ANOVA (F) TEST

→ Anova test is statistical test used to determine if there is statistical difference between two or more, categorical groups by testing for differences of means using variance.

→ when to use Anova test?

1) It is only conducted when there is no relationship between the subjects in each sample. This means that subjects in the first group cannot also be in the second group. ie independant samples between groups.
2) groups must have equal sample size

Example: 1: A study conducted to see the effect of Green, pepermint and No Tea on weight loss.

Green Tea	No Tea	Peppermint Tea.
10	1	1
12	2	4
13	3	5
15	9	8
18	10	9
20	10	5
22	11	10

Note: Anova test considers Two types of variation
1) variation within the groups
2) variation betwen the groups.

$$F\ value = \frac{\text{Variation b/w groups}}{\text{Variation within groups}}$$

from above ratio it can be concluded that as variation between group increases so does the F-value increase.

ie $\boxed{\text{F-value} \uparrow \Rightarrow \text{variation b/w groups} \uparrow}$

ie there is difference in means of groups.

Note: There is standard null and alternate hypothesis. in anova test:

H_0 = Means of groups are equal.

H_1 = Means of groups are not equal.

Note: For Large F value. the variation between groups will be large.

ie H_0 will be rejected.

- In above Tea Example it is clearly visible that there variation in means of 3 Groups.

⇒ for this example null hypothesis will be rejected.

⇒ Some more examples where Anova test can be applied.

1) Group of psychiatric patients trying three different therepies.

ie
1) counselling
2) medication
3) Group sessions

} Aim is to see which therepy is better.

2) A light bulb is manufactured using 2 processes.

1) Process 1
2) process 2

Aim is to see which process is better.

3) Students of different colleges taking CAT Exam

1) college 1
2) college 2
3) College 3

Aim is to see, students of which college performs better than other college students

* Types of Anova test

① one-way Anova test

It is used when there is one independant variable with atleast two levels.

Ex: I.V 1 — levels

Brand of cereal → corn flakes.
→ Lucky charms
→ Raisin Bran.

② Two-way Anova test

It is used when there are two independant variable with atleast two levels.

Ex: IV 1 — Levels

Brand of cereal — cornflakes
— Lucky charms
— Raisin Bran

IV. 2 — levels.

calories → Sweetened
→ un-sweetened.

* Procedure to conduct Anova test.

Ex: consider 4 independant groups.

	Group 1	Group 2	Group 3	Group 4
Sample size	n_1	n_2	n_3	n_4
sample mean	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$
sample Std.	S_1	S_2	S_3	S_4

① Hypothesis

$H_0 \Rightarrow \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$

$H_1 \Rightarrow$ means are not equal.

② Decision rule (α, dof).

$df_1 = K-1$ $\quad$ $K \equiv$ No. of Independent groups

$df_2 = N-K$ $\quad$ $N \equiv$ Total no. of observation

$N = n_1 + n_2 + n_3 + n_4 + \cdots + n_k$

$\alpha \equiv$ significance value.

F value is taken from F-table using df_1, df_2 and α.

Note: 1) if $F_{test\ statistic} \geq F_{value}$, H_0 is rejected.

2) $F_{test\ statistic} < F_{value}$, H_0 cannot be rejected.

③ Test statistic

$$\boxed{F = \frac{\sum n_j (\bar{x}_j - \bar{x})^2 / (k-1)}{\sum\left(\sum (x - \bar{x}_j)^2\right) / (N-k)}}$$

where

n_j = sample size of jth Group $\{j = \{1, 2, 3, \dots k\}\}$
$\bar{x}_j$ = sample mean of jth Group
$\bar{x}$ = overall mean

k = No. of independent groups.
N = Total No. of observations.
(N does not represent population size, but it represents total sample size)
ie $N = n_1 + n_2 + n_3 + \dots + n_k$

Anova table

Source of variation	Sums of squares (SS)	dof (df)	Mean Squares (MS)	F value
B/w Group	$SSB = \sum n_j (x_j - \bar{x})^2$	$k-1$	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSE}$
Error (Residual)	$SSE = \sum\left(\sum (\bar{x} - \bar{x}_j)^2\right)$	$N-k$	$MSE = \frac{SSE}{N-k}$	–
Total	$SST = \sum\left(\sum (x - \bar{x})^2\right)$	$N-1$	–	–

where x = individual observation
$\bar{x}_j$ = sample mean of jth Group.
$\bar{x}$ = overall sample mean
k = No. of groups
N = Total No. of observations.

