→ Inferential Statistics :-

↳ Hypothesis testing

↳ 1) (t-test / Z-test)

2) Anova

3) Chi²-Square

→ [Confidence Interval] → (Add on for this session)

⇒ Hypothesis – ?

1 – Null Hypothesis →

2 – Alternate hypothesis

2 – friends which are having a debate on any topic

↳ difference in opinion

P–1 ↳ "Statement"

⇒ Try to explain him via facts (google, show some data, etc)

P–2 → will oppose.

1) $p < 0.05 \Rightarrow$ We accept the alternate hypothesis

2) $p > 0.05 \rightarrow$ We failed to reject the (Null Hypothesis)

$H_0 \Rightarrow$ Assumption, thought process, Status Quo.

$H_a \Rightarrow$ Opposing the thought process

* α = critical value / threshold value / significance threshold.

$p > \alpha$ = We failed to reject the Null hypo (H_0)

{ $p < \alpha$ ↳ accept Alter hypo } Simple words (→ we accept H_0)

⇒ Eg:

factory → Ice-cream (Dairy product)

↓

{Milk}

└ Salmonella Bacteria

> 0.03 mg/L

↓

(Milk is not fit for any consumption)

⇒ Inspector → (fssai) →

$$\begin{cases} H_0: & \beta > 0.03 \\ H_a: & \beta < 0.03 \end{cases}$$

Plant manager

Samples (collect from Random Batches of milk)

↓

Lab test

↳ (Reading) →

⇒ t-statistics value → p-value → compare with α-value → Make final decision

---

J-D - {Statistical Modelling}

↳ {Hypothesis, Anova, Chi-Sq, Linear, Logistic Regression}

↓

{Experiments (Lab-testing}

{Bio-tech}

⇒ Steps - Involved :-

1→ formulate hypothesis eqⁿ

2→ check which test to apply

3→ Calculate t/z statistical value

4) using t-statistical value → calculate p-value

5) then check for the threshold / critical value (α);
this value value is given by SME (Subject Matter Expert) / if not
then we go with 0.05 ⇒ α-value.

6) → then p value is compared with α-value

↳ 6.1) → $p < \alpha$ ⇒ we accept the alternate hypothesis

6.2) → $p > \alpha$ ⇒ we failed to reject the Null hypothesis.

⇒ How α-value is calculated?

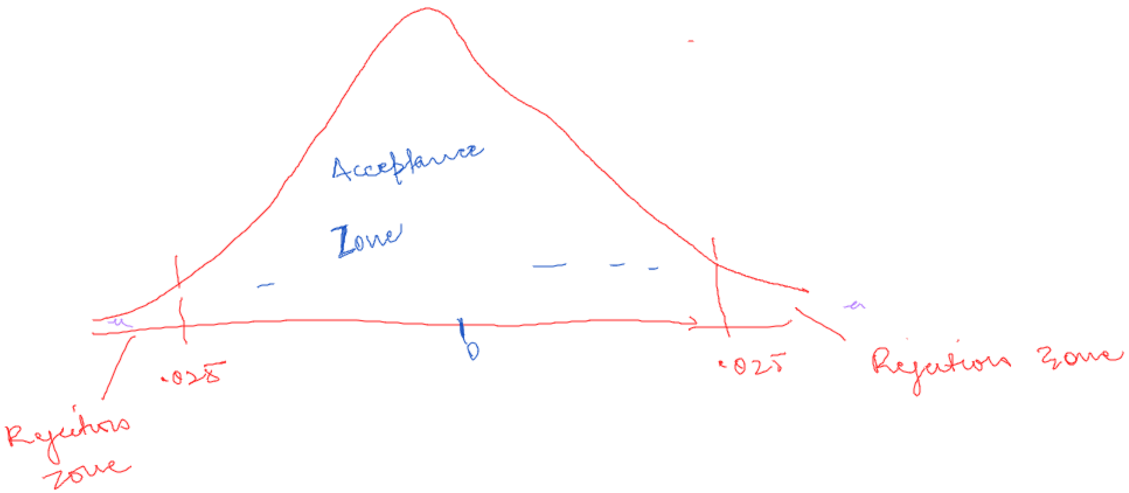
95% ⇒ .95

α = 1 - Confidence Interval.

⇒ $\alpha = 1 - .95$

$= \underline{0.05}$

$\alpha = 0.05$



$(p < 0.05) \rightarrow$ Acceptance zone

$p > 0.05 \Rightarrow$ Rejection Zone

⇒ Difference between t-test & Z-test (p-test)

As per theory

t-test	Z-test
1) Sample set is less than 30	1) When sample size > 30
2) It is used when sample std deviation is given along with the sample mean	2)- It is used when (population std dev) is given

⇒ In reality pop-std dev is absolutely impossible

∴ we always use (t-test)

⇒ f-test ⇒ test of variance — f-ratio

$\Downarrow$

Anova

⇒ How the t-statistical value is calculated?

$$t\text{-st} = \frac{x-\mu}{\sigma/\sqrt{n-1}}$$

x = Reference Value

μ = sample mean

σ = sample std deviation

& after this you use t→p conversion table for p-value.

