Hypothesis Testing

→ What, Why, Null hypothesis, Alternative hypothesis, 1 tailed test, 2 tailed test (Pair test)

→ Type I & Type II error

→ Numerical problem

$\mu = 50$ cm - null

$\mu \neq 50$ cm $P \leq 0.05$ = Alternative

Z = -ve 5%

Z = +ve

$P > 0.05$

$CL < 95\%$.

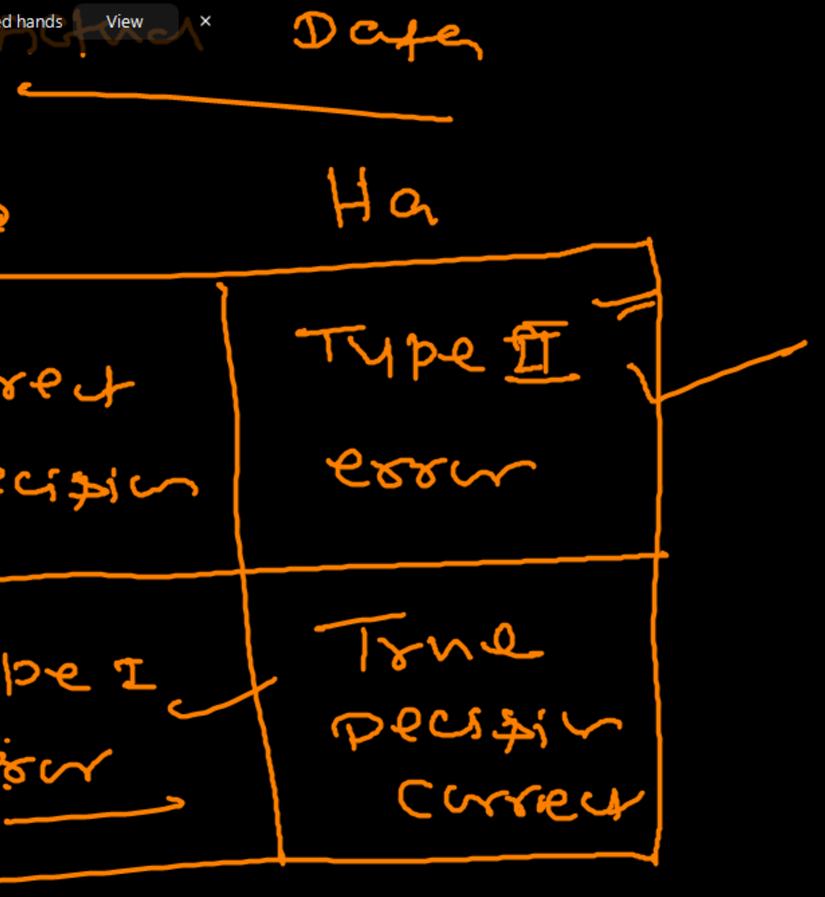
Truth about the Population

Decision Based Sample	H_0 True	H_0 false (H_a)
Reject H_0 (Alternate H_a)	Type I error	correct decision
Accept H_0	Correct decision	Type II error

Actual value

Predicted	Ho ✓ covid 19	covid-19 ✗
Ho ✓ covid 19	True	False Type II error
covid 19 ✗	False Type I	True —

Decision Based on Sample



(1) P_value ≤ 0.01 – indicate strong evidence against the null hypothesis → Accept Alternative

(2) $0.1 \leq P \leq 0.05$ – Accept Null hypothesis

Students - underfed - midday - meal

committee $\leq$ $p \leq 0.05$  $0.08 \geq$ NVLL Hypn

-5.1.

Avg weight of students of age 10 is 32 kg with S.D = 9 kg. A sample of 25 students were selected and avg found 29.5 kg

$H_0: \mu = 32$

$H_a: \mu < 32$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{29.5 - 32}{9/\sqrt{25}} = \frac{-2.5}{9} \times 5$$

$$= -\frac{12.5}{9} = -1.39$$

Chi-Square Test – Categorical test

- Goodness of fit test — 1 category
- test for independence — 2 category

Titanic

P_Class	Survived
1	0
2	1
3	0

1	2	3	4	5	
10	10	10	10	10	— uniform dice
5	15	20	2	18	— Not uniform

ANOVA - Analysis of Variance

Sir Ronald Fisher

$m = 3$

	C1	C2	C3
	A	B	C
R1	1	3	5
R2	2	4	6
R3	3	5	7

$n = 3$

$\bar{X}$ (mean) = 2 = 4 = 6

Grand mean (mean of mean) $= \frac{1+2+3+3+4+5+5+6+7}{9} = \frac{36}{9} = 4$ Grand mean

