"ANOVA"

"Analysis Of Variance"

→ Hypothesis test of 3 or More means.

For example suppose a teacher wishes to test the effects of five → factors or treatment (no. of levels or) diff types of plant food on the growth of certain plants. The one independent → qualitative variable (independents) variable is the plant food while the dependent is the plant's growth. → quantitative variable (dependent) (response variable) numbers ki form mein response dega ye.

Other factor such as type of soil, water, temperature are held constant.

Jaha 2 se zyada means test krne ho waha ANOVA lagayenge. 2 se kam k liye hypo testing.

→ Qualitative variable k change se quantitave variable par effect ayega.

→ 3 ya 3 se zyada factors honge qualitative variable k

"Want to study the effect of one or more qualitative variables on a quantitative outcome variable."

Assumptions:

→ All samples are independent (har ek population ek diff factor ko use krrhi hogi.)

→ Populations are normally distributed.

→ The variances of the populations must be equal.

→ Independent variable can be one or greater than 1.

→ For 1 we use One Way ANOVA

→ For 2 we use Two Way ANOVA → [jese mein plant food ke saath soil ko bhi consider krleta]

→ more than 2 it's very complicated so not used.

* "ANOVA is a tool to test equality of more than two mean simultaneously."

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ we cannot tell which two are not equal.

H_1: At least two means are not equal

→ **F-test** is used in ANOVA, to determine the significance difference among three or more means

"One Way ANOVA"

qualitative — levels/treatments

Your independent variable is social media use, and you assign groups to low, medium & high levels of social media use to find out if there is a diffe in hrs of sleep/night. → quantitative dependent

Null Hypothesis: (H_0)

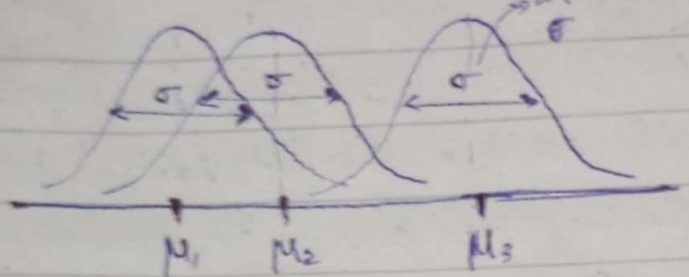
$H_0: \mu_1 = \mu_2 = \dots \mu_k$ (hamesha yehi hoga)

Alternate Hypothesis: (H_1)

H_1: At least one of the mean is diff

This is always a one tailed test i.e a Right tail

all have same σ.



F - Distribution:

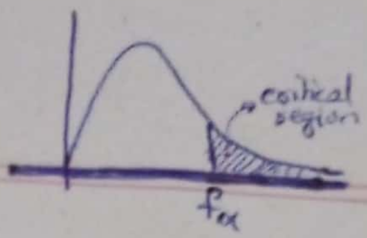
$U \rightarrow$ has ν_1 degree of freedom
$V \rightarrow$ has ν_2 degree of freedom

$$F = \frac{U/\nu_1}{V/\nu_2}$$

where, U & V are independent random variables having chi-squared distributions with ν_1 & ν_2 degrees of freedom, respectively.

if 'x' is an F random variable with u numerator and v denominator degrees of freedom then PDF of x is;

$$h(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2} x^{(u/2)-1}}{\Gamma(u/2)\,\Gamma(v/2)\left[\left(u/v\right)x+1\right]^{(u+v)/2}}$$



→ Sensitivity of F-Statistics

↳ The F-statistic is sensitive to differences among a set of sample mean.

↳ The greater the variation among sample means, the larger is the value of the test statistics

↳ The smaller the variation among the sample mean, the smaller the value of test statistics.

