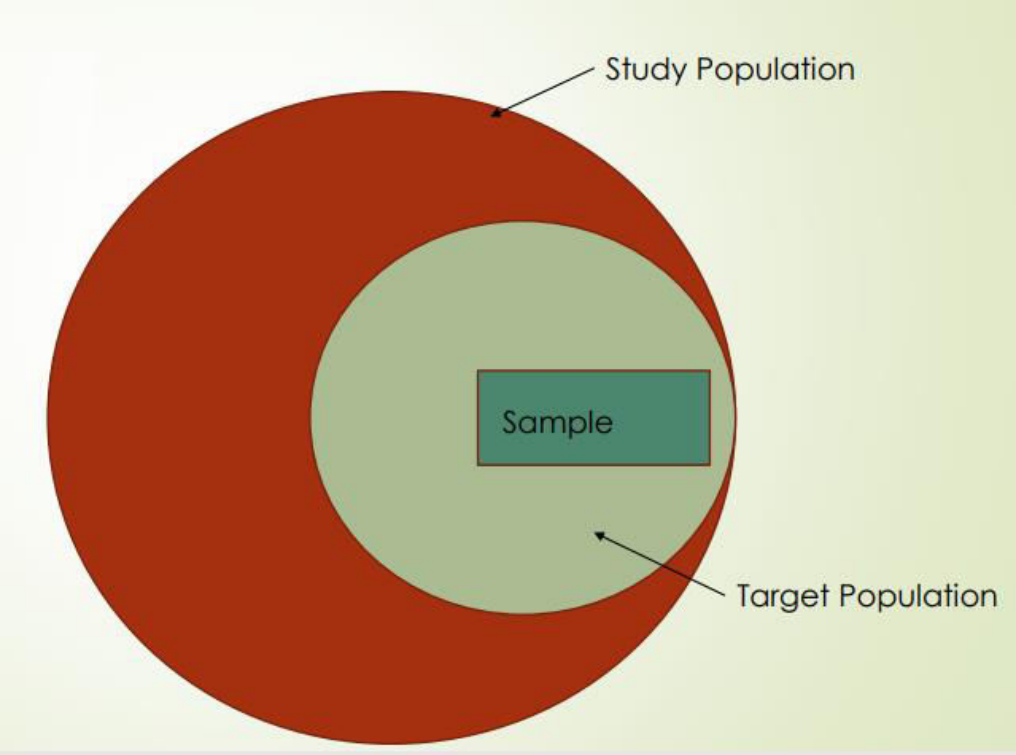
One way to answer these questions is to go around and collect data on every single individual in the population of interest.

However, this is typically too costly and time-consuming which is why researchers instead take a **sample** of the population and use the data from the sample to draw conclusions about the population as a whole.

➔ **A population is the entire group that you want to draw conclusions about.**

➔ **A sample is the specific group that you will collect data from.**

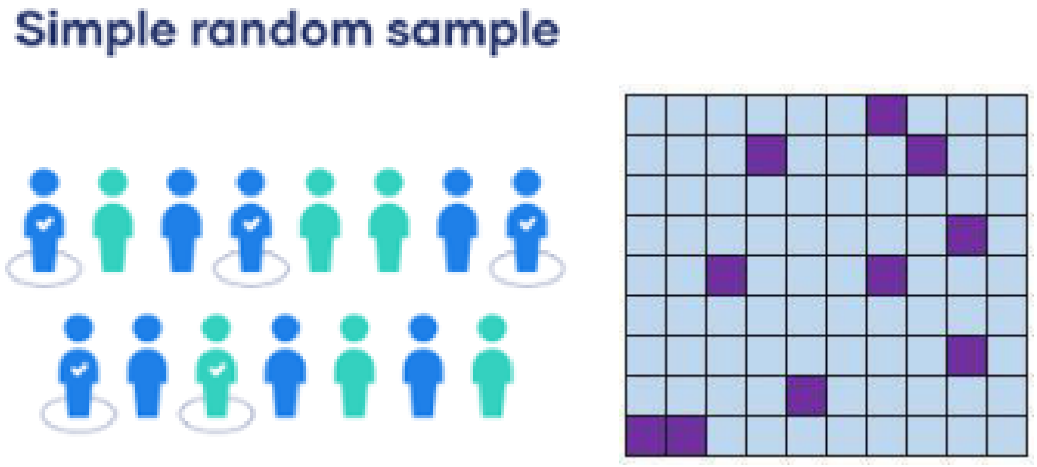
A well-chosen sample will contain most of the information about a particular population parameter.



