Module 6: Hypothesis Testing II

Small sample tests- Student's t-test, F-test- chi-square test- goodness of fit – independence of attributes- Design of Experiments - Analysis of variance – One way-Two way-Three way classifications - CRD-RBD-LSD.

Test of Hypothesis – Small Sample

Small Sample (i.e sample size is $<= 30$) test is done by using **F-test** and **t-test**.

F-Test

To test the significant difference btn population variances.

$$\boxed{F = \frac{\text{Greater variance}}{\text{Smaller variance}}}$$

Eg: $F = \frac{\sigma_1^2}{\sigma_2^2}$ (if $\sigma_1^2 > \sigma_2^2$)

or

$F = \frac{\sigma_2^2}{\sigma_1^2}$ (if $\sigma_2^2 > \sigma_1^2$).

If σ_1^2 and σ_2^2 are unknown

Thus $\underbrace{\sigma_1^2}_{\text{Population variance}} = \dfrac{n_1 \overbrace{s_1^2}^{\text{Sample variance}}}{n_1 - 1}$ and $\underbrace{\sigma_2^2}_{\text{Population variance.}} = \dfrac{n_2 s_2^2}{n_2 - 1}$

Degree of freedom, $\nu = (n_1 - 1, n_2 - 1)$ if $\sigma_1^2 > \sigma_2^2$

and $\nu = (n_2 - 1, n_1 - 1)$ if $\sigma_2^2 > \sigma_1^2$.

Problems

1. A sample of size 13 gave an essential population variance of 3.0, while another sample of size 15 gave an estimate of 2.5. Could both samples be from population with the same variance?

Prob:-

① Given:

Sample 1	Sample 2
$n_1 = 13$	$n_2 = 15$
$\sigma_1^2 = 3$	$\sigma_2^2 = 2.5$

Step 1: $H_0: \sigma_1^2 = \sigma_2^2$ (Same population).

$H_1: \sigma_1^2 \neq \sigma_2^2$ (Different population).

Step 2: LOS, $\boxed{\alpha = 5\%}$ and $\gamma = (n_1 - 1, n_2 - 1)$

$\boxed{\gamma = (12, 14)}$

$\boxed{\text{Since } \sigma_1^2 > \sigma_2^2}$

Step 3: Test statistic, $F = \dfrac{\sigma_1^2}{\sigma_2^2} = \dfrac{3}{2.5} \Rightarrow \boxed{F = 1.2}$

Step 4: Table value

F_α at $\nu = (12, 14) = F_{5\%}$ at $\nu = (12, 14)$

$= 2.53$

Step 5: Conclusion

Here, $F < F_{5\%}$ at $\nu = (12, 14)$

$\therefore$ H_0 is accepted at 5% L.o.S.

2. Two random samples gave the following data:

	Size	Mean	Variance
Sample I	**8**	**9.6**	**1.2**
Sample II	**11**	**16.5**	**2.5**

Can we conclude that the two samples have been drawn from the same normal population?

(2)

	Size	Mean	Variance
Sample 1	8	9.6	1.2
Sample 2	11	16.5	2.5

$x_1 \rightarrow$ Sample 1

$x_2 \rightarrow$ Sample 2.

Given:

$n_1 = 8$ | $n_2 = 11$

$\bar{x}_1 = 9.6$ | $\bar{x}_2 = 16.5$

$s_1^2 = 1.2$ | $s_2^2 = 2.5$

First we need to find σ_1^2 and σ_2^2.

W.K.T, $\sigma_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{(8)(1.2)}{7} \Rightarrow \boxed{\sigma_1^2 = 1.37}$

$$\sigma_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{(11)(2.5)}{10}$$

$$\Rightarrow \boxed{\sigma_2^2 = 2.75}$$

Here, $\sigma_2^2 > \sigma_1^2$.

<u>Step 1:</u> $H_0: \sigma_1^2 = \sigma_2^2$ (Same population).

$H_1: \sigma_1^2 \neq \sigma_2^2$ (Different population).

