

Inferential Statistics : Making Predictions from Data

What is inferential statistics?

Inferential statistics involves using a sample (a small subset of data) to make generalization about a larger population. Instead of analyzing entire populations - often impractical due to size or cost

Population vs sample

Term	Definition	Example
Population	The complete set of all individuals in terms of interest	All USSR citizens in a voting survey
sample	A smaller group selected from the population	2000 randomly chosen US citizens surveyed about voting fairness
Sampling Bias	Occurs when the sample does not accurately represent the population	surveying only Florida residents about voting fairness

Examples of inferential statistics in Action

"Classroom Test Scores"

Population : All students in the class
 sample : 10 students sitting in the front row
 Inference : class performance on the last test

Problem :- Front-row students tend to be more engaged and perform better than the average student, making the sample unrepresentative

Key Takeaways

- Inferential statistics allow us to make generalization about a population from a sample.
- Random sampling is crucial to avoid bias and ensure accuracy.
- Bias in sample selection can lead to misleading conclusions.
- Larger samples typically provide more accurate inferences.

Simple Random Sampling : Ensuring fair selection

What is Simple Random Sampling?

- => Simple Random Sampling (SRS) is a method where:
- Every member of the population has an equal chance of being selected.
 - The selection of one member is independent of the selection of others.
 - The sample is chosen purely by chance (eg drawing names from a hat, using a random number generator).

Example: Twin Study Analysis

Population: All twins listed in the National Twin Registry

Sample: Twins with last names starting with "Z" and a subset of those with "B"

Issues with sampling:

Not Simple Random Sampling

- The sample selection was based on last names, which does not give every twin an equal chance of being chosen.
- Overrepresentation of certain ethnic groups (eg surnames starting with "Z")

Potential Bias:

- Ethnic bias (some cultures may have more surnames starting with Z or B)
- Behavioral bias (eg people with last names starting with "Z" may be more patient)

Why Sample Size Matters in Inferential Statistics

Random Samples vs Representative Samples

A random sample means every member of the population has an equal chance of being selected.

A representative sample accurately reflects the characteristics of the population

small random samples may not always be representative due to chance fluctuations.

Example Gender Imbalance in Small Samples

Imagine a population where 50% are male and 50% are female. If we take a random sample of 20 people, there's still a 6% probability (0.06) that 70% or more of the sample could be female.

- This does not mean the sampling method was flawed but rather that small samples are more likely to produce skewed results due to random chance.

How Large Samples Reduce Bias

Larger samples tend to be more representative because they reduce the impact of randomness.

Inferential statistics adjust for sample size when making predictions about the population.

Key Rule:

Small samples \rightarrow More variability \rightarrow less reliable conclusions
Larger samples \rightarrow less variability \rightarrow More reliable conclusions

More complex sampling methods in inferential statistics

- In some cases, simple random sampling isn't practical due to logistical challenges, such as: -
- people moving in and out of the population
- lack of contact information (eg no phone, not registered to vote).
- Constant population changes (new residents, out of city matters)

Random Assignment in Inferential Statistics

In experimental research, random assignment is the process of randomly dividing a sample into different experimental conditions (eg treatment vs control)

Why is it important?

- Ensures that pre-existing differences between participants are evenly distributed across groups.
- Eliminates selection bias, making results valid
- Increases the likelihood that differences between groups are due to the treatment not other factors.

Stratified Sampling in Inferential Statistics

Stratified sampling is a method used when a population consists of distinct subgroups (strata). Instead of randomly sampling the entire population, researchers

- (1) Identify subgroups (Strata) in the population.
- (2) Ensure each subgroup is proportionally represented in the sample.
- (3) Randomly select individuals from each sub group

Example Capital Punishment Views at Urban University
Population: All students at the University
Strata:

- Day students (70%) → Avg age 19
- Night students (30%) → Avg age 39

Stratified sampling approach

- Total sample size: 200 students
- ∴ 140 Day students (70%)
- 60 Night students (30%)

Why use stratified sampling?

- ⇒ Ensures Representation: Each subgroup is proportionally included
Increases Accuracy: Reduces sampling bias compared to simple random sampling.