ANOVA

Dr K Saraswathi

Anova

Answer: ANOVA is used to compare the statistical significance between the means of *three or more independent (unrelated) groups*.

ANOVA

1) Developed by R.A. Fisher in 1920.
2) Also knows as F-Test, which is based on *F-distribution.*
3) It **compares the means** between the groups/levels and determines whether any of those means are statistically different from each other.

Basic Terms:

Experimental unit – the object on which a measurement (or measurements) is taken.

A factor is an independent variable whose values are controlled and varied by the experimenter.

A level is the intensity setting of a factor.

A treatment is a specific *combination of factor* levels.

Example

Example: A group of people is randomly divided into an *experimental and a control group*. The *control group* is given an aptitude test after *having eaten a full breakfast*. The *experimental group* is given the same test *without having eaten any breakfast*. What are the **factors, levels, and treatments in this experiment?**

Solution:

The **experimental units** are the people on which the response (test score) is measured.

Factor: described as "meal" and has *two levels*: "breakfast" and "no breakfast."

Treatments: Since MEAL is the only factor controlled by the experimenter, the two levels—"breakfast" and "no breakfast"—also represents the ***treatments*** in the experiment.

Factors	
Breakfast	**No-breakfast**
23	24
25	21
24	32

YouTube Like Subscribe

Example

Example 3: Consider an experimental design of the teaching method to examine the MARKS of the students in a respective subject via, the online mode, offline mode and video lecture based mode.

Factor: Teaching Method (Independent Variables):
Group/levels: the online mode, offline mode and video lecture based mode.

Teaching Method		
Online learning	Offline Leaning	Video Based learning
12	21	23
34	23	32
35	25	34
43	27	39
47	35	45

1 Factor
3 Levels

ANOVA

The responses that are generated in an experimental situation always exhibit a certain amount of ***variability***.

In an **ANALYSIS OF VARIANCE (ANOVA)**, our task to ***measure such variability*** and then examine whether there is any significant difference between the different sample means or not.

ANOVA is classified into the various types as

1) **Completely Randomized Design (CRD)**: One Way Classification/Factor/Levels
2) **Randomized Block Design (RBD)**: Two Way Classification/Factor/Levels

Completely Randomized Design (CRD):

One Way Classification/Factor/Levels

In it, random samples are *selected independently from each of k populations*. This design involves ***only one factor,*** and hence called as One-way ANOVA.

Group	Observations				Mean
1	x_{11}	x_{12}	...		$\overline{x_1}$
2	x_{21}	x_{22}	...		$\overline{x_2}$
...	...	...			
k	x_{k1}	x_{k2}	...		$\overline{x_k}$
	OVERALL MEAN				$\bar{x}$

Example:

	Teaching Methods	Observations				
1	Online learning	12	34	35	43	
2	Offline learning	21	23	25		
3	Video based learning	23	32	34	39	45

Three kinds of variations in ANOVA:

Between Groups: variation from one group to another.

$$\sum n_i(\bar{x}_i - \bar{x})^2$$

Within Groups: variation among the observations of each specific group.

$$\sum\sum(x_{ij} - \bar{x}_i)^2$$

Total: variations among all the observations (*which is nothing but the sum of Between Groups & Within Groups)*

$$\sum\sum(x_{ij} - \bar{x})^2$$

Assumptions:

1) **Random Selection:** Samples are drawn randomly
2) **Normal distribution:** Population from which samples are drawn follows normal distribution.
3) **Homogeneity of variance:** All sub-populations have the same variance, i.e., $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$
4) **Additivity of variance:** Total variance should be equal to sum of between variance and within variance.