Step 2: LOS, $\boxed{\alpha = 5\%}$ and $\nu = (n_2-1, n_1-1)$ $[\because \sigma_2^2 > \sigma_1^2]$

$\boxed{\nu = (10, 7)}$

Step 3: Test statistic,

$$F = \frac{\sigma_2^2}{\sigma_1^2} = \frac{2.75}{1.37} = 2.007$$

$$\Rightarrow \boxed{F = 2.007}$$

Step 4: Table value:

$$F_\alpha \text{ at } \nu = (10, 7) = F_{5\%} \text{ at } \nu = (10, 7)$$

$$= 3.64.$$

Step 5:

$$F < F_\alpha \text{ at } \nu = (10, 7)$$

$\therefore H_0$ is accepted at 5% l.o.s.

3. The nicotine contents in two random samples of tobacco are given below:

Sample I	21	24	25	26	27	
Sample II	22	27	28	30	31	36

Can we conclude that the two samples have been drawn from the same population?

(3)

Sample 1	21	24	25	26	27	
Sample 2	22	27	28	30	31	36.

Given: $x_1 \to$ Sample 1 $\Rightarrow n_1 = 5$

$x_2 \to$ Sample 2 $\Rightarrow n_2 = 6$

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{123}{5} \Rightarrow \boxed{\bar{x}_1 = 24.6}$$

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{174}{6} \Rightarrow \boxed{\bar{x}_2 = 29}$$

$$s_1^2 = \frac{\sum x_1^2}{n_1} - (\bar{x}_1)^2 \Rightarrow \boxed{s_1^2 = 4.24}$$

$$s_2^2 = \frac{\sum x_2^2}{n_2} - (\bar{x}_2)^2 \Rightarrow \boxed{s_2^2 = 18}$$

Now,

$$\sigma_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{(5)(4.24)}{4} \Rightarrow \boxed{\sigma_1^2 = 5.3}$$

$$\sigma_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{(6)(18)}{5} \Rightarrow \boxed{\sigma_2^2 = 21.6}$$

Here, $\sigma_2^2 > \sigma_1^2$

Step 1: $H_0: \sigma_1^2 = \sigma_2^2$ (Same popl.)

$H_1: \sigma_1^2 \neq \sigma_2^2$ (Different popl).

Step 2: LOS, $\boxed{\alpha = 5\%}$ and $\nu = (n_2 - 1, n_1 - 1)$

Step 3: Test Statistic,

$$F = \frac{\sigma_2^2}{\sigma_1^2} = \frac{21.6}{5.3}$$

$$\Rightarrow \boxed{F = 4.07}$$

Step 4: Table value

$F_{5\%}$ at $\nu = (5, 4) = 6.26$.

Step 5: $F < F_{5\%}$ at $\nu = (5, 4)$

$\therefore H_0$ is accepted at 5% L.O.S.

(4)

Sample	Size	Sample mean	Sum of squares of deviation from the mean
1	10	15	90
2	12	14	108

Sol:-

Sample 1	Sample 2
$n_1 = 10$	$n_2 = 12$
$\bar{x}_1 = 15$	$\bar{x}_2 = 14$
$\sum(x_1 - \bar{x}_1) = 90$	$\sum(x_2 - \bar{x}_2) = 108$

$$s_1^2 = \frac{\sum(x_1-\bar{x}_1)^2}{n_1-1} = \frac{90}{9} = 10$$

$$s_2^2 = \frac{\sum(x_2-\bar{x}_2)^2}{n_2-1} = \frac{108}{11} = 9.82$$

$\longrightarrow$ Solve it!

t-test

<u>There are three types under t-test :</u>

- Single mean

(1) Student t- distribution

To test the significant difference btn sample mean and population mean.

* Test statistic, $\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}}$ → Single mean

* Degree of freedom = $n-1$

(i.e) $\boxed{\nu = n-1}$

Critical Value for T-test and F-test

- **T-test:**

$t_{\alpha}\ (at\ d.o.f\ n-1)$

$t_{\frac{\alpha}{2}}\ (at\ d.o.f\ n-1)$

Where, $d.o.f$ is the degree of freedom

Problems

1. A machinist is expected to make engine parts with axle diameter of 1.75cm. A random sample of 10 parts shows a mean diameter 1.85cm with a S.D. of 0.1cm. On the basis of this sample, would you say that the work of the machinist is inferior?

① Solution:

Given: $\bar{x} = 1.85$ and $n = 10$ and $s = 0.1$

$\mu = 1.75$

Step 1:

H_0: The work of machinist is not inferior

(i.e) $\boxed{\bar{x} = \mu}$

H_1: The work of machinist is inferior.

