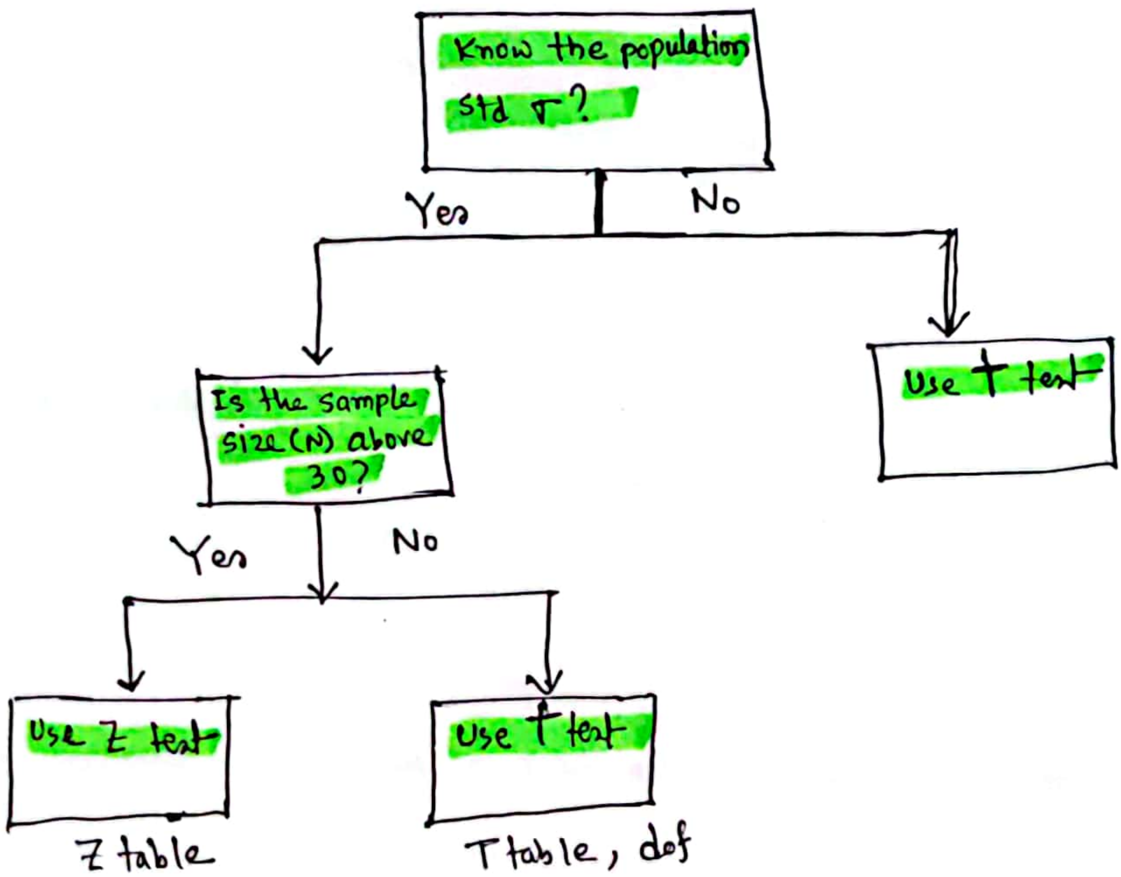
Advance Statistics - 04

Topics

1. When to use T test VS Z Test
2. Type 1 and Type 2 error
3. Confidence Interval & Margin of error
4. Bayes Theorem

① When to use T test vs Z-Test



② **Type 1 and Type 2 error:**

In reality: Null hypothesis is True or False

Decision (Conclusion): Null hypothesis can be True or False

Outcome 1: We reject null hypothesis when in reality it is False.

→ Good scenario

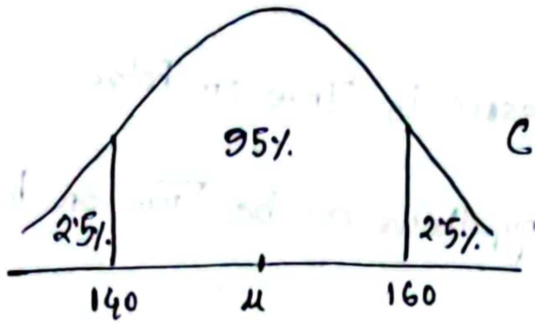
Outcome 2: We reject null hypothesis when in reality it is True

→ Type 1 error

Outcome 3: We retain or accept null hypothesis when in reality it is False → Type 2 error

Outcome 4: We retain or accept null hypothesis when in reality it is True → Good scenario.