I. Probability Sampling Methods

Type of Probability Sampling Methods: -

1. Simple random sampling



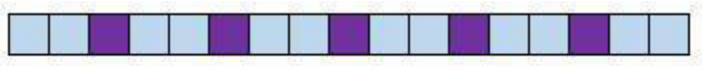
Example: Simple random sampling

You want to select a simple random sample of 10 employees of Company X. You assign a number to every employee in the company database from 1 to 100, and use a random number generator to select 100 numbers.

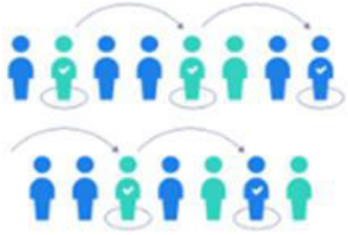
The output may be 2,4,8,12,29,33,45,55,87,99

Definition: Every member of a population has an equal chance of being selected to be in the sample.

2. **Systematic random sample** is similar to simple random sampling, but instead of randomly generating numbers, individuals are chosen at regular intervals.



Definition: Put every member of a population into some order. Choosing a random starting point and select every n^{th} member to be in the sample.



Example: Systematic sampling

One commonly used sampling method is **systematic sampling**, which is implemented with a simple two-step process:

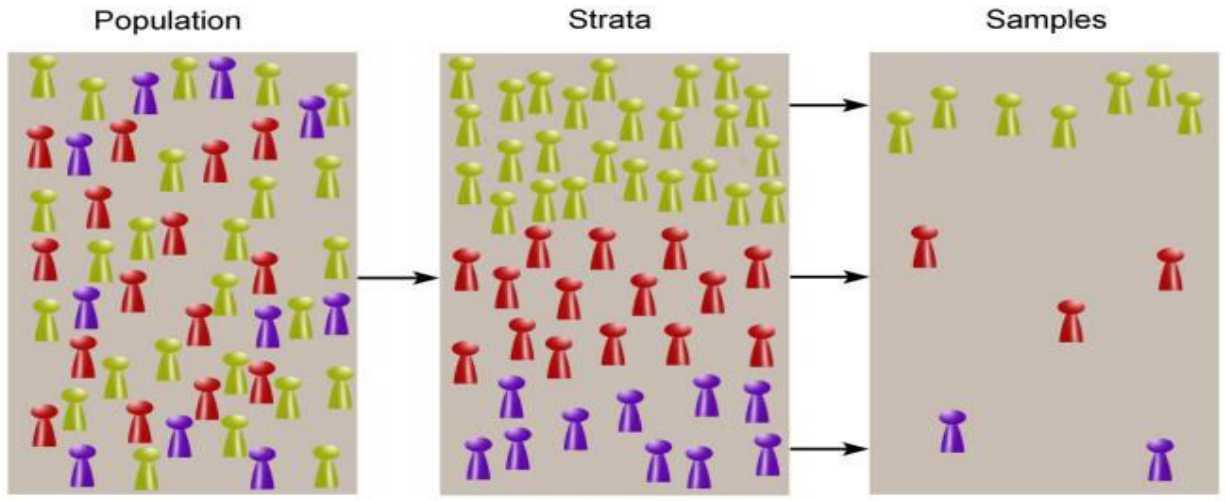
1. Place each member of a population in some order.

2. Choose a random starting point and select every n^{th} member to be in the sample. **If 10th person on the list is selected (6, 16, 26, 36, and so on)**

It is important to make sure that there is no hidden pattern in the list.

3. Stratified random sample

Split a population into groups. Randomly select some members from each group to be in the sample.



Divide the population into subgroups (called strata) based on the relevant characteristic (e.g. gender, age range, income bracket, job role).