(i.e.) $\boxed{\bar{x} \neq \mu}$ (Two-tailed test)

Step 2: LOS, $\alpha = 5\%$ and $\nu = n-1 = 9$.

Step 3: Test statistics, $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

$$t = \frac{1.85 - 1.75}{0.1/\sqrt{10}}$$

Step 4: Table value of 't'

t_α at d.o.f 9

$= t_{5\%}$ at d.o.f $9 = t_{2.5\%}$ at d.o.f 9 (∵ two-tailed test)

$= 2.26.$

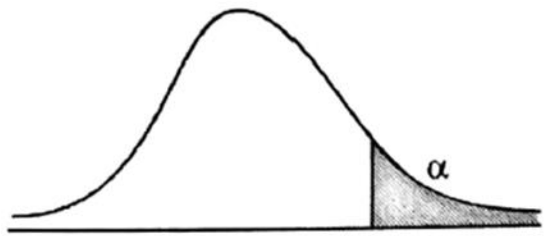
Given:

$\alpha = 2.5\%$

$\boxed{\alpha = 0.025}$

$\boxed{\nu = 9} \rightarrow d.f.$

Table 9: Values of "Students" *t* Distribution



d.f. ↓ / α →	0.1	0.05	0.025 $= \alpha$	0.01	0.005	0.001	0.0005	*d.f.*
1	3.078	6.314	12.706	31.821	63.657	318.300	637.000	1
2	1.886	2.920	4.303	6.965	9.925	22.330	31.600	2
3	1.638	2.353	3.182	4.541	5.841	10.210	12.920	3
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	4
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	5
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	6
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	8
9 $= \nu$	1.383	1.833	$\boxed{2.262}$	2.821	3.250	4.297	4.781	9

$t_{2.5\%}(\text{at } \nu = 9) = 2.262$

Step 5: Conclusion

Here, $|t| = 3 > t_{2.5\%}$ at d.o.f 9

$|t| > t_{2.5}(\nu = 9)$

$\therefore$ H_0 is rejected at $\alpha = 5\%$.

x

2. A certain injection administered to each of 12 patients resulted in the following increases of blood pressure: 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the injection will be, in general, accompanied by an increase in B.P.?

(2) Solution:

x	5	2	8	-1	3	0	6	-2	1	5	0	4

Given:- $n = 12$, $\bar{x} = \frac{\sum x}{n} = \frac{31}{12} = 2.58$

and $s = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2} \Rightarrow s = 2.96$

Step 1:- $H_0: \bar{x} - \mu = 0 \Rightarrow \boxed{\bar{x} = \mu} \Rightarrow \boxed{\mu = 0}$ (No change in B.P)

$H_1: \bar{x} > \mu$ (Increase in B.P) one-tailed test

Step 2:- LOS, $\boxed{\alpha = 5\%}$ and $\nu = n-1 \Rightarrow \boxed{\nu = 11}$

Step 3:- Test statistic, $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{2.58 - 0}{2.96/\sqrt{11}}$

Step 4: Table value of 't'

t_α at d.o.f 11 = $t_{5\%}$ at d.o.f 11 [$\because$ one tail test]

=1.8

Step 5:
Conclusion

Here $|t| = 2.89 > t_{5\%}$ at d.o.f $11 = 1.8$

(i.e) $|t| > t_{5\%}$ $(\nu = 11)$

$\therefore$ H_0 is rejected at $\alpha = \cancel{10\%}$ 5%.

3. The mean lifetime of a sample of 25 bulbs is found as 1550 hours with a S.D. of 120 hours. The company manufacturing the bulbs claims that the average life of their bulbs is more than 1600 hours. Is the claim acceptable at 5% level of significance?

③ Solution:

Given:- $n = 25$, $\bar{x} = 1550$, $s = 120$

$\mu = 1600$

Step 1: H_0: claim is not acceptable (i.e.) xbar <= mu

H_1: claim is acceptable (i.e.) xbar > mu (one-tailed test)

Step 2: LOS $\boxed{\alpha = 5\%}$, $\boxed{\nu = 24}$

Step 3:- Test statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1550 - 1600}{120/\sqrt{24}}$$

Step 4:- Table value of 't'

$$t_\alpha(\nu = 24) = t_{5\%}(\nu = 24) \text{ [one-tailed]}$$

$$t_\alpha = 1.71$$

Step 5:-

Conclusion

$|t| = 2.04.$

$t_{5\%}(\nu = 24) = 1.71$

$\therefore |t| > t_{5\%}(\nu = 24)$

H_0 is rejected at $\alpha = 5\%$.

