Chapter 2

ANOVA MODELS

One-way Classification Model

The term one-factor analysis of variance refers to the fact that a single variable or factor of interest is controlled and its effect on the elementary units is observed. In other words, in one-way classification the data are classified according to only one criterion. Suppose we have independent samples of $n_1, n_2, ..., n_k$ observations from k populations. The population means are denoted by $\mu_1, \mu_2, ..., \mu_k$. The one-way analysis of variance is designed to test the null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

i.e., the arithmetic means of the population from which the k samples are randomly drawn are equal to one another. The steps involved in carrying out the analysis are:

1. Calculate the Variance Between the Samples

Sum of squares is a measure of variation. The sum of squares between samples is dented by SSC. For calculating variance between samples, we take the total of the square of the variations of the means of various samples from the grand average and divide this total by the degrees of freedom. Thus the steps in calculating variance between samples will be:

a. Calculate the mean of each sample, ie., $\bar{X}_1, \bar{X}_2, ..., \bar{X}_k$;

b. Calculate the grand average $\bar{\bar{X}}$. Its value is obtained as follows: $\bar{\bar{X}} = \dfrac{\bar{X}_1 + \bar{X}_2 + ... + \bar{X}_k}{N_1 + N_2 + ... + N_k}$

c. Take the difference between the means of the various samples and the grand average;

d. Square these deviations and obtain the total which will The Analysis of Variance Table

Since there are several steps involved in the computation of both the between and within sample variances, the entire set of results may be organised into an analysis of variance (ANOVA) table. This table is summarized and shown below.

Source of variation	Sum of Squares SS	Degrees of Freedom	Mean Square MS	Variance Ratio F
Between Samples	SSC	$c-1$	$MSC=\dfrac{SSC}{c-1}$	$F=\dfrac{MSC}{MSE}$
Within Samples	SSE	$n-c$	$MSE=\dfrac{SSE}{n-c}$	
Total	SST	$n-1$		

Shortcut method

To use ANOVA table it is convenient to use the shortcut computation formula. The steps are given below.

1. Assume the means of the populations from which the k samples are randomly drawn are equal.
2. Compute Mean squares between the samples, say MSC and Mean square within the samples say MSE.

For computing MSC and MSE, following calculations are made,

a. T = sum of all the observations in rows and columns.

b. SST = sum of squares of all observations $-T^2/n$.
Here T^2/n is called correction factor.

c. SSC $= \dfrac{(\Sigma x_1)^2}{n_1}+\dfrac{(\Sigma x_2)^2}{n_2}+\ldots-T^2/n$ where $\Sigma x_1, \Sigma x_2, \ldots$ are the column totals.

d. $SSE = SST - SSC$

e. Then, $MSC=\dfrac{SSC}{n-c}$

f. Calculate: $MSE=\dfrac{SSE}{n-c}$

3. Calculate F-ratio $= \dfrac{MSC}{MSE}$

4. Obtain the table value of F for $(c-1, n-c)$ degrees of freedom. If the calculated value of F<table value accept the hypothesis that the sample means are equal. That is, the factors influence inthe same manner.

Example 1

Below are given the yield (in kg) per acre for 5 trial plots of 4 varieties of treatment.

	Treatment			
Plot No.	1	2	3	4
1	42	48	68	80
2	50	66	52	94
3	62	68	76	78
4	34	78	64	82
5	52	70	70	66

Carry out an analysis of variance and state your conclusions.

Solution

$I(x_1)$	$II(x_2)$	$III(x_3)$	$IV(x_4)$
42	48	68	80
50	66	52	94
62	68	76	78
34	78	64	82
52	70	70	66
240	330	330	400

T = Sum of all observations = 42 + 50 + ...+ 66 = 1300

$$\text{Correction factor} = \frac{T^2}{n} = \frac{(1300)^2}{20} = 84500$$

SST = Sum of the squares of all the observations $-T^2/n$.

$$= \left(42^2 + 50^2 + 62^2 + \dots + 66^2\right) - 84500 = 4236$$

$$\text{SSC} = \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \dots - T^2/n$$

$$= \frac{(240)^2}{5} + \frac{(240)^2}{5} + \frac{(330)^2}{5} + \frac{(330)^2}{5} + \frac{(400)^2}{5} - 84500 = 2580$$

$$\text{SSE} = SST - SSC = 4236 - 2580 = 1656$$

$$\text{MSC} = \frac{SSC}{c-1} = \frac{2580}{3} = 860,\ MSE = \frac{SSE}{n-c} = \frac{1656}{20-4} = 103.5$$

The degree of freedom = $(c-1, n-c) = (3, 16)$

The Analysis of Variance Table

Source of variation	Sum of Squares SS	Degrees of Freedom	Mean Square MS
Between Samples	SSC = 2580	$c-1 = 3$	$MSC = 860$
Within Samples	SSE = 1656	$n-c = 16$	$MSE = 103.5$
Total	SST = 4236	$n-1 = 19$	

$$F = \frac{MSC}{MSE} = \frac{860}{103.5} = 8.3$$

The table value of F at 5% level of significance for (3, 16) degrees of freedom is 3.24. The calculated value of F is more than the table value of F. Therefore the null hypothesis is rejected. $\therefore$ The treatments do not have same effect.

