Chapter:13

Analysis of Variance (ANOVA)

INTRODUCTION

Analysis of variance is a powerful tool in the test of significance. In the test of significance 't-test' is used to test the equality of two population means. Whereas 'F-test' is used to test the equality of several means. So F-test plays a vital role in the analysis of variance.

Definition

According to Prof. R.A. Fisher, "separation of variances ascribable to one group of causes from the variance ascribable to another group of causes" is known as analysis of variance.

The following are the important assumptions of analysis of variance.

1. All the observations are independent.
2. The parent population from which observations are drawn should be a normal population.
3. The various class effects are additive in nature.

Uses

1. Analysis of variance is used to test the equality of several means.
2. It is used to have a linearity regression lines.
3. It is used to test the significance of correlation ratio.

These are two types of analysis of classification.

1. One-way classification

   If the data is divided into some classes according to one factor of causes, then the classification is called one way-classification.

2. Two-way classification

   If the data is divided into some classes according to two factors of causes, then the classification is called two-way classification.

ANALYSIS OF VARIANCE–ONE-WAY CLASSIFICATION

n units of the variable X are divided into k classes $A_1, A_2, \ldots, A_k$ of sizes $n_1, n_2, \ldots, n_k$. Such that $\sum_{i=1}^{k} ni = n$. n_i

Let Y_{ij} denote the value of jth unit in the ith class ($i = 1, 2, \ldots, k; j = 1, 2, \ldots, n_i$). These values can be put in the tabular form.

A_1	A_2		A_i		A_k
x_{11}	x_{21}		x_{i1}		x_{k1}
x_{12}	x_{22}		x_{i2}		X_{k2}
x_{1j}	x_{2j}		x_{ij}		x_{kj}
x_{1n1}	x_{2n2}		x_{ini}		x_{knk}
$x_{1.}$	$x_{2.}$		$x_{i.}$		$x_{k.}$

Mathematical Linear Model

The observations x_{ij} follow mathematical linear model as

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

where $i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, n_i$

μ = General mean effect

α_i = The mean effect of ith class

ϵ_{ij} = Error follows $N(0, \sigma^2)$

Assumptions

1. The sample observation x_{ij} should be independent.
2. Different class effects are additive in nature.
3. ϵ_{ij} follows $N(0, \sigma^2)$

Testing of Hypothesis

Now we set up null hypothesis as

$$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_k$$

i.e. $\quad H_0$: All the class effects are equal

Steps of Calculation

Calculate total G of all the observations is all samples

$$G = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 + \ldots + \Sigma x_k$$

Correction factor is (CF) = $\frac{(G)^2}{N}$

Calculate the total sum of squares (TSS) = $\Sigma\Sigma X_{ij}^2 - \frac{(G)^2}{N}$

Sum of squares of treatments

$$(\text{SST}) = \left[\frac{\sum X_1^2}{n_1} + \frac{\sum X_2^2}{n_2} + \frac{\sum X_3^2}{n_3} + \ldots + \frac{\sum X_k^2}{n_k}\right] - \frac{(G)^2}{N}$$

Error sum of squares (SSE) = TSS – SST

Mean sum of squares of treatments (MST) = $\frac{\text{SST}}{k-1}$

Mean sum of squares of errors (MSE) = $\frac{\text{SSE}}{N-k}$

Calculated F value = $\frac{\text{MST}}{\text{MSE}}$

Table F value = $F_{(k-1),\,(N-k)}$

Analysis of Variance Table

Source	*Degrees of freedom*	*Sum of squares*	*Mean sum of squares*	*F-calculated*	*F-table*
Treatments	$k-1$	SST	$\text{MST} = \dfrac{\text{SST}}{k-1}$	$F = \dfrac{\text{MST}}{\text{MSE}}$	$F_{(k-1),(N-K)}$
Error	$N-k$	SSE	$\text{MSE} = \dfrac{\text{SSE}}{N-k}$		
Total	$N-1$	TSS			

At α% level of significance if calculated value ≤ table value then we accept H_0.
If calculated value > table value then we reject H_0.

EXAMPLE 13.1: The following data relating to the prices of commodities in different months in five cities.

A	*B*	*C*	*D*	*E*
29	31	26	19	28
24	35	28	21	28
27	28	24	23	25
26	28	25	29	26
21	22	20	25	20

Test whether the difference between mean prices of commodities in cities is significant or not.

Solution: H_0: There is no significant difference between the mean prices of commodities in five cities.