Chi-Square Test

Parametric test:

If the information about population is completely known by its parameters, then the test is called parametric test.

Eg: Z-test, t-test and F-test.

Chi-square test:

It is a non-parametric test used to find out the attributes are associated (or) not.

Note: It is not a small sample test.

Applications:

1) To test if there is any association between attributes. (i.e) To test the independence of attributes.

2). To test the "goodness of fit".

S.No.	Type of $\chi^2 - test$	Degree of Freedom
1.	Independence (Using Contingency table) $r - no. of\ rows$ $c - no. of\ columns$	$(r-1)(c-1)$
2.	Goodness of fit – Binomial Distribution	$n-2$
3.	Goodness of fit – Poisson Distribution	$n-2$

χ^2-test for independence of attributes:

Attributes is a characteristic (or) a quality which may be present among the members of a population.

Eg: Beauty, heathy, good etc.,

Contigency table:

Let A and B be two attributes.

Table values corresponds to observed frequencies

B \ A	A_1	A_2	Total
B_1	(A_1B_1)	(A_2B_1)	(B_1)
B_2	(A_1B_2)	(A_2B_2)	(B_2)
Total	(A_1)	(A_2)	N

(B_1) = (A_1B_1)-value + (A_2B_1)-value.

(A_1) = (A_1B_1) value + (A_1B_2) value.

N = Total value in the cells.

Here, $\sum_i (A_i) = \sum_i (B_i) = N$.

We are testing whether A and B are independent (or) not.

Procedure:

1) H_0: Attributes A and B are independent

H_1: " " " " are not independent.

2) Choose, α.

Degree of freedom, $\gamma = (r-1)(c-1)$

Ci

Row size

Column size.

3) Find expected frequency to each cell,

$$E = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}.$$

4) $\chi^2 = \boxed{\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$

5) **Table Value:**

$\chi^2{}_{\alpha}$ at $d.f\ (\gamma = (r-1)(c-1))$

6) Conclusion:

$\chi^2 < \chi^2_{\alpha} \Rightarrow$ Accept H_0.

$\chi^2 > \chi^2_{\alpha} \Rightarrow$ Reject H_0.

Problem

1. A total number of 3759 individuals were interviewed in a public opinion survey on a political proposal and we have the following data

	Favoured	Opposed	Undecided
Men	1154	475	243
Women	1103	442	342

Do you justify or contradict the hypothesis that there is no association between gender and attitude?

Solution:

	Favoured	Opposed	Undecided	Row total
Men	1154	475	243	1872
Women	1103	442	342	1887
Column total	2257	917	585	3759

3759 → Total no. of people.

Step 1: H_0: Gender and attitude are independent.

H_1: Gender and attitude are not independent.

Step 2: L.O.S, $\boxed{\alpha = 5\%}$ and $\nu = (r-1)(c-1)$

$\nu = (2-1)(3-1) = (1)(2)$

$\boxed{\nu = 2}$

Step 3: Test statistic

O	E (rounded)	$(O-E)^2/E$
1154	$\frac{1872 \times 2257}{3759} = 1124$	0.80
475	$\frac{1872 \times 917}{3759} = 457$	0.71
243	$\frac{1872 \times 585}{3759} = 291$	7.92
1103	$\frac{1887 \times 2257}{3759} = 1153$	0.79
442	$\frac{1887 \times 917}{3759} = 460$	0.70
342	$\frac{1887 \times 585}{3759} = 294$	7.84

$$\boxed{E = \frac{\text{row total} \times \text{column total}}{\text{Grand total.}}}$$

$$\chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right] = 0.80 + 0.71 + 7.92 + 0.79 + 0.70 + 7.84.$$

$$\boxed{\chi^2 = 18.76}$$

Step 4: Table value.

$\chi^2_{5\%}$ at $\nu = (2) = 5.99.$

Step 5: Conclusion

$\chi^2 = 18.76 > \chi^2_{5\%}$ at $\nu = 2.$

$\therefore$ H_0 is rejected at 5% l.o.s.

Goodness of Fit

Binomial distribution

$$P(X = x) = nC_x \, p^x q^{n-x}$$

Expected frequency : $E = N \times (q+p)^n$

Expected frequencies are given by the terms of the binomial expansion.