Coding Method

Coding method is, in fact, a furtherance of the short cut method. This method is based on a very important property of the variance ratio or the F-coefficient that its value does not change if all the given item values are either multiplied or divided by a common figure or if a common figure is either added or subtracted from each of the given item values. The main advantage of this method is that big figures are reduced in magnitude by division or subtraction and work is simplified without any disturbance on the variance ratio. This method should be used specially when given figures are big or otherwise inconvenient. Once the given figures are converted with the help of some common figure or the common factor, then all the steps of the shortcut method outlined above may be adopted for obtaining and interpreting the variance ratio.

Example 2

For the data given in example 1 carry out the analysis of variance technique using coding method.

Solution

Applying the coding method let us subtract 20 from each observation. Then the coded data is obtained as shown below:

X_1	X_2	X_3	X_4
0	+5	+4	+3
−1	+3	0	0
+1	+1	+2	0

To compute diffrent quantities, let us make the following table:

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
0	0	+5	25	+4	16	+3	9
−1	1	+3	9	0	0	0	0
+1	1	+1	1	+2	4	0	0
$\Sigma X_1=0$	$\Sigma X_1^2=2$	$\Sigma X_2=9$	$\Sigma X_2^2=35$	$\Sigma X_3=6$	$\Sigma X_3^2=20$	$\Sigma X_4=3$	$\Sigma X_4^2=9$

T = Sum of all observations = 0 + 5 + 4 + 3 + ... + 0 = 18

$$T^2/n = \frac{18 \times 18}{12} = 27$$

SST = Sum of squares of all observations $-T^2/n$

$$= \left(0^2 + 5^2 + 4^2 + \ldots + 0^2\right) - 27 = 66 - 27 = 39$$

$$\text{SSC} = \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} + \frac{(\Sigma x_4)^2}{n_4} - T^2/n$$

$$= \frac{0^2}{3} + \frac{9^2}{3} + \frac{6^2}{3} + \frac{3^2}{3} - 27 = 42 - 27 = 15$$

$$\text{SSE} = SST - SSC = 39 - 15 = 24$$

$$\text{MSC} = \frac{SSC}{c-1} = \frac{15}{3} = 5,\ MSE = \frac{SSE}{n-c} = \frac{24}{8} = 3$$

Analysis of Variance Table

Source of variation	Sum of Squares SS	Degrees of Freedom	Mean Square MS	Variance Ratio F
Between Samples	15	$4-1=3$	$MSC = \frac{15}{3} = 5$	$F = \frac{5}{3} = 1.67$
Within Samples	24	$12-4=8$	$MSE = \frac{24}{8} = 3$	
Total	39	$12-1=11$		

Table value of F for (3, 8) at 5% level of significance is 4.07. Since the calculated value of F is less than the table value, the null hypothesis is accepted. Therefore we can infer that the average life time of different brands of bulbs are equal.

Two-way Classification Model

In one-factor analysis of variance explained above the treatments constitute different levels of a single factor which is controlled in one experiment. There are, however many situations in which the response variable of interest may be affected by more than one factor. For example, sales of Maxfactor cosmetics, in addition to being affected by the point of sale display, might also be affected by the price charged, the size and/or location of the store or the number of competitive products sold by the store. Similarly petrol mileage may be affected by the type of car driven, the way it is driven, road conditions and other factors in addition to the brand of petrol used.

Thus with the two-factor analysis of variance, we can test two sets of hypothesis with the same data at the same time.

The procedure for analysis of variance is somewhat different from the one followed while dealing with problems of one-way classification.

The steps are as follows:

1. (a) Assume that the mean of all columns are equal. Choose the hypothesis as $H_0 = \alpha_1 = \alpha_2 = ... = \alpha_c$

   (b) Assume that the means of all rows are equal. Choose the hypothesis as $H_0 = \beta_1 = \beta_2 = ... = \beta_r$
2. Find T, the sum of all observations
3. Calculate SST = Sum of squares of all observations $-T^2 / n$

4. Find SSR $= \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \ldots - T^2 / n$ where

$\sum x_1, \sum x_2, \ldots$ are row totals.

5. Find SST $= \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \ldots - T^2 / n$ where

$\sum x_1, \sum x_2, \ldots$ are column totals.

