Inferential stats

1. Central limit theoram (CLT)
2. Hypothesis testing (Null hypothesis, Alternate hypothesis)
3. CI (Confidence interval)
4. Significance value (Critical value)
5. two and one fail test (acceptance and rejection area)
6. Standard error
7. P-value
8. Sampling technique
9. Point estimate
10. type 1 and type 2 error
11. z-test (t-test, Chisqure, F-test, ANOVA)

Sampling techniques → With respect to Population

1. Simple Random Sample
2. Stratified sampling
3. Systematic Sampling
4. cluster Sampling

What is Population.

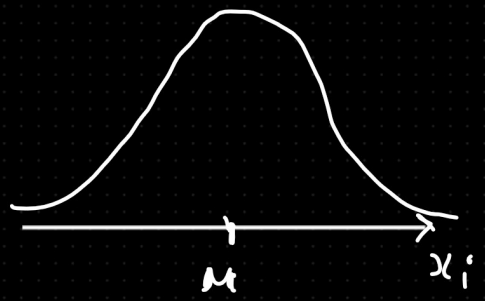
→ City
→ Country
→ world

Complete data set wrt any entity

Sample ⇒ it is a subset or part of the population

(1) CLT ⇒ Population.

$$X \simeq N(\mu, \sigma)$$

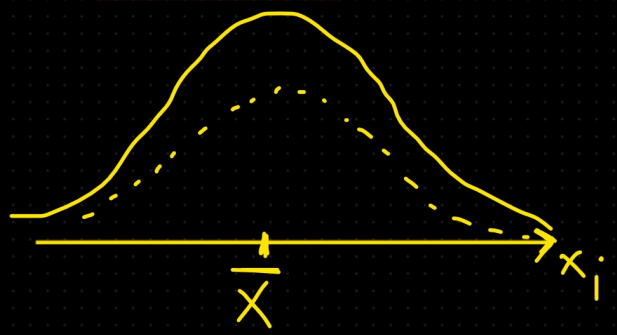


n = Sample size ⇒ any value

$n \geq 30$

$$S_1 = \{x_1\ x_2\ x_3\ \dots\ x_n\} = \bar{x}_1$$
$$S_2 = \{x_1\ x_2\ x_3\ \dots\ x_n\} = \bar{x}_2$$
$$S_3 = \{x_1\ x_2\ x_3\ \dots\ x_n\} = \bar{x}_3$$
$$\vdots$$
$$S_n = \{x_1\ x_2\ x_3\ \dots\ x_n\} = \bar{x}_n$$

Normally Distributed



(Sampling Distribution)

⇐ Sample mean Distribution

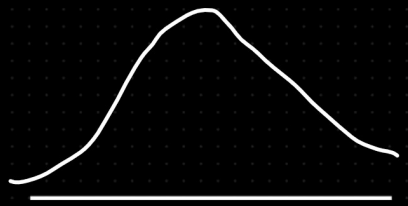
$$\therefore X = N(\mu, \sigma)$$

$$\bar{X} = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Population size

Standard error

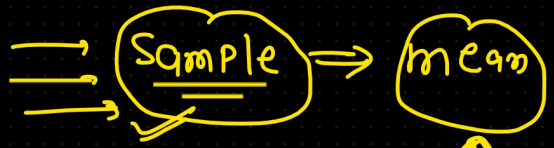
Sampling Distribution of the mean



n = Sample size

$(n \geq 30)$

Point estimate ⇒ = Population



— Sample Size ⇒ $n \geq 30$

Point estimate

= no. of sample ⇒ $k = 1, 2, 3, \ldots n$

1 ⇒ 35
2 ⇒ 30
3 ⇒ 37

⇒ Same

30
30
30

hypothesis testing

1. Null hypothesis (H_0)
2. alternative hypothesis (H_a)

$p \leq 0.1$

90-1.

95-1.

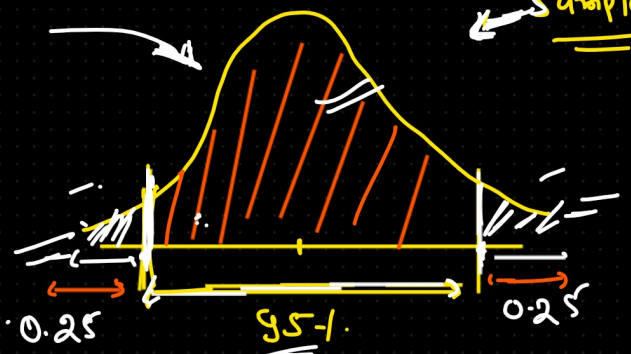
$p \leq 0.01$

Sample

CI 95-1.

