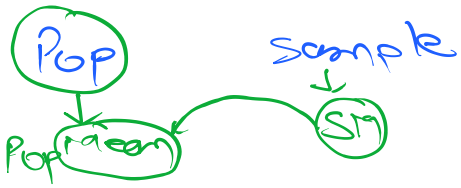
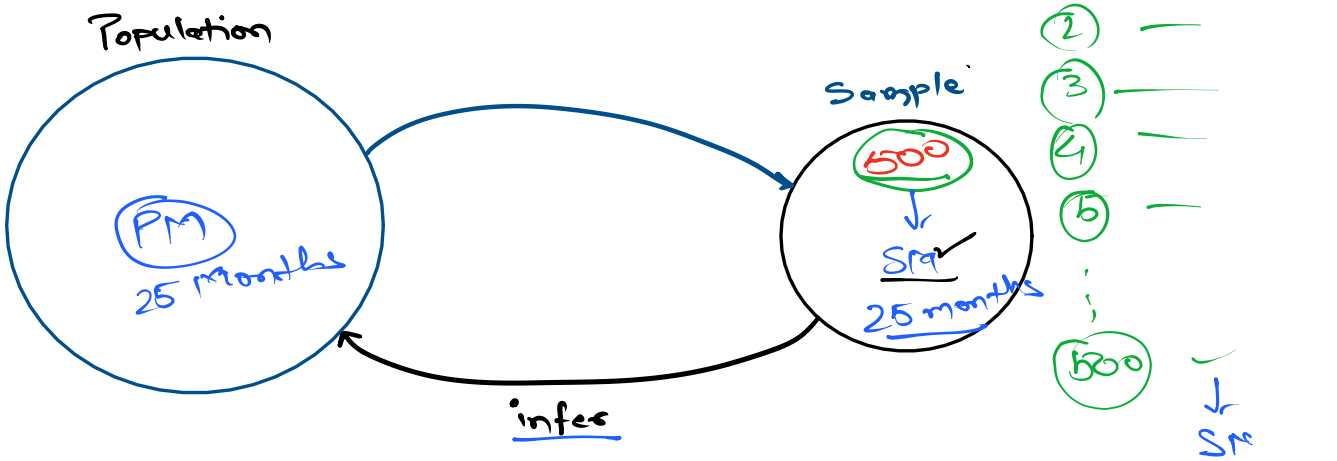
Inferential Statistics

Monday, 9 June 2025 19:04



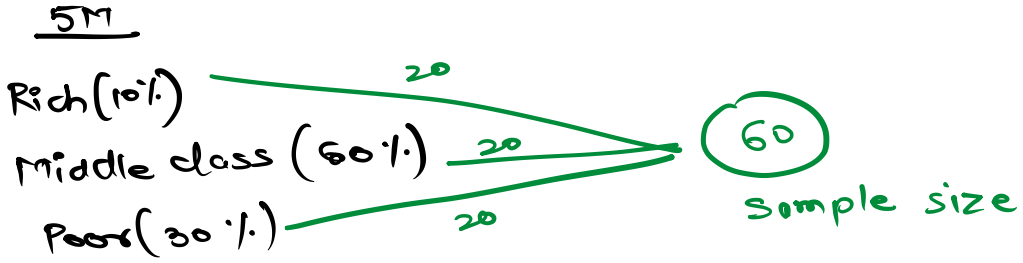
Inferential statistics →

Involves making predictions or inferences about a popula-tion based on the sample data taken from that population.

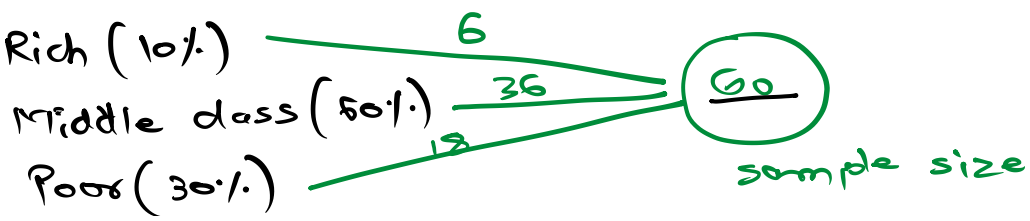


→ We use sample mean to infer population mean.

→ The sample should be representative of the popula-tion.

