TRIBHUVAN UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY
AMRIT SCIENCE CAMPUS

Statistics - II

Lab Report

SUBMITTED BY:
Name: Sasank Lama
Roll: 13/079
Date: 2024/05/16

SUBMITTED TO:
Om Narayan Pradhan
Department of Statistics

Internal Teacher's Signature

External's Signature

TABLE OF CONTENTS

Unit 1: Sampling Distribution and Estimation, Parametric test

Experiment 1 (Calculation of confidence intervals)

Find 99% and 95% confidence interval for mean for the given set of data age of patients recently visited to a Government Hospital. Age in years:
44,42,25,46,52,21,23,33,44,12,21,22,27,29,81,35,34,36,37,71,48,10,11,20,29,30,40,44,40,50,54,45,52,61,44,54,56,57,58,69,70,60,70,30,38,48,47,69,50,33,24,45,45,45,45,54,57,74,51,23,34,42,33,41,20,28,27

Working formula:

(100-α) % C.I. for mean(μ) = sample mean ($\bar{x}$) ± $Z_{\alpha/2}$ /S.E. ($\bar{x}$)

Where, S.E. ($\bar{x}$) = σ/√n,

for 95% C.I. $Z_{\alpha/2}$ = 1.96

for 99% C.I., $Z_{\alpha/2}$ = 2.58

Computation:

Age in Years (x)	x^2	Measurements	Formulae	Values
44	1936	n	COUNTA(A2:A69)	67
42	1764	$\bar{x}$	AVERAGE(A2:A69)	41.94029851
25	625	σ	STDEV.S(A2:A69)	16.17225514
46	2116	$\alpha_1/2$	5/200	0.025
52	2704	Z_1	NORM.S.INV(1-E5)	1.959963985
21	441	SE($\bar{x}$)	E4/SQRT(E2)	1.975755424
23	529	95% C.I. of μ	E3 & "±" & (E6*E7)	41.9402985074627±3.87240947386967
33	1089	μ_{Lower}	E3-E6*E7	38.06788903
44	1936	μ_{Upper}	E3+E4*E7	73.89271932
12	144	$\alpha_2/2$	1/200	0.005
21	441	Z_2	NORM.S.INV(1-E6)	2.575829304
22	484	99% C.I. of μ	E3 & "±" & (E12*E7)	41.9402985074627±5.089208718534
27	729	μ_{Lowe}r	E3-E12*E7	36.85108979
29	841	μ_{Upper}	E3+E12*E7	47.02950723
…	…			

Conclusion:

95% C.I. of mean μ = (39.58059335, 48.77562302) and

99% C.I. of mean μ = (39.282453, 41.77637053)

Experiment 2 (Test of significance of two proportions)

Two separate colleges have 500 and 1000 students, if the colleges have passed results 400 and 900 respectively. Test whether there is significant difference between the pass proportion.

Step 1: Setting of Hypothesis:

Null Hypothesis; H_0: P_1=P_2, there is no significant difference between two proportions of pass students in two colleges.

Alternative Hypothesis; H_1:$P_1 \neq P_2$, there is significant difference between two proportions of pass students in the two colleges.

Step 2: Test statistics

$$Z_{cal} = \frac{p1-p2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \hat{P} = \frac{n_1p_1+n_2p_2}{n_1+n_2}, \hat{Q} = 1 - \hat{P}$$

Measurements	College A			College B		
	Symbols	Formulae	Values	Symbols	Formulae	Values
No of Students	n_1		500	n_2		1000
No of Passed Sample	x_1		400	x_2		900
Proportions	p_1	D4/D3	0.8	p_2	G4/G3	0.9
Level of Significance	α	5/100	0.05			
Population Proportions	**P**	(D4+G4)/(D3+G3)	0.866667			
	Q	1-D7	0.133333			
Test Statistics	Z_{cal}	(G5-D5)/SQRT(D7*D8*((1/D3)+(1/G3)))	5.370862			
	P_{value}	1-NORM.S.DIST(D9,1)	0			
Decision		IF(2*D10<=D6,"Reject H0","Do not reject H0")	Reject H0			

Step 3: Level of significance and critical value:

Take level of significance(**α**)= 5%

Critical value is $Z_{tab}(0.05)=1.96$

Step 4: Decision
2p <= α, Reject H_0
Step 5: Conclusion
There is a significant difference between two proportions of pass students in the two colleges.