$p = 0.05$

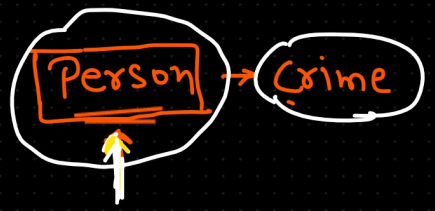
$p \leq 0.05$



court

hypothesis testing

(Any statement or any assumption)



test

hypothesis (Null hypothesis) H_0 ⇒ Person has not done any crime.

(evidence ⇒ DNA test, fingerprint, CCTV footage)

test → fail or Pass →

reject the null hypothesis and accept alternative hypo.

accept null hypothesis and we reject alternative hypothesis

Statement ⇒ H_0 (null hypothesis) (Population — hypothesis)

OR

Assumption

Sample ← test (2-test, t-test, chisquer, anova, f-test)

(evidence)

(Point estimation, CI, α → Pvalue, 2 tail test, one tail test)

P-value ⇒ Probability value for the Null hypothesis to be true

95% CI ⇒ $P \leq 0.05$

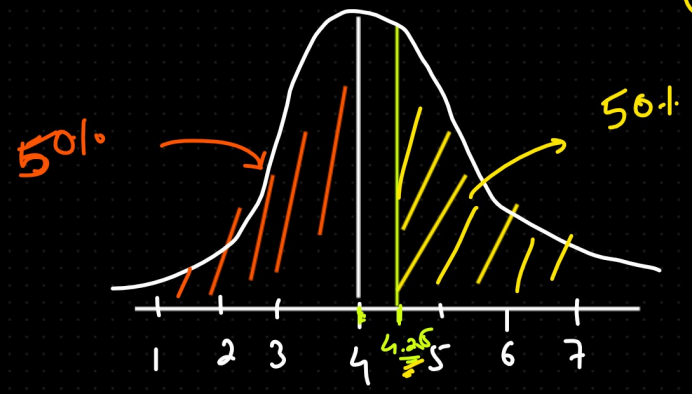
99.1% CI ⇒ $P \leq 0.01$

90% CI ⇒ $P \leq 01$

Question $X \sim N(4, 1)$

μ σ

100% → 1



$x_i = 4.25$

$$\text{Z-score} = \frac{\text{D.P.} - \text{mean}}{\text{std}}$$

$$= \frac{4.25 - 4}{1} = 0.25$$

my DP ⇒ how many std dev from the mean

→ Z-table

4 (4.25) ← Z-score (Standard score) = 0.59871

0.25

Z-table

1 - 0.59871 = 0.4013 = 40.13%

3.5 4 4.5

Data = 3.5 & 4.5 → 0.30854

Z-score = $\frac{3.5 - 4}{1} = -0.5$

Z-score = $\frac{4.5 - 4}{1} = +0.5$ → 0.69146

0.69146 - 0.30854 = 0.38292

= 38.29%

Q. In india the avg IQ is 100, with the std dev of 15. What is the percentage of the population would you expect to have an IQ lower than 85.

$\mu = 100$, $\sigma = 15$

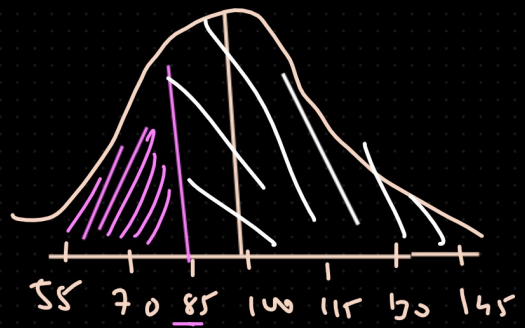
$$2\text{-score} = \frac{85 - 100}{15} = \frac{-15}{15}$$

$$= -1$$

AUC = 0.15866

↓

15.86% < 85



IQ ≥ 85 → $1 - 0.15866$ ⇒ 84.13%

$= 0.8413$

1. 2-tail
2. t-tail
3. Anova, f-tail
4. Chi-square

CI | P value | two-one tail