Que: 1

A clinical trial is run to compare wt loss program. Outcome of interest is wt. loss = $(wt_{start} - wt_{end})$

Three programs are considered 1) low calorie diet. 2) low fat diet 3) low carbohydrate diet 4) control/placebo group (simply participating without any diet plan). A total of 20 participants agree to participate. ie 5 per group.

Differences of weight loss is given in following table. Positive value indicates wt loss whereas Negative value indicates weight gain.

low calorie	low fat	low carb	control group
8	2	3	2
9	4	5	2
6	3	4	-1
7	5	2	0
3	1	3	-3

Is there any statistical difference in mean weight loss among 4 groups? consider significance level of 0.05.

Step 1 Hypothesis

$H_0 \Rightarrow \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 \Rightarrow$ Means are not equal.

Step 2 F test $\because$ 4 different groups are given.

$\alpha = 0.05$

$$\boxed{F = \frac{MSB}{MSE}}$$

step 3 Decision rule

$df_1 = k-1 = 4-1 = 3$

$df_2 = N-k = 20-4 = 16$

$\alpha = 0.05$

$\therefore$ from f table (socr.ucla.edu - f tables).

critical value = 3.2389

if $F \geq 3.2389$, reject H_0.

step 4 test statistic

①

	Low cal	low fat	low carb	control
Group mean $\bar{x}_j$	6.6	3	3.4	1.2
n	5	5	5	5

overall mean, $\bar{x}$

$$\bar{x} = \frac{(8+9+6+7+3)+(2+4+3+5+1)+(3+5+4+2+3)+(2+2+(-1)+0+3)}{20}$$

$= 3.6$

② $SSB = \Sigma n_j (\bar{x}_j - \bar{x})^2$

$SSB = 5(6.6-3.6)^2 + 5(3-3.6)^2 + 5(3.4-3.6)^2 + 5(1.2-3.6)^2$

$= 45 + 1.8 + 0.2 + 28.8$

$= 75.8$

③ $SSE = \Sigma\left(\Sigma (x - \bar{x}_j)^2\right)$

low cal	$(X-6.6)$	$(X-6.6)^2$
8	1.4	1.96
9	2.4	5.76
6	-0.6	0.36
7	0.4	0.16
3	-3.6	12.96
Total	0	21.2

$$\Sigma (x-\bar{x}_1)^2 = 21.4$$

low fat	$(X-3)$	$(X-3)^2$
2	-1	1
4	1	1
3	0	0
5	2	4
1	-2	4
Total	0	10

$$\Sigma (x-x_2)^2 = 10$$

low carb	$(X-3.4)$	$(X-3.4)^2$
3	-0.4	0.16
5	1.6	2.56
4	0.6	0.36
2	-1.4	1.96
3	-0.4	0.16
Total	0	5.2

$$\Sigma (x-x_3)^2 = 5.2$$

control	$(X-1.2)$	$(X-1.2)^2$
2	0.8	0.64
2	0.8	0.64
-1	-2.2	4.84
0	-1.2	1.44
3	1.8	3.24
Total	0	10.8

$$\Sigma (x-x_4)^2 = 10.8$$

$$SSE = \Sigma\left(\Sigma (x-\bar{x}_j)^2\right) = 21.4 + 10 + 5.2 + 10.8$$

$$SSE = 47.4$$

Annova table

Source of variation	Sum of squares (SS)	dof	Mean squares (MS)	F value
B/w groups	75.8	4-1=3	75.8/3 = 25.3	25/2.96 = 8.44
Error (Residual)	47.4	20-4=16	47.4/16 = 2.96	–
Total	123.2	16+3=19	–	–

Step 5 conclusion

F value = 8.44

decision rule, $F \geqslant 3.24$, reject Ho

since $8.43 > 3.24$, Ho is rejected.

ie there is difference in mean of weight losses amoung 3 test groups and 1 placebo group.

