Sunday, April 14, 2024 9:14 AM

ANOVA

ANALYSIS OF VARIANCE

→ extension of z-test & t-test

→ f-statistic

↳ is used to compare variances among the groups

↳ $f = \frac{\text{variance of } 1^{st} \text{ group}}{\text{variance of } II^{nd}} = \frac{SD_1^2}{SD_2} \Rightarrow SD_1 > SD_2$

ANOVA → n-way ANOVA

ANOVA → One way ANOVA, two way ANOVA

ANOVA:

One Way ANOVA

Q1: To assess the significance of possible variation in performance in a certain test between the convent schools of a city, a common test was given to a number of students taken at a random from the 5th class of the 3 schools concerned. The result is given as follows:

A	B	C
9	13	14
11	12	13
13	10	17
9	15	7
8	5	9

5-1 = 4 ; 5-1 = 4 ; 5-1 = 4 ⇒ 12

degrees of freedom (df_1) = c − 1 (↳ columns)

$df_2 = (5-1)+(5-1)+(5-1)$

$= 15 - 3$

$= n - c$ (↳ columns)

Make the Analysis of Variance of the given data. (Null Hypo: No Significance Variation in the schools).

Solution:

Null Hypothesis = No variation between schools

Alt. Hypothesis = There is variation between schools

Source of Variation	Sum of Square	Degrees of freedom	Mean Square	F
Between the Sample	SSC	(c-1) = 2	MSC = SSC/df1	F = MSC/MSE
Within the sample	SSE	(n-c) = 12	MSE = SSE/df2	

$\bar{X}_A = 10$

$\bar{X}_B = 11$

$\bar{X}_C = 12$

$\bar{\bar{X}} = 11 = \frac{\bar{X}_A + \bar{X}_B + \bar{X}_C}{3}$

SSC

SSC

$\bar{X}_A - \bar{\bar{X}}$	$(\bar{X}_A - \bar{\bar{X}})^2$	$\bar{X}_B - \bar{\bar{X}}$	$(\bar{X}_B - \bar{\bar{X}})^2$	$\bar{X}_C - \bar{\bar{X}}$	$(\bar{X}_C - \bar{\bar{X}})^2$
$10 - 11 = -1$	1	$11 - 11 = 0$	0	$12 - 11 = 1$	1
$10 - 11 = -1$	1	$11 - 11 = 0$	0	$12 - 11 = 1$	1
$10 - 11 = -1$	1	$11 - 11 = 0$	0	$12 - 11 = 1$	1
$10 - 11 = -1$	1	$11 - 11 = 0$	0	$12 - 11 = 1$	1
$10 - 11 = -1$	1	$11 - 11 = 0$	0	$12 - 11 = 1$	1
	5		0		5

$$SSC = 5 + 0 + 5 = 10$$

$$MSC = \frac{SSC}{df_1} = \frac{10}{2} = 5$$

$$df_1 = C - 1 = 3 - 1 = 2$$

SSE

$A - \bar{X}_A$	$(A - \bar{X}_A)^2$	$(B - \bar{X}_B)$	$(B - \bar{X}_B)^2$	$(C - \bar{X}_C)$	$(C - \bar{X}_C)^2$
$9 - 10 = -1$	1	$13 - 11 = 2$	4	$14 - 12 = 2$	4
$11 - 10 = 1$	1	$12 - 11 = 1$	1	$13 - 12 = 1$	1
$13 - 10 = -3$	9	$10 - 11 = -1$	1	$17 - 12 = 5$	25
$9 - 10 = -1$	1	$15 - 11 = 4$	16	$7 - 12 = -5$	25
$8 - 10 = -2$	4	$5 - 11 = -6$	36	$9 - 12 = -3$	9
	16		58		64

$$SSE = 16 + 58 + 64 = 138$$

$$df_2 = n - C = 15 - 3 = 12$$

$$MSE = \frac{SSE}{df_2} = \frac{138}{12} = 11.5$$

$F_{cal} = \frac{MSC}{MSE} = \frac{5}{11.5} = 0.435$

f_{tab}, Df1 $(\nu_1) = 2$

Df2 $(\nu_2) = 12$

$f_{tab} = 3.89$

Compare f_{tab} with f_{cal}

$f_{tab} > f_{cal}$ Failed to Reject H_0

2-WAY ANOVA

The following data represents the number of Units of Tablet production (in thousands) per day by five different technicians by using 4 different machines.

a. Tell whether the mean productivity of the different machines are same?
b. Test whether the 5 technicians differ w.r.t. the mean productivity?