Experiment 3 (Test of single mean)
A sample of size 400 laptop cells produced by a company is found to be 1570hrs, with S.D. of 150 hrs. Test the hypothesis that the mean lifetime of the laptop cells produced by the company is 1600hrs against the alternative hypothesis that it is greater than 1600hrs at 5% level of significance.

Step 1: Setting of Hypothesis
Null Hypothesis; H_0: μ = 1600hrs, there is no significant difference between battery life of laptops with 1600hrs.
Alternative Hypothesis; H_1: μ>1600hrs, the battery life of laptop is more than 1600 hrs.

Step 2: Test statistics

$$Z_{cal} = \frac{\bar{x}-\mu}{\sigma\sqrt{n}}$$

Measurements	Symbols	Formulae	Values
Sample Size	n		400
Population Mean	μ		1600
Population S.D	σ		150
Level of Significance	α	5/100	0.05
Sample Mean	X̄		1570
Test Statistics	Z_{cal}	(D6-D3)/(D4*SQRT(D2))	-4
	P_{value}	1-NORM.S.DIST(D9,1)	3.16712E-05
	Z_{tab}	NORM.S.INV(1-D5)	1.644853627
Decision		IF(ABS(D7)>D9,"Reject H0","Do not reject H0")	Reject H0

Step 3: Level of significance and critical value
Take level of significance($\boldsymbol{\alpha}$)= 5%
Critical value is Z_{tab} (0.05) =1.645

Step 4: Decision

Since Z_{calc} > Ztab, reject H_0.

Step 5: Conclusion

There is no significant difference between battery life of laptops with 1600hrs.

Experiment 4 (Test of difference between two means)

Two random samples of Nepalese people taken from rural and urban region gave the following data of income:

Sample	Size	Average monthly income	S.D.
Rural	150	800	50
Urban	100	1250	40

Step 1: Setting of Hypothesis

Null Hypothesis; H_0: $\mu_1 = \mu_2$, there is no significance difference between income of rural and urban area

Alternative Hypothesis; H_1: $\mu_1 \neq \mu_2$, there is significant difference between the income of rural and urban area.

Step 2: Test statistics

$$Z_{cal} = \frac{\overline{x}_1 - \overline{x}_2}{\left(\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}\right)}$$

Measurement	Rural			Urban	
	Symbols	Formulae	Values	Symbols	Values
Sample Size	n_1		150	n_2	100
Sample Mean	$\bar{X}_1$		800	$\bar{X}_2$	1250
Sample SD	s_1		50	s_2	40
Level of Significance	α		0.05		
Calculated Z	Z_{cal}	ABS((D4-G4)/SQRT((POWER(D5,2)/D3)+(POWER(G5,2)/G3)))	78.7336		
Calculated P	P	NORM.S.DIST(D7,1)	1		
Tabulated Z	Z_{tab}	NORM.S.INV(1-D6/2)	1.959964		
Decision		IF(ABS(D7)>D9,"Reject H_0","Do not reject H_0")	Reject H_0		

Step 3: Level of significance and critical value

Take level of significance(α)= 5%

Critical value is Z_{tab} (0.05) = 1.96

Step 4: Decision

Reject H_0 since $Z_{cal} > Z_{tab}$

Step 5: Conclusion

There is a significant difference between the income of rural and urban areas.

Experiment 5 (t-test, single sample mean)

The time in minutes, spent by 12 using randomly selected customers internet in a cybercafé is as follows: 33,35,41,45,48,71,58,89,51,54,66,48. Can you conclude that the average time spent by the customers is more than 50 minutes? Test at 5% level of significance.

Step 1: Setting of Hypothesis

Null Hypothesis; H_0: μ = 50min, there is no significant difference between the average time spent by customers with 50 min.

Alternative Hypothesis; H_1: μ> 50, the average time spent by customers is more than 50 min.