F- ratio / ANOVA:

In ANOVA, we compute F-ratio, which is defined as

$$F = \frac{Mean\ Sum\ of\ Square\ Between\ Factors}{Mean\ Sum\ of\ Squates\ Within}$$

Procedure to compute the ANOVA Table:

2nd Method: (Preferable)

Group	Observations				Total
1	x_{11}	x_{12}	...		T_1
2	x_{21}	x_{22}	...		T_2
...	...	...			
k	x_{k1}	x_{k2}	...		T_k
	GRAND TOTAL				$G = T_1 + T_2 + \dots T_k$

Step 1: Compute Correction Factor

$$C = \frac{G^2}{N}$$

Here, N is total number of elements

Step 2:

SS total:

$$\sum\sum x_{ij}^2 - C$$

SS Between Group

$$\sum_{i=1}^{k} \frac{T_i^2}{n_i} - C$$

SS Within group (Error)

= SS Total - SS Between Group

Example 1: In an experiment to determine the **effect of nutrition** on the attention spans of elementary school students, *a group of 15 students* were randomly assigned to each of three meal plans: **no breakfast, light breakfast, and full breakfast**. Their attention spans (in minutes) were recorded during a morning reading period and are shown as

no breakfast	8	7	9	13	10
light breakfast	14	16	12	17	11
full breakfast	10	12	16	15	12

Construct the analysis of variance table for this experiment

Solution:

						Total
no breakfast	8	7	9	13	10	47
light breakfast	14	16	12	17	11	70
full breakfast	10	12	16	15	12	65
						G = 182

Correction Factor: $C = \frac{G^2}{N} = \frac{(182)^2}{15} = 2208.2667$

SS Total = $\sum\sum x_{ij}^2 - C$

$= 8^2 + 7^2 + \dots\dots\dots + 15^2 + 12^2 - 2208.2667$

$= 129.7333$

SS between Group

$= \sum_{i=1}^{3} \frac{T_i^2}{n_i} - C$

$= \frac{47^2}{5} + \frac{70^2}{5} + \frac{65^2}{5} - 2208.2667$

$=58.5333$

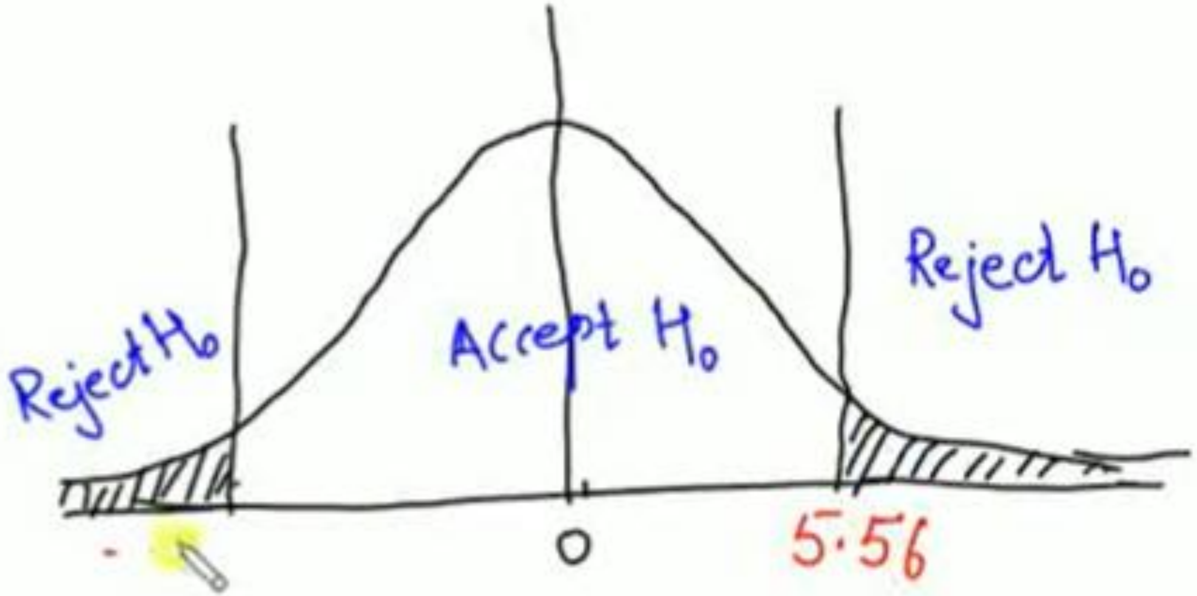
Thus, ANOVA table is

	Df (1)	SS (2)	MSS = SS/df (2)/(1)	F-ratio
Between Group		58.5333		
Within Error				
Total		129.7333		