Poisson distribution

$$P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, 2, \cdots,$$

where, $\lambda = \frac{\Sigma f x}{N}$ = Mean.

Expected frequency: ~~E(x)~~

$$E = N \times [P(X=x)].$$

Expected frequencies are given by the each value of x in E.

Conditions for the validity of χ^2-test:

(i) The sample observations should be independent.

(ii) The sample size, $n \geq 50$.

(iii) The theoretical cell frequency should be at least 10

(i.e) Expected frequency $\geq$ 10

(iv) The constraints on the cell frequencies should be linear. $\boxed{\text{Eg: } \sum O_i = \sum E_i}$

$O_i \rightarrow$ Observed frequency

$E_i \rightarrow$ Expected frequency.

S.No.	Type of $\chi^2 - test$	Degree of Freedom
1.	Independence (Using Contingency table) $r - no.of\ rows$ $c - no.of\ columns$	$(r-1)(c-1)$
2.	Goodness of fit – Binomial Distribution	$n-2$
3.	Goodness of fit – Poisson Distribution	$n-2$

Problem

1. Fit a Poisson distribution for the following data and also test the goodness of fit.

x	0	1	2	3	4	5
f	142	156	69	27	5	1

Solution:

x	0	1	2	3	4	5
O.F	142	156	69	27	5	1

Step 1: H_0 : Poisson distb is a good fit.

H_1 : Poisson distb is not a good fit

Step 2: L.O.S, $\boxed{\alpha = 5\%}$ and ν = yet to find.

Eg 3:

x	0	1	2	3	4	5
f	142	156	69	27	5	1
xf	0	156	138	81	20	5

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{400}{400} \Rightarrow \boxed{\bar{x} = 1}$$

Expected frequency,

$$\boxed{E = N \cdot [P(X=x)]}$$

Here $N = 400$.

$P(X=x) \rightarrow$ Poiss dist for $x = 0, 1, 2, 3, 4, 5$.

Here, $\lambda = \text{mean} \Rightarrow \boxed{\lambda = 1}$

$$P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

x	$E = N[P(X=x)]$ (Rounded)
0	$400\left[\frac{e^{-1}\cdot(1)^0}{0!}\right] = 147$
1	$400\left[\frac{e^{-1}(1)^1}{1!}\right] = 147$
2	$400\left[\frac{e^{-1}\cdot(1)^2}{2!}\right] = 74$
3	$400\left[\frac{e^{-1}(1)^3}{3!}\right] = 25$
4	$400\left[\frac{e^{-1}(1)^4}{4!}\right] = 6$
5	$400\left[\frac{e^{-1}(1)^5}{5!}\right] = 1$

O	142	156	69	27	5	1
E	147	147	74	25	6	1

< 10.

O	142	156	69	33
E	147	147	74	32

Combined last three column so that the value is > 10.

Now, degree of freedom

$$\nu = n - 2 = 4 - 2 \Rightarrow \boxed{\nu = 2}$$

$$\chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$= \frac{5^2}{147} + \frac{9^2}{147} + \frac{5^2}{74} + \frac{1^2}{32}$$

$$\boxed{\chi^2 = 1.09}$$

Step 4: Table value.

$\chi^2_{5\%}$ at $(\nu=2) = 5.99$.

Step 5:

$\chi^2 < \chi^2_{5\%}$ at $(\nu=2)$

$\therefore H_0$ is accepted at 5% l.o.s.

x

Design of Experiments

The sequence of steps taken to ensure a scientific analysis leading to a valid inferences about the hypothesis is called "design of experiment".

Origin: Agricultural research
— R.A. Fisher.

Eg: In any situation (or) problems there are parameters (or) variables involved.
↳ Directly
→ Indirectly.

Agricultural experiment:

A particular manure will increase the yield of paddy (or) not.

Variables → Experimental – Quantity of manure and yield.

Variables → Extraneous – Feritility of soil, rainfall etc.

Objective:

1) To control the extraneous variables.

2) Then the results (yields) could be attributed only to the experimental variables.

Factors: Independent experimental variables.

- Quantitative: Takes a real number as value.
- Qualitative: ~~where~~ The case in which a real number can't be assigned.

Basic principles of design of experiment:

1) Randomization.
2) Replication.
3) Local control.

Basic designs of experiments:

1) CRD : Completely randomised design
2) RBD : Randomised block design
3) LSD : Latin square design.