Step 2: Test statistics

$$t_{cal} = \frac{\bar{x}-\mu}{s\sqrt{n}}$$

where $\bar{x} = \frac{\sum x}{n}$, $s = \frac{\sum(x-\bar{x})^2}{n-1}$

```
T-TEST
  /TESTVAL=50
  /MISSING=ANALYSIS
  /VARIABLES=x
  /CRITERIA=CI(.95).
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Time Spent	12	53.25	15.955	4.606

One-Sample Test

	Test Value = 50					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Time Spent	.706	11	.495	3.250	-6.89	13.39

Here,

Measurements	Symbols	Formulae	Value
Mean	X̄	from SPSS	53.25
S.D.	σ	from SPSS	15.955
Size	n	from SPSS	12
Level of Sig.	α	from SPSS	0.05
Pop. Mean	μ	from SPSS	50
S.E. (X̄)		from SPSS	4.606
t_{cal}		ABS((D2-D6)/D7)	0.705601389
p-value		from SPSS	0.495
t_{tab}		T.INV(1-D5,11)	1.795884819
Decision		IF(D8<=D10,"Do not reject H_0", "Reject H_0")	Do not reject H_0

Step 3: Level of significance and critical value:

Take level of significance(**α**)= 5%

Critical value is $t_{tab}(0.05,n-1)$

Step 4: Decision

Since $t_{calc} < t_{tab}$, do not reject H_0.

Step 5: Conclusion

Hence, there is no significant difference between the average time spent by customers with 50 minutes.

Experiment 6 (t-test , difference between two sample means)

Two types of manure were applied to sixteen one-hectare plot, other conditions remaining same, the yield in quintals are given in the table below.

Manure I	18	20	36	54	48	74	45	51	47
Manure II	28	29	25	36	30	29	51		

Can we conclude that the first manure gives more yield? Teat at 5% level of significance.

Step 1: Setting of Hypothesis:

Null Hypothesis; H_0: $\mu_1 = \mu_2$, there is no significant difference between the average yield given by two manures

Alternative Hypothesis; H_1: $\mu_1 > \mu_2$, the average yield given by first manure is more.

Step 2: Test statistics

$$t_{cal} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $sp = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$, $s_1 = \frac{\Sigma\left(x_1 - \overline{x_1}\right)^2}{n_1-1}$, $s_2 = \frac{\Sigma\left(x_2 - \overline{x_2}\right)^2}{n_2-1}$

Manure I	Manure II	Measurements	Symbols	Formulae	Values
18	28	Size	n_1	COUNTA(A2:A10)	9
20	29		n_2	COUNTA(B2:B8)	7
36	25	Mean	$\bar{X}_1$	AVERAGE(A2:A10)	43.66666667
54	36		$\bar{X}_2$	AVERAGE(B2:B8)	32.57142857
48	30	S. D	s_1	STDEV.S(A2:A10)	17.28438602
74	29		s_2	STDEV.S(B2:B8)	8.772250621
45	51		S^2	((F2-1)*(F6^2) + (F3-1)*(F7^2))/(F2+F3-2)	203.6938776
51		t_{calc}		ABS(F5-F4)/SQRT(F8*((1/F3)+(1/F2)))	1.542616321
47		Level of Sig.	α		0.05
		t_{tab}		T.INV(1-F10,F2+F3-2)	1.761310136
		Decision		IF(F9<F11,"Do not Reject H0", "Reject H0")	Do not Reject H0

Step 3: Level of significance and critical value:

Take level of significance(α)= 5%

Critical value is t_{tab}(0.05,n1+n2-2)

Step 4: Decision

Do not reject H_0 as $t_{calc} < t_{tab}$.

Step 5: Conclusion

There is no significant difference between the average yield given by two manures.

Experiment 7 (paired t-test)

Memory capacity of ten students was tested before and after training, state whether the training was effective or not from the following scores.

Roll	1	2	3	4	5	6	7	8	9	10
Before Training	12	14	11	8	7	10	3	0	5	6
After Training	15	16	10	7	5	12	10	2	3	8

Step 1: Setting of Hypothesis: Null Hypothesis; H_0: $\bar{d}= 0$, the training is insignificant to increase the memory capacity of students.
