ANOVA

⇒ Two-way Analysis of Variance

When each observation is classified according to two criteria (or variables) of classification simultaneously. The classified data are recorded in a table, in which the columns represent one criterion of the classification and rows represent the other criterion.

Rows	Columns 1	2	...	j	...	c	Total	Mean
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1c}	$T_{1\cdot}$	$\bar{X}_{1\cdot}$
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2c}	$T_{2\cdot}$	$\bar{X}_{2\cdot}$
⋮	⋮							
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ic}	$T_{i\cdot}$	$\bar{X}_{i\cdot}$
⋮								
r	X_{r1}	X_{r2}	...	X_{rj}	...	X_{rc}	$T_{r\cdot}$	$\bar{X}_{r}$
Totals	$T_{\cdot 1}$	$T_{\cdot 2}$		$T_{\cdot j}$		$T_{\cdot c}$	$T_{\cdot\cdot}$	
Mean	$\bar{X}_{\cdot 1}$	$\bar{X}_{\cdot 2}$		$\bar{X}_{\cdot j}$		$\bar{X}_{\cdot c}$		$\bar{X}_{\cdot\cdot}$

⇒ There are now two null hypothesis, one corresponding to the problem that all the r-rows mean are equal and the other corresponding to the problem that all the c-column means are equal. Thus

$$H_0' : \mu_{1\cdot} = \mu_{2\cdot} = \cdots \mu_{r\cdot}$$

$$H_0'' : \mu_{\cdot 1} = \mu_{\cdot 2} = \cdots = \mu_{\cdot c}$$

H_1' : Not all $\mu_{i\cdot}$ are equal

H_1'' : Not all $\mu_{\cdot j}$ are equal

ANOVA - Table

Source of variation	d.f	Sum of squares	Mean Square (MS)	F
Between Rows	$r-1$	$SSR = \sum_i \frac{T_{i\cdot}^2}{c} - \frac{T_{\cdot\cdot}^2}{rc}$	$S_1^2 = \frac{SSR}{r-1}$	$F_1 = \frac{S_1^2}{S_3^2}$
Between columns	$c-1$	$SSC = \sum_j \frac{T_{\cdot j}^2}{r} - \frac{T_{\cdot\cdot}^2}{rc}$	$S_2^2 = \frac{SSC}{c-1}$	$F_2 = \frac{S_1^2}{S_3^2}$
Error (within)	$(r-1)(c-1)$	$SSE = SST - SSR - SSC$	$S_3^2 = \frac{SSE}{(r-1)(c-1)}$	
Total	$rc-1$	$SST = \sum\sum X_{ij}^2 - \frac{T_{\cdot\cdot}^2}{rc}$		

$F_1 \Rightarrow v_1 = r-1 \qquad v_2 = (r-1)(c-1)$

$F_2 \Rightarrow v_2 = c-1 \qquad v_2 = (r-1)(c-1)$

$\Rightarrow$ when ANOVA reject we move toward.
LSD (Least significant difference test)
Duncan's test.

Example 20.5. Four experimenters determine the moisture content of samples of a powder, each man taking a sample of six consignments. Their assessments are:

Observers	Consignments					
	1	2	3	4	5	6
1	9	10	9	10	11	11
2	12	11	9	11	10	10
3	11	10	10	12	11	10
4	12	13	11	14	12	10

Perform a two-way analysis of variance on these data and discuss whether there is any significant difference between consignments or between observers. (P.U., B.A/B.Sc. (Hons.) 1970)

(i) We set up the two null hypotheses corresponding to the problems that

(a) there is no significant difference between consignments, and

(b) there is no significant difference between observers, as

$$H_0' : \mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} = \mu_{\cdot 4} = \mu_{\cdot 5} = \mu_{\cdot 6}\text{, and}$$

$$H_0'' : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}.$$

The corresponding alternative hypotheses would be

H_1' : Not all $\mu_{\cdot j}$ are equal,

H_1'' : Not all $\mu_{i\cdot}$ are equal,

(ii) We choose the level of significance at $\alpha = 0.05$.

(iii) The test-statistics to use are

$$F_1 = \frac{\text{estimated variance from "Between Consignments } SS\text{"}}{\text{estimated variance from "Error } SS\text{"}} = \frac{s_1^2}{s_3^2},$$

and $$F_2 = \frac{\text{estimated variance from "Between Observers } SS\text{"}}{\text{estimated variance from "Error } SS\text{"}} = \frac{s_2^2}{s_3^2}.$$

which have F-distributions with $\upsilon_1=5$, $\upsilon_2=15$ and $\upsilon_1=3$ $\upsilon_2=15$ *d.f.* respectively, when the null hypotheses are true.

(iv) **Computations.** The necessary sums of squares are computed as shown in the table below:

Obser-vers	Consignments (Figures in brackets are the squares of X_{ij})								
	1	2	3	4	5	6	$T_{i\cdot}$	$T_{i\cdot}^2$	$\sum_j X_{ij}^2$
1	9 (81)	10 (100)	9 (81)	10 (100)	11 (121)	11 (121)	60	3600	604
2	12 (144)	11 (121)	9 (81)	11 (121)	10 (100)	10 (100)	63	3969	667
3	11 (121)	10 (100)	10 (100)	12 (144)	11 (121)	10 (100)	64	4096	686
4	12 (144)	13 (169)	11 (121)	14 (196)	12 (144)	10 (100)	72	5184	874
$T_{\cdot j}$	44	44	39	47	44	41	259	16849	2831
$T_{\cdot j}^2$	1936	1936	1521	2209	1936	1681	11219		↑
$\sum_i X_{ij}^2$	490	490	383	561	486	421	2831	←	Check

Now Total SS $= \sum_i \sum_j X_{ij}^2 - \frac{T_{..}^2}{rc}$

$$= 2831 - \frac{(259)^2}{24}$$

$$= 2831 - 2795.04 = 35.96,$$

Between Consignments SS $= \sum_j \frac{T_{\cdot j}^2}{r} - \frac{T_{..}^2}{rc}$

$$= \frac{11219}{4} - 2795.05 = 9.71,$$

Between Observers SS $= \sum_i \frac{T_{i\cdot}^2}{c} - \frac{T_{..}^2}{rc}$

$$= \frac{16849}{6} - 2795.04 = 13.13, \text{ and}$$

Error $SS = 35.96 - (9.71 + 13.13) = 13.12$

The ANOVA - Table is

Source of Variation	$d.f.$	Sum of Squares	Mean Square	Computed F
Between Consignments SSC	5	9.71	1.94	$F_1 = \frac{1.94}{0.87} = 2.23$
Between Observers SSR	3	13.13	4.38	$F_2 = \frac{4.38}{0.87} = 5.03$
Error SSE	15	13.12	0.87	---
Total SST	23	35.96	---	---

(v) The critical regions are (a) $F \geq F_{0\cdot05}\,(5, 15) = 2.90$,

(b) $F \geq F_{0\cdot05}\,(3, 15) = 3.29$.

(vi) **Conclusion.** Since the computed value of $F_1 = 2.23$ does not fall in the critical region but the computed value of $F_2 = 5.03$ falls in the critical region, so we accept the hypothesis relating to the consignments, and we reject the hypothesis corresponding to the fact that there is no significant difference between observers.