8360944207

Sampling Chapter 2 Module A

The foundation of statistical inference enables us to draw meaningful insights from limited data, making informed decisions in business and research contexts.

iibf.info/app

What is Sampling?

Sampling involves selecting a representative subset from a larger population to study characteristics without examining every individual.

Cost-Effective

Significantly reduces research expenses and resource allocation

Time-Efficient

Enables rapid data collection and analysis processes

Reliable Results

Provides accurate estimates when executed with proper methodology

Why Sampling Matters

Population Study Efficiency

Enables comprehensive analysis of large populations without exhaustive enumeration, making research practically feasible

Resource Optimisation

Dramatically reduces time, financial investment, and human resources required for data collection and analysis

Data-Driven Decision Making

Provides statistically sound foundation for strategic business decisions and policy formulation

AB AC AD
BC BD CD

Two Major Sampling Types

A B C D

$\frac{1}{6}$

$P(A) = P(B) = P(C) = P(D)$

$\frac{3}{6} = \boxed{\frac{1}{2}}$

$^4C_2 = \frac{4!}{2!\,2!} = \frac{4 \times 3 \times \cancel{2 \times 1}}{\cancel{2} \times 1 \times \cancel{2 \times 1}}$ $= 6$

Judgement Sampling

biasness

Non-random selection based on researcher expertise and discretion

Random Sampling ✓

Probability-based selection ensuring equal opportunity for all members

The choice between sampling methods depends fundamentally on research objectives, population characteristics, and desired inference validity.

Judgement or Non-Random Sampling

1

Researcher Discretion

Selection based on expert knowledge and professional judgement of population characteristics

2

Subjective Criteria

Predetermined criteria applied consciously, introducing potential selection bias into the sample

3

Practical Application

Example: selecting experienced bankers with 10+ years tenure for specialised financial research

Random Sampling

Equal Probability

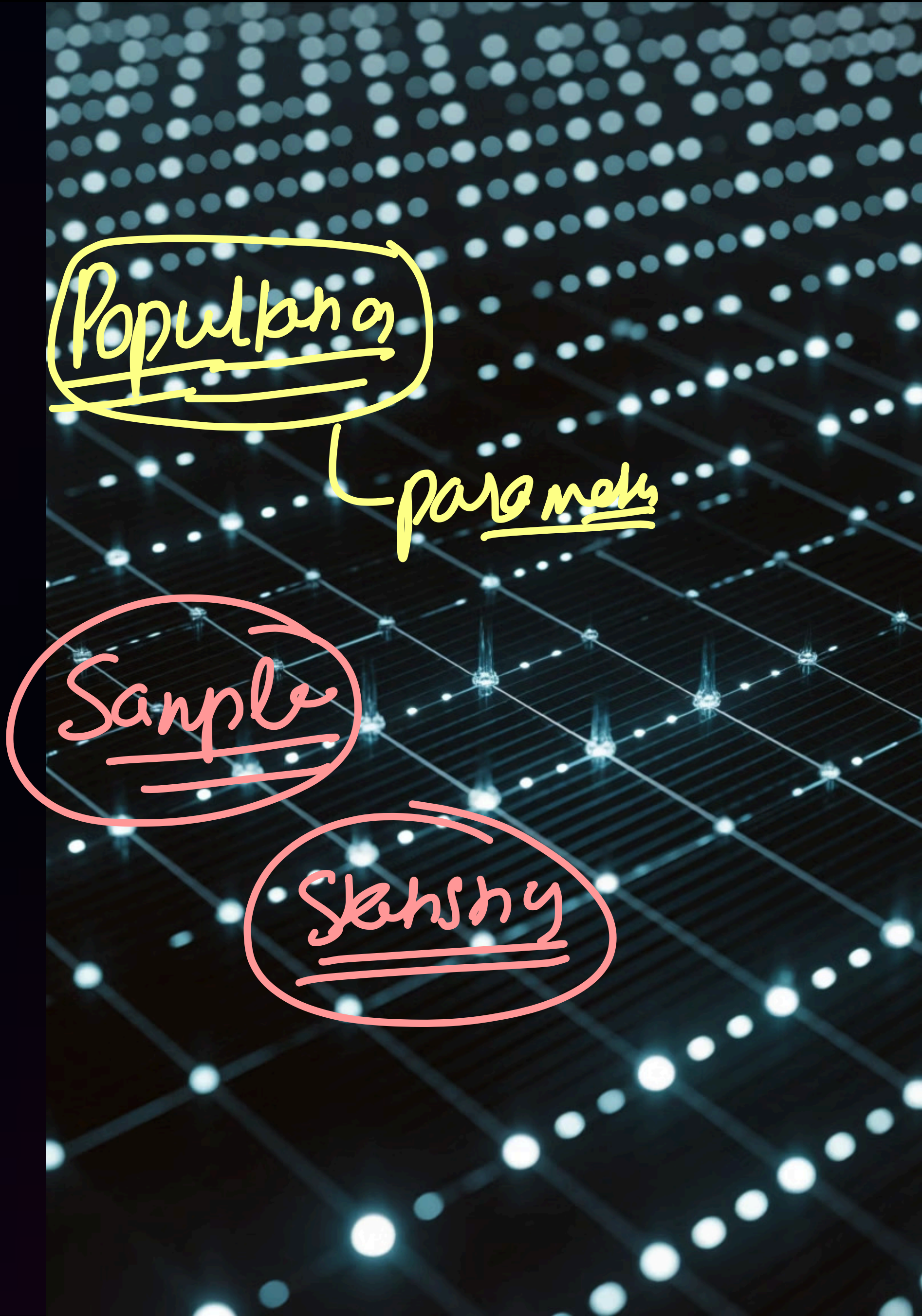
Every population member has identical chance of selection

