F – Test

It's a statistical test that is used to determine if the samples are homogeneous or drawn from the same population sometimes referred to as the variance ratio i.e used to test the homogeneity of different sets of data.

Now, suppose s_1^2 and s_2^2 are variances of two different samples drawn from the same population, then the hypothesis will be stated as

$$H_0: s_1^2 = s_2^2 \ vs \ s_1^2 \neq s_2^2$$

or

$$H_0: There\ is\ no\ significant\ difference\ in\ the\ samples\ i.e\ the\ are\ drawn\ from\ the\ same\ population$$
$$against$$
$$H_1: There\ is\ a\ significant\ difference\ in\ the\ samples\ i.e\ the\ samples\ are\ not\ drawn\ from\ the\ same\ p$$

The test statistic F will be given as

$$F_c = \frac{s_1^2}{s_2^2}\ if\ s_1^2 > s_2^2\ otherwise\ F_c = \frac{s_2^2}{s_1^2}\ if\ s_1^2 < s_2^2$$
$$such\ that\ F_c > 1$$

The critical value is

$$F_t = F_{\alpha(v_1, v_2)}$$
$$where\ \alpha - the\ level\ of\ significance,$$
$$v_1 - \ degrees\ of\ freedom\ of\ numerator$$

Properties of an F – distribution

i) The shape of the curve depends upon the number of degrees of freedom

ii) It is positively skewed and its skewness decreases with increase in the degrees of freedom

iii) The value of F must always be positive

Assumptions of an F – distribution

i) All samples must be randomly selected and independent

ii) The population from which the sample is drawn must be normally distributed

iii) The ratio between the variances must equal or greater than one

iv) All F – distributions are unimodal and positively skewed

v) Total variation of the various sources of variation should be additive

i.e Total sum of squares = sum of squares between the groups + sum of squares within the groups or

Total variation = variation due to assignable factors + variation due to chance factors

Example

Two different drugs are under scrutiny by the ministry of health to determine if they have the same effect in curing a certain disease. Drug A is locally manufactured and drug B is imported. 10 patients were subjected to drug A and the variance in their recovery rate was found to be 7.285 and the variance in the recovery rate of 11 patients subjected to drug B was 6.145. Do you think the drugs have the same effect at 1% level of significance?

Solution

$$H_0: S_1^2 = S_2^2 \; versus \; H_1: S_1^2 \neq S_2^2$$

$$F_c = \frac{S_1^2}{S_2^2} = \frac{7.285}{6.145} = 1.1855$$

$$F_t = F_{\alpha(V_1,V_2)} = F_{0.01(9,10)} = 4.94$$

$we\ realize\ that\ F_c < F_t$ Hence we accept the null hypothesis implying that the drugs have the same effects i.e there is no significant difference in the effect of the drugs in curing the disease

<u>Analysis of Variance (ANOVA)</u>

- The analysis of variance refers broadly to a collection of experimental situations and statistical procedures for the analysis of quantitative responses from experimental units.
- It refers to a statistical procedure of using population means and variance to test uniformity or homogeneity of data i.e a population that is not homogeneous has a very large variance. If samples are drawn from the same population then any variations will be purely environmental caused by random factors i.e
    (a) Variation within: - refers to unexplained variation
    (b) Variation among is the variation explained by the treatment
- ANOVA is a procedure adapted for the analysis and interpretation of data collected from a random experiment. The procedure essentially consists of portioning the experiment into components ascribable to different sources of variation. It is used to measure the variability or effects among different assignable factors and the error factor.

Example

Three different nutrients: proteins, carbohydrates and vitamins were given to 6 weeks old babies in the same neighborhood. The increase in weight of the babies was monitored for a period of three months. The following information was recorded.

Treatment (Nutrients) increase in weight of babies (Kgs)

Protein	5	4	6	
Carbohydrates	3	6	5	4
Vitamins	7	3		

Test whether the nutrients differ significantly at 1% level of significance.

Solution

$$H_0: t_1 = t_2 = t_3 \; Vs \; H_1: t_1 \neq t_2 \neq t_3$$

t_1	5 4 6	15
t_2	3 6 5 4	18
t_3	7 3	10
		43

$$G = Grand\ total = \sum_i \sum_j y_{ij} = 43$$

$$\sum_i \sum_j y_{ij}^2 = 5^2 + 4^2 + 6^2 + 3^2 + 6^2 + 5^2 + 4^2 + 7^2 + 3^2 = 221$$

$$n = 9\ implying\ degrees\ of\ freedom\ is\ n - 1 = 8$$

$$totat\ sum\ of\ squares = total\ variation = SS_T = \sum_i \sum_j y_{ij}^2 - \frac{G^2}{n} = 221 - \frac{43^2}{9} = 16$$

$$sum\ of\ variation\ due\ to\ treatment = ss_t = \sum \frac{t_i^2}{n_j} - \frac{G^2}{n} = \left(\frac{15^2}{3} + \frac{18^2}{4} + \frac{10^2}{2}\right) - \frac{43^2}{9} = 1$$

$$since\ SS_T = ss_t + ss_e\ we\ ss_e = ss_T - ss_t = 15$$

ANOVA Table

Sources of variation	Degrees of freedom	Sum of squares (SS)	Mean sum of squares (MSS)	F_c