A	29	24	27	26	21	127
B	31	35	28	28	22	144
C	26	28	24	25	20	123
D	19	21	23	29	25	117
E	28	28	25	26	20	127
Grand Total						638

No. of cities (treatments) $(k) = 5$
No. of columns $(n) = 5$
Total values or observation in table $(N) = 25$
Grand total $(G) = 638$

$$\text{Correction factor (CF)} = \frac{G^2}{N} = \frac{(683)^2}{25} = 16281.76$$

$$\text{TSS} = \Sigma\Sigma X_{ij}^2 - \text{CF}$$

$$\text{TSS} = [(29)^2 + (24)^2 + (27)^2 + \ldots + (25)^2 + (26)^2 + (20)^2] - 16281.76$$

$$= 16628 - 16281.76 = 346.24$$

$$\text{SST} = \left[\frac{(127)^2}{5} + \frac{(144)^2}{5} + \frac{(123)^2}{5} + \frac{(117)^2}{5} + \frac{(127)^2}{5}\right] - 16281.76 = 80.64$$

$$\text{SSE} = 346.24 - 80.64 = 265.6$$

$$\text{MST} = \frac{80.64}{5-1} = 20.16$$

$$\text{MSE} = \frac{265.6}{25-5} = 13.28$$

$$F = \frac{20.16}{13.28}\ 1.518072$$

F table value at 5% level of significance $F_{(5-1),(25-5)} = F_{(4, 20)} = 2.87$

Anova Table

Source	*Degrees of freedom*	*SS*	*MSS*	*F-calculated*	*F-table value*
Treatment	4	80.64	20.16	1.518072	2.87
Error	20	256.6	13.28		
Total	24	346.24			

Calculated value < table value then we accept H_0.

We conclude that there is no significant difference between mean price of commodities in five cities.

EXAMPLE 13.2: Analyse the following data test whether all the class effects are equal or nc

Class

A	B	C	D
1.60	1.40	1.50	2.10
1.81	1.80	1.60	2.30
1.63	1.64	1.80	2.40
1.65	1.91	–	2.50
1.70	1.81	–	–
1.71	2.01	–	–
1.81	–	–	

Solution: H_0: All the class effects are equal

Number of class (k) = 4
Total number of values (N) = 20

A	1.60	1.81	1.63	1.65	1.70	1.71	1.81	11.91
B	1.40	1.80	1.64	1.91	1.81	2.01	–	10.57
C	1.50	1.60	1.80	–	–	–	–	4.90
D	2.10	2.30	2.40	2.50				9.30
Grand Total								36.68

$$\text{Correction factor (CF)} = \frac{(36.68)^2}{20} = 67.27112$$

$$\text{TSS} = [(1.60)^2 + (1.81)^2 + (1.63)^2 + \ldots + (2.40)^2 + (2.50)^2] - 67.27112 = 1.64848$$

$$\text{SST} = \left[\frac{(11.92)^2}{7} + \frac{(10.57)^2}{6} + \frac{(4.09)^2}{3} + \frac{(9.30)^2}{4}\right] - 67.27112 = 1.239544$$

$$\text{SSE} = 1.64848 - 1.239544 = 0.408936$$

$$\text{MST} = \frac{1.239544}{4-1} = 0.413181$$

$$\text{MSE} = \frac{0.413181}{20-4} = 0.0255585$$

$$F = \frac{0.413181}{0.0255585} = 16.1660896$$

F table value at 5% level of significance $F_{(4-1),\,(20-4)} = F_{3,\,16} = 3.24$

Anova Table

Source	*Degrees of freedom*	*SS*	*MSS*	*F-calculated*	*F-table value*
Treatment	3	1.239544	0.413181	16.1660896	3.24
Error	16	0.408936	0.0255585		
Total	19	1.64848			

Calculated value > table value then we reject H_0.

We conclude that all the class effects are not equal.

ANALYSIS OF VARIANCE–TWO-WAY CLASSIFICATION

Let A and B denote two factors. Factor A is divided into k class like $A_1, A_2, \ldots, A_k$ and factor B is divided in to n classes $B_1, B_2, \ldots, B_n$.

Let X_{ij} denote the value of jth unit of factor B in the ith class of factor A. These values are arranged in the two-way table.

A/B	B_1	B_2		B_j		B_n	*Total*
A_1	X_{11}	X_{21}		X_{1j}		X_{1n}	X_1
A_2	X_{21}	X_{22}		X_{2j}		X_{2n}	X_2
A_i	X_{i1}	X_{i2}		X_{ij}		X_{in}	X_i
A_k	X_{k1}	X_{k2}		X_{kj}		X_{kn}	X_k
Total	X_1	X_2		X_j		X_n	$X = G$

Mathematical Linear Model

The observations x_{ij} follows mathematical linear model as

$$X_{ij} = \mu + \alpha_i + \beta_j + \in_{ij} \quad \text{where } i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n$$

where μ = General mean effect

α_i = Mean effect of ith class

β_j = Mean effect of jth class

$\in_{ij}$ = Error follows $N(0, \sigma^2)$.

Assumptions

1. The sample observation x_{ij} should be independent.
2. Different class effects are additive in nature.
3. $\in_{ij}$ follows $N(0, \sigma^2)$
4. There is no interaction effect between two factors A and B.

Testing of Hypothesis

H_{01}: All the class effects of factor A are equal.
H_{02}: All the class effects of factor B are equal.

Steps of Calculation

Calculate total G of all the observations is all samples

$$G = \Sigma x_{11} + \Sigma x_{21} + \Sigma x_{31} + \ldots + \Sigma x_{kn}$$

$$\text{Correction factor (CF)} = \frac{(G)^2}{N}$$

$$\text{Calculate total sum of squares (TSS)} = \Sigma\Sigma X_{ij}^2 - \frac{(G)^2}{N}$$

$$\text{Sum of squares of treatments (SST)} = \left[\frac{\sum X_{1.}^2}{n_1} + \frac{\sum X_{2.}^2}{n_2} + \ldots + \frac{\sum X_{k.}}{n_k}\right] - \frac{(G)^2}{N}$$

$$\text{Sum of squares of blocks (SSB)} = \left[\frac{\sum X_{.1}^2}{k_1} + \frac{\sum X_{.2}^2}{k_2} + \ldots + \frac{\sum X_{.n}^2}{k_n}\right] - \frac{(G)^2}{N}$$

Error sum of squares (SSE) = TSS – SST – SSE

Mean sum of squares of treatments (MST) = $\dfrac{\text{SST}}{k-1}$

Mean sum of squares of blocks (MSB) = $\dfrac{\text{SSB}}{n-1}$

Mean sum of squares of errors (MSE) = $\dfrac{\text{SSE}}{(k-1)\,(n-1)}$

Calculated F treatment value = $\dfrac{\text{MST}}{\text{MSE}}$

Table F (treatment) value = $F_{(k-1),\,(k-1)(n-1)}$

Calculated F blocks value = $\dfrac{\text{MSB}}{\text{MSE}}$

Table F (blocks) value = $F_{(n-1),\,(k-1)(n-1)}$

Analysis of Variance Table

Source	*Degrees of freedom*	*Sum of squares*	*Mean sum of squares*	*F-calculated*	*F-table value*
Factor A Treatment	$k-1$	SST	MST	$F = \dfrac{\text{MST}}{\text{MSE}}$	$F_{(k-1),\,(k-1)(n-1)}$
Factor B Blocks	$n-1$	SSB	MSB	$F = \dfrac{\text{MSB}}{\text{MSE}}$	$F_{(n-1),\,(k-1)(n-1)}$
Error	$(n-1)(k-1)$	SSE	MSE		
Total	$nk-1$	TSS			

At $\alpha\%$ level of significance if calculated value $\le$ table value then we accept H_{01}.
If calculated value $>$ table value then we reject H_{01}.

At $\alpha\%$ level of significance if calculated value $\le$ table value then we accept H_{02}.
If calculated value $>$ table value then we reject H_{02}.

EXAMPLE 13.3: Two drugs were simultaneously administered to patients to control body temperature. The number of days required to come to normal temperature were recorded and presented. Test the significant difference between doses of two drugs A and B using two-way analysis of variance technique at 5 per cent level of significance.

Drug A	*Drug B*		
	50 mg	*100 mg*	*150 mg*
50 mg	10	8	7
100 mg	9	6	5
150 mg	8	4	3
200 mg	7	3	2

Solution: H_{01}: In drug 'A' all doses are equal in their effectiveness.
H_{02}: In drug 'B' all doses are equal in their effectiveness.

