Module II

Testing of Hypothesis

04 : ANOVA MODELS

→ 1-way classification Model =

* Here, the data are classified according to only 1 criterian.
* Suppose we have independent samples of $n_1, n_2, \ldots n_k$ from k population.
* The population means are denoted by $\mu_1, \mu_2 \ldots \mu_k$.
* 1-way analysis of variance (Anova) is designed to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$.
* Steps involved in carrying out the analysis–

a) calculate V(x) b/w the samples.

b) Sum of squares is a measure of variation & The sum of squares b/w samples is denoted by SSC

* Steps in cal. variance b/w samples –
* calculate mean of each samples, $\bar{x}_1, \bar{x}_2 \ldots \bar{x}_k$
* calculate grand avg $\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_k}{n_1 + n_2 + \cdots + n_k}$.
* find $\bar{x}_1 - \bar{\bar{x}}, \bar{x}_2 - \bar{\bar{x}}, \ldots \bar{x}_k - \bar{\bar{x}}$.
* Square this deviation & obtain the total which will give sum of □'s b/w the samples & ÷ the total obtained by the degrees of freedom: (DF)

$$\frac{(\bar{x}_1 - \bar{x})^2 + (\bar{x}_2 - \bar{x})^2 + \cdots + (\bar{x}_k - \bar{\bar{x}})^2}{k-1}$$

k → no. of samples.

* cal V(x) with in the samples.

V(x) [sum of □'s] within samples measures those inter sample diffrnce that arise due to chance only. Denoted by SSE

* Steps in calculating V(x) within the sample will be –

a) cal. mean value of each sample (i.e) $\bar{x}_1, \bar{x}_2 \ldots \bar{x}_k$

b) Deviation of variation of various items in a sample from mean values of samples. (i.e) $x_1 - \bar{x}, x_2 - \bar{x} \ldots x_k - \bar{x}$

c) □ this deviations & obtain the total which gives the sum of □'s within samples.

d) ÷ this total obtained in step c by DF

e) cal. F ratio, $F^* = \frac{V(x) \text{ b/w sample}}{V(x) \text{ within the sample}}$

(or)

$$F = \frac{S_1^2}{S_2^2}$$

f) Compare the cal. value of F with the table value of F for gvn DF at certain critical level (generally we take 5% level of significance).

* If calculated value of F is < than table value of F, indicates that the diffrence in sample means is significant. (accepted)

* If caln value of F is > then table value of F, the diffrence is not significant (rejected).

Q
1) As head of a deptmnt of a consumr's research organization, you have responsibility for testing & comparing lifetimes of light bulbs for 4 brands of bulbs. Suppose you test the lifetime of 3 bulbs of each of the 4 brands. yr test data are shown below, each entry representing the lifetime of a bulb, measured in 100's hr.

Brand.

A	B	C	D
20	25	24	23
19	23	20	20
21	21	22	20.

Can we infer that the mean life time of the four brands are equal.

A) The null hypothesis, the avg lifetime of 4 brands of bulb are equal.,

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

let x_1, x_2, x_3 & x_4 denote the mean life time of brand A, B, C, D & $\bar{\bar{X}}$ be overall grand mean.

x_1	x_2	x_3	x_4
20	25	24	23
19	23	20	20
21	21	22	20
60	69	66	63

$\bar{x}_1 = \frac{60}{3} = 20$

$\bar{x}_2 = \frac{69}{3} = 23$

$\bar{x}_3 = \frac{66}{3} = 22$

$\bar{x}_4 = \frac{63}{3} = 21$

grand mean

$$\bar{\bar{X}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4}{4}$$

$$= \frac{20 + 23 + 22 + 21}{4} = 21.5$$

$\bar{X}$	$(\bar{X} - \bar{\bar{X}})$	$(\bar{X} - \bar{\bar{X}})^2$
20	$20 - 21.5 = -1.5$	2.25
23	1.5	2.25
22	0.5	0.25
21	-0.5	0.25
		5

(k) sample size = 4

$k-1 = 3$

$$S_{\bar{X}}^2 = \frac{\Sigma(\bar{X} - \bar{\bar{X}})^2}{k-1}$$

$$= 5/3$$

$\sigma^2 = n S_{\bar{X}}^2$

$n = 3$

$\sigma^2 = S_1^2$ (v.b/w samples) $= 3 \times 5/3 = 5$

S_2^2 (v within samples),

A		B		C		D	
x	$(x-\bar{x})^2$	x	$(x-\bar{x})^2$	x	$(x-\bar{x})^2$	x	$(x-\bar{x})^2$
20	0	25	4	24	4	23	4
19	1	23	0	20	4	20	1
21	1	21	4	22	0	20	1

A

$\bar{x} = 20$, $S_1^2 = \frac{\Sigma(x - \bar{x})^2}{n-1} = \frac{2}{2} = 1$

B $\bar{x} = 23$

$s_2^2 = \frac{8}{2} = 4$

C $\bar{x} = 22$

$s_3^2 = \frac{8}{2} = 4$

D $\bar{x} = 21$

$s_4^2 = \frac{6}{2} = 3$

$\therefore s^2 = \frac{s_1^2 + s_2^2 + s_3^2 + s_4^2}{4} = \frac{1+4+4+3}{4} = \frac{12}{4} = 3$

F ratio $= \frac{V(x) \text{ b/w samples}}{V(x) \text{ within samples}}$

$F = \frac{s_1^2}{s_2^2} = \frac{5}{3} = 1.67$ ~~(from table)~~ ~~1.67~~

from f-table, the table value of $(k-1, n-k) = (4-1), (12-4) = (3, 8)$ DF at 5% level of significant

	3
8	4.07

$F = 4.07$

Computed value of F is < than table value of F, we accept our null hypothesis.

