Analysis Of Variance (ANOVA)

Defn: ANOVA is a statistical method used to compare the means of 2 or more groups.

ANOVA

1. Factors (variable)
2. Levels

① Factors (variable)

② Levels

Eg: Medicine (Factor)

[Dosage] 5mg 10mg 15mg → Levels

Mode of payment (Factor)

[Levels] GPAY PHONEPE IMPS NET-T

Analysis Of Variance (ANOVA)

Assumptions in ANOVA

1. Normality of Sampling Distribution of Mean

   The distribution of sample mean is normally distributed

2. Absence of Outliers

   Outlying score need to be removed from the dataset

3. Homogenity of Variance

③ Homogenity of Variance

Population variance in different levels of each independent variable are equal

$$[\sigma_1^2 = \sigma_2^2 = \sigma_3^2]$$

④ Samples are independent and random.

Types of ANOVA (3 Types)

① One Way ANOVA : One factor with atleast 2 levels | these levels are independent

Eg: Doctor wants to test a new medicaton to decrease headache. They split the participants in 3 conditions [10mg, 20mg, 30mg] Doctor ask the participants to rate the headache [1-10]

Medication → Factor

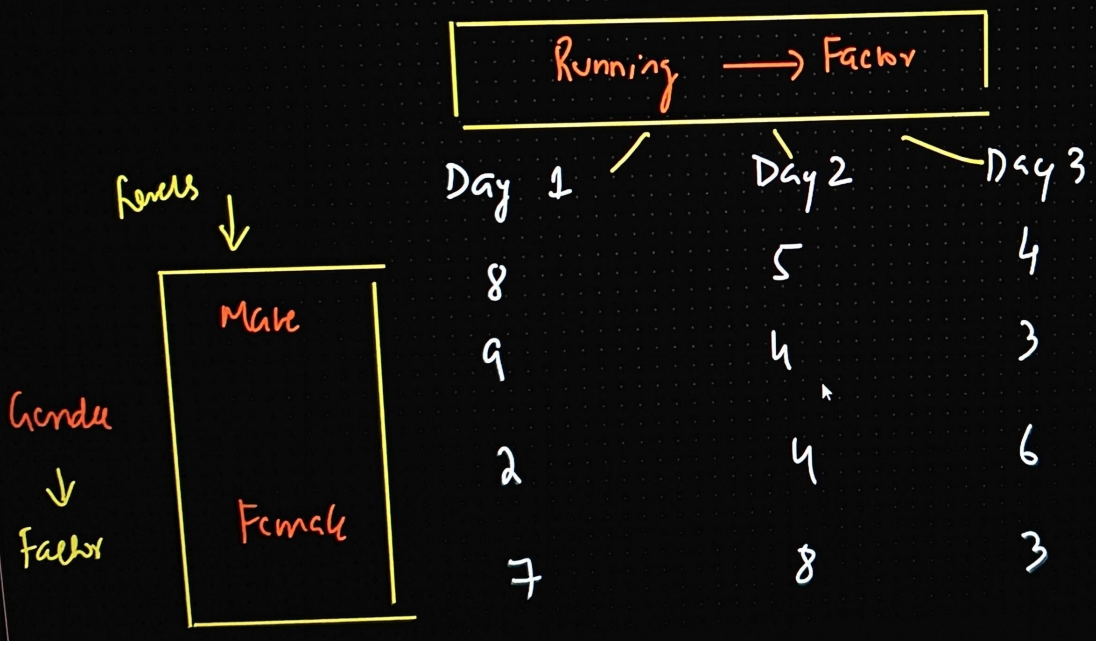
10mg	20mg	30mg
5	7	2
3	4	6
—	—	—
—	—	—

② Repeated Measures Anova : One factor with atleast 2 levels, levels are dependent.