→ MODEL FOR ONE WAY ANOVA :

both are error terms

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$\alpha_1 + \alpha_2 + \alpha_3 - \alpha_n = 0$

Y_{ij} = response variable (i = konsa sample, j = us sample ka konsa member hai)

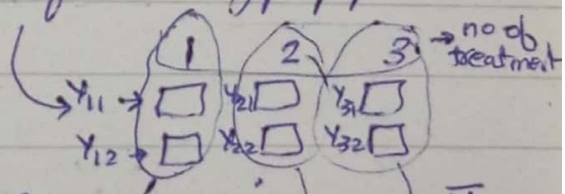
μ = Grand mean of all μ_i, $\mu = \frac{1}{K}\sum_{i=1}^{k} \mu_i$

↳ Notations :

k = no. of groups/population/levels of treatment (in plant example that was 5)

n_i = the sample size taken from group i (can be = or ≠)

Y_{ij} = the jth response sample from ith grp/population.



$\overline{Y}_{i.}$ = the sample mean of responses from the ith grop → iska mean lenge to $\overline{Y}_{1.}$

$\overline{Y}_{2.}$ $\overline{Y}_{3.}$

$\overline{Y}_{..}$ = grand mean of all responses

↳ Tamam values ko add kardeng or divide by total no. of values irrespective of their treatments.

N = The total samples irrespective of groups
S_i = Sample Std-dev from i^{th} group/level/treatment
$\bar{Y}_{ij}$ =

→ The Analysis of variance is derived from partitioning of total variability into its component parts.

Measure of total variability

$$SS_T = SS_{treatments} + SSE$$

Total corrected sum of squares — Sum of sq. b/w treatments (variation b/w treatments) — Some of sq. due to error (within treatments)

Σ

har ek value ko sample ki, grand mean se minus krenge to total variance mil jayga

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \text{total sum of sq.}$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i} - \bar{y}_{..})^2 = \text{treatment sum of sq.}$$

har ek treatment ke mean - grand mean.
(it tells variation b/w diff means).

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i})^2 = \text{Error sum of sq.}$$

har ek treatment ki value ko usi ke sample mean se minus krenge.
(tells variability within treatment)

Sum-of-Squares Identity

$$\boxed{SST = SSA + SSE}$$

Source of Variation	Sum of Sqs	Degree of Freedom	Mean Sq.	Computed f
Treatments	SSA	$k-1$	$S_1^2 = SSA/k-1$	$\frac{S_1^2}{S_2}$
Error	SSE	$k(n-1)$	$S_2^2 = SSE/k(n-1)$	
Total	SST	$(kn-1)$		

Q: A researcher wishes to try three diff techniques to lower the blood pressure of individuals diagnosed with high blood pressure. The subject are randomly assigned to three groups; the first grp takes medication, the 2nd grp exercises & 3rd grp follows a special diet. After four weeks, the reduct in each person's blood pressure is recorded. At $\alpha = 0.05$, test the claim that there is no difference among the means.

Medication	Exercise	Diet
10	6	5
12	8	9
9	3	12
15	0	8
13	2	4

→ response variable = blood pressure
→ independent variable =

Step# 01 : $\boxed{\bar{Y}_1 = 11.8}$, $\boxed{\bar{Y}_2 = 3.8}$, $\boxed{\bar{Y}_3 = 7.6}$

(medication) (exercise): (diet)

$\boxed{S_1^2 = 5.7}$, $\boxed{S_2^2 = 10.2}$, $\boxed{S_3^2 = 10.3}$

Step# 02 : $\bar{Y}_{..} = 7.733$ (Grand mean)

Step# 03 : $SSA = 5(11.8-7.733)^2 + 5(3.8-7.733)^2 + 5(7.6-7.733$

$\boxed{SSA = 160.13}$

Step# 04 : $SSE = \sum_{i=}^{k} (n_i - 1)(S_i)^2$ (another formula)

$= (5-1)S_1^2 + (5-1)S_2^2 + (5-1)S_3^2$

$= 4 \times 5.7 + 4 \times 10.2 + 4 \times 10.3$

$\boxed{SSE = 104.8}$

Step# 05 : Calculating degree of freedom.