Final ANOVA table is

	df (1)	SS (2)	MSS = SS/df (2)/(1)	F-ratio
Between Group	2	58.5333	29.2666	4.9326
Within Error	12	71.2000	5.9333	
Total	14	129.7333		

Given that F(3,14) = 5.56 or $F_{3,14}(0.01) = 5.56$



2nd Method (More easier in terms of Calculation)

Brand 1	20	23	18	17	
Brand 2	19	15	17	20	16
Brand 3	21	19	20	17	16
Brand 4	15	17	16	18	
					=

Subtract 20 (**or any Number**) from each number

						Total
Brand 1	0	3	-2	-3		-2
Brand 2	-1	-5	-3	0	-4	-13
Brand 3	1	-1	0	-3	-4	-7
Brand 4	-5	-3	-4	-2		-14
					G =	-36

$$C = \frac{G^2}{N} = \frac{(-36)^2}{18} = 72$$

SS Total $= \sum\sum x_{ij}^2 - C$

$= 0^2 + 3^2 + \cdots + (-4)^2 + (-2)^2 - 72$

$= 82$

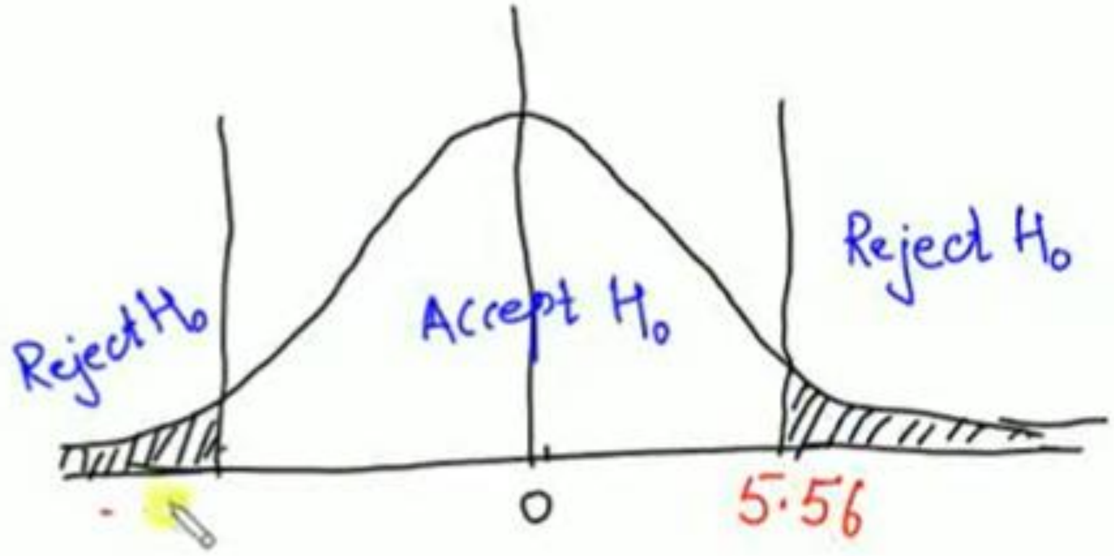
SS Between Brands $= \sum \frac{T_i^2}{n_i} - C$

$= \frac{(-2)^2}{4} + \frac{(-13)^2}{5} + \frac{(-7)^2}{5} + \frac{(-14)^2}{4} - 72$

$=21.6$

	df (1)	SS (2)	MSS = SS/df (2)/(1)	F-ratio
Between Brands	3	21.6	7.2	1.67
Within Error	14	60.4	4.31	
Total	17	82		

Given that F(3,14) = 5.56 or $F_{3,14}(0.01) = 5.56$



Two Way ANOVA

Example: An agricultural experiment was conducted to compare the yields of *three varieties* of rice applied by *two types of fertilizers*.

