Introduction to Analysis Of Variance (ANOVA)

<u>Defn</u> :- ANOVA Is a statistical method used to compare the means of 2 or more groups

<u>ANOVA</u>

① Factors (variables)

② Levels

Eg : Gender (↗ Variable) → Male, → Female } → Levels

② Medicines → 0mg, 5mg, 10mg } 3 levels

↳ Dosage

<u>Types of ANOVA</u>

① <u>One Way ANOVA</u> : One factor with atleast <u>2 levels</u>, levels are <u>independent</u>

Eg : Dosage of a medicine

		0mg → p.	50mg	100mg
S1	→	9	7	4
S2	→	8	6	2

② <u>Repeated Measure ANOVA</u> : One factor with atleast 2 levels, level are <u>dependent</u>

Running

	Day 1	Day 2	Day 3	Day 4
→	25km	10km	2km	5km

③ FACTORIAL ANOVA : Two or more factors (each of which atleast 2 levels), levels can be either independent, dependent, or both (mixed)

Running

	Day 1	Day 2	Day3
Men	9 8 7	7 6 5	4 3 2
Female	8 7 10	7 8 7	3 4 3

① One way ANOVA

Researchers wants to test a new anti-anxiety medication. They split participants into 3 Conditions (0mg, 50mg, 100mg), then ask them to rate their anxiety levels on a scale 1-10. Are there any differences between the condition using alpha = 0.05?

$(\sum a_i)^2$

0mg	50mg	100mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

0mg = 9+8+7+8+8
+9+8 = 57

50mg = 7+6+6+
7+8+7+6 = 47

100mg = 4+3+2+3+4+
3+2 = 21

① Define Null and Alternate Hypothesis

$H_0 = \mu_{0mg} = \mu_{50mg} = \mu_{100mg}$

H_1 = not all μ's are equal

② State alpha $\alpha = 0.05$

a = No. of levels

③ Calculate degree of freedom

$N = 21$ $\quad n = 7 \quad a = 3$ {↑ No. of levels}

$$\left\{\begin{aligned} df_{Between} &= a-1 = 3-1 = 2 \\ df_{within} &= N-a = 21-3 = 18 \end{aligned}\right\}$$

N = Total no of datpoint

n = No. of sample

$df_{total} = N-1 = 21-1 = 20$

④ State Decision Rule : F test = 2 Different degree of freedom

$df_{Between} = 2 \qquad df_{within} = 18$

$(2, 18)$

[If F test is greater than 3.5546, reject the Null Hypothesis] ✓

⑤ Calculate F Test Statistic

Mean Square

	SS		df	MS	F
Between	98.67	÷	2	49.34	$\boxed{86.56}$ ✓
Within	10.29	÷	18	0.57	
Total	108.95		20		

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

0mg	57 ✓
50mg	47 ✓
100mg	21 ✓
T	125

$$= \frac{(57)^2 + (47)^2 + (21)^2}{7} - \frac{(125)^2}{21}$$

$$SS_{between} = \underline{\underline{98.67}}$$

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

$$= 853 - \left[\frac{(57)^2 + (47)^2 + (21)^2}{7}\right] = \underline{\underline{10.29}}$$

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

$$= \underline{\underline{853}}$$

③ $SS_{total} = \sum y^2 - \frac{T^2}{N}$

$= 853 - \frac{(125)^2}{21}$

$= 108.95$

$\alpha = 0.10$

$F = \frac{MS_{between}}{MS_{within}} = \frac{49.34}{0.57} = 86.50$

$86.50 > 3.55$, Reject Null Hypothesis

① Compare Mean between 2 groups

Population sd , $n \geq 30$ ⟶ Z test

② Compare Mean between 2 groups

≠ population sd, n ⟶ t test

③ Compare mean between categorical value ⟶ Chisquare

{6000 + 18% GST} → Lifetime

1:07 → 1000+ Cores

1. Permutation
2. Bessel's Corrections
3. Power Law
4. Central Limit theorem.
5. Bernoulli's distribution.
6. Probability density Function

DRONE PROGRAMMING

Micro Arduino, Raspberry pie,

↳ Jetson NANO, Jetson Xavier,

Google Coral

{KRISH10}

7. Permutation And Combination

Zoo : { Tiger, Monkey, Zebra, deer, Snake, Lion }

↳ Student

↳ $\underline{6} * \underline{5} * \underline{4} = \underline{\underline{120}}$

Tiger Monkey Lion

Lion Monkey Tiger

Order matters

n = Total no. of animal = 6

r = 3

→

→ Permutation $\quad {}^{n}P_{r} = \dfrac{n!}{(n-r)!}$

→

$= \dfrac{6!}{(6-3)!} = \dfrac{6 \times 5 \times 4 \times \cancel{3!}}{\cancel{3!}}$