⇒ t-test ⇒ 2-samples (Continuous variables)

⇒ Anova ⇒ >2-samples (Continuous Variable)

⇒ Chi-Square ⇒ when you have categorical variables

⇛ Chi-Square :

Eg - problem statement

⇒ Is there any relationship between

Marital Status & Education

χ^2 tabular > χ^2 calculated ⇒ failed to reject Null hypo

χ^2 cal > χ^2 tab ⇒ we accept Alternate hypothesis

⇒ What is Pearson Correlation Coefficient?

$$\text{Variance}(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$\bar{x}$ = Mean of the sample

x_i = data point

n = sample size

Co - Variance

$$(\text{Cov}(x,y)) = \sum_{i=1}^{n} \left(\frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \right)$$

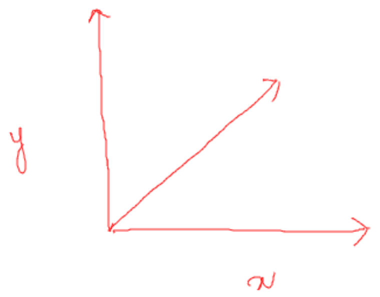
$$(a-b)^2 = (a-b)(a-b)$$

⇒ 3-types of Relationships $(-1, 1)$

-ve - -ve relation

+ve = +ve relation

+ve relationship

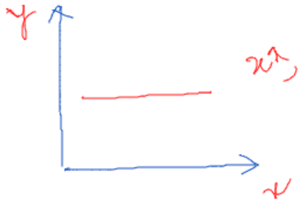


$y\uparrow, x\uparrow$

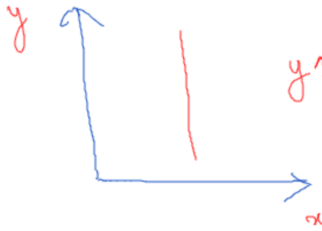
-ve relationship

$x\uparrow, y\downarrow$

0 / No relationship



$x\uparrow, y\times$



$y\uparrow, x\times$

⇒ Pearsonꞌs Co-efficient / ratio (r)

$$\Rightarrow \frac{Cov(x,y)}{(\text{Std of } x) * (\text{Std of } y)} = (-1, 1)$$

⇒ t-test

↳ 1-> 1-tail test

2-> 2-tail test

3-> 1 sample 1-tail test

4) 2-sample 2-tail test

5) 2-proportion Test

Sample ⇒ No. of datasets used for analysis

1-tail test → in the problem statement -

>, <, tall short, high low,

1 tail is considered

2-tail test → equal to / Not equal, diff, No diff

Eg = [2=2
2≠2]

→ <u>Anova</u>

1 - way Anova ⇒ 2-Independent Variables

2 - way Anova ⇒ >2-Independent Variables

→ Degree of freedom in Anova

→ Total Variables = 3

12

A	B	C

⇒ D of freedom

1→ Inter group = $3-1 = (2)$

2→ Intra group = $3 \times (12-1) = 3(11)$

(for intra group variance) $\boxed{33}$

$$\text{f-Ratio} = \frac{\text{variance b/ween the group}}{\text{Variance within the group}}$$

$Z = 95\% = 1.96$

⇒ Confidence Interval :

⇒ $CoI = \bar{x} \pm Z \frac{S}{\sqrt{n}}$

$\bar{x}$ = Mean Value | point estimate (Red line)

Z = confidence , S = Std dev , n = sample size

Eg → Weather forecast

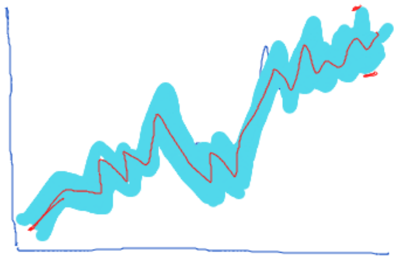
⇓

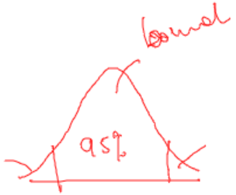
Min temp & Max temp , X Never tell 1 temp for day

↓

Range

upper bound & lower bound } ⇒ decided basis confidence interval





⇒ Red line = Absolute values

⇒ Blue zone = The area within which your value might oscillate.