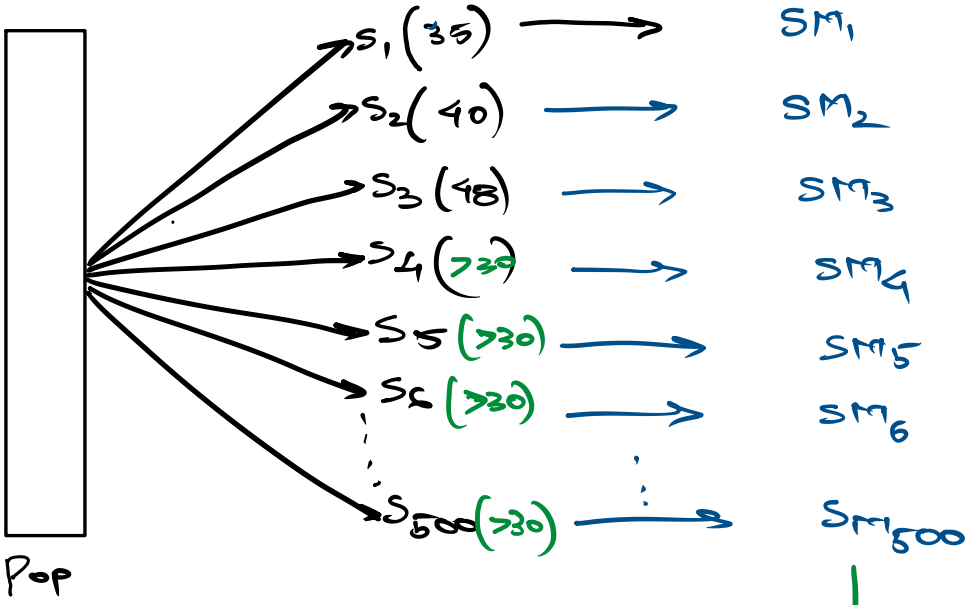


Disproportionate sampling.



Proportionate Sampling.

Central Limit Theorem → Sample Means



Take all the sample means & plot these value.

(sns.kdeplot or sns.displot)

↓

It will always come out to be a "Normal Distribution".

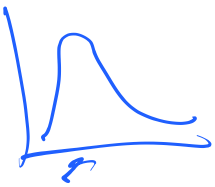
sample 500 Means

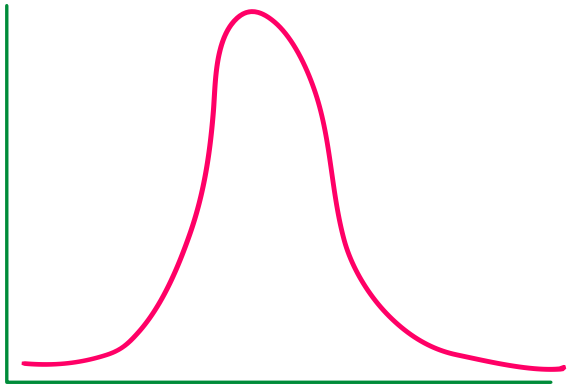
↓

Num

→ Histogram

→ Kdeplot





Pop → 6 rm → 730 / 40

Pop → 1 sample (S6 > 30) → ND

Note: Sample sizes > 30.

Takeaway: If we take one sample only from any population, we can safely assume that the sample will be 'Normally Distributed', given that the sample size > 30..

Sample = 50 Tyres

↓

Test these

1. 24.3 months.
2. 29.8 months
3. 30.5 months.
4. 28.9 months.

⋮

50. 31.9 months.

→ Plot these (kdeplot)

↓

Normal Distribution

Population Mean = Sample Mean

↓

Point Estimate

↓

Mostly incorrect

With some adjustments, we can say →

Population Mean = Sample Mean ± Margin

= 24.7k ± 1k

= 24.7k + 1k to 24.7k − 1k

PM = 23.7k to 25.7k

Interval / Range

90% confidence ⟶ 24.7k ± 1k

95% confidence ⟶ 24.7k ± 1.5k

99% confidence ⟶ 24.7k ± 2k

↓

Confidence Interval.

Formula to calculate the Confidence Interval →

$$CI = \bar{x} \pm z^* \times \frac{s}{\sqrt{n}}$$

$\bar{x}$: Sample Mean

s : standard deviation of the sample.

n : Sample size

z^* : z-score for a certain confidence level.

z^*	confidence level
1.65	90%
1.96	95%
2.58	99%

Ex: Estimate whether the mean lead content in maggi packets is within the allowed range or not?

Statistics:

1. Descriptive Statistics
2. Inferential Statistics → Hypothesis Testing — Populations & samples