ANOVA (ANALYSIS OF VARIANCE)

PRESENTED BY
DR. VANDANA PANDEY
ASSOCIATE PROFESSOR
DEPARTMENT OF COMMERCE

INTRODUCTION

- *The statistical technique used to compare the means of variation of more than two samples for population is called analysis of variance (ANOVA).*
- *ANOVA is based on two types of variation:*

➢ *Variations existing within the sample*

➢ *Variations existing between the samples.*

The ratio of these two variations is denoted by "F".

ANOVA

ANALYSIS OF VARIANCE (ANOVA)

- Powerful statistical tool for tests of significance.
- The test of significance using t-distribution is an adequate procedure only for testing the significance of the difference between two sample means.
- When we have three or more samples to consider at a time, an alternative procedure is needed for testing the hypothesis that all the samples are drawn from the same population i.e., they have the same means.
- PURPOSE OF ANOVA: To test the homogeneity of several means.

ANOVA

- ANOVA – A hypothesis testing procedure that is used to evaluate the mean differences between two or more treatments (or) populations.
- Invented by R.A. FISHER – Fisher's ANOVA.
- Similar to t-test and z-test – used to compare means and also the relative variance between them.
- It is used to test the statistical significance of the relationship between a dependent variable (Y) and a single or multiple independent variable (X).
- Helps to figure out if the null hypothesis has to be rejected or accepted.

HISTORY

- ***The concept of ANOVA was introduced by R.A.Fisher in 1920.***

ANOVA

TERMS USED IN ANOVA

- DEGREE OF FREEDOM (DF) – The number of independent conclusions that can be drawn from the data.
- SS_{factor} – It measures the variation of each group mean to the overall mean across all groups.
- Ss_{error} – It measures the variation of each observation within each factor level to the mean of the level.
- Main effect – The effect where the performance of one variable considered in isolation by neglecting other variables in the study.

ANOVA

- Interaction – This effect occurs where the effect of one variable is different across levels of one or more other variables.
- Mean square error (MSE) – Divide the sum of squares of the residual error by the degrees of freedom.
- F- test – The null hypothesis that the category means are equal in the population is tested by F statistic based on the ratio of mean square related to X and MSE.
- P – value – It is the smallest level of significance that would lead to rejection of the null hypothesis (H_0).

ANOVA

ASSUMPTIONS

1. *All ANOVA require* ***random sampling***
2. *The items for analysis of variance should be* ***independent*** *to each other or mutually exclusive.*
3. ***Equality or homogeneity*** *of variances in a group of sample is important.*
4. *The residual components should be* ***normally distributed.***

ANOVA

NULL HYPOTHESIS

- *It tests the null hypothesis*
- *H_o:There is no significant difference between the means of all groups.(all groups are same)*
- $H_o=\mu_1=,\mu_2=,\mu_3=....=\mu_K$
- *Where μ=group mean ,K=no. of group*

ALTERNATIVE HYPOTHESIS

- *HA:There are at least two groups means that are statistically significantly different from each other.*
- $H_A:\mu_1\neq\mu_2\neq\mu_3\neq...\neq\mu_K$

ANOVA

SUM OF SQUARES (SS)

- *It is the Sum of squares is computed as the sum of squares of deviation of the values for mean of the sample. It means:*

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

Here,

SS = represents sum of squares

$\bar{x}$ *= represent mean of sample*

ANOVA

MEAN SQUARES (MS)

- *Mean square (MS) is the mean of all the sum of squares. It is obtained by dividing SS with appropriate degree of freedom (d.f.)*

MS = sum of squares/degree of freedom= ss/df

ANOVA

DEGREE OF FREEDOM

- *The degree of freedom is equal to the sum of the individual degrees of freedom for each sample since each sample has degree of freedom equal to one less than sample size and there are k- sample, the total degrees of freedom is k less than the total sample size.*

$$df=N-k$$

ANOVA

TYPES OF ANOVA

- *Analysis of variance are classified into two groups:-*

1. One way Analysis of variance

- *In one way ANOVA only **one source of variation** (or factor) is investigated. ANOVA is used to test the null hypothesis that three or more treatments are equally effective.*
- *Examples:- To compare the effectiveness of two or more drugs in controlling or curing a particular disease.*

ANOVA

CALCULATION OF ONE WAY ANOVA

1. *Computation of one-way analysis of variance (ANOVA)*

A) Population variance between the groups.