Machines …… Technicians	A	B	C	D
P	54	48	57	46
Q	56	50	62	53
R	44	46	54	42
S	53	48	56	44
T	48	52	59	48

Mid value $= \frac{42+62}{2} = 52 \& 50$

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of squares	F
Between the columns	SSC = 338.8	df = c-1 = 4-1 = 3	MSC= SSC/(c-1) =	MSC/MSE =
Between the rows	SSR = 158	df = r-1 = 5-1 = 4	MSR = SSR/(r-1)	MSR/MSE
Residual Errors	SSE = 67.2	df = (c-1)(r-1) = 3×4 = 12	MSE = SSE/(c-1)(r-1)	
Total Sum of Square	SST = 564	df = n-1		

$MSC = \frac{3388}{3} = 112.93$

$MSR = 158/4 = 39.5$

Mid value = 50

$MSE = \frac{67.2}{12} = 5.6$

Grand Total:

M/cs → Tech ↓	A	B	C	D	Total
P	54−50 = 4	48−50 = −2	57−50 = 7	46−50 = −4	5 = ΣP
Q	56−50 = 6	50−50 = 0	62−50 = 12	53−50 = 3	21 = ΣQ
R	44−50 = −6	46−50 = −4	54−50 = 4	42−50 = −8	−14 = ΣR
			56−50 = 6	44−50 = −6	1 = ΣS

K	$44-50=-6$	$46-50=-4$	$56-50=6$	$44-50=-6$	$1=\Sigma S$
S	$53-50=3$	$48-50=-2$	$59-50=9$	$48-50=-2$	$7=\Sigma T$
T	$48-50=-2$	$52-50=2$			$\Rightarrow 20 \Rightarrow$ Grand total
	$\Sigma A=5$	$\Sigma B=-6$	$\Sigma C=38$	$\Sigma D=-17$	

Explored Correction factor $= \frac{(GT)^2}{N} = \frac{20^2}{20} = 20$

SSC $\Rightarrow \frac{(\Sigma A)^2}{n_A} + \frac{(\Sigma B)^2}{n_B} + \frac{(\Sigma C)^2}{n_C} + \frac{(\Sigma D)^2}{n_D}$ - correction factor

$\Rightarrow \frac{(5)^2}{5} + \frac{(-6)^2}{5} + \frac{(38)^2}{5} + \frac{(-17)^2}{5} - \frac{20^2}{20}$

$\Rightarrow 338.8$

SSR $\Rightarrow \frac{(\Sigma P)^2}{n_P} + \frac{(\Sigma Q)^2}{n_Q} + \frac{(\Sigma R)^2}{n_R} + \frac{(\Sigma S)^2}{n_S} + \frac{(\Sigma T)^2}{n_T}$ - correction factor

$\frac{(5)^2}{4} + \frac{21^2}{4} + \frac{(-14)^2}{4} + \frac{1^2}{4} + \frac{7^2}{4} - \frac{20^2}{20}$

$= 158$

SST $\Rightarrow$ Sum of squared of all residual — correction factor

$\Rightarrow 4^2 + 6^2 + (-6)^2 + \text{- - - -} + (-6)^2 + (6)^2 \rightarrow \frac{20^2}{20}$

$\Rightarrow 564$

SSE $\Rightarrow$ SST $-$ (SSC + SSR)

$\Rightarrow 564 - (330.8 + 158)$

$= 67.2$

$\alpha = 0.05$

Between the columns

$f_{cal} = \frac{MSC}{MSE} = \frac{112.93}{5.6} = 20.16$

$df_1 = \nu_1 = 3$ $\quad f_{tab} = 3.49$

$df_2 = \nu_2 = 12$

$f_{cal} > f_{tab}$ Reject H_0

Between the rows

$F_{cal} = \frac{MSR}{MSE}$

$f_{cal} = \frac{39.5}{5.6} = 7.05$

$df_1 = \nu_1 = 4$ $\quad F_{tab} = 3.26$

$df_2 = \nu_2 = 12$

$F_{cal} > f_{tab}$

Reject H_0