Confidence Interval and Margin of Error:



C.I = 95%

Point Estimate: A value of any statistics that estimates the value of an unknown population parameter.

$$\vec{x} \longrightarrow \mu$$

$$\bar{x} = 2.95 \qquad \mu = 3$$

Confidence Interval: We construct a confidence interval to help estimate what the actual value of the unkown population mean is.

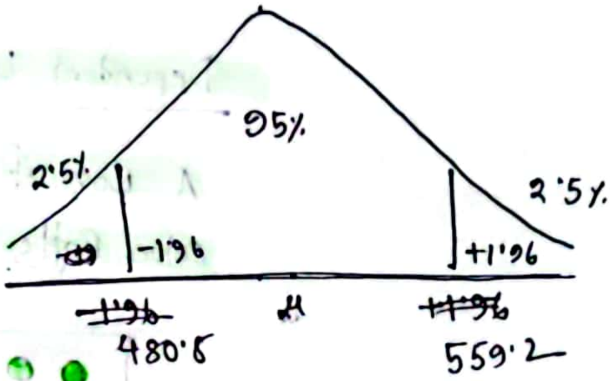
Point estimate $\pm$ Margin of error

For, Z test,

Margin of error: $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Problem: On the verbal section of CAT exam, the std is known to be 100. A sample of 25 test takers has a mean of 520. Construct a 95% C.I about the mean?

ans: $\bar{x} = 520$, $\sigma = 100$, $n = 25$, C.I = 0.95, $\alpha = 0.05$



$$MOE = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

(C.I = Confidence Interval)

$$\text{lower CI} = 520 - (1.96) * \frac{100}{\sqrt{25}}$$

$$= 480.8$$

$$\text{higher CI} = 520 + (1.96) * \frac{100}{\sqrt{25}}$$

$$= 559.2$$

Answer: I am 95% confident that the mean CAT score lies between 480.8 and 559.2.

Baye's Theorem:

Bayesian Statistics is an approach to data analysis and parameter estimation based on Bayes' Theorem.

Probability → Independent Events
→ Dependent Events

Independent Events

(i) Rolling a Dice

$\{1,2,3,4,5,6\}$

$Pr(1) = \frac{1}{6}, Pr(2) = \frac{1}{6} \ldots$

Same probability for all case

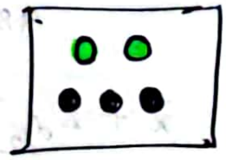
(ii) Tossing a coin

$Pr(\text{Head}) = 0.5, \ Pr(\text{tail}) = 0.5$

Same probability for all case

Dependent Events

A Box of 2 different color Balls



$Pr(\text{Green}) = \frac{2}{5}$

Suppose 1 ball taken out

$Pr(\text{black}) = \frac{3}{4}$

Here one event of taking out one ball from the box is impacting the other event. Because for reducing the ball Pr calculation becomes different

$Pr(\text{Green and Black})$

$= Pr(\text{Green}) \times Pr(\text{Black} \mid \text{Green})$

$= \frac{2}{5} \times \frac{3}{4} = \frac{6}{20}$

Now we can say,

$$Pr(A \text{ and } B) = Pr(B \text{ and } A)$$

$$\Rightarrow Pr(A) \times Pr(B/A) = Pr(B) \times Pr(A/B)$$

$$\Rightarrow \boxed{Pr(B/A) = \frac{Pr(B) \times Pr(A/B)}{Pr(A)}} \rightarrow \text{Baye's Theorem}$$

It also can be written as →

$$\boxed{Pr(A/B) = \frac{Pr(A) \times Pr(B/A)}{Pr(B)}} \rightarrow \text{Baye's Theorem}$$

A, B = Events

P(A|B) = Probability of A Given B is true

P(B|A) = Probability of B Given A is true

P(A), P(B) = Independent probabilities of A and B

Suppose we have a dataset:

Size	Room Count	Location	Price
x_1	x_2	x_3	y

In terms of Baye's theorem the probability we can write →

$$Pr(y|x_1, x_2, x_3) = \frac{Pr(y) \times Pr(x_1, x_2, x_3 | y)}{Pr(x_1, x_2, x_3)}$$