Unbiased Representation

Eliminates systematic selection bias from sampling process

Statistical Foundation

Enables valid inference from sample to population parameters



Simple Random Sampling

100 students

Pure Randomisation

The most straightforward probability sampling method where selection occurs entirely by chance.

- **Equal Selection Probability**

  Each population member has identical likelihood of inclusion

- **Lottery Mechanism**

  Comparable to drawing names from a hat or using random number tables

- **Unbiased Approach**

  Eliminates researcher influence and selection prejudice

Systematic Random Sampling

01

Calculate Sampling Interval

Determine k = N/n where N is population size and n is desired sample size

02

Random Starting Point

Randomly select first element between 1 and k

03

Select Every kth Element

Continue selecting at regular intervals throughout the sampling frame

Example: Selecting every 10th bank account record from an ordered database of 10,000 accounts to obtain a sample of 1,000

Stratified Sampling

5-15 Homogeneous 16-25 26-35 35+

V. Imp

Ensuring Proportional Representation

G1 ↔ G4

Population divided into homogeneous subgroups (strata) based on key characteristics before sampling occurs.

1

Identify Strata

Divide population by relevant characteristics: age, income, region, education level

2

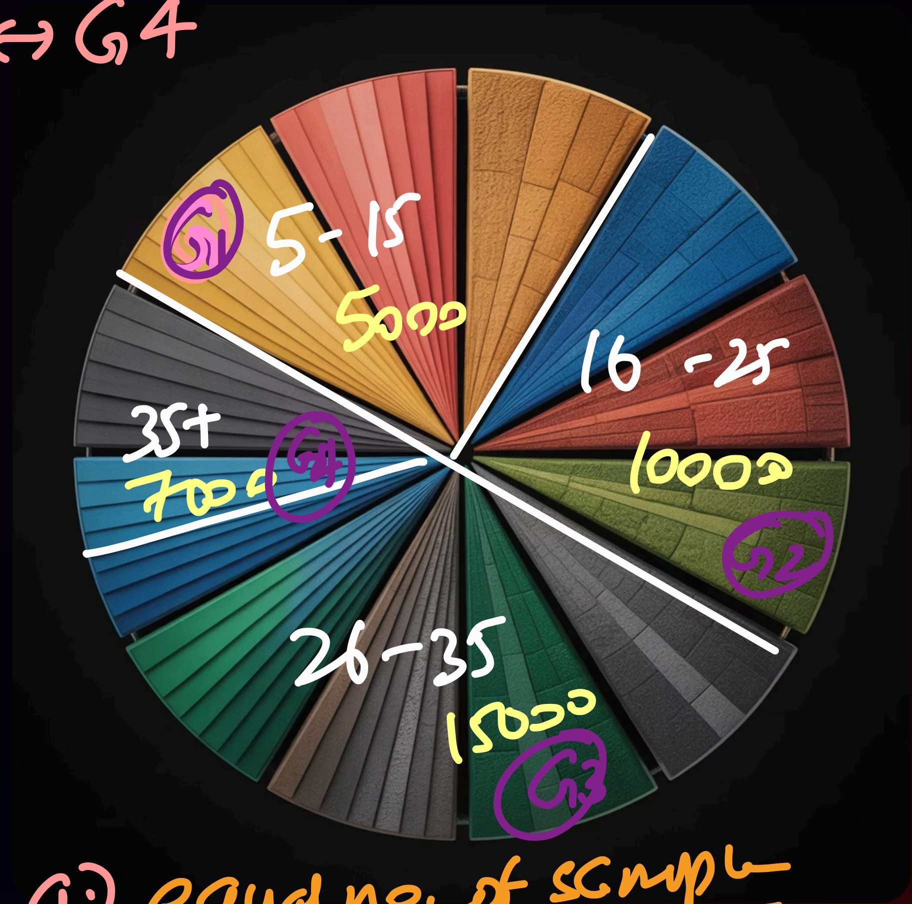
Sample Each Stratum

Apply random sampling within each subgroup independently

3

Combine Results

Aggregate data ensuring all population segments represented



a:) equal no. of sample

weightage $\frac{5000}{37000}$

b) $\frac{5000}{37000} \times 10000$

Heterogenous groups

$C_1 \leftrightarrow C_2$

Cluster Sampling

1	2	3

Geographic Division

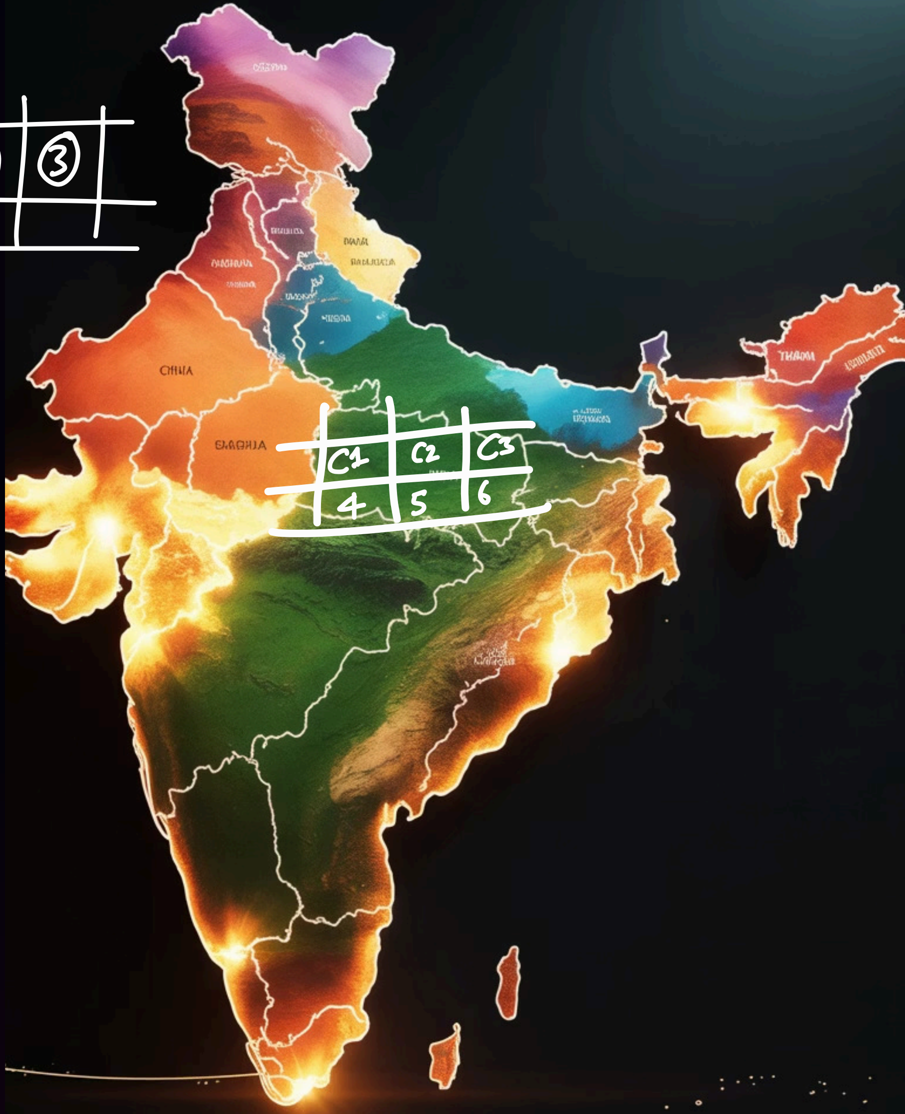
Population naturally divided into clusters (cities, districts, branches) with internal heterogeneity

Random Cluster Selection

Entire clusters chosen randomly rather than individual members

Cost Efficiency

Dramatically reduces travel and administrative costs for geographically dispersed populations



Sampling Distribution

The probability distribution of a sample statistic (such as the mean) obtained from all possible samples of a specific size.