SST (Sum of Square Total)

$$\sum_{i=1}^{n} (\text{G.m} - \text{Ind value})^2$$

$$SST = (4-1)^2 + (4-2)^2 + (4-3)^2 + (4-3)^2 + (4-4)^2 + (4-5)^2 + (4-5)^2 + (4-6)^2 + (4-7)^2$$

$$SST = 9 + 4 + 1 + 1 + 0 + 1 + 1 + 4 + 9$$

$$\therefore SST = 30$$

Degree of freedom (SST) = Total $- 1 = 9 - 1 = 8$

SSW (Sum of Square within) $\sum_{i=1}^{n}$ (A.M - Ind. value (obs))^2

$= (2-1)^2 + (2-2)^2 + (2-3)^2 + (4-3)^2 + (4-4)^2 + (4-5)^2 +$

$+ (6-5)^2 + (6-6)^2 + (6-7)^2$

$= 1 + 0 + 1 + 1 + 0 + 1 + 1 + 0 + 1$

SSW = 6

degree of freedom (within) $= m(n-1)$

$= 3 * 2 = 6$

SSB (Sum of Square Between)

$$\sum_{i=1}^{k} n(G.m - A.M_i)^2$$

$$= 3\{(4-2)^2 + (4-4)^2 + (4-6)^2\}$$

$$= 3\{4 + 0 + 4\} = 3 * 8$$

$$\therefore SSB = 24$$

Degree of Freedom (Between) $= m - 1 = 3 - 1 = 2$

ANOVA

$SST = SSB + SSW$

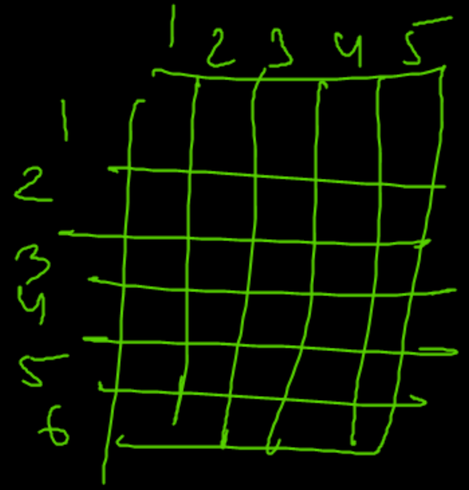
$SSB = SST - SSW$

$SSW = SST - SSB$

$30 = 24 + 6$

$DOF(T) = DOF(B) + DOF(W)$

$8 = 2 + 6$



$DOF(W) = m * (n-1)$

$= 5 * 5 = 25$

$$F\text{-Stats} = \frac{\frac{SSB}{m-1}\ DOF\ 1}{\frac{SSW}{m(n-1)}\ DOF\ 2} = \frac{24}{\frac{6}{6}}$$

$$F\text{-Stats} = 12$$

$DOF\ 1 = 2$

$DOF\ 2 = 6$

Calculate > Tabulate

$12 > 5.5$ — Alternative hypothesis

Low Noise		Medium Noise		Loud Noise	
emp	Perform	emp	Perf	emp	Perf
1	10	5	8	9	4
2	9	6	4	10	3
3	6	7	6	11	6
4	7	8	7	12	4

10 observations

SST, SSW, SSB — Avg-Tot, Grp-Avg

DOF = Total − 1

DOF(W) = obs × (k − 1) ← column

DOF(B) = obs − 1 ← Row.

H_0 (NULL) :- No significant effect of noise on no. of questions solved

H_a (Alternate) :- Significant effect of noise on no. of questions solved

$\alpha = 0.05$

$P \leq 0.05$

Cal > Tab

$$SST = \sum_{i=1}^{n} (\text{Grand mean} - \text{Actual value})^2$$

$$DOF(T) = \text{Total} - 1$$

$$SSW = \sum_{i=1}^{n} (\text{Avg value} - \text{Actual value})^2$$

$$DOF(W) = \text{Total column} \,(\text{Total row} - 1)$$

$$SSB = \sum_{i=1}^{n} \text{Total row} (\text{Grand mean} - \text{Adv means})^2$$

$$DOF(B) = \text{Total column} - 1$$

cal < Tab = NU

F-dist table

cal > Tab = Alte

F-Steps

$$= \frac{\frac{SSB}{n-1}}{\frac{SSW}{n(n-1)}}$$