Sampling Techniques

1. What is a population in statistics, and how is it different from a sample?

Population:

A population is the entire set of individuals, items, or data that you want to study.

Sample:

A sample is a subset of the population selected for analysis.

Difference:

- Population includes all members, while a sample includes only some members.
- Population parameters (mean, variance) are often unknown, while sample statistics are used to estimate them.

Example:

- Population: All students in a university.
- Sample: 100 students selected from that university.

2. Define sampling in statistics. Why is it necessary?

Sampling:

Sampling is the process of selecting a subset (sample) from a larger population to gather information.

Why it's necessary:

1. Cost-effective: Studying the entire population may be expensive or time-consuming.
2. Practical: Sometimes it's impossible to collect data from everyone.
3. Faster results: Analysis on a sample is quicker.
4. Reliable estimates: Proper sampling can provide accurate estimates of population parameters.

Example: Surveying 500 voters to estimate the preferences of all voters in a country.

3. What is the difference between probability sampling and non-probability sampling?

Feature	Probability Sampling	Non-Probability Sampling
Definition	Every member of the population has a **known, non-zero chance** of being selected.	Members are selected based on **judgment, convenience, or other non-random methods**.
Selection Method	Random selection.	Non-random selection.
Bias	Less bias; results are more **generalizable**.	Higher risk of bias; results may **not represent** the population.
Examples	Simple random sampling, stratified sampling, systematic sampling.	Convenience sampling, judgmental sampling, quota sampling.

4. Explain simple random sampling with an example.

Simple Random Sampling (SRS):

A simple random sample is a sampling method where every member of the population has an equal chance of being selected, and each selection is independent.

Example:

- A school has 200 students, and we want to select 20 for a survey.
- Using SRS, we could randomly assign numbers to all students and then use a random number generator to pick 20 students.
- Each student has an equal probability (20/200 = 0.1) of being selected.

5. What is systematic sampling? How is the sampling interval calculated?

Systematic Sampling:

Systematic sampling selects every k-th member from a population after choosing a random starting point.

How to calculate the interval:

- Divide the total population by the number of samples you want. This gives the step size (how often to pick a member).

Example:

- A company has 500 employees and wants a sample of 50.
- Pick a random starting employee (e.g., #4), then select every 10th employee: 4, 14, 24, … until 50 are chosen.

6. Describe stratified sampling and when it is useful.

Stratified Sampling:

Stratified sampling divides the population into distinct groups (strata) based on a characteristic (e.g., age, gender, department), and then samples are taken from each group.

When it's useful:

- When the population has different subgroups and you want each subgroup to be represented proportionally.
- It reduces sampling bias and improves the accuracy of results.

Example:

- A university has 60% undergraduates and 40% postgraduates.
- To survey 100 students, take 60 from undergraduates and 40 from postgraduates.

7. Explain cluster sampling with a real-world example.

Cluster Sampling:

Cluster sampling divides the population into groups (clusters), usually based on location or naturally occurring divisions, and then randomly selects entire clusters for the sample.

Example:

- A city has 50 schools. To survey students, instead of picking individual students from all schools, randomly select 5 schools (clusters) and include all students from those schools in the sample.

When used:

- Useful when the population is large and spread out, making it difficult to sample individuals directly.

8. What are the main differences between stratified sampling and cluster sampling?

Feature	Stratified Sampling	Cluster Sampling
Grouping	Population divided into **homogeneous subgroups (strata)** based on a characteristic.	Population divided into **heterogeneous groups (clusters)**, often naturally occurring.
Selection	Random samples are taken **from each stratum**.	Entire clusters are randomly selected, and **all members** of chosen clusters are included.
Purpose	Ensures **representation** of each subgroup.	Reduces **time and cost** when population is large and spread out.
Example	Surveying 60% undergraduates and 40% postgraduates by sampling proportionally from each group.	Selecting 5 schools randomly and surveying **all students** in those schools.

9. What is convenience sampling, and why is it not always reliable?

Convenience Sampling:

Convenience sampling selects samples that are easiest to access rather than randomly.

Why it's not always reliable:

- High risk of bias because it may not represent the whole population.
- Results cannot be generalized confidently.

Example:

- Surveying only friends or nearby people because they are easy to reach, instead of sampling the entire city.

10. Define quota sampling. How is it different from stratified sampling?

Quota Sampling:

Quota sampling selects samples to meet specific quotas for certain groups, but the selection within each group is non-random.

Difference from Stratified Sampling:

- Stratified Sampling: Random selection within each subgroup (stratum).
- Quota Sampling: Non-random selection to fill predetermined quotas.

Example:

- A survey requires 50% men and 50% women. Interviewers choose any available men and women until quotas are filled, rather than randomly selecting them.

11. What is snowball sampling, and in which type of research is it most useful?

Snowball Sampling:

Snowball sampling is a non-probability sampling method where existing study participants recruit or refer other participants from their network.

Most useful in:

- Research involving hard-to-reach or hidden populations where a complete list of members is not available.

Example:

- Studying people who use rare drugs or a niche online community, where participants refer others they know.

12. Explain judgmental/purposive sampling with an example.

Judgmental (Purposive) Sampling:

Judgmental sampling is a non-probability method where the researcher selects participants based on their knowledge or judgment, believing they are most relevant to the study.

Example:

- A researcher studying expert opinions on AI ethics chooses only AI professionals rather than randomly selecting people.

Use:

- Useful when specific expertise or characteristics are required for the research.

13. How does the sample size affect the accuracy and reliability of results?

Effect of Sample Size on Accuracy and Reliability:

1. Larger Sample Size:
    - More accurate estimates of population parameters.
    - Reduces sampling error.
    - Results are more reliable and generalizable.
2. Smaller Sample Size:
    - Higher chance of error or bias.
    - Less reliable; may not represent the population well.

Example:

- Surveying 5 people about city traffic may give inconsistent results, but surveying 500 people provides a more accurate picture.

14. What is sampling bias? Give an example of how it can occur.

Sampling Bias:

Sampling bias occurs when a sample does not represent the population accurately, leading to distorted or untrustworthy results.

Example:

- Surveying only people at a gym to estimate the city's average fitness level.
    - People at the gym are more likely to be fit, so the sample is biased and does not reflect the whole population.

15. Why is it important to ensure that a sample is representative of the population?

Importance of a Representative Sample:

- Ensures that results accurately reflect the population.
- Reduces bias and errors in estimates.
- Makes findings generalizable to the entire population.
- Increases credibility and reliability of the research.

Example:

- If a survey on students' study habits only includes top-performing students, it won't represent the habits of all students.