

## 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, peppermint 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 between the groups.

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

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

i.e  $F\text{-value} \uparrow \Rightarrow \text{variation b/w groups } \uparrow$

i.e 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.

i.e  $H_0$  will be rejected.

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

$\Rightarrow$  for this example null hypothesis will be rejected.

$\Rightarrow$  Some more examples where Anova test can be applied.

1) Group of \* psychiatric patients trying three different therapies.

i.e 1) counselling      } Aim is to see which  
          2) medication     } therapy is better.  
          3) Group sessions

2) A light bulb is manufactured using 2 processes.

1) Process 1      } Aim is to see  
2) process 2      } which process is better.

3) Students of different colleges taking CAT exam

1) college 1      } Aim is to see, students  
2) college 2      } of which college performs  
3) college 3      } 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

Brand of cereal

levels

→ 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: I.V 1

Brand of cereal

Levels

corn flakes

lucky charms

Raisin Bran

I.V. 2

calories

levels

→ 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

$$H_0 \Rightarrow \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$H_1$  means are not equal.

② Decision rule ( $\alpha$ , dof).

$$df_1 = k - 1 \quad k \equiv \text{No. of Independent groups}$$

$$df_2 = N - k \quad N \equiv \text{Total no. of observation}$$

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

$\alpha$   $\equiv$  Significance value.

F value is taken from F-table using  
 $df_1$ ,  $df_2$  and  $\alpha$ .

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

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

### ③ Test statistic

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

where

$n_j$  = sample size of  $j$ th group  $\{j = \{1, 2, 3, \dots, k\}\}$

$\bar{x}_j$  = sample mean of  $j$ th 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)

$$\text{i.e. } N = n_1 + n_2 + n_3 + \dots + n_k$$

### Anova table

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

where  $x$  = individual observation

$\bar{x}_j$  = sample mean of  $j$ th Group

$\bar{x}$  = overall sample mean

$k$  = No. of groups

$N$  = Total No. of observations.

Ques 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

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

$H_1$ : Means are not equal.

Step 2 F test : 4 different groups are given.

$$\alpha = 0.05$$

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

### Step 3 Decision rule

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

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

$$\alpha = 0.05$$

∴ from f table (socr.ucla.edu - f tables),  
critical value = 3.2389

if  $F > 3.2389$ , reject  $H_0$ .

### Step 4 Test statistic

①

	low cal	low fat	low carb	control
Group mean	6.6	3	3.4	1.2
n	5	5	5	5

overall mean,  $\bar{X}$

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

$$= 3.6$$

$$\textcircled{2} \quad SSB = \sum n_j (\bar{X}_j - \bar{X})^2$$

$$\begin{aligned} SSB &= 5(6.6 - 3.6)^2 + 5(3 - 3.6)^2 + 5(3.4 - 3.6)^2 + 5(1.2 - 3.6)^2 \\ &= 45 + 1.8 + 0.2 + 28.8 \\ &= 75.8 \end{aligned}$$

$$\textcircled{3} \quad SSE = \sum (\sum (x - \bar{x}_j)^2)$$

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

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

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

$$\sum (x - \bar{x}_2)^2 = 10$$

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

$$\sum (x - \bar{x}_3)^2 = 5.2$$

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

$$\sum (x - \bar{x}_4)^2 = 10.8$$

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

$$SSE = 47.4$$

## Anova table

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

Step 5 conclusion

$$F \text{ value} = 8.44$$

decision rule,  $F \geq 3.24$ , reject  $H_0$

since  $8.43 > 3.24$ ,  $H_0$  is rejected.

i.e there is difference in mean of weight losses among 3 test groups and 1 placebo group.

Ques 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

$$H_0 \Rightarrow \mu_{\text{normal}} = \mu_{\text{osteopenia}} = \mu_{\text{osteoporosis}}$$

$H_1 \Rightarrow$  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

$$\text{critical value} = 2.69517$$

$\Rightarrow$  Decision rule

if  $F > 2.69517$ ,  $H_0$  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	$\bar{x}_1 = 938.33$	$\bar{x}_2 = 800$	$\bar{x}_3 = 715$

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

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

$$SSB = 6(938.33 - 817.78)^2 + 6(800 - 817.78)^2 + 6(715 - 817.78)^2 \\ = 87193.81 + 1896.77 + 63382.37$$

$$SSB = 152472.95$$

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

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

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

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

$$\sum (x - \bar{x}_2)^2 = 240000$$

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

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

$$SSE = 130083.34 + 240000 + 449750 = 819833.34$$

## 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	15+2=17	-	-

### Step 4 conclusion

$$F\text{value} = 1.3948$$

decision rule  $F > 2.69517$ ,  $H_0$  is rejected

since  $1.3948 < 2.69517$ ,  $H_0$  cannot be rejected

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