Alternative hypothesis; H_1: $\bar{d}\neq 0$, the training is significant. $\bar{d}> 0$, the training increases the memory capacity. $\bar{d}< 0$, the training decreases the memory capacity.

Step 2: Test statistics

$$t_{cal} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

Where, $s_d = \frac{\sum(d-\bar{d})^2}{n-1}$

Before training	After training	d	Measurements	Symbols	Formulae	Values
12	15	3	Mean	$\bar{X}_d$	AVERAGE(C2:C11)	1.2
14	16	2	Standard Dev.	S_d	STDEV.S(C2:C11)	2.780887149
11	10	-1	Size	n	COUNTA(A2:A11)	10
8	7	-1	t_{calc}		G2/(G3/SQRT(G4))	1.364576478
7	5	-2	Level of Sig.	α		0.05
10	12	2	t_{tab}		T.INV.2T(G6,G4-1)	2.262157163
3	10	7	Decision		IF(G5<G7,"Do not Reject H_0", "Reject H_0")	Do not Reject H_0
0	2	2				
5	3	-2				
6	8	2				

Step 3: Level of significance and critical value:
Take level of significance(α)= 5%
Critical value is t_{tab}(0.05,n1)
Step 4: Decision

Reject H_0 if $t_{cal} > t_{tab}$, otherwise accept H_0.

Step 5: Conclusion

The training significantly increases the memory capacity.

Unit 3: Non-Parametric Tests

Experiment 8: (run test with small sample and Binomial Test)

On tossing a coin 30 times outcomes of head and tail are found as; Head, Head, Tail, Head, Tail, Head, Head, Tail, Tail, Head, Tail, Head, Head, Tail, Tail, Head, Head, Head, Tail , Head, Tail, Head ,Head , Tail, Tail, Head, Tail, Tail, Tail, Head.

(i) Are outcomes in random order?

(ii) Is coin unbiased?

Using 1% level of significance

 # To test randomness

Step 1: Setting up hypothesis

Null Hypothesis; H_0: Samples are in random order

Alternative Hypothesis; H_1: Samples are not in random order

Step 2: Test statistic

Values	Measurements	Symbol	Formulae	Values
Head	Number of Heads	n_1	SUMPRODUCT(--(ISNUMBER(SEARCH("Head",A2:A31))))	16
Head	Number of Tails	n_2	COUNTA(A2:A31)-E2	14
Tail	Number of runs	r		19
…				

r =19.

Step 3: Level of significance and Critical Value

Level of Significance= α = 1% =0.01

From Table, with d.f. n_1, n_2: $\overline{r} = 10, \underline{r} = 22$

Step 4: Decision

Accept H_0 since $r \epsilon \left(\overline{r}, \underline{r}\right)$

Step 5: Conclusion
The Samples are in random order.

 # To test unbiasedness
Step 1: Setting up Hypothesis.
Null Hypothesis; H_0: P = 1/2
Alternative Hypothesis; H_1: P ≠ 1/2

Step 2: Test statistic

$$Z = \frac{(x_0 \pm 0.5) - np}{\sqrt{npq}}$$

Where $x_0 = min(n_1, n_2)$

$n = n_1 + n_2$, use +0.5 if x_0 <np and -0.5 otherwise.

$p = prob(Z > |Z_{calc}|)$

Measurements	Symbol	Formulae	Values
Number of Heads	n_1	SUMPRODUCT(--(ISNUMBER(SEARCH("Head",A2:A31))))	16
Number of Tails	n_2	COUNTA(A2:A31)-E2	14
Number of runs	r		19
X_0		MIN(E2,E3)	14
Probability of Head	p	E2/(E2+E3)	0.533333333
	q	1-E6	0.466666667
n			30
np		E8*E6	16
Z_{calc}		(E5+0.5-E9)/SQRT(E9*E7)	-0.548943791
Level of sig.	α		0.01
Z_{tab}		NORM.S.INV(1-E11/2)	2.575829304
Decision		IF(ABS(E10)<E12,"Do not Reject H_0", "Reject H_0")	Do not Reject H_0

Step 3: Level of Significance and Critical Values
Level of significance = α = 0.01

Step 4: Decision

Do not reject H_0 as $Z_{calc} < Z_{tab}$.

Step 5: Conclusion

Thus, the Coin is unbiased.

Experiment 9: (Run test, One Sample Kolmogorov Smirnov test, chi square test)

Marks secured by a sample of 32 students in Final examination of Statistics I are found as 43, 52, 34, 56, 28, 12, 46, 38, 10, 51, 49, 38, 46, 24, 36, 44, 38, 46, 27, 35, 41, 11, 23, 35, 42, 52, 49, 20, 35, 43, 37.