Here number of drug $A(k) = 4$
Number of drug $B(n) = 3$

Drug A	*Drug B*			
	50 mg	*100 mg*	*150 mg*	*Total*
50 mg	10	8	7	25
100 mg	9	6	5	20
150 mg	8	4	3	15
200 mg	7	3	2	12
Total	34	21	17	72

$$\text{Correction factor (CF)} = \frac{(72)^2}{12} = 432$$

$$TSS = [(10)^2 + (8)^2 + (7)^2 + \ldots + (7)^2 + (3)^2 + (2)^2] - 432 = 506 - 432 = 74$$

$$\text{SS Drug } A = \left[\frac{(25)^2}{3} + \frac{(20)^2}{3} + \frac{(15)^2}{3} + \frac{(12)^2}{3}\right] - 432 = 32.6667$$

$$\text{SS drug } B = \left[\frac{(34)^2}{4} + \frac{(21)^2}{4} + \frac{(17)^2}{4}\right] - 432 = 39.5$$

$$ESS = 74 - 32.6667 - 39.5 = 1.8333$$

$$\text{MSS drug } A = \frac{32.6667}{4-1} = 10.8889$$

$$\text{MSS drug } B = \frac{39.5}{3-1} = 19.75$$

$$MSE = \frac{1.8333}{(3-1)(4-1)} = 0.30555$$

$$F^A = \frac{10.8889}{0.30555} = 35.63705$$

$$F^B = \frac{19.75}{0.30555} = 64.63754$$

At 5% level of significance the table value $F_{(4-1),\,(3-1)(4-1)} = F_{3,\,6} = 4.76$

$$F_{(3-1),\,(3-1)(4-1)} = F_{2,\,6} = 5.14$$

Analysis of Variance Table

Source	*Degrees of freedom*	*SS*	*MSS*	*F-calculated value*	*F-table value*
Drug *A*	$4 - 1 = 3$	32.6667	10.8889	$F^A = 35.63705$	$F^A = 4.76$
Drug *B*	$3 - 1 = 2$	39.5	19.75		
Error	$(4 - 1)(3 - 1) = 6$	1.8333		$F^B = 64.63754$	$F^B = 5.14$
Total	$12 - 1 = 11$	74			

At 5% level of significance if calculated value > table value then we reject H_{01}.
If calculated value > table value then we reject H_{02}.

We conclude that there is significant difference between doses of drug *A* and there is significant difference between doses of drug *B*.

EXAMPLE 13.4: A farmer applied three types of fertilizers on 4 separate plots. The figure of yield per acre are tabulated below.

Fertilizer	*Yield*			
	A	*B*	*C*	*D*
Nitrogen	6	4	8	6
Potash	7	6	6	9
Phosphates	8	5	10	9

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

Solution: H_{01}: There is no significant difference between the fertilizers plots.
H_{02}: There is no significant difference between the yields.

Here $k = 4$ and $n = 3$
$nk = 4 \times 3 = 12$

Fertilizers plots	*Yield*				
	A	*B*	*C*	*D*	*Total*
Nitrogen	6	4	8	6	24
Potash	7	6	6	9	28
Phosphates	8	5	10	9	32
Total	21	15	24	24	84

$$\text{Correction factor (CF)} = \frac{(84)^2}{12} = 588$$

12

$$\text{TSS} = [(6)^2 + (4)^2 + (8)^2 + \ldots + (10)^2 + (9)^2] - 588 = 624 - 588 = 36$$

$$\text{SSF} = \left[\frac{(21)^2}{3} + \frac{(15)^2}{3} + \frac{(24)^2}{3} + \frac{(24)^2}{3}\right] - 588 = 606 - 588 = 18$$

$$\text{SSY} = \left[\frac{(24)^2}{4} + \frac{(28)^2}{4} + \frac{(32)^2}{4}\right] - 588 = 596 - 588 = 8$$

$$\text{ESS} = 36 - 18 - 8 = 10$$

$$\text{MSF} = \frac{18}{4-1} = 6$$

$$\text{MSY} = \frac{8}{3-1} = 4$$

$$\text{MSE} = \frac{10}{(4-1)(3-1)} = 1.6667$$

F-calculated value (Fert) $= \dfrac{6}{1.6667} = 3.6667$

F-table value $= F_{(4-1),\,(4-1)(3-1)} = F_{3,\,6} = 4.76$

F-calculated value (Yield) $= \dfrac{4}{1.6667} = 2.4$

F-table value $= F_{(3-1),\,(4-1)(3-1)} = F_{2,\,6} = 5.14$

Analysis of Variance Table

Source	*Degrees of freedom*	*SS*	*MSS*	*F-calculated*	*F-table value*
Treatment	$4 - 1 = 3$	18	6	3.6667	4.76
Blocks	$3 - 1 = 2$	8	4		
Error	$(4 - 1)(3 - 1) = 6$	10	1.6667	2.4	5.14
Total	$12 - 1 = 11$	36			

At 5% level of significance if calculated value < table value then we accept H_{01}.
if calculated value < table value then we accept H_{02}.

We conclude that the plots are equally effective and the fertilizers have the same effect.

END