⇒ Analysis V(x) table =

source of variation	sum of □'s (SS)	DF	mean square (MS)	variance ratio (F)
b/w samples	SSC	K-1	MSC = $\frac{SSC}{K-1}$	$F = \frac{MSC}{MSE}$
within samples	SSE	n-K	MSC = $\frac{SSE}{n-K}$	

1) Assume the means of population from which the K-samples are randomly drawn are equal

2) compute mean □'s b/w the samples & mean □s within the sample for computing MSC & MSE,

a) T = some of all the observation

b) SST = some of □'s of all observation $- T^2/n$.

Here T^2/n → correction factor

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

where $\Sigma x_1, \Sigma x_2$ → column total

d) SSE = SST − SSC

e) $MSC = \frac{SSK}{n-k}$

f) $MSE = \frac{SSE}{n-k}$

3) F = MSC/MSE

4) obtain table value of F (k-1) n-1 DF. If the calculated value of F < than table value accept hypothesis that the sample means are equal

P

1) Below are gvn the yield (in kg) per acre for 5 trial plots of 4 varieties of treatmnt.

Treatment.

plat no	1	2	3	4.
1	42	48	68	80
2	50	66	52	94
3	62	68	76	78
4	34	18	64	82
5	52	70.	70	66

Carry out an analysis of variance & state you conclusion?..

A) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$.

$T = 42+50+62+34+52+ \ldots\ldots$
$80+94+78+82+66$

$T = 1300$

Correction factor $= T^2/n$.

$= \frac{(1300)^2}{20}$

$= \frac{1690000}{12}$

$= 84500$

$SST = 42^2+50^2+62^2+34^2+52^2+\ldots\ldots$
$80^2+94^2+78^2+82^2+66^2 -$
84500

$= 4236$

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

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

$= 2580$

$SSE = SST - SSC$

$= 4236 - 2580$

$= 1656$

$MSC = \frac{SSC}{K-1}$

$= \frac{2580}{4-1}$

$= \frac{2580}{3} = 860$

$MSE = \frac{SSE}{n-k}$

$n = 20$
$K = 4$

$= \frac{1656}{20-4}$

$= \frac{1656}{16}$

$= 103.5$

$F = \frac{MSC}{MSE} = \frac{860}{103.5} = 8.3091$

we hue to find table value,

$DF\ (k-1, n-k) = (3, 16)$

The table value of F at 5% level of significant for ordered pairs (3,16) is 3.24

The calculated value is more than the table value of F

∴ The null hypothesis is rejected that is the treatment's do not have same effect.

→ Coding method =

main adv of this method is that big figures are reduced in mag by ÷ or subtract & work is simplified without any disturbance on variance ratio
This method should be used specially when gvn fig are big / otherwise inconvinent

Q1) For the data gvn below,

x_1	x_2	x_3	x_4
20	25	24	23
19	23	20	20
21	21	22	20

carry out analysis of variance technique using coding method.

A) Apply the coding method let us subtract 20 from each observation. then coded data is obtained as

x_1	x_2	x_3	x_4
0	5	4	3
-1	3	0	0
1	1	2	0
0	9	8	3

Sum of all obs^r, $T = 0+(-1)+1+5+3+1+4+0+2+3$
$= 18$

Correction factor, $\frac{T^2}{n} = \frac{(18)^2}{12} = 27$

SST = square of all obs^r − C.factor
$= 0^2+(-1)^2+1^2+5^2+\dots 0^2 - 27$
$= 66 - 27 = 39$

$$SSC = \frac{\Sigma x_1^2}{n_1} + \frac{\Sigma x_2^2}{n_2} + \frac{\Sigma x_3^2}{n_3} + \frac{\Sigma x_4^2}{n_4} - CL$$

$$= \frac{0}{3} + \frac{9^2}{3} + \frac{6^2}{3} + \frac{3^2}{3} = 81 + 36 + 9$$

$$= \frac{126}{3} = 42$$

$$= 42 - CF$$

$$= 42 - 27$$

$$= 15$$

$$SSE = SST - SSC$$

$$= 39 - 15 = 24$$

$$MSC = \frac{SSC}{k-1} = \frac{SSC}{4-1} = \frac{15}{3} = 5$$

$$MSE = \frac{SSE}{n-k} = \frac{24}{12-4} = \frac{24}{8} = 3$$

$$f \text{ ratio} = \frac{MSC}{MSE} = \frac{5}{3}$$

table value, ~~(k-1) (n-k) = 3~~

$(k-1, n-k) = (3, 8)$ at 5% of level of significant is 4.07

cal value of F is < than table value

∴ null hypothesis is accepted

~~⇒ 2-way classification model~~

~~when 2 independent factor~~

Source of variation	Sum of □'s (SS)	DF	mean □ (MS)	variance ratio F.
Btw 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		