- *The variation due to the interaction between the samples, denoted SSB for sum of squares between groups. There are k samples involved with one data volume for each sample (the sample mean) so there are k – 1 degrees of freedom.*
- *The variance due to the interaction between the samples, denoted MSB for mean square between groups. This is the between group variation divided by its degree of freedom.*

ANOVA

CONTII..

B) Population variance within the group

- ***The variation due to differences within the individual samples, denoted SSW for sum of squares within groups. Each sample is considered independently, The degree of freedom is equal to the sum of individual degrees of freedom for each sample. Since each sample has degree of freedom equal to one less than their sample size, and there are k samples, the total degrees of freedom is k less than the total sample size***
- $Df = N - k$
- ***The variance due to the difference within individual samples, denoted MSW for mean square within groups. This is the within group variation divided by its degree of freedom.***

ANOVA

PROCEDURE FOR CALCULATION OF ONE WAY ANOVA

- *Procedure for calculation of one way ANOVA*

1. *Obtain the mean of each sample.*
2. *Take out the average mean of the sample means.*
3. *Calculate sum of squares for variance between the sample SSB(SB^2)*
4. *Obtain variance for mean square (MS) between samples.*
5. *Calculate sum of squares for variance within samples SSW(SW^2)*
6. *Obtain the variance or means square (MS) within samples.*
7. *Find sum of squares or deviation for total variance*
8. *Finally, find F- ratio*

ANOVA

TABLE :- ANOVA TABLE: FOR ONE WAY CALCULATION.

S.No.	Source of variation	Sum of squares (SS)	Degree of freedom (df)	Mean of squares (MS)	Test statistic
1	Between	$SSB(SB^2)$	K - 1	$MSB = \frac{SSB}{K-1}$	
2	Within	$SSW(SW^2)$	N- k	$MSW = \frac{SSW}{N-K}$	$F= \frac{MSB}{MSW}$

ANOVA

EXAMPLE

Q.1.) The yield of 3 varieties of wheat A,B,C in 4 separated fields is show in following table. Test of significance of different in the yields of 3 varieties of wheat.

WHEAT	FIELDS			
VARIETY	FIELD 1	FIELD 2	FIELD 3	FIELD 4
A	12	18	14	16
B	19	17	15	13
C	14	16	18	20

ANOVA

SOLUTION:-

Fields	Wheat variety		
	A	B	C
1	12	19	14
2	18	17	16
3	14	15	18
4	16	13	20
TOTAL	$\Sigma x_1 = 60$	$\Sigma x_2 = 64$	$\Sigma x_3 = 68$

ANOVA

STEP 1:- SAMPLE MEAN

$\overline{X_1} = 60/4 = 15$

$\overline{X_2} = 64/4 = 16$

$\overline{X_3} = 68/4 = 17$

STEP 2 :- GRAND SIMPLE MEAN

$$\overline{X} = \frac{\overline{X1} + \overline{X2} + \overline{X3}}{3}$$

$= 15+16+17/3$

$=16$

ANOVA

STEP 3:- SSB = SUM OF SQUARES BETWEEN THE SAMPLE

$$= \Sigma ni\,(\bar{X}_1 - \bar{x})^2$$

$$= n_1(\bar{X}_1 - \bar{x})^2 + n_2(\bar{X}_2 - \bar{x})^2 + n_3(\bar{X}_3 - \bar{x})^2$$

$$= 4(15-16)^2 + 4(16-16)^2 + 4(17-16)^2$$

$$= 4 + 0 + 4$$

$$= 8$$

Step 4:- DEGREE OF FREEDOM

$$\text{Degree of freedom} = k - 1$$

$$= 3-1 = 2$$

$$MSB = SSB / (k - 1)$$

$$= 8 / 2 = 4$$

ANOVA

STEP 5:- CALCULATION OF SUM OF SQUARES WITHIN THE SAMPLE (SSW)

Wheat variety sample 1		Wheat variety sample 2		Wheat variety sample 3	
x1	$(x_1 - \overline{x_1})^2$	x2	$(x_2 - \overline{x_2})^2$	x3	$(x_3 - \overline{x_3})^2$
12	$(12\text{-}15)^2 = 9$	19	$(19\text{-}16)^2=9$	14	$(14\text{-}17)^2=9$
18	$(18\ \text{-}15)^2=9$	17	$(17\text{-}16)^2=1$	16	$(16\text{-}17)^2=1$
14	$(14\text{-}15)^2=1$	15	$(15\text{-}16)^2=1$	18	$(18\text{-}17)^2=1$
16	$(16\text{-}15)^2=1$	1	$(13\text{-}16)^2=9$	20	$(20\text{-}17)^2=9$
Total	20		20		20