= 120

Combination { Unique combination }

{ Tiger Lion Monkey
Lion Tiger Monkey } → 1

Permutation > Combination

Order

$$\left\{ {}^{n}C_{r} = \frac{n!}{r!(n-r)!} \right\}$$

$$= \frac{6!}{3!(6-3)!} = \frac{6!}{3! \times 3!}$$

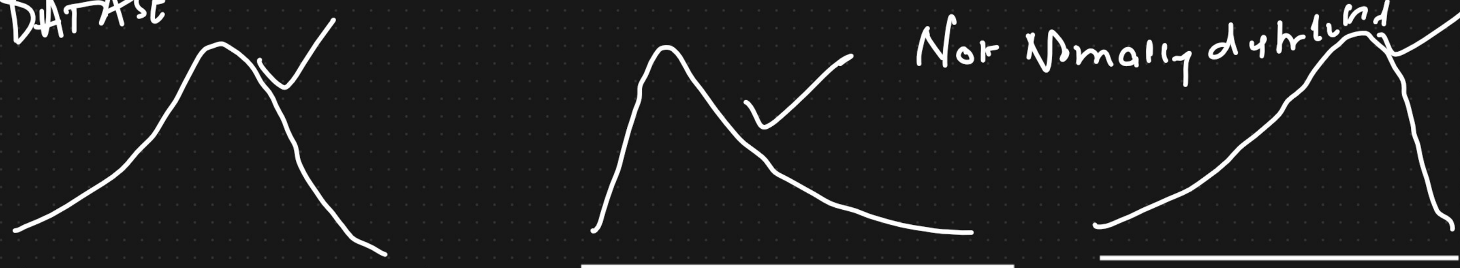
$$= \frac{\cancel{6}^{2} \times 5 \times 4 \times \cancel{3!}}{\cancel{3\times 2} \times 1 \times \cancel{3!}} = 20$$

120 > 20

② Central Limit theorem

DATASET →

May be Normally or Not Normally distributed



$n \geq 30$

⇓

→ sample mean

Sample mean of sample data $n \geq 30$

$S_1 - x_1, x_2, \ldots - x_n = \overline{S_1}$ $\quad n \geq 30$

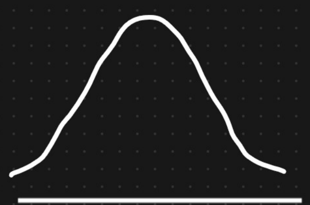
$S_2 = x_2, x_3 \ldots x_n = \overline{S_2}$ $\quad n = 10$ More Normally Distribute

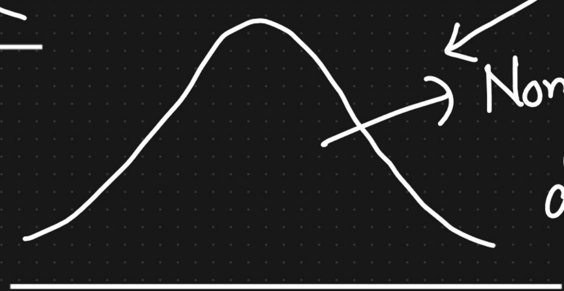
$S_3 = \ldots = \overline{S_3}$

$\vdots$

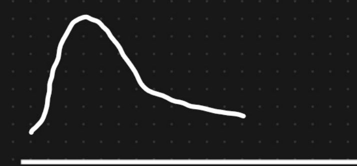
$\overline{S_m}$

Less Normally distributed





Normally distributed



$n : \underline{1}$ ✓✓✓

More Normally Distributed

Less Normally distri

$S_1 \quad \overline{S_1}$ $\quad n \geq 30$

$S_2 \quad \overline{S_2}$

$S_3 \quad \overline{S_3}$

100 data points?

Dataset = { 10, 12, 13, 15, 20, 40, 70, ... }

$n \geq 30$ ↳ $S_1 = \{ \quad \} \rightarrow \overline{S_1}$

$S_2 = \{ \quad \} \rightarrow \overline{S_2}$

$S_3 = \{ \quad \} \rightarrow \overline{S_3}$

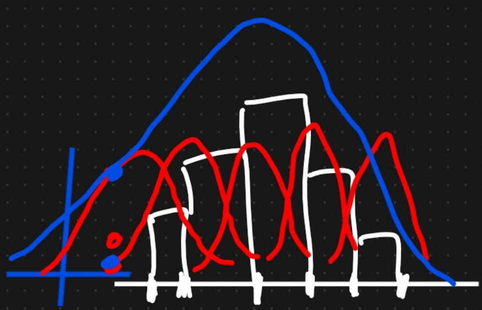
$\vdots \quad \vdots$

S_m : { } $\rightarrow \bar{S}_m$

$n \geq 30$

it is following normality

③ Probability Density function (Kernel density Estimator)



④ Bessel's Correction

80% $\longrightarrow$ 20%

$\dfrac{}{n-1}$

{why n-1}

n

n-1

n-2

Degree of freedom = $\boxed{n-1}$

Bernoulli and Binomial Distribution

Cummulative Density Function (cdf)

Adding probability Commulatively