Running → Factor

Levels →	Day 1	Day 2	Day 3
	8	5	4
	7	4	9
	-	-	-

③ Factorial ANOVA : Two or more factors (each of which with atleast) 2 levels, levels can be independent and dependent



Analysis Of Variance (ANOVA)

Hypothesis Testing In ANOVA (Partitioning Of Variance In The Anova)

Null Hypothesis H_0 : $\mu_1 = \mu_2 = \mu_3 \cdots \mu_K$

Alternate Hypothesis H_1 : Atleast one of the sample mean is not equal

~~$\mu_1 \neq \mu_2 \neq \mu_3 \cdots \mu_K$~~

Test Statistics

$$F = \frac{\text{Variance between Samples}}{\text{Variance within Samples}}$$

F^2

Variance between Samples

Variance within Samples

x_1	x_2	x_3
1	6	5
2	7	6
4	3	3
5	2	2
3	1	4
$\bar{x}_1 = 3$	$\bar{x}_2 = 19/5$	$\bar{x}_3 = 4$

H_0 $\bar{x}_1 = \bar{x}_2 = \bar{x}_3$

H_1 Atleast one sample mean is not equal

One Way ANOVA

One factor with atleast 2 levels, levels are independent

① Doctors want to test a new medication which reduces headache. They split the participant into 3 condition [15mg, 30mg, 45mg]. Later on the doctor ask the patient to rate the headache between [1-10]. Are there any differences between the 3 conditions

alpha = 0.05 ?

15 mg	30mg	45mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

① Define Null and Alternate Hypothesis?

H_0 : $\mu_{15} = \mu_{30} = \mu_{45}$

H_1 : not all μ are equal

② Significance $\alpha = 0.05$ C.I = 95

③ Calculate Degree of freedom

$N = 21$

Calculate Degree of freedom

$N = 21 \quad a = 3 \quad n = 7$

$df_{between} = a - 1 = 3 - 1 = 2$

$df_{within} = N - a = 21 - 3 = 18$

$df_{total} = N - 1 = 20$

$df_1 \quad df_2$

$(2, 18)$

⇓

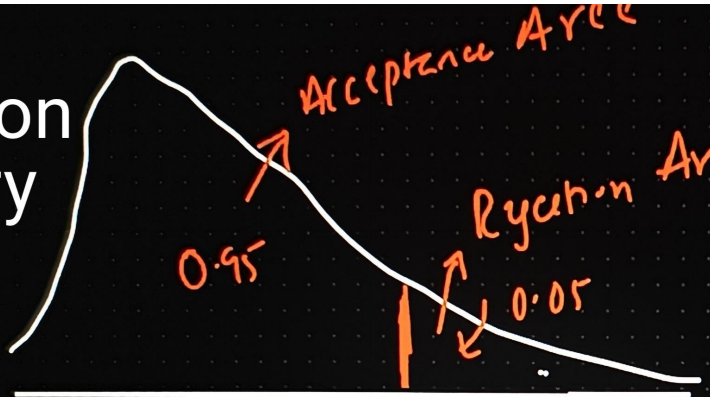
F table

$\alpha = 0.05$

⇓

Critical value

4) Decision Boundary



3.5546

Decision Rule

If F is greater than 3.5546, reject the Null hypothesis

⑤ Calculate F Test Statistics

$$F = \frac{\text{Variance between Samples}}{\text{Variance within Sample}}$$

	SS	df	MS	F
Between	98.67			
Within	10.29			
Total	108.96			

① $SS_{between} = \sum (\sum a_i)^2 - \frac{T^2}{\cdot}$

	15 mg	30mg	45mg
n	9	7	4

① SS between $= \frac{\sum (\sum a_i)^2}{n} - \frac{T^2}{N}$

15 mg	30mg	45mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

15mg : 9+8+7+8+8+9+8 = 57

30mg : 7+6+6+7+8+7+6 = 47

45mg : 4+3+2+3+4+3+2 = 21

$= \frac{57^2+47^2+21^2}{7} - \left[\frac{57^2+47^2+21}{21}\right]^2$

$$= \boxed{98.67}$$

② $$SS_{within} = \sum y^2 - \frac{\sum (\sum a_i)^2}{n}$$

$$\sum y^2 = 9^2 + 8^2 + 7^2 + 8^2 + 8^2 + \dots$$

$$= 853$$

$$= 853 - \left[\frac{57^2 + 47^2 + 21^2}{7}\right]$$

$$= \boxed{10.29}$$

	SS	df	MS	F
Between	98.67	2	49.34	
Within	10.29	18	0.54	
Total	108.96	20		

$$F_{test} = \frac{MS_{Between}}{MS_{within}}$$

$$F = \frac{\text{Variance between Samples}}{\text{Variance within Sample}}$$

MS_{within}

$$F = \frac{49.34}{0.54} = 86.56$$

If F is greater than 3.5546, Reject the H_0

$86.56 > 3.5546$ Reject the H_0