ANOVA - Analysis of Variance.

*) To test the significance of the difference among more than two sample means.

Eg: If we want to compare the mileage achieved by five different brands of petrol, we can use ANOVA.

(H_0) **Null-hypothesis:** There is no significante difference between means of the population.

Test statistic: $\boxed{F\text{-test}}$

→ Used to test any significant difference b/w the two variances exist (or) not.

Types of ANOVA:

1) **CRD:** One-way classification (or) one factor ANOVA.

2) **RBD:** Two-way classification (or) two factor ANOVA.

3) **LSD:** Three factor ANOVA.

Following assumptions are made inorder to make use of ANOVA

*) The samples are drawn from normal populations.

*) The samples are independently drawn from the population.

*) All the populations will have the same variance.

Completely Randomized Design

CRD – ANOVA

— one factor ANOVA.

① Compute: Q, Q_1, and Q_2

② $Q = \sum_j \sum_i (x_{ij}^2) - \frac{T^2}{N}$

③ $Q_1 = \sum_i \left(\frac{T_i^2}{n_i}\right) - \frac{T^2}{N}$

④ $Q_2 = Q - Q_1$

Next, we need to form the ANOVA-table

Where,
'T' represent the total value of observation.
'N' represent the total no.of observation.

ANOVA table: (CRD).

S.V.	S.S	d.f.	M.S	F-test
Between classes	Q_1	$h-1$	$MSB = \frac{Q_1}{h-1}$	$F = \frac{\text{Greater variance}}{\text{Smaller variance}}$
		h is the no.of classes		$F = \frac{MSB}{MSW}$ if $MSB > MSW$. (or)
Within Classes	Q_2	$N-h$	$MSW = \frac{Q_2}{N-h}$	$F = \frac{MSW}{MSB}$ if $MSW > MSB$.
Total	Q	$N-1$	—	

F-test: (From the table values).

* $F_{5\%}$ (at d.f $= (h-1, N-h)$) if $MSB > MSW$.

(or)

$F_{5\%}$ (at d.f $= (N-h, h-1)$) if $MSW > MSB$.

Conclusion for CRD

(I) If $F < F_{5\%}$ then accept H_0.

$\boxed{\text{Hypothesis}}$:-

H_0: The classes (treatments) do not differ significantly.

H_1: The classes (treatments) differ significantly.

Problem:

Question :-

A car rental agency, which uses 5 different brands of tyres in the process of deciding the brand of tyre to purchase as standard equipment for its fleet, finds that each of 5 tyres of each brand last the following number of kilometers (in thousands):

Tyre brands

A	B	C	D	E
36	46	35	45	41
37	39	42	36	39
42	35	37	39	37
38	37	43	35	35
47	43	38	32	38

Test the hypothesis that the five tyre brands have almost the same average life.

Solution :-

Step ① H_0 : five tyre brands donot differ significantly in their lives.

H_1 : five tyre brands differ significantly in their lives.

we shift the origin to 40 and workout with the new values of x_{ij}.

Step ②:-

Tyre brand	x_{ij}					$T_{i\cdot}$	$n_{i\cdot}$	$\frac{T_{i\cdot}^2}{n_{i\cdot}}$	$\sum_{j=1}^{5} x_{ij}^2$
A	-4	-3	2	-2	7	0	5	0	82
B	6	-1	-5	-3	3	0	5	0	80
C	-5	2	-3	3	-2	-5	5	5	51
D	5	-4	-1	-5	-8	-13	5	33.8	131
E	1	-1	-3	-5	-2	-10	5	20	40
			Total			-28	25	58.8	384

$$T = \sum_i T_{i\cdot} = -28$$

$$\sum\sum x_{ij\cdot}^2 = \sum_i \left(\sum_j x_{ij\cdot}^2 \right) = 384$$

$$Q = \sum\sum x_{ij\cdot}^2 - \frac{T^2}{N} = 384 - \frac{(-28)^2}{25}$$

$$= 352.64$$

$$Q_1 = \sum \frac{T_{i\cdot}^2}{n_i} - \frac{T^2}{N} = 58.8 - 31.36$$

$$= 27.44$$

$$Q_2 = Q - Q_1 = 352.64 - 27.44$$

$$= 325.20$$

Step (3) ANOVA Table

S.V	S.S	d.f	M.S	F-test
Between tyre brands	$Q_1 = 27.44$	$h-1$ $= 5-1 = 4$	$MSB = \frac{Q_1}{h-1} = 6.86$	$F = \frac{MSB}{MSW}$ If $MSB > MSW$ (or)
within tyre brands	$Q_2 = 325.20$	$N-h$ $= 25-5$ $= 20$	$MSW = \frac{Q_2}{N-h}$ $= 16.26$	$F = \frac{MSW}{MSB}$ If $MSW > MSB$ Now, $F = \frac{16.26}{6.86}$ $= 2.37$
	$Q = 352.64$	$N-1$ $= 24$		

Step ④

From the F-tables,

Degree of freedom at 5 %

(ie) $F_{5\%} (V_1 = 20, V_2 = 4) = 5.80$

Here, $F_{cal} < F_{5\%}$

$2.37 < 5.80$.

Step ⑤

Hence, H_0 is accepted (five tyre brands have almost the same average life is accepted).

(ie) The five tyre brands do not differ significantly in their lives

Practice Problems

Problem 1: A completely randomized design experiment with 10 plots and 3 treatments gave the following results. Analyse the results for treatment effects.

Plot.no.	1	2	3	4	5	6	7	8	9	10
Treatment	A	B	C	A	C	C	A	B	A	B
Yield	5	4	3	7	5	1	3	4	1	7

Problem 2: In order to determine whether there is significant difference in the durability of 3 brands of computers, sample of size 5 are selected from each brand and the frequency of repair during the first year of purchase is observed. The results are given in the following table. In the view of above data, what conclusion can we draw?

	Brands		
	A	**B**	**C**
Frequency of repair	5	8	7
	6	10	3
	8	11	5
	9	12	4
	7	4	1

Randomized Block Design

RBD – ANOVA

– Two factor ANOVA.

① Compute: Q, Q_1, Q_2 and Q_3.

② $$Q = \sum_j \sum_i (x_{ij}^2) - \frac{T^2}{N}.$$

③ $$Q_1 = \frac{1}{c}\left(\sum_i T_i^2\right) - \frac{T^2}{N}.$$

④ $$Q_2 = \frac{1}{r}\left(\sum_j T_j^2\right) - \frac{T^2}{N}.$$

⑤ $$Q_3 = Q - Q_1 - Q_2.$$

Where,
'c' represent the total no.of columns.
'r' represent the total no.of rows.

∴ Next, we need to form the ANOVA-table.

Where,
'T' represent the total value of observation.
'N' represent the total no.of observation.

ANOVA table (RBD)

S.V	S.S	d.f	M.S	F-test
Between rows	Q_1	$r-1$	$MSR = \frac{Q_1}{r-1}$	$F^R = \frac{MSR}{MSE}$ → $MSR > MSE$ (or) $\frac{MSE}{MSR}$ → $MSE > MSR$.
Between columns	Q_2	$c-1$	$MSC = \frac{Q_2}{c-1}$	$F^C = \frac{MSC}{MSE}$ → $MSC > MSE$. (or) $\frac{MSE}{MSC}$ → $MSE > MSC$.
Residual (Error)	Q_3	$(c-1)(r-1)$	$MSE = \frac{Q_3}{(c-1)(r-1)}$	
Total	Q	$rc-1$		

F-test (From the table values):

(*) $F^R_{5\%}$ (at d.f $= (r-1, (c-1)(r-1))$) if $MSR > MSE$

(or)

$F^R_{5\%}$ (at d.f $= ((c-1)(r-1), r-1)$) if $MSE > MSR$.

(*) $F^C_{5\%}$ (at d.f $= (c-1, (c-1)(r-1))$) if $MSC > MSE$.

(or)

$F^C_{5\%}$ (at d.f $= ((c-1)(r-1), c-1)$) if $MSE > MSC$.

Remember:

$$F = \frac{\text{Greater variance}}{\text{Smaller variance}}$$

Conclusion for RBD.

(R) If $F^R < F^R_{5\%}$ then accept H_0 for the row.

(R) If $F^C < F^C_{5\%}$ then accept H_0 for the column.

Hypothesis:

Rows:

H_0: The difference between the rows is not significant.

H_1: The difference between the rows is significant.

Column

H_0: The difference between the columns is not significant.

H_1: The difference between the columns is significant.

RBD

(1) Row → H_0 is accepted
Column → H_0 is rejected.

(2) Row → H_0 is rejected
Column → H_0 is accepted.

(3) Row → H_0 is reject / accepted.
Column → H_0 is reject / accepted

Problem: Four doctors each test 4 treatments for a certain disease and observe the number of days each patient takes to recover. The results are displayed in the following table (recovery time in days). Discuss the difference between doctors and treatments.

	Treatments				
Doctors		A	B	C	D
	I	10	14	19	20
	II	11	15	17	21
	III	9	12	16	19
	IV	8	13	17	20

Sol:-

Step 1:- $\boxed{\text{Doctors}}$

H_0: The difference between the doctors is not significant.

H_1: The difference between the doctors is significant.

$\boxed{\text{Treatements}}$

H_0: The difference between the treatments is not significant.

H_1: The difference between the treatments is significant.

Sol:-

We subtract 15 from the given values, so that it is easy to compute the parameters.

Doctor	Treatments 1	2	3	4	T_i	T_i^2/c	$\sum_j x_{ij}^2$
A	-5	-1	4	5	3	2.25	67
B	-4	0	2	6	4	4	56
C	-6	-3	1	4	-4	4	62
D	-7	-2	2	5	-2	1	82
T_j	-22	-6	9	20	$\boxed{T=1}$	$\sum(T_i^2/c) = 11.25$	267
T_j^2/r	121	9	20.25	100	$\sum(T_j^2/r) = 250.25$		
$\sum_i x_{ij}^2$	126	14	25	102	267		

$T = \sum_i (T_i)$.

$$Q = \sum_j \sum_i (x_{ij}^2) - \frac{T^2}{N} = 267 - \frac{1}{16} = 266.94$$

$$Q_1 = \frac{1}{c} \sum (T_i^2) - \frac{T^2}{N} = 11.25 - 0.0625 = 11.19$$

$$Q_2 = \frac{1}{r} \sum (T_i^2) - \frac{T^2}{N} = 250.25 - 0.0625 = 250.19$$

$$Q_3 = Q - Q_1 - Q_2 = 266.94 - 11.19 - 250.19$$

$$\boxed{Q_3 = 5.56}$$

Step 8: Anova table.

S.V.	S.S.	d.f	M.S.	F-test
Between rows (Doctors)	$Q_1 = 11.19$	$r-1=3$	$MSR = \frac{Q_1}{r-1} = 3.73$	$F^R = \frac{MSR}{MSE} = 6.02$
Between Columns (Treatments)	$Q_2 = 250.19$	$C-1=3$	$MSC = \frac{Q_2}{C-1} = 83.40$	$F^C = \frac{MSC}{MSE} = 134.52$
Residual	$Q_3 = 5.56$	$(r-1)(C-1) = 9.$	$MSE = \frac{Q_3}{(C-1)(r-1)} = 0.62$	
Total	$Q = 266.94$	$rC-1 = 15$	–	–

Step 4:- Table value.

$F^{R}_{5\%}$ (at d.f $= (r-1, (c-1)(r-1))$

$= F^{R}_{5\%}$ (at $\gamma = (3, 9)) = 3.86$.

$F^{C}_{5\%}$ (at d.f $= (c-1, (c-1)(r-1))$

$= F^{C}_{5\%}$ (at $\gamma = (3, 9)) = 3.86$.

Step 5:- $\boxed{\text{Conclusion}}$

Since $F^{R} > F^{R}_{5\%}$ and $F^{C} > F^{C}_{5\%}$

we reject H_0 in doctors case

and also reject H_0 in treatment

case.

Practice Problems

Problem 1: The following data represent the number of units of production per day turned out by 5 different workers using 4 different types of machines. ***(i)*** Test whether the 5 workers differ with respect to mean productivity and ***(ii)*** whether the mean productivity is the same for the four different machines.

		Machine Types			
		A	**B**	**C**	**D**
Workers	**I**	44	38	47	36
	II	46	40	52	43
	III	34	36	44	32
	IV	43	38	46	33
	V	38	42	49	39

Problem 2: The following data represents the final grades obtained by five students in Mathematics, English, Physics and Chemistry. Test the hypothesis that the courses are of equal difficulty using the P-value in your conclusions and discuss your findings.

	Subjects			
	Mathematics	**English**	**Physics**	**Chemistry**
Students	68	57	73	61
	83	94	91	86
	72	81	63	59
	55	73	77	66
	92	68	75	87

The End