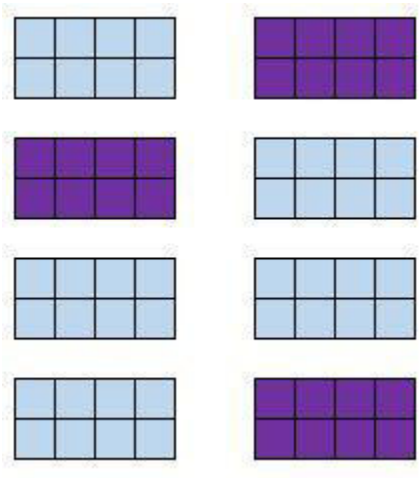
Then can use random or systematic sampling to select a sample from each subgroup.

Example: Stratified sampling

Benefit: Stratified random samples ensure that members from each group in the population are included in the survey.

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.



Example: A company that gives **whale watching tours**, wants to survey its customers. Out of ten tours they give one day, they randomly select four tours and ask every customer about their experience.

Benefit: Cluster random samples get every member from some of the groups, which is useful when each group is reflective of the population as a whole.

Difference between stratified sampling and cluster sampling: -

- **In cluster sampling it is not mandatory to select from all the sub groups. We can select from our choice either from 2 group of from more or all groups.**
- **In stratified sampling it is mandatory to select from all the sub-groups.**

II Non-probability Sampling Methods

Not every member in a population has an equal probability of being selected to be in the sample.

This type of sampling method is sometimes used because it's **much cheaper and more convenient** compared to probability sampling methods.

It's often used when researchers simply want to gain an initial understanding of a population.

However, that result from these sampling methods cannot be used to draw inferences about the populations they came from because they typically aren't representative of the entire population.

Type of Non Probability sampling:-

(1)Convenience sample

Example: A researcher stands in front of a Shopping Mall during the day and polls people.

Drawback: Location and time of day will affect the results. More than likely, the sample will suffer from under coverage bias since certain people (e.g. those who work during the day) will not be represented as much in the sample.

(2) Voluntary response sample

Definition: A researcher puts out a request for volunteers to be included in a study and members of a population voluntarily decide to be included in the sample or not.

Example: A radio host asks listeners to go online and take a survey on his website.

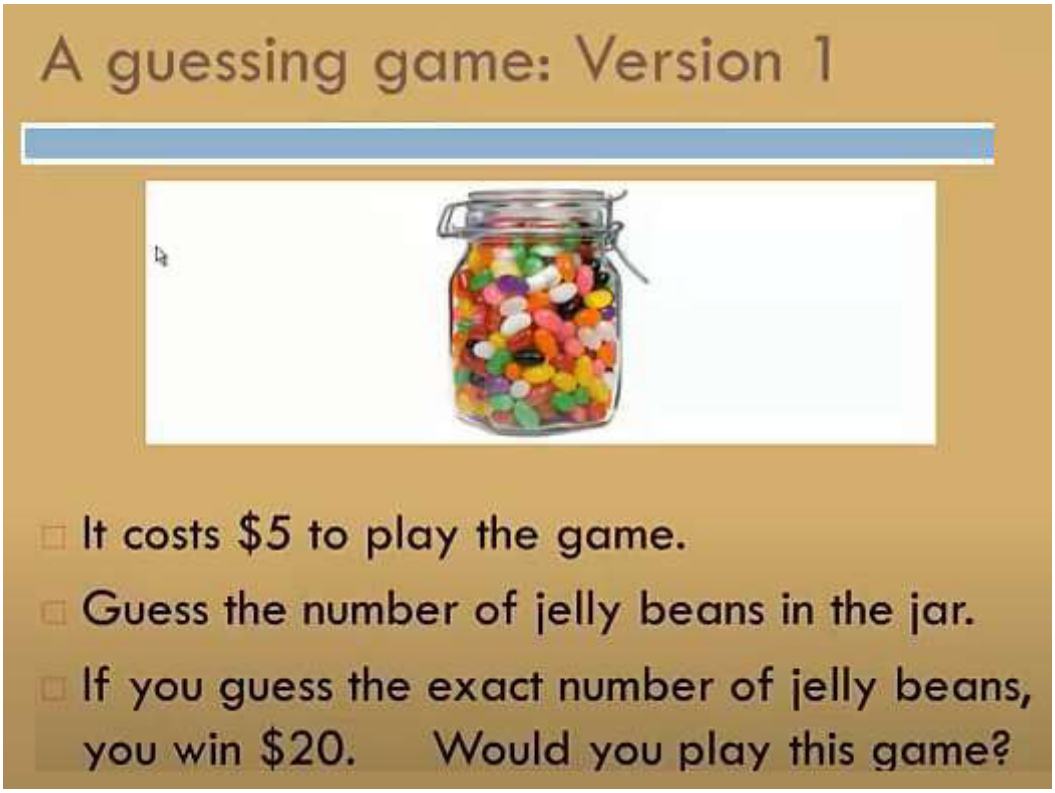
Drawback: People who voluntarily respond will likely have stronger opinions (positive or negative) than the rest of the population, which makes them an unrepresentative sample.