⇒ 2-way classification model =

when 2 independent factors have an effect on response variable of interest, it is possible to design the test so that a analysis of variance can be used to test for the effect of 2 factors. such a test → 2 factor analysis of var. or 2 way anova.

Procedure:

1) a) Assume that the mean of all colums are =.

b) choose the hypothesis.

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_c$

c) Assume that mean of all rows are =

d) Choose the hypothesis.

$H_0 : \beta_1 = \beta_2 = \dots = \beta_r$

2) find T (sum of all obs^n)

3) cal SST = sum of □'s of all obs^n $- T^2/n$

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

where $\Sigma x_1, \Sigma x_2 \dots$ are row total

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

where $\Sigma x_1, \Sigma x_2 \dots$ are colum total.

6) $SSE = SST - SSC - SSR$.

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

$MSR = \frac{SSR}{r-1}$

c → no. of colums

r → no. of rows

8) $F_c = \frac{MSC}{MSE}$ (column) & $F_r = \frac{MSR}{MSE}$ (row)

Df for $F_c = [(c-1), (c-1)(r-1)]$.

Df for $F_r = [(r-1), (c-1)(r-1)]$.

9) Determine table value of F.

If F_c < table value → accept, $H_0: \alpha_1 = \alpha_2 = \dots \alpha_c$.

If F_r < table value → ~~reject~~ accept

$H_0: \beta_1 = \beta_2 = \dots \beta_r$

Analysis of table of V(?) =

Source of variation	Sum of □'s (SS)	DF	mean □'s (MS)	varience ratio F
B/w col	SSC	$c-1$	MSC	$F_c = MSC/MSE$
B/w rows	SSR	$r-1$	MSR	$F_r = MSR/MSE$
residual	SSE	$(c-1)(r-1)$	MSE	
Total	SST	$n-1$		

Q

1) follow represents the no of units of production per day turn out by 4 diffrnt workers using 5 diff. ~~stypes~~ types of machine

machine types

workers	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 can it be conclude that

a) mean productivity is the same for diffrnt machines.

b) mean productivity is diffrnt w.r.t diffrnt workers.

A) let us take the hypothesis that mean productivity of diffrnt machines is same

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$ (5 machines A,B,C,D,E)

mean productivity of diffrnt ~~machines~~ workers is same

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$ (workers - 4 1,2,3,4)

$T = 112$.

$CF \to \frac{T^2}{n} = \frac{112^2}{20} = 627.2$

$SST = 16 + \cancel{25} + 36 + 49 + 9 + 25 + 64 + 36 + 25 + 9 + 36 + 49 + 16 + 49 + 25 + 64 + 64 + 36 + 16 + 64 + 4 - CF$

$= 692 - 627.2$

$= 64.8$

(row) $SSR = \frac{25^2}{5} + \frac{29^2}{5} + \frac{36^2}{5} + \frac{22^2}{5} - CF$

$= \frac{625 + 841 + 1296 + 484}{5} - CF$

$= \frac{3246}{5} - CF = 649.2 - 627.2$

$= 22$

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

$= \frac{400 + 576 + 400 + 784 + 400}{4} - CF$

$= 640 - 627.2 = \cancel{115.2}\ 12.8$

$SSE = SST - SSC - SSR$

$= 64.8 - 12.8 - 22 = \underline{30}$

$c \rightarrow$ column $\downarrow$ 5.

$MSC = \frac{SSC}{c-1} = \frac{12.8}{5-1}$

$= \frac{12.8}{4} = \underline{3.2}$

$r \rightarrow$ row $\downarrow$ 4.

$MSR = \frac{SSR}{r-1} = \frac{22}{3}$

$= \underline{7.33}$

$MSE = \frac{SSE}{(c-1)(r-1)} = \frac{30}{4 \cdot 3} = \frac{30}{12} = \underline{2.5}$

$f_c = \frac{MSC}{MSE} = \frac{3.2}{2.5} = \underline{1.28}$

$f_r = \frac{MSR}{MSE} = \frac{7.33}{2.5} = \underline{2.93}$

table value of F,

Df for f_c = ~~$(c-1)(r-1) = 4, 3$~~

$\hookrightarrow [(c-1), (c-1)(r-1)] = (4, 12)$

$[(r-1), (c-1)(r-1)] = (3, 12)$

table value of f_c = 3.26.

< than cal. value → accepted

no significant diff on productivity of 5 machine

table value of f_r = 3.49 → accepted

hence no s. difference in mean productivity of different workers.