Illustrative Example

Sample	Mean Height (cm)
Sample A	159
Sample B	162
Sample C	160
Sample D	163
Population Mean (μ)	**161**

$\frac{159+162}{2}$

160.5 159.5

AB AC AD

BC BD CD

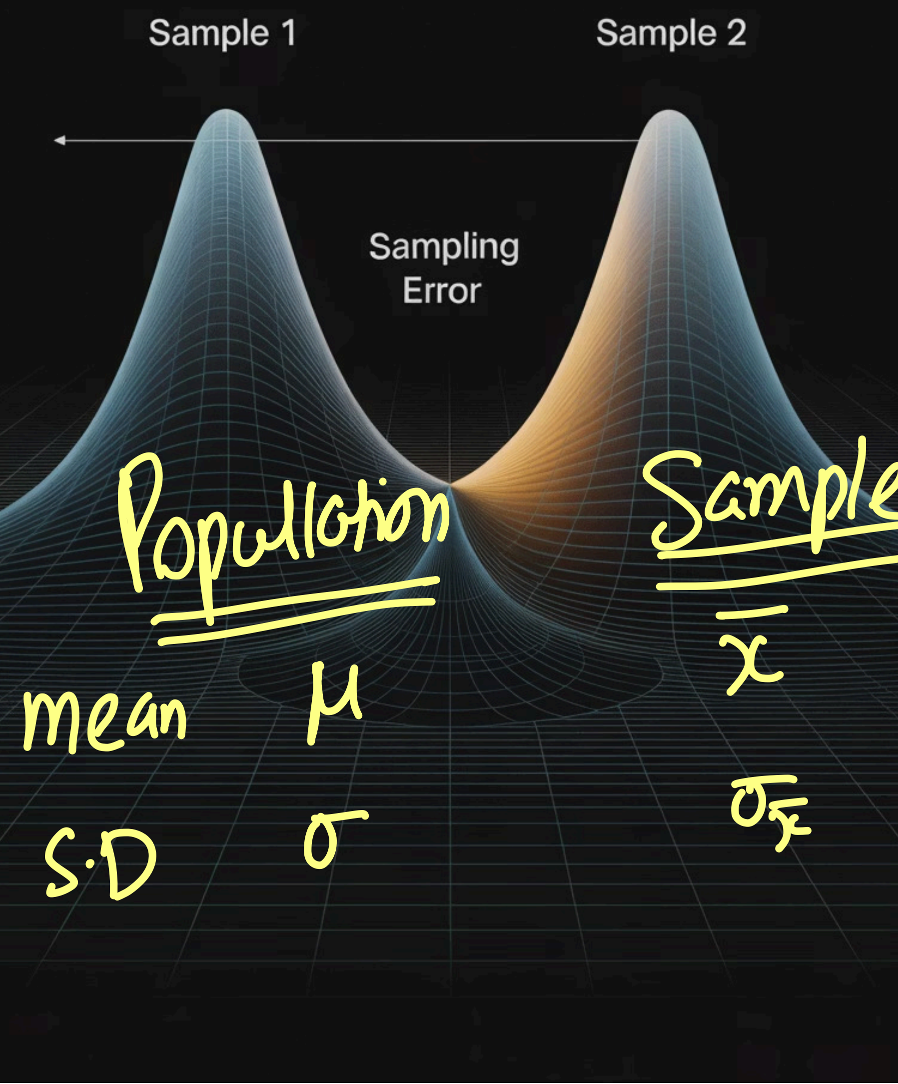
$^4C_2 = \boxed{6}$

1000? 50?

Demonstrates natural variation amongst sample means, all clustering around the true population parameter.