Using this sampling method, the sample is likely to suffer from nonresponsive bias – certain groups of people are simply less likely to provide responses.

Which sampling method is best?

The method of sampling best to use will depend on the nature of the analysis and the data being used. In general, simple random sampling is widely used, but in specific research application stratified sampling can produce a more accurate sample relative to the population under study.

Confidence Level and Margin of Error

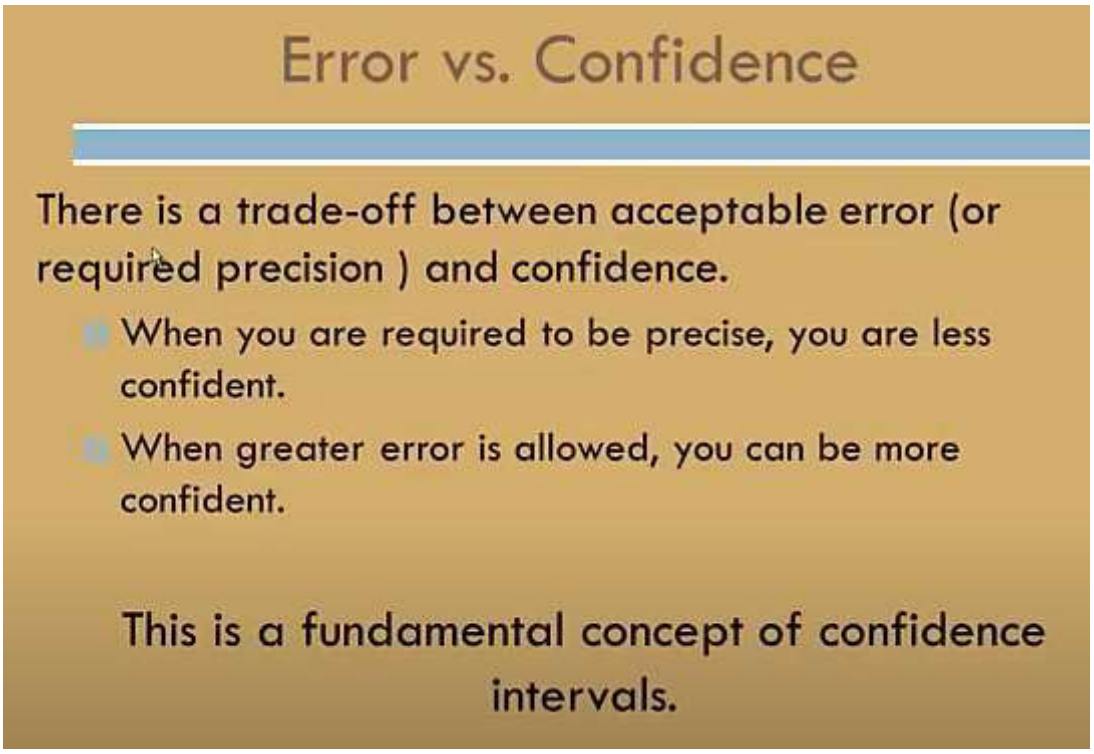


Option 1:- You Guess within 5 no. of actual no. of jelly.

Option 2:- You Guess within 25 no. of actual no. of jelly.

Option 3:- You Guess within 50 no. of actual no. of jelly.