6. SSE = SST − SSC − SSR.

7. MSC $= \frac{SSC}{c-1}, MSR = \frac{SSR}{r-1}, MSE = \frac{SSE}{(c-1)(r-1)}$

where c = number of columns, r = number of rows

8. $F_c = \frac{MSC}{MSE}$ and $F_r = \frac{MSR}{MSE}$

9. Determine the table value of F

If F_c < table value, accept H_0 as $\alpha_1 = \alpha_2 = \ldots = \alpha_c$

If F_r < table value, accept H_0 as $\beta_1 = \beta_2 = \ldots = \beta_r$

The Analysis of Variance Table

Source of variation	Sum of Squares SS	Degrees of Freedom	Mean Square MS	Variance Ratio F
Between Columns	SSC	$c-1$	MSC	F_c = MSC/MSE
Between rows	SSR	$r-1$	MSR	F_r = MSR/MSE
Residual	SSE	$(c-1)(r-1)$	MSE	
Total	SST	$n-1$		

Example 3

The following represent the number of units of production per day turned out by 4 different workers using 5 different types of machines.

Machine Types

Worker	A	B	C	D	E	Total
1	4	5	3	7	6	25
2	6	8	6	5	4	29
3	7	6	7	8	8	36
4	3	5	4	8	2	22
Total	20	24	20	28	20	112

On the basis of this information, can it be concluded that (a) the mean productivity is the same for different machines, (b) the mean productivity is different with respect to different workers.

Solution

Let us take the hypothesis that (a) the mean productivity of diffrent machines is same, ie., $H_0 ; \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$ and that (b) the 4 workers don't differ in respect of productivity. to carryout the analysis of variance. ie $H_0 ; \beta_1 = \beta_2 = \beta_3 = \beta_4$

Correction factor $= T^2 / n = (112)^2 / 20 = 627.2$

Sum of squares between machines

$$SSC = \frac{(20)^2}{4} + \frac{(24)^2}{4} + \frac{(20)^2}{4} + \frac{(28)^2}{4} + \frac{(20)^2}{4} - C.F.$$

$= 100+144+100+196 \quad +100 - 627.2 = 640 - 627.2 = 12.8$

d.f. $= (c-1) = (5-1) = 4$

Sum of squares between workers

$$SSR = \frac{(25)^2}{5} + \frac{(29)^2}{5} + \frac{(36)^2}{5} + \frac{(22)^2}{5} - C.F.$$

$$= 125+168.2+259.2+98.8 - 627.2$$

$$= 649.2 - 627.2 = 22$$

d.f.= $(r-1) = (4-1) = 3$

Total sum of squares = SST

$$= 4^2 + 6^2 + 7^2 + 3^2 + 5^2 + 8^2 + 6^2 + 5^2 + 3^2 + 6^2 + 7^2 + 4^2 +$$

$$7^2 + 5^2 + 8^2 + 8^2 + 6^2 + 4^2 + 8^2 + 2^2 - 627.2$$

$$= 692 - 627.2 = 64.8$$

SSE = SST − SSC − SSR = 64.8 − 12.8 − 22.0 = 30

df = $(c-1)(r-1) = (5-1)(4-1) = 12$

Source of variation	Sum of Squares SS	d.f.	Mean Square MS	Variance Ratio F
Between machines	12.8	4	3.20	$F_C = \frac{3.2}{2.5} = 1.28$
Between workers	22.0	3	7.33	$F_r = \frac{7.33}{2.5} = 2.93$
Residual	30.0	12	2.50	
Total	64.8	19		

For d.f. (4, 12) = $F_{0.05}$ = 3.26. The calculated value of F_c is less than the table value. Our hypothesis is true. There is no significant difference in the mean productivity of five machines. The table values of F_r for (2, 12) d.f; at 5% level is 3.49.

The calculated value of F is less than table value. Our hypothesis is true. Hence there is no significant difference in the mean productivity of different workers.

EXERCISES

Very Short Answer Questions

1. Give a practical situation where ANOVA can be applied.
2. What is the use of ANOVA technique?
2. State three assumptions of ANOVA technique.
4. What is one-way classification in ANOVA.
5. What do you mean by two-way classification model?

Short Essay Questions

6. What is 'analysis of variance' and where is it used? Give two suitable examples.
7. State some applications of the analysis of variance.
8. In order to determine whether there are significant differences in the durability of three makes of computer, samples of size n = 5 are selected from each make and the frequency of repair during the first year of purchase is observed. The results are as follows. In view of the above data, what conclusion can you draw?

Make A	:	5	6	8	9	7
MakeB	:	8	10	11	12	4
Make C	:	7	3	5	4	1

Long Essay Questions

9. The following figures relate to production in kg. of three varieties A, B and C of wheat sown in 12 plots

   A 14, 16, 18

   B 14, 13, 15, 22

   C 18, 16, 16, 19, 20

   Is there any significant difference in the production of 3 varieties?

10. Set a table of analysis of variance for the following data

Plots /Variety	A	B	C	D

1	200	230	250	300
2	190	270	300	270
3	240	150	145	180

Test whether varieties are different.

11. Following table gives the number of refrigerators sold by 4 salesmen in three months:

	Salesman			
Month	A	B	C	D
May	50	40	48	39
June	46	48	50	45
July	39	44	40	39

a. Determine whether there is any difference in the average sales made by four salesmen.

b. Determine whether the sales differ with respect to different month.