MODULE-5
ANALYSIS OF VARIANCE

Analysis of Variance

- Analysis of Variance is a technique developed by **R.A.Fisher**
- It is used to test for the significance of the difference among ***more than two sample means***
- To make inferences about whether such samples are drawn from the populations having the same mean.

Analysis of Variance

Analysis of Variance is a method of splitting the total variation of a data into constituent parts which measures different sources of variances.

The total variation is split into two components:

(a)Variation within the subgroups of samples

(b)Variation between the subgroups of the samples

Assumptions

- Each of the samples is a simple random sample.
- Populations from which the samples are selected is normally distributed .
- Each of the samples is independent of the other samples.
- Each one of the populations has the same variance and identical means.
- The effect of various components are additive.

Classification of ANOVA

One way ANOVA

Two way ANOVA

One way ANOVA

One way classification model is designed to study the effect of one factor in an experiment.

Example

- The following table gives the yields on 15 sample fields under three varieties of seeds (viz. A, B, C):

Yields		
A	**B**	**C**
5	**3**	**10**
6	**5**	**13**
8	**2**	**7**
1	**10**	**13**
5	**0**	**17**

- Test at 5% level of significance

Solution

- Hypothesis

$$H_0 : \sigma_1^{\,2} = \sigma_2^{\,2}$$

$$H_1 : \sigma_1^{\,2} \neq \sigma_2^{\,2}$$

Calculation of mean of each sample and grand average

$$\overline{X_1} = (5+6+8+1+5)/5 = 25/5 = 5$$

$$\overline{X_2} = (3+5+2+10+0)/5 = 20/5 = 4$$

$$\overline{X_3} = (10+13+7+13+17)/5 = 60/5 = 12$$

$$\overline{\overline{X}} = (5+4+12)/3 = 7$$

Calculation of sum of squares between the samples (SSB)

$A = \left(\overline{X_1} - \overline{\overline{X}}\right)^2$	$B = \left(\overline{X_2} - \overline{\overline{X}}\right)^2$	$C = \left(\overline{X_3} - \overline{\overline{X}}\right)^2$
$(5-7)^2 = 4$		
$(5-7)^2 = 4$		
$(5-7)^2 = 4$		
$(5-7)^2 = 4$		
$(5-7)^2 = 4$		
TOTAL= 20	45	125

SSB= 20+45+125 = 190

Calculation of sum of squares within the samples (SSW)

$A=\left(X_1-\overline{X_1}\right)^2$	$B=\left(X_2-\overline{X_2}\right)^2$	$C=\left(X_3-\overline{X_3}\right)^2$
TOTAL= 26	58	56

SSW=26+58+56 = 140

Contd….

- Prepare ANOVA table as follows:

Source of Variation	Sum of Squares	Degree of freedom	Mean Squares	Computed value of F	Table value of F
Between Samples	SSB = 190	c-1 = 3-1 = 2	MSB = SSB/c-1 = 190/2 = 95	F = MSB/MSW= 95/11.67 = 8.14	3.88
Within Samples	SSW = 140	n-c = 15-3 = 12	MSW= SSW/n-c = 140/ 12 = 11.67		
Total	SST= 330	n-1= 15-1 =14			

INTERPRETATION

- Since the computed value of F is greater than the table value of F, we reject the null hypothesis and conclude that there is significant difference in the variances. Hence, the average yield of land under different varieties of seed show significant differences.

Two way Classification Model

Two way classification model is designed to study the effects of two factors simultaneously in the same experiment.

ANOVA Table

S.V	Sum of Squares	Degree of freedom	Mean Squares	Variance Ratio
Between Columns	SSC	c-1	MSC = SSC/c-1	F1 = Greater variance/ Smaller Variance
Within Samples	SSR	r-1	MSR = SSR/r-1	F2 = Greater variance/smaller Variance
Residual/Error	SSE	(c-1) (r-1)	MSE = SSE /(c-1) (r-1)	
Total	SST	rc-1		

$$\text{SSC} = \frac{\text{sum of squares of total of each columns}}{\text{no. of items in each column}} - \text{correction factor}$$

$$\text{SSR} = \frac{\text{sum of squares of total of each row}}{\text{no. of items in each row}} - \text{correction factor}$$

$$\text{SST} = \text{sum of squares of all the obserations - correction factor}$$

$$\text{SSE} = \text{SST - (SSC + SSR)}$$

$$correction\ factor = \left(\frac{T^2}{N}\right)$$

Question

- The following table gives per hectare yield for three varieties of wheat each grown on fie plots:

Per hectare yield (in tons)			
Plot of land	Varieties of wheat		
	A	B	C
1	5	3	10
2	6	5	13
3	8	2	7
4	1	10	13
5	5	0	17

- Test at 5% level of significance

Solution

$$H_0 : \mu 1 = \mu 2 = \mu 3$$

$$H_1 : \text{At least two of the population means are unequal}$$

Calculation of sum of observations of each row and each column and their grand total

A	B	C	Row Total
5	3	10	18
6	5	13	24
8	2	7	17
1	10	13	24
5	0	17	22
Column Total=25	20	60	105 (Grand Total)

Correction factor

$$\left(\frac{T^2}{N}\right) = \frac{(25+20+60)^2}{15} = 735$$

SSC (Sum of squares between column)

$$SSC = \frac{\text{sum of squares of totals of each column}}{\text{no. of items in each column}} - correctionfactor$$

$$SSC = \left[\frac{(25)^2}{5} + \frac{(20)^2}{5} + \frac{(60)^2}{5}\right] - 735 = 190$$

SSR (sum of squares between rows)

$$\text{SSR} = \frac{\text{sum of squares of total of each row}}{\text{no. of items in each row}} - \text{correction factor}$$

$$SSR = \left[\frac{(18)^2}{3} + \frac{(24)^2}{3} + \frac{(17)^2}{3} + \frac{(24)^2}{3} + \frac{(22)^2}{3}\right] - 735 = 14.66$$

SST (Total Sum of Squares)

- SST=sum of squares of all the observations-correction factor

$$SST = \left[(5)^2 + (6)^2 + (8)^2 + (1)^2 + (5)^2 + \ldots\ldots + (17)^2\right] - 735 = 330$$

$$SSE = SST - (SSC + SSR) = 330 - (190 + 14.66) = 125.34$$

Solution

S.V	*Sum of Squares*	*Degree of freedom*	*Mean Squares*	*Computed value of F*	*Table value of F*
Between Columns	SSC=190	2	MSC = 95	F1 =95/15.67 = 6.06	4.46
Within Samples	SSR=14.7	4	MSR = 3.665	F2= 15.67/3.665=4.28	6.04
Error	SSE=125.3	8	MSE =15.67		
Total	SST=330	14			

1) ***6.06>4.46. Reject H0. There is significant difference in the variances between varieties of wheat***
2) ***4.28<6.04 . Accept H0. There is no significant difference in the variances between plots of land.***

Question

- The following table gives per hectare yield for three varieties of wheat each grown on fie plots:

Per hectare yield (in tons)			
Plot of land	Varieties of wheat		
	A	B	C
1	30	26	38
2	24	29	28
3	33	24	35
4	36	31	30
5	27	35	33

- Test at 5% level of significance