ANOVA

STEP :- 6 CALCULATION OF SUM OF SQUARES WITHIN THE SAMPLE (SSW)

$$SSW = \Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2 + \Sigma(x_3-\bar{x}_3)^2$$

$$= 20 + 20 + 20$$

$$= 60$$

$$d.f. = N - k$$

$$= 12-3$$

$$= 9$$

$$MEAN\ SQUARE = SSW + (N-k)$$

$$= 60/9$$

$$= 6.7$$

ANOVA

STEP :- 7 ANOVA TABLE: ONE WAY ANOVA TABLE

S.No.	Source of variation	Sum of square	Degree of freedom	Mean square	Test statistic
1	Between sample	8	2	4	F = 6.7/4 = 1.57
2	Within sample	60	9	6.7	
TOTAL		68 (SST)	N-1=11		

ANOVA

The F - Distribution with $\alpha = 0.05$								
$\nu_2 \backslash \nu_1$	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
2	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18

ANOVA

CONCLUSION

Here, F(calculated) is $<$ F(tabulated) at 0.05 (2.19)

i.e. $F(calculated) = 1.67 < 4.26$

The value of F is less than 4.26 at 5% level with degree of freedom being $V_1 = 2$ and $V_2 = 9$ and hence could have arisen due to chance. There is no significant difference in the yield of three wheat varieties.

Hence, this analysis supports the the null hypothesis of no difference in sample mean.

ANOVA

2. Two way Analysis of variance

- *In this there are two independent variables are present. In ANOVA impact of two different factors on the variation in a specific variable is tested .*
- *Example:- Productivity of several nearby fields influenced by seed varieties and types of manure used.*

ANOVA

COMPUTATION (PROCEDURE) OF TWO WAY ANALYSIS OF VARIANCE

a. Sum of squares between the column = SSC

b. Sum of square between the rows = SSR

c. Sum of square for the residuals or error = SSE

Therefore, sum of square of total variance= SST

i) $SST = SSC + SSR + SSE$

ii) Correlation factor $(CF) = \frac{T2}{N}$

iii) $SST = \Sigma X_1^2 + \Sigma x_2^2 + \ldots + \Sigma x_k^2 - (T2/N)$

ANOVA

(iv) $$SSC = \frac{\Sigma(\Sigma X_c)^2}{n_c} - \frac{T^2}{N}$$

n_c is the number of units in each column

(v) $$SSR = \frac{\Sigma(\Sigma X_r)^2}{n_r} - \frac{T^2}{N}$$

n_r is the number of units in each row

(vi) SSE = SST − (SSC + SSR)

d.f between column = C-1

- ***d.f between row. = r- 1***
- ***d.f between residuals = (C-1)(r-1)***

ANOVA

FIG :- TWO WAY ANOVA TABLE

S.No.	Source of variation	Sum of squares SS	Degree of freedom d.f	Mean square MS	Total Statistic
1	Between column	SSC	C-1	MSC=SSC+C-1	$F = \frac{MSC}{MSE}$
2	Between Rows	SSR	r-1	MSR=SSR+ r-1	$F = \frac{MSR}{MSE}$
3	Residual	SSE	(C-1)(r-1)	MSE=SSE	
TOTAL		SST	N-1		

ANOVA

EXAMPLE

Q.1) In the the following table the crop yield of three varieties of crop from 4 different fields is shown, test whether (i) the mean yield of these varieties are equal and so also, (ii) the quality of field mean.

Variety of crops	Field			
	I	II	III	IV
A	9	5	9	7
B	7	9	5	5
C	7	8	4	9

ANOVA

SOLUTION

STEP 1:- Substract 5 from each value and arrange the data into columns and rows showings fields or blocks and varieties that represent factors.