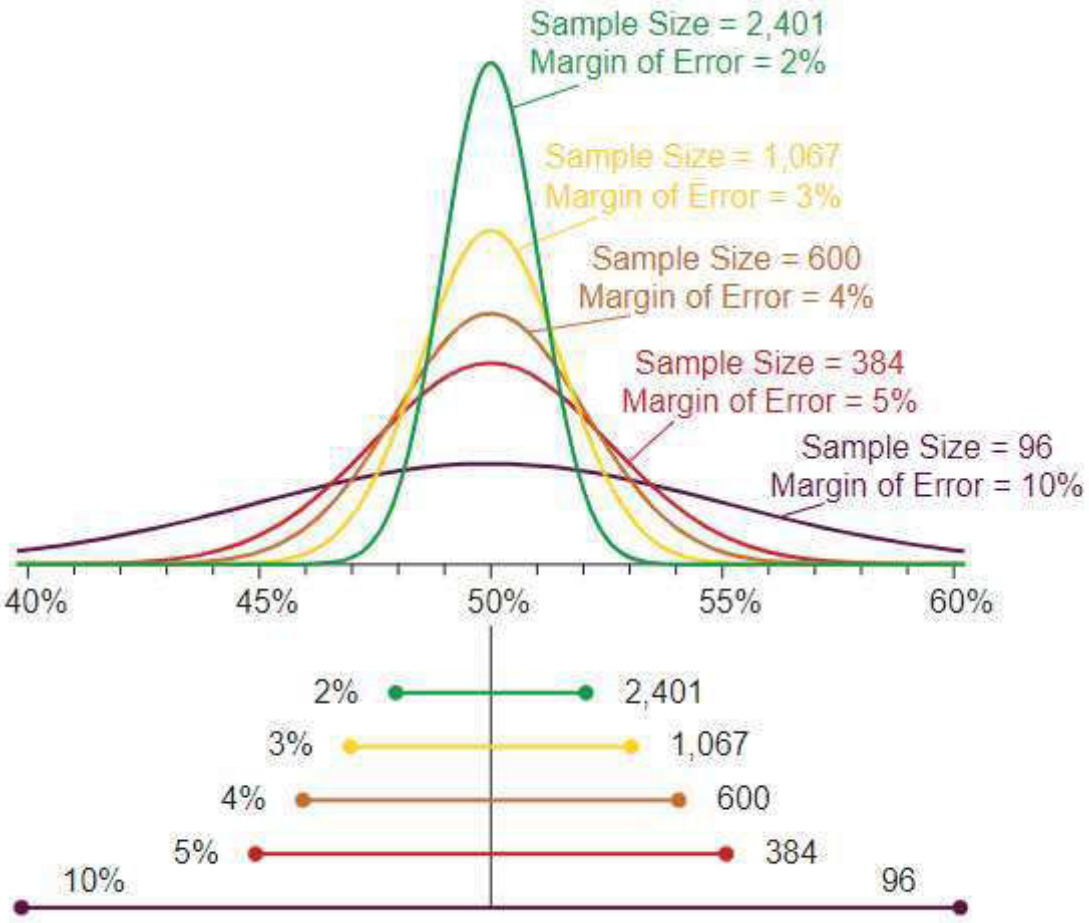
Option 4:- You Guess within 100 no. of actual no. of jelly.



An increase in confidence level results in the increase in the margin of error. We have to choose between the precision and confidence.

Calculating Sample Size

- **To determine a sample size that will provide the most meaningful results, researchers first determine the preferred margin of error (ME) or the maximum amount they want the results to deviate from the statistical mean.**
- **It's usually expressed as a percentage, as in plus or minus 5 percent. Researchers also need a confidence level, which they determine before beginning the study.**
- **This number corresponds to a Z-score, which can be obtained from tables. Common confidence levels are 90 percent, 95 percent and 99 percent, corresponding to Z-scores of 1.645, 1.96 and 2.576 respectively.**
- **Researchers express the expected standard of deviation (SD) in the results. For a new study, it's common to choose 0.5.**



Having determined the margin of error, Z-score and standard deviation, researchers can calculate the ideal sample size by using the following formula:

$$(\text{Z-score})^2 \times \text{SD} \times (1\text{-SD})/\text{ME}^2 = \text{Sample Size}$$

This formula does not depend on the size of the population, only on the size of the sample.

Effects of Small Sample Size

- **In the formula, the sample size is directly proportional to Z-score and inversely proportional to the margin of error.**
- **Consequently, reducing the sample size reduces the confidence level of the study, which is related to the Z-score.**
- **Decreasing the sample size also increases the margin of error.**
- **In short, when researchers are constrained to a small sample size for economic or logistical reasons, they may have to settle for less conclusive results.**
- **Whether or not this is an important issue depends ultimately on the size of the effect they are studying.**