Assumptions:

The *three assumptions* for a two factor ANOVA, when there is only one observed measurement at each combination of levels of the two factors are as follows:

1) Normal distribution: Population at each factor level combination is normally distributed.

2) Homogeneity of variance: All sub-populations have the same variance, i.e., $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$

3) The effect of one factor is the same at all levels of the other factor
   a. It means that there is *no interaction* between the two factors.

TWO-WAY ANOVA test is used when the *NUMBER of observation in the subclasses are EQUAL.*

i.e., When number of observation in each row are equal and in column are equal then *only* Two-Way ANOVA can be used.

Example 1-

A	23	21	45	42	
B	25	24	36		
C	28	27	33	40	36

Example 2-

	P	Q	R	S
A	23	21	45	42
B	25	24	36	23
C	28	27	33	40

HYPOTHESIS

Hypothesis for factor 1:

H_0: There is no significance between the *means of* the row factor

H_1: there is a significance between the *means of* the row factor

Hypothesis for factor 2:

H_0: There is no significance between the *means of* the column factor

H_1: there is a significance between the *means of* the column factor

Source Table of Two-Way ANOVA

Source of Variation	df	SS	MSS = SS/df	F-ratio
Between Row	r-1	SSR	MSSR	MSSR/MSSE
Between Column	c-1	SSC	MSSC	MSSC/MSSE
Within Error	(r-1)(c-1)	SSE	MSSE	
Total	rc-1	SST		

r: No. Of rows; c = No. Of columns; df: Degree of freedom ;

SS: Sum of Squares ; MSS : Mean sum of squares

DEGREE OF FREEDOM

(1) Between Rows = No. Of Rows - 1

(2) Between Column = No. Of Column - 1

(3) Total = Total No. Of Elements - 1

Within Error = (3) – (1) – (2)

Correction Term: $C = \frac{G^2}{N}$

(1) (SSR) Between Row: $\sum \frac{T_i^2}{n_i} - C$

(2) (SSC) Between Column: $\sum \frac{T_j^2}{n_j} - C$

(3) (SST) Total $= \sum \sum x_{ij}^2 - C$

Within Error = (3) - (1) - (2)

Example: A farmer applied three types of fertilizers on 4 separate plots. The figures on yield per acre are tabulated below:

Fertilizers Plots →	Yield			
	A	B	C	D
Nitrogen	6	4	8	6
Potash	7	6	6	9
Phosphates	8	5	10	9

Find out if plots are materially different in fertility, as also, if the three fertilizers make any material difference in yields.

Perform a 2-way ANOVA on the data given below

Teachers	Students				
	I	II	III	IV	V
A	30	24	33	36	27
B	26	29	24	31	35
C	38	28	35	30	35

(a) Shift the origin to 30. Perform the ANOVA for the transformed data.

(b) How do the results compare with those obtained for the original data?

Solution:

Source of Variation	df	SS	MSS = SS/df	F-ratio
Between Teachers	2			
Between Students	4			
Within Error				
Total	14			

Teachers	Students					Total
	I	II	III	IV	V	
A	0	-6	3	6	-3	0
B	-4	-1	-6	1	5	-5
C	8	-2	5	0	5	16
Total	4	-9	2	7	7	G=11

Correction Factor: $C = \frac{G^2}{N} = \frac{(11)^2}{15}$

SS Total = $\sum\sum x_{ij}^2 - C$

$= 0^2 + (-6)^2 + \cdots \ldots + 0^2 + 5^2 - \frac{11^2}{15}$

$= 278.93333$

Teachers	Students					Total
	I	II	III	IV	V	
A	0	-6	3	6	-3	0
B	-4	-1	-6	1	5	-5
C	8	-2	5	0	5	16
Total	4	-9	2	7	7	G=11