$\mu = 161$ $\quad$ $\bar{x} = 159.5$

161.5



	Population	Sample
mean	μ	$\bar{x}$
S.D	σ	$\sigma_{\bar{x}}$

Sampling Error

Definition

The difference between a sample statistic and the corresponding population parameter

$$\text{Sampling Error} = \bar{x} - \mu$$

Natural Occurrence

Arises inevitably due to chance variation in random selection processes

Reduction Strategy

Increasing sample size (n) systematically reduces sampling error magnitude

$\mu = \bar{x}$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Normal Distribution

N.D

$\boxed{n > 30}$

#1 **Symmetrical Shape**

Perfectly balanced around the central mean value

#2 **Central Tendency**

Mean = Median = Mode = μ at the distribution's peak

Statistical Foundation

Underpins parametric inferential statistics and hypothesis testing

Continuous Range

Extends infinitely in both directions, approaching but never touching zero

N.D.

CLT $n \geq 30$

Central Limit Theorem (CLT)

$\mu = \bar{x}$

The Foundation of Inference

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Sample Size Increases

As n grows larger (typically n ≥ 30)

Normality Emerges

Sampling distribution of means becomes approximately normal

Mean Converges

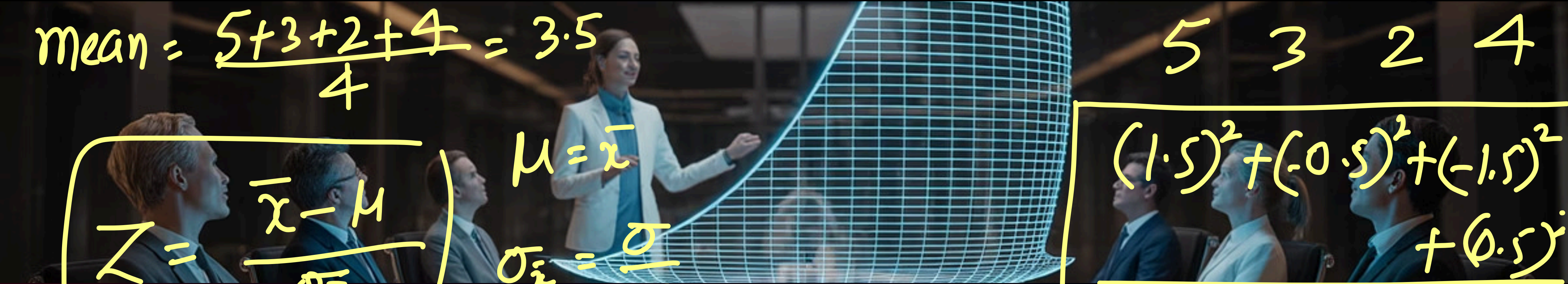
Mean of sampling distribution equals population mean (μ)

σ^2

$S.D = \sqrt{\text{variance}}$

$= \frac{\sum (x-\bar{x})^2}{n}$

This remarkable property holds regardless of the original population's distribution shape.



$$\text{mean} = \frac{5+3+2+4}{4} = 3.5$$

5 3 2 4

$$\boxed{Z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}}}$$

$$\mu = \bar{x}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{\frac{(1.5)^2+(-0.5)^2+(-1.5)^2+(0.5)^2}{4}}$$

Why CLT Matters

Enables Statistical Inference

Allows us to make probability statements about population parameters using sample data

Universal Applicability

Works effectively even when population distribution is unknown or non-normal

Hypothesis Testing Core

Forms the theoretical foundation for confidence intervals and significance tests

Application: Bank Teller Earnings

$0 \le P(\bar{x}) \le 1$

$\mu = \bar{x}$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2000}{\sqrt{30}} = 365.16$

Given Parameters

- Population mean: **μ = ₹19,000**
- Population standard deviation: **σ = ₹2,000**
- Sample size: **n = 30**

$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$

$Z = \frac{19750 - 19000}{365.16} = 2.05$

Question

What is the probability that the sample mean exceeds ₹19,750?