Variety of crop	Fields				
	I	II	III	IV	Total of rows
A	4	0	4	2	$R_1 = 10$
B	2	4	0	0	$R_2 = 6$
C	2	3	-1	4	$R_3 = 8$
Total of columns	$C_1 = 8$	$C_2 = 7$	$C_3 = 3$	$C_4 = 6$	$R = 24$

ANOVA

Here, T= 24

$N = 12$

$CF = T^2/N = (24)^2/12$

$= 576/12$

$= 48$

STEP 2:- SSC = SUM OF SQUARES BETWEEN COLUMNS

$$= \frac{(8)^2}{3} + \frac{(7)^2}{3} + \frac{(3)^2}{3} + \frac{(6)^2}{3} - \frac{(24)^2}{12}$$

$= 21.3 + 16.3 + 3 + 12 - 48$

$= 52.6 - 48 = 4.6$

$d.f. = (C-1) = 4-1$

$= 3$

ANOVA

STEP 3:- SSR= SUM OF SQUARES BETWEEN VARIETIES (ROWS)

$$= \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(8)^2}{4} - \frac{T^2}{N}$$

$= 25+9+16 -48$

$=50-48$

$=2$

$d.f. = (r-1)$

$= 3-1$

$= 2$

ANOVA

<u>***STEP 4 :- SST = Total Sum of square***</u>

$$SST = \{4^2+0^2+4^2+2^2+2^2+4^2+0^2+0^2+2^2+3^2+(-1)^2+4^2\} - 48$$
$$= (16 + 0 + 16 + 4 + 4 + 16 + 0 + 0 + 4 + 9 + 1 + 16) - 48$$
$$= 86 - 48$$
$$= 38$$

Degree of freedom $= N - 1$
$= 12 - 1$
$= 11$

<u>***STEP 5 :- SSE = SST – SSC + SSR***</u>

$$= 38 - 4.6 + 2$$
$$= 31.4$$

Degree of freedom $= (C - 1)\,(r - 1)$
$= (4\text{-}1)\,(3\text{-}1)$
$= 3 \times 2$
$= 6$

ANOVA

STEP :- 6 ANOVA TABLE: TWO WAY ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREE OF FREEDOM	MEAN SQUARE (MS)	VARIANCE RATIO
BETWEEN THE COLUMN (VARIETIES)	2	2	2/2 = 1.00	F= 5.23/1.00 = 5.23
BETWEEN THE ROW (FIELDS)	4.6	3	4.6/3 = 1.53	F = 5.23 / 1.53 = 3.42
RESIDUAL	31.4	6	31.4/6 = 5.23	
TOTAL	38.0	11		

ANOVA

The F - Distribution with $\alpha = 0.05$								
$\nu_2 \backslash \nu_1$	2	3	4	5	6	7	8	9
2	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18

ANOVA

CONCLUSION

F(tabulated) is compared with F(calculated) at 5% for (C-1), (r-1)(C-1) i.e. F(3,6) and for (r-1), (r-1)(C-1) i.e. F(2,6).

F(calculated) for F(3,6) = 0.29

F(calculated) for F (2,6) = 0.19

And, F(tabulated) for F(3,6)= 4.76

F(tabulated) for F(2,6)= 5.14

F (calculated) < F(tabulated)

Hence, F(calculated) is less than F(tabulated), so the null hypothesis is accepted.

ANOVA

Merits and Demerits of One-Way ANOVA

- **Merits**

  Layout is very simple and easy to understand.

  Gives maximum degrees of freedom for error.

- **Demerits**

  Population variances of experimental units for different treatments need to be equal.

  Verification of normality assumption may be difficult.

ANOVA

Two-way ANOVA or Randomized block design

- In two-way ANOVA a study variable is compared over three or more groups, controlling for another variable.
- The grouping is taken as one factor and the control is taken as another factor.
- Grouping factor - Treatment.
- Control factor - Block.
- The accuracy of the test in two -way ANOVA is considerably higher than that of the one-way ANOVA, as the additional factor, block is used to reduce the error variance.

ANOVA

- The total variation present in the observations can be split into the following three components:
- (i) The variation between treatments (groups)
- (ii) The variation between blocks.
- (ii) The variation inherent within a particular setting or combination of treatment and block.

ANOVA

Merits and Demerits of two-way ANOVA

- **Merits**

  Any number of blocks and treatments can be used.

  Number of units in each block should be equal.

  It is the most used design in view of the smaller total sample size since we are studying two variable at a time.

- **Demerits**

  If the number of treatments is large enough, then it becomes difficult to maintain the homogeneity of the blocks.

  If there is a missing value, it cannot be ignored. It has to be replaced with some function of the existing values and certain adjustments have to be made in the analysis. This makes the analysis slightly complex.

ANOVA

Comparison between one-way ANOVA and two-way ANOVA

Basis of comparison	ANOVA	
	One-way	Two-way
Independent variable	One	Two
Compares	Three or more levels of one factor	Three or more levels of two factors, simultaneously
Number of observations	Need not be same in each treatment group	Need to be equal in each treatment group

Thank you