(i) Are samples selected in random order?
(ii) Are marks uniformly distributed? Use Kolmogorov Smirnov test
(iii) Are marks uniformly distributed? Use chi square test
Using 5% level of significance.

 # To test randomness

Find median then to sample values assign symbol say A if sample value less than median and symbol say B if sample value more than median and omit if sample value is equal to median.

Step 1: Setting up Hypothesis.

Null Hypothesis; H_0: Samples are in random order.

Alternative Hypothesis: H_1: Samples are not in random order.

Step 2: Test statistic

Values	Symbols	Measurements	Symbols	Formulae	Values
43	B	Median		MEDIAN(A2:A33	38
52	B	Total Data	N	COUNTA(A2:A33)	32
34	A	Number of A	n_1	SUMPRODUCT(--(ISNUMBER(SEARCH("A",B2:B33))))	14
56	B	Number of B	n_2	SUMPRODUCT(--(ISNUMBER(SEARCH("B",B2:B33))))	15
28	A	total n	n	F4+F5	29
12	A	Runs	r		16
46	B				
38	0				
10	A				

Runs: r =16.

Step 4: Decision

Accept H0 since $r\epsilon\left(\overline{r}, \underline{r}\right)$

Step 5: Conclusion

The Samples are in random order.

 # To test uniformity

Step 1: Setting up Hypothesis.

Null Hypothesis H_0: Marks are uniformly distributed.

Alternative Hypothesis; H_1: Marks are not uniformly distributed Kolmogorov Smirnov test.

Step 2: Test statistic

$D_0 = Max\left|F_e - F_o\right|$

Where $F_e = \frac{cf_e}{N}, F_o = \frac{cf_o}{N}$

Eval	Symbol	Formulae	Values	x	f	cf_0	F_o	f_e	cf_e	F_e	$\|F_e-F_o\|$
Median		MEDIAN(A2:A33	38	43	2	2	0.062	1.454	1.454	0.045	0.017
Total Data	N	COUNTA(A2:A33)	32	52	2	4	0.125	1.454	2.909	0.090	0.034
Nor of A	n_1	SUMPRODUCT(--ISNUMBER(SEARCH("A",B2:B33)))	14	34	1	5	0.156	1.454	4.363	0.136	0.019
No of B	n_2	SUMPRODUCT(--ISNUMBER(SEARCH("B",B2:B33)))	15	56	1	6	0.187	1.454	5.818	0.181	0.005
total n	n	F4+F5	29	28	1	7	0.218	1.454	7.272	0.227	0.008
Runs	r		16	12	1	8	0.25	1.455	8.727	0.272	0.022
Unique x	n_{unique}		22	46	3	11	0.343	1.454	10.181	0.318	0.025
D_o		MAX(N2:N23)	0.093	38	3	14	0.437	1.454	11.636	0.363	0.073
D_{tab}	$D_{32,0.05}$		0.234	10	1	15	0.468	1.454	13.090	0.409	0.059
Decision		IF(F9<F10,"Accept H_0", "Reject H_0")	Accept H_0	51	1	16	0.5	1.454	14.545	0.454	0.045
				…	…	…	…	…	…	…	…

Step 3: Level of Significance and Critical Values.
Critical value At α level of significance critical value is $D_{n,\alpha}$

Step 4: Decision
Since $D_o < D_{tab}$, accept H_0.

Step 5: Conclusion
Marks are uniformly distributed.

Chi Square (not appropriate), so same as Kolmogorov test.

Experiment 10: (Median Test, Mann Whitney U test and Two sample Kolmogorov test)

Following are marks secured by 14 students of section A and 15 students of section B of DWIT in final examination of Digital logic are found as

A	34	48	21	52	31	43	29	37	24	52	49	34	40	48	
B	11	53	27	38	47	50	26	38	44	33	27	33	41	10	28

Are median marks of section A and section B identical at 5% level of significance using?

(i) Median test
(ii) Mann Whiteny U test
(iii) Kolmogorov Smirnov test

 # For Median Test

Step 1: Setting up Hypothesis
Null Hypothesis; H_0: $Md_A = Md_B$
Alternative Hypothesis; H_1: $Md_A \neq Md_B$
Using median test: Combine n_1 and n_2 such that $N = n_1+n_2$ and obtain median of N observation. Find number of observations in $x_i \leq Md$ and denote by a. Find number of observations in $y_i \leq Md$ and denote by b. Also find the number of observations in $x_i > Md$ an $y_i > Md$ and denote by c and d respectively.