0.4798

Probability Dist Table

0.4798

$0.5 - 0.4798$

0.0202

Solution

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{2000}{\sqrt{30}} = 365.15$$

$$Z = \frac{19750 - 19000}{365.15} = 2.05$$

Result: $P(\bar{x} > ₹19{,}750) = 2.02\%$

Standard Error Concept

Definition

Standard error (SE) measures the precision of the sample mean as an estimate of the population mean

$$SE = \frac{\sigma}{\sqrt{n}}$$

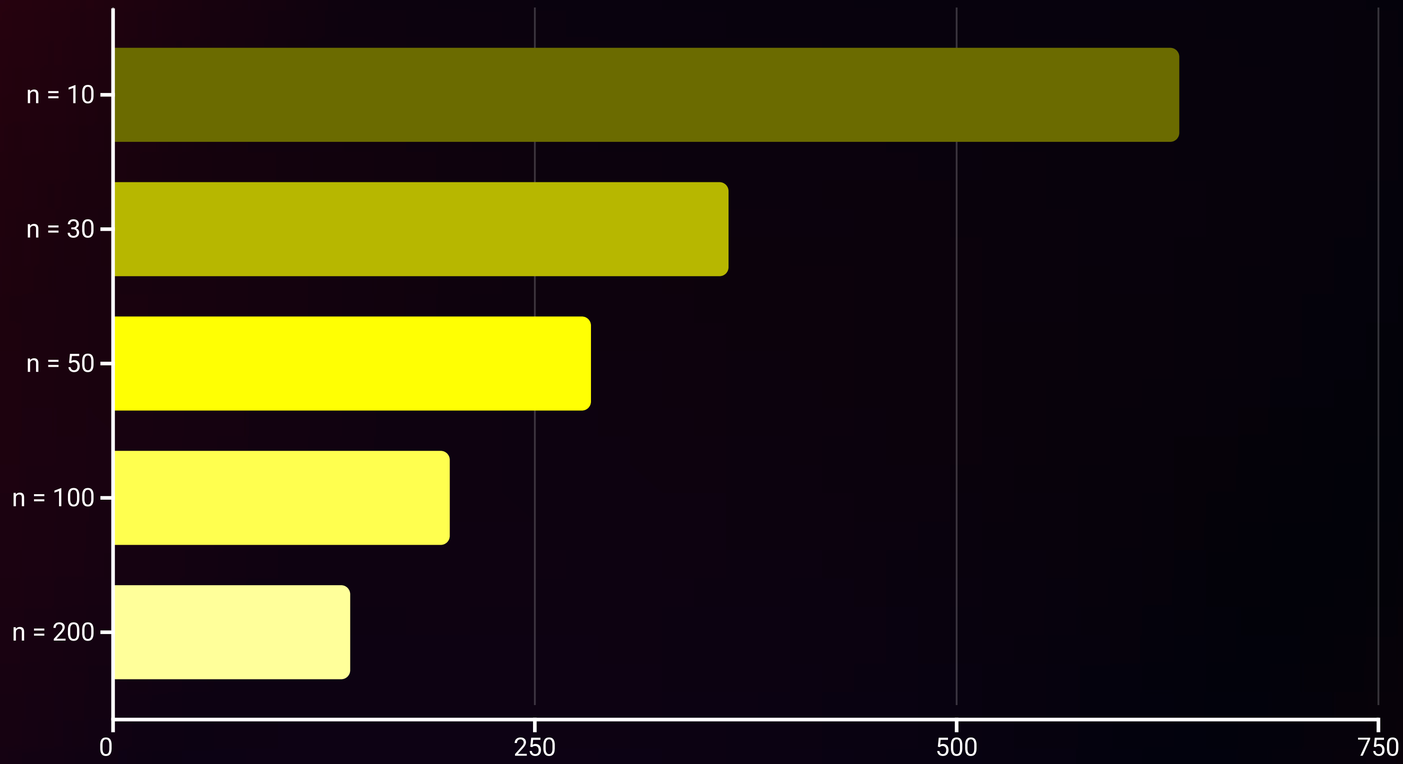
Interpretation

Lower SE indicates sample mean is a more reliable estimator of population mean

Practical Impact

Determines width of confidence intervals and power of hypothesis tests

Relationship Between Sample Size & Standard Error



4.5×

SE Reduction

When sample size increases from 10 to 200

$\sqrt{n}$

Inverse Relationship

SE decreases proportionally to square root of sample size

↑

Precision Gain

Larger samples yield dramatically improved estimation accuracy

Finite Population Multiplier

Adjustment for Sampling Without Replacement

When sampling from finite populations without replacement, the standard error formula requires modification.

Adjusted Standard Error Formula

$$SE = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

Where N = population size and n = sample size

Practical Rule: The finite population correction can be ignored when $n/N < 0.05$ (sample is less than 5% of population)