Treatment	V – 1 = 3 – 1 =2	$ss_t = 1$	$MSS_t = \frac{ss_t}{v-1}$ $= \frac{1}{2} = 0.5$	$F_c = \frac{MSS_t}{MSS_e}$ $= \frac{0.5}{2.5} = 0.2$
Error	n – v =9 – 3 = 6	$ss_e = 15$	$MSS_e = \frac{ss_e}{n-v}$ $= \frac{15}{6} = 2.5$	
Total	n – 1 =9 – 1 = 8	$ss_T = 16$		

$$Therefore\ F_c = 0.2$$

$$F_t = F_{\alpha(V_1,V_2)} = F_{0.01(2,6)} = 10.92$$

$we\ realize\ that\ F_c < F_t$ we accept the null hypothesis implying there is no significant difference in the nutrients.

Example

Four different drugs have been developed to control the effect a certain disease. The drugs are tried on patients in three different hospitals. The result shown gives the number of recoveries from the disease per a group of 30 patients given a particular drug in a given facility

	HOSPITALS			
Drugs	Government (H_1)	Private (H_2)	Mission (H_3)	Total (t_i)
A	22	15	26	63
B	18	23	16	57
C	20	26	27	73
D	21	13	17	51
Total (b_j)	81	77	86	244

At 5% levels of significance confirm if the drugs differ and also whether the hospitals differ.

Solution

	Blocks	

treatments	b_1	b_2	b_3	Total (t_i)
t_1	22	15	26	63
t_2	18	23	16	57
t_3	20	26	27	73
t_4	21	13	17	51
Total (b_j)	81	77	86	244

$$Treatments\ are\ the\ drugs\ and\ blocks\ the\ hospitals$$

$$H_0: t_1 = t_2 = t_3 = t_4\ Vs\ H_1: t_1 \neq t_2 \neq t_3 \neq t_4$$

$$H_0: b_1 = b_2 = b_3\ Vs\ H_1: b_1 \neq b_2 \neq b_3$$

$$G = Grand\ total = 244$$

$$\sum_i \sum_j y_{ij}^2 = 5198$$

$$totat\ sum\ of\ squares = total\ variation = SS_T = \sum_i \sum_j y_{ij}^2 - \frac{G^2}{n} = 5198 - \frac{244^2}{12}$$

$$= 236.7$$

$$sum\ of\ variation\ due\ to\ treatment = ss_t$$

$$= \sum \frac{t_i^2}{b_j} - \frac{G^2}{n} = \left(\frac{63^2}{3} + \frac{57^2}{3} + \frac{73^2}{3} + \frac{51^2}{3}\right) - \frac{244^2}{12} = 88$$

$$sum\ of\ variation\ due\ to\ block = ss_b = \sum \frac{b_j^2}{n_i} - \frac{G^2}{n} = \left(\frac{81^2}{4} + \frac{77^2}{4} + \frac{86^2}{4}\right) - \frac{244^2}{12} = 10.2$$

$$since\ SS_T = ss_t + ss_b + ss_e\ \ then\ it\ follows\ that\ \ ss_e = ss_T - ss_t - ss_b = 138.5$$

<u>ANOVA Table</u>

Sources of variation	Degrees of freedom	Sum of squares (SS)	Mean sum of squares (MSS)	F_c

Treatment (Drugs)	V – 1 = 3	$ss_t = 88$	24.3	$F_c = \frac{MSS_t}{MSS_e} = 1.27$
Blocks (Hospitals)	b – 1 = 2	$ss_b = 10.2$	5.1	$F_c = \frac{MSS_b}{MSS_e} = 0.02$
Error	(v-1)(b-1) = 6	$ss_e = 138.5$	23.1	-
Total	bv – 1 = 11	$ss_T = 236.7$	-	-

$$Note\ that\ F_t = F_{0.05(3,6)} = 4.76\ and\ F_{0.05(2,6)} = 5.14$$

$$in\ both\ cases\ F_c < F_t$$

$$so\ we\ accept\ the\ null\ hypothesis\ (H_0) implying\ that\ there\ is\ no\ significant$$

(i) in the treatments (drugs)
(ii) in the blocks (hospitals)