Correction Factor: $C = \frac{G^2}{N} = \frac{(11)^2}{15}$

SS Total $= \sum\sum x_{ij}^2 - C$

$= 0^2 + (-6)^2 + \cdots \ldots + 0^2 + 5^2 - \frac{11^2}{15}$

$= 278.93333$

SS between Teachers $= \sum \frac{T_i^2}{n_i} - C$

$= \frac{0^2}{5} + \frac{(-5)^2}{5} + \frac{16^2}{5} - \frac{11^2}{15}$

$= 48.13333$

SS between Students $= \sum \frac{T_j^2}{n_j} - C$

$= \frac{4^2}{3} + \frac{(-9)^2}{3} + \frac{2^2}{3} + \frac{7^2}{3} + \frac{7^2}{3} - \frac{11^2}{15}$

$= 58.26667$

Example 1: Construct the ANOVA table for the following information

Drivers\Cars	1	2	3	4
a	18 A	21 B	25 C	11 D
b	22 B	12 C	15 D	19 A
c	15 C	20 D	23 A	24 B
d	22 D	21 A	10 B	17 C

Drivers\Cars	1	2	3	4	
a	18 A	21 B	25 C	11 D	
b	22 B	12 C	15 D	19 A	
c	15 C	20 D	23 A	24 B	
d	22 D	21 A	10 B	17 C	

Subtract 19 from each entry

Drivers\Cars	1	2	3	4	Total
a	-1 A	2 B	6 C	-8 D	-1
b	3 B	-7 C	-4 D	0 A	-8
c	-4 C	1 D	4 A	5 B	6
d	3 D	2 A	-9 B	-2 C	-6
					G=-9

Solution

Drivers\Cars	1	2	3	4	Total
a	18 A	21 B	25 C	11 D	
b	22 B	12 C	15 D	19 A	
c	15 C	20 D	23 A	24 B	
d	22 D	21 A	10 B	17 C	
Cars					

Subtract 19 from each entry

Drivers\Cars	1	2	3	4	Total
a	-1 A	2 B	6 C	-8 D	-1 ✓
b	3 B	-7 C	-4 D	0 A	-8
c	-4 C	1 D	4 A	5 B	6
d	3 D	2 A	-9 B	-2 C	-6
Cars	1	-2	-3	-5	G=-9

Totals of A, B, C, D = 5, 1, -7, -8

Correction Factor: $C = \frac{G^2}{N} = \frac{(-9)^2}{16}$

SS between Rows (Drivers) $= \sum \frac{T_i^2}{n_i} - C$

$= \frac{(-1)^2}{4} + \frac{(-8)^2}{4} + \frac{(6)^2}{4} + \frac{(-6)^2}{4} - \frac{(-9)^2}{16}$

$= 29.1875$

SS between Columns (Cars) $= \sum \frac{T_j^2}{n_j} - C$

$= \frac{(1)^2}{4} + \frac{(-2)^2}{4} + \frac{(-3)^2}{4} + \frac{(-5)^2}{4} - \frac{(-9)^2}{16}$

$= 4.6875$

SS between Treatments (petrol) $= \sum \frac{T_k^2}{n_k} - C$

$= \frac{5^2}{4} + \frac{1^2}{4} + \frac{(-7)^2}{4} + \frac{(-8)^2}{4} - \frac{(-9)^2}{16}$

$= 29.6875$

Total SS $= \sum \sum x_{ij}^2 - C$

$= (-1)^2 + (2)^2 + \cdots + (-9)^2 + (-2)^2 - \frac{(-9)^2}{16}$

$= 329.9375$

Source of Variations	df	SS	MSS	F-ratio
Between Row	3	29.1875	9.72317	0.21901 = Row/Error
Between Column	3	4.6875	1.56250	0.03519
Between Treatments	3	29.6875	9.89583	0.22290
Within Error	6	$\boxed{266.375}$	44.39583	
Total	15	329.9375		