Statistics Day 04

Bayes Theorem

Bayes' Theorem is a fundamental concept in probability theory and statistics. It calculates the probability of an event based on prior knowledge of related conditions. The formula for Bayes' Theorem is:

$$\boxed{P(A/B) = \frac{P(A) \cdot P(B/A)}{P(B)}}$$

Bayes' Theorem

Events → Dependent Events
Events → Independent Events

Independent Events

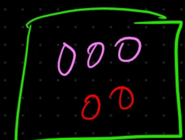
Tossing a coin → H, T

$P(H) = \frac{1}{2}$, $P(T) = \frac{1}{2}$

Rolling a dice $= \{1, 2, 3, 4, 5, 6\}$

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

Dependent Events



1 Marble drawn

$5 \rightarrow 4$

$P_R = \frac{2}{5}$

$P(\text{pink}) = \frac{3}{4}$

Multiplicative Rule

$P(A)$, $P(B)$

conditional prob.

$$P(A \cap B) = P(A/B) \times P(B)$$

Prob. of A, given B has occurred

Derivation

$$P(A \cap B) = P(B \cap A)$$

$$P(A/B) \times P(B) = P(B/A) \times P(A)$$

$$\boxed{P(A/B) = \frac{P(A) \times P(B/A)}{P(B)}}$$

$P(A/B) \rightarrow$ Prob. of A, given B has occurred

$P(B/A) \rightarrow$ Prob. of B, given A has occurred

$P(A)$, $P(B) \rightarrow$ Independent Event of A and B

Bayesian statistics $\rightarrow$ It is an approach to - data analysis and parameter estimation based on Baye's Theorem

Inferential statistics

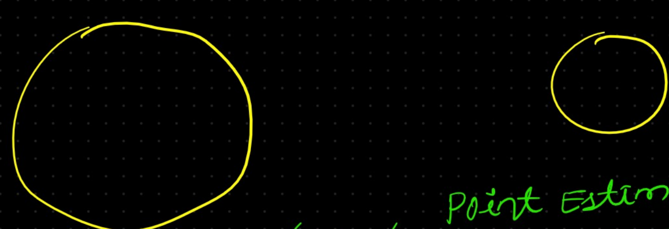
Estimate: It is an observed numerical values used to estimate a unknown population parameter.

	Estimator	parameter
Mean	$\bar{x}$	μ
Variance	s^2	σ^2

Estimate

- Point Estimate
  - ↓
  - A single value
  - eg: 1, 5, 105.67, 0.32
- Confidence Interval
  - An Interval
  - (1, 5)
  - (12, 35)
  - (−1, +2)

65



60

65 → 60

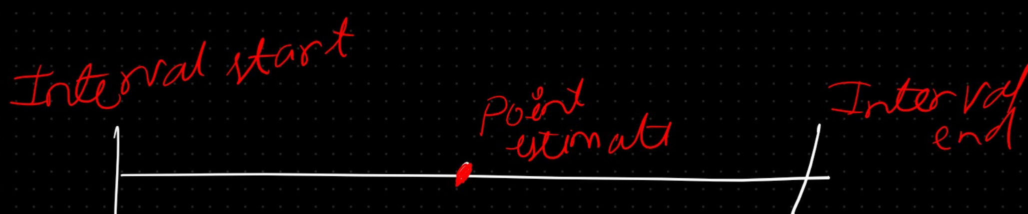
65 $\xleftrightarrow{\text{Point Estimate}}$ 60

Interval → Range of values used to estimate the unknown population parameter

$\boxed{55 - 65}$

→ Intervals estimate of population parameters are called confidence intervals

→ C.I are more precise than P.E. That's why they are preferred when making inferences.

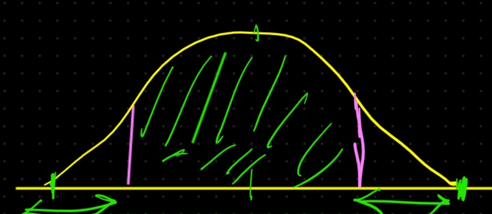


A confidence interval is an interval within which we are confident (with a certain percentage of confidence) the population parameter will fall

Margin of Error

$[\bar{x} - M.E , \bar{x} + M.E]$

$[\bar{x} \pm M.E]$



M.E = reliability factor * $\frac{\text{std}}{\sqrt{\text{sample size}}}$

Z score

$$\boxed{Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}}$$

$\to \alpha \to 1\,0$

$\alpha \to$ significance value

$(1-\alpha) \to$ level of confidence

$\alpha \to 0.01,\ 0.05,\ 0.1$

$100 - 95 = 5\%$

$= 0.05$

$1 - 0.05 = 95\%$

Hypothesis Testing

Inference

$\downarrow$

Conclusion

sample data

population data

$\Downarrow$

Conclusion
II Hypothesis Testing

A hypothesis is " an idea than can be tested "

→

It is a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.