Que: 2 calcium is essential mineral for bones. Daily recommended intake is 1000-1200 mg/day for mean & women. people take calcium supplement to overcome calcium deficiency.

A study is designed to test whether there is difference in mean daily intake of calcium in adults with 1) Normal bone density (NBD) 2) Osteopenia (low bone density which may lead to osteoporosis 3) osteoporosis. Adults with 60 years of age having above 3 conditions are selected at random from hospital record. and are invited to participate in study. Each participants calcium intake is measured and is reported/ tabulated in the following table.

normal bone density	osteopenia	osteoporosis
1200	1000	890
1000	1100	650
980	700	1100
900	800	900
750	500	400
800	700	350

Is there any significant statistical difference in mean calcium intake in patients with normal bone density as compared to patients with other two conditions. (Assume $\alpha = 0.05$)

Step 1 Hypothesis

$H_0 \Rightarrow \mu_{Normal} = \mu_{osteopenia} = \mu_{osteoporosis}$

$H_1 \Rightarrow$ Above mentioned means are not equal.

Step 2 Decision rule (F-test)

$df_1 = K-1 = 3-1 = 2$

$df_2 = N-K = 18-3 = 15$

$\alpha = 0.05$

critical value at above conditions

critical value $= 2.69517$

$\Rightarrow$ Decision rule

if $F \geq 2.69517$, H_0 is rejected.

Step 3 Test statistic

①

	Normal bone density	osteopenia	osteoporosis
sample size	$n_1 = 6$	$n_2 = 6$	$n_3 = 6$
Sample mean	$\bar{X}_1 = 938.33$	$\bar{X}_2 = 800$	$\bar{X}_3 = 715$

$\boxed{\bar{X} = 817.78}$ ie summation of all samples / Total samples

(2) Now $SSB = \Sigma n_j (\bar{x}_j - \bar{X})^2$

$SSB = 6(938.33 - 817.78)^2 + 6(800 - 817.78)^2 + 6(715 - 817.78)^2$

$= 87193.81 + 1896.77 + 63382.37$

$\boxed{SSB = 152472.95}$

(3) $SSE = \Sigma (\Sigma (x - \bar{x}_j)^2)$

NBD	$x - 938.33$	$(x - 938.33)^2$
1200	261.67	68471.19
1000	61.67	3803.19
980	41.67	1736.39
900	-38.33	1469.19
750	-188.33	35468.19
800	-138.33	19135.19
Total	0	130083.34

$\Sigma (x - \bar{x}_1)^2 = 130083.34$

ostopenia	$x - 800$	$(x - 800)^2$
1000	200	40000
1100	300	90000
700	-100	10000
800	0	0
500	-300	90000
700	-100	10000
Total	0	240000

$\Sigma (x - \bar{x}_2)^2 = 2400000$

osteoporosis	$(x - 715)$	$(x - 715)^2$
890	175	30625
650	-65	4225
1100	385	148225
900	185	34225
400	-315	99225
350	-365	133225
Total	0	449750

$\Sigma (x - \bar{x}_3)^2 = 449750$

$SSE = 130083.34 + 240000 + 449750 = 819833.34$

Anova table

SOV	SS	dof	MS	Fvalue
B/w Group	152472.95	3-1=2	152472.95/2 = 76236.47	76236.47/ 54655.56 = 1.3948
Error	819833.34	18-3=15	819833.34/15 = 54655.56	-
Total	972306.29	18-1=17	-	-

step 4 conclusion

Fvalue = 1.3948

decision rule $F \geq 2.69517$, Ho is rejected

since $1.3948 < 2.69517$, Ho cannot be rejected

ie. There is no statistical difference in mean calcium intake in patients with normal bone density as compared to patients with ostreopenia and osteoporosis.