SSA → $k-1 = 3-1 = 2$

SSE → $(3)(5-1) = 12$

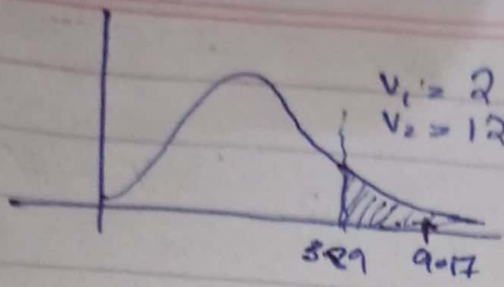
no. of treatments. ↳ har group mein kitna data hai

Mean Sq = $MS = \frac{SSA}{df} = \frac{160.13}{2} = \boxed{80.07}$

$= \frac{SSE}{df} = \frac{104.8}{12} = \boxed{8.73}$

Step# 06 :

$F = \frac{80.07}{8.73} = 9.17$



as $9.17 > 3.89$ mean it lies in critical region
therefore $\boxed{\text{Reject } H_0}$

Unequal SAMPLE Size

Example 13-2:

Step#01: $n_1 = 20$, $n_2 = 9$, $n_3 = 9$, $n_4 = 7$

$\bar{y}_1 = 73.0125$, $\bar{y}_2 = 48.93$, $\bar{y}_3 = 93.6$, $\bar{y}_4 = 101.06$

$s_1^2 = 662.26$ $s_2^2 = 2219.78$ $s_3^2 = 2168.43$, $s_4^2 = 946.03$

Step# 02: $\bar{y}_{..} = 76.693$ (grand mean).

Step # 03: $SSA = 20(73.0125 - 76.693)^2 + 9(48.93 - 76.693)^2 + 9(93.6 - 76.693)^2 + 7(101.06 - 76.693)^2$

$SSA = 13935.02$

$SSE = (20-1)602.26 + (9-1)2219.78 + (9-1)2168.43 + (7-1)946.03$

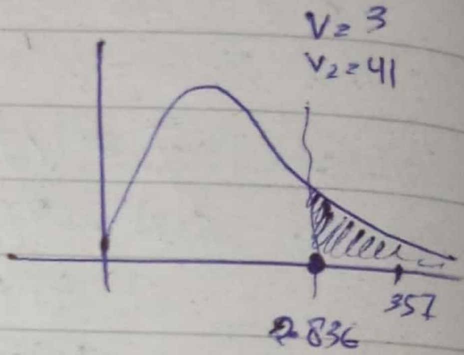
$SSE = 53376$

Step#04:

$$MS = \frac{SSA}{df} = \frac{13935.02}{3} = 4646$$

$$= \frac{SSE}{df} = \frac{53376}{45-4} = 1302$$

$$F = \frac{4646}{1302} = 3.57$$



$\boxed{\text{Reject } H_0}$

Ex: 13

Q13.1. Step#01: $n = 4$, $k = 6$

$\bar{y}_1 = 17.2$ | $\bar{y}_2 = 17.175$ | $\bar{y}_3 = 18.175$ | $\bar{y}_4 = 17.75$ | $\bar{y}_5 = 18.425$ | $\bar{y}_6 = 18.0$

$S_1^2 = 1.366$ | $S_2^2 = 2.709$ | $S_3^2 = 3.769$ | $S_4^2 = 7.216$ | $S_5^2 = 3.155$ | $S_6^2 = 2.66$

Step#02: $\bar{y}_{..} = 17.792$

Step#03: $SSA = 4(17.2-17.792)^2 + 4(17.175-17.792)^2 + 4(18.175-17.792)^2 + 4(17.75-17.79)^2 + 4(18.425-17.792)^2 + 4(18.025-17.792)^2$

$SSA = 5.338336$

Step#04: $SSE = 3(1.366) + 3(2.709) + 3(3.769) + 3(7.216) + 3(3.155) + 3(2.6625)$

$SSE = 62.6325$

Step#5: $MS = \frac{SSA}{df} = \frac{5.338336}{6-1} = 1.068$ 5

$\frac{SSE}{df} = \frac{62.6325}{6(4-1)} = 3.4796$ 18

Step#06:

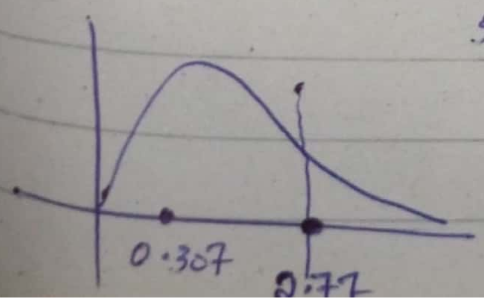
$F = \frac{1.068}{3.4796} = 0.307$

P-Value test

$P(F < 0.307)$

$0.902 > \alpha$

(accept Ho)



as F value donot lie in CR

So (Donot Reject Ho)

Q13.3: $n = 8$, $k = 3$

$\bar{Y}_1 = 81$ | $\bar{Y}_2 = 90.875$ | $\bar{Y}_3 = 84.625$

$S_1^2 = 13.1429$ | $S_2^2 = 6.9821$ | $S_3^2 = 21.125$

Step#02: $\bar{y}_{..} = 85.5$

Step#03: $SSA = 8(81-85.5)^2 + 8(90.875-85.5)^2 + 8(84.625-85.5)^2$

$SSA = 399.25$

$SSE = 7(13.1429) + 7(6.9821) + 7(21.125)$

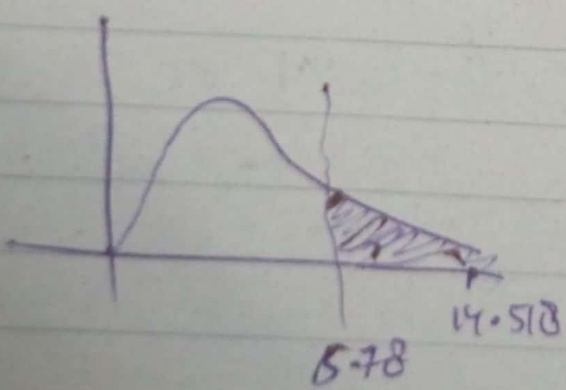
$SSE = 288.75$

Step#04:

MS, $\frac{SSA}{df} = \frac{SSA}{2} = 199.625$

$\frac{SSE}{df} = \frac{288.75}{21} = 13.75$

Step#05: $F = \frac{199.625}{13.75} = 14.518$



$\boxed{\text{Reject } H_0}$

$k = 4$

Q13.5 $n_1 = 4$, $n_2 = 4$, $n_3 = 4$, $n_4 = 9$

$\bar{Y}_1 = 11.2275$ $\bar{Y}_2 = 11.115$ $\bar{Y}_3 = 10.875$ $\bar{Y}_4 = 8.769$

$S_1^2 = 0.417$ $S_2^2 = 0.939$ $S_3^2 = 0.218$ $S_4^2 = 1.724$

Step#02, $\bar{Y}_{..} = 10.086$

Step#03, $SSA = 4(11.2275 - 10.086)^2 + 4(11.115 - 10.086)^2 + 4(10.875 - 10.086)^2 + 9(8.769 - 10.086)^2$

$= 27.54938$

$SSE = 3(0.417) + 3(0.939) + 3(0.258) + 8(1.724)$

$SSE = \cancel{15.763}\ 18.634$

Step#04, MS, $\frac{SSA}{df} = \frac{27.54938}{3} = 9.183$

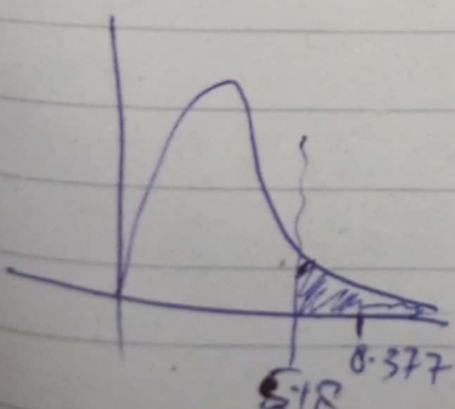
$\frac{SSE}{df} = \frac{18.634}{17} = 1.0962$

$F = \frac{9.183}{1.0962} = 8.377$

P-Value Test

$P(F < 8.377) = 0.0012$

$\boxed{P < \alpha}$ Reject Ho



as value lies in CR so

$\boxed{\text{Reject Ho}}$