Hypothesis Testing Mechanism

1. Null Hypothesis (H_0)

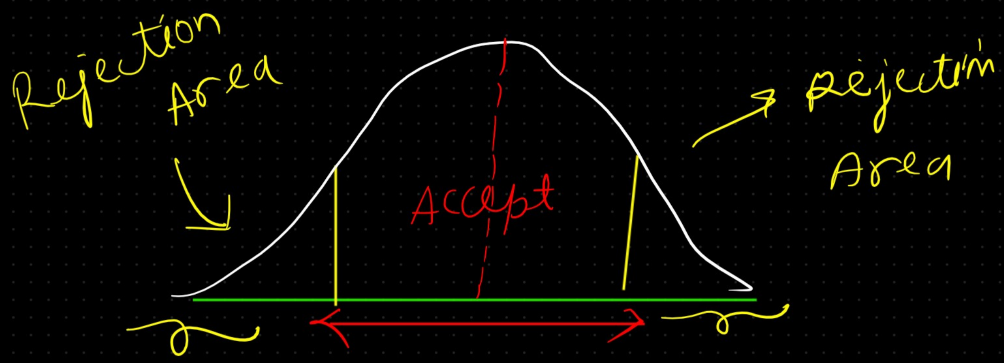
   — the assumption you are beginning with → evidence

2. Alternate Hypothesis (H_1)

   — Opposite of Null Hypothesis

3. Statistical Analysis / Experiment

4. Accept the H_0 or Reject the H_0

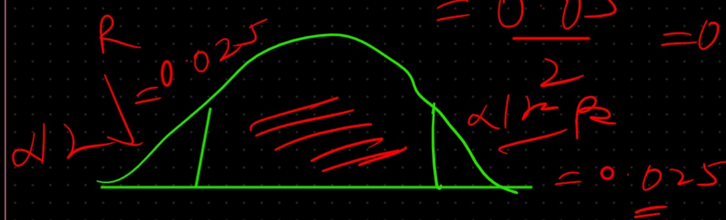


Level of significance (α)

→ The prob. of rejecting a null Hypothesis that is true; the prob. of making this error.

Two - Tailed test

$\alpha/2 = \frac{0.05}{2} = 0.025$



One-Tail

P-value

α

The p-value is the smallest level of significance at which we can still reject the null hypothesis, given the observed sample statistic

Eg → coin is fair or Not

{100 TOSSES}

H, T

$p(H) = 0.5$

$p(H) = 0.6$

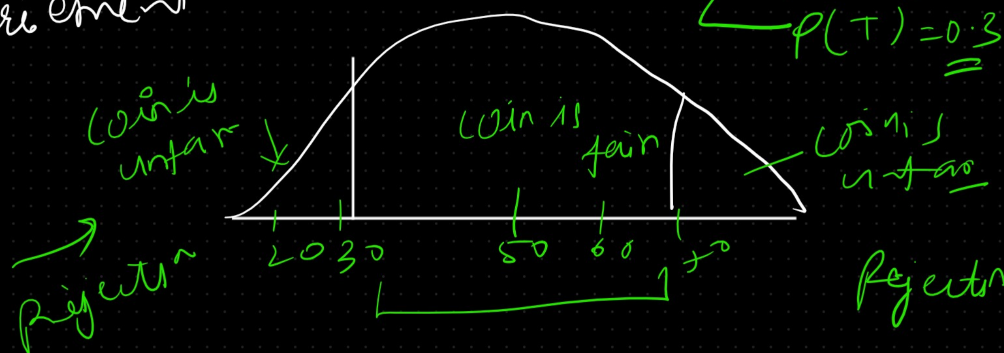
unfair $p(H) = 0.7$

$p(T) = 0.3$

① H_0 → coin is fair

② H_1 → coin is not fair

③ Experiment



$\alpha = 0.05$

$C.I = 1 - \alpha = 1 - 0.05$
$= 0.95$

Conclusion

$0.01 < 0.05$

We reject the Null Hypothesis

else

We Accept H_0

→ z-score, p value

1. z-test → } Average (mean)
2. t-test →
3. Chi-Square → categorical data
4. Annova → variance

Suppose a chocolate company claims that the mean weight of their chocolate bars is 168 grams. You, as a quality control inspector, want to test this claim. You take a random sample of 36 chocolate bars and measure their weights. Your sample has a mean weight of 169.5 grams, with a standard deviation of 3.9

Sol) $\mu = 168$, $\sigma = 3.9$, $\bar{x} = 169.5$

(i) Null Hypothesis H_0 $\mu = 168$

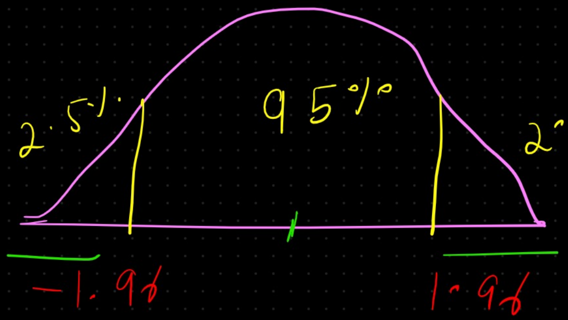
(ii) Alternate Hypothesis H_1 $\mu \neq 168$

(iii) $CI = 0.95$ $\alpha = 0.05$ $(1 - CI)$

https://www.math.arizona.edu/~rsims/ma464/standardnormaltable.pdf

z-test

(Z) → table



$1 - 0.025$

$= 0.9750$

$$Z\ score = \frac{x - \mu}{\sigma}$$

Statistical Analysis

$$Z\ test = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= \frac{169.5 - 168}{3.9 / \sqrt{36}} = 2.31$$

→ If Z-test value is less than −1.96 or greater than +1.96 we Reject the H_0

Else

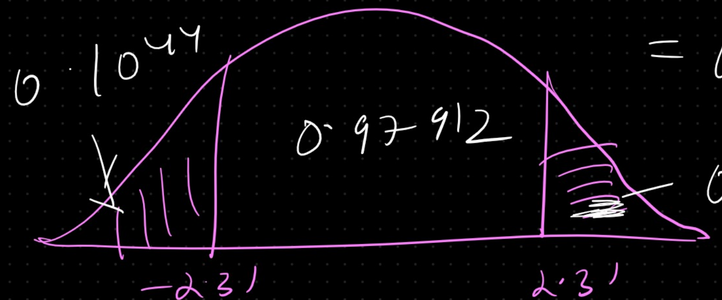
we Accept Null Hypothesis

↓ Z score

→ 2.31 > 1.96 {We Reject the H_0}

II- p value

$1 - 0.98956 = 0.01044$



$$p = 0.1044 + 0.1044 = \underline{\underline{0.02088}}$$

p-value $< \alpha$

$0.02088 < 0.05$

{Reject the H_0}

https://www.socscistatistics.com/pvalues/normaldistribution.aspx

student t distribution

Z-score → we need population std

ztest

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

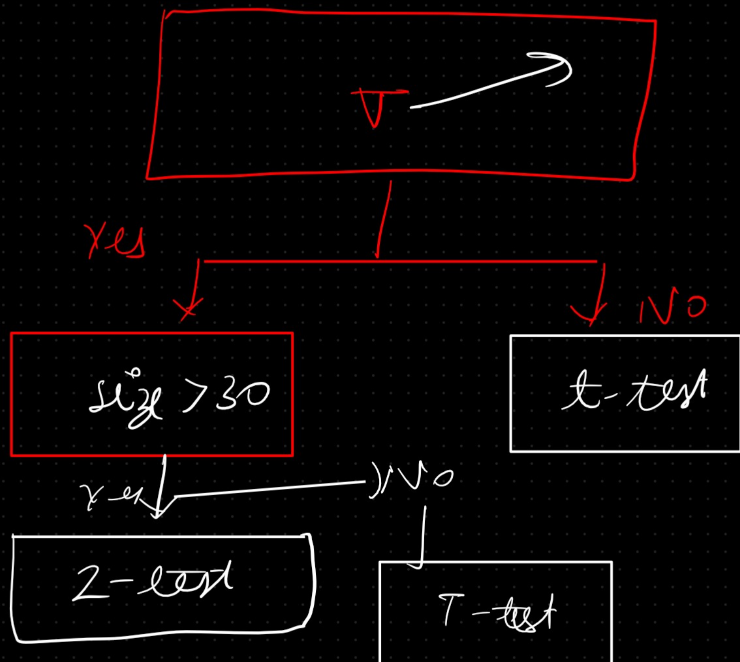
ztest

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Degree of freedom

$\boxed{dof = n - 1}$

When To use T- test Vs Z-test



Type I and Type 2 Error

Reality → H_0 → True / False

Decision → H_0 → True / False

⇓

Conclusion

Outcome I: We Reject the H_0 when in Reality False → Good

Outcome 2 : We Reject the H_0, when in Reality is True

$\Downarrow$

Type I Error

Outcome 3 : We retain the H_0, when in Reality False —

Type 2 Error

Outcome 4 : We retain the H_0, when in Reality it is True

$\downarrow$

Good