	No of obs. ≤ Md	No of obs. > Md	Total
Sample x	a	c	a + c
Sample y	b	d	b + d
Total	a + b	c + d	a + b + c + d

Step 2: Test statistics

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)} \sim \chi^2_{(1)}$$

If any cell frequency is less than 5 then

$$Corrected\chi^2 = \frac{N\left(|ad-bc|-\frac{N}{2}\right)^2}{(a+c)(b+d)(a+b)(c+d)} \sim \chi^2_{(1)}$$

Sample x	Sample y	Measurements	Symbols	Formulae	Values
34	11	Sample Size 1	n_1	COUNTA(A2:A15)	14
48	53	Sample Size 2	n_2	COUNTA(B2:B16)	15
21	27	Combined Size	N	J4	29
52	38	Median	Md	MEDIAN(A2:B15, B16)	37
31	47	Chi -Squared Calc	χ^2_{calc}	(J4 *(H2*I3 -H3*I2)^2)/(J2*J3*I4*H4)	0.032222222
43	50	Level of sig.	α	5/100	0.05
29	26	Chi-Squared Tab	χ^2_{tab}	CHISQ.INV.RT(F7,1)	3.841458821
37	38	Decision		IF(F6<F8, "Do not Reject H_0", "Reject H_0")	Do not Reject H_0
…	…				

	No of obs. ≤ Md	No of obs. > Md	Total
Sample x	7	7	14
Sample y	8	7	15
Total	15	14	29
Formula	COUNTIF(A2:A15, "<=" & F5)		

Step 3: Level of Significance and Critical Values

$\alpha = 0.05 \; and \; \chi^2_{tab} = 3.841$

Step 4: Decision

Since $\chi^2_{calc} \leq \chi^2_{tab}$, Do not Reject Null Hypothesis.

Step 5: Conclusion

The Medians of Sample A and Sample B are significantly identical.

 # For Mann Whitney U Test

Step 1: Setting up Hypothesis

Null Hypothesis; H_0: $Md_A = Md_B$

Alternative Hypothesis; H_1: $Md_A \neq Md_B$

Using Mann Whitney U test: Combine n_1 and n_2 such that $n_1+n_2= n$ and rank these n observations in ascending order. If two or more observations are equal, then assign average rank and is called tied. Sum the ranks of sample of sizes n_1 and n_2 separately to get R_1 and R_2. If two sample sizes are unequal, then the smaller one is n_1. Obtain U_1 and U_2 as

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \, , \; U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$U_0 = min\{U_1, U_2\}$$

Step 2: Test statistic

$$Z = \frac{U_0 - \mu}{\sigma_n} = \frac{U_0 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

It is used even if tied occurs within sample. If tied occurs between samples then standard deviation is corrected as

$$\sigma_u = \sqrt{\frac{n_1 n_2}{n(n-1)\{\frac{n^3 - n}{12} - \sum \frac{t_i^3 - t_i}{12}\}}}$$

t_i = number of times i^{th} rank repeated between samples.

Sample x	Rank 1	Sample y	Rank 2	Combined	Measurements	Symbols	Formulae	Values
34	13.5	11	2	34	Sample Size 1	n_1	COUNTA(A2:A15)	14
48	23.5	53	29	48	Sample Size 2	n_2	COUNTA(B2:B16)	15
21	3	27	6.5	21	Combined Size	N	M4	29
52	27.5	38	16.5	52	Sum of Rank 1	R_1	SUM(B2:B15)	233
31	10	47	22	31	Sum of Rank 2	R_2	SUM(D2:D16)	202
43	20	50	26	43	U_1	U_1	I2*I3 + ((I2 *(I2 +1))/2) -I5	82
29	9	26	5	29	U_2	U_2	I2*I3 + ((I3 *(I3 +1))/2) -I6	128
37	15	38	16.5	37	U_0	U_0	MIN(I7:I8)	82
24	4	44	21	24	Corrected S.D	σ_u	SQRT(((I2*I3)/(I4*(I4-1)))*(((POWER(I4,3)-I4)/12)-6*((POWER(2,3)-2)/12)))	22.89594152
52	27.5	33	11.5	52			I2*I3/2	105

49	25	27	6.5	49	Z_{calc}	Z_{calc}	(I9-I11)/I10	-1.004544844
34	13.5	33	11.5	34	Z_{tab}	Z_{tab}	NORM.S.INV(1-I14)	1.644853627
40	18	41	19	40	Level of sig.	α	5/100	0.05
48	23.5	10	1	48	Decision		IF(ABS(I12)<I13, "Do not Reject H_0", "Reject H_0")	Do not Reject H_0

Step 3: Level of Significance and Critical Value

$\alpha = 0.05 \, and \, Z_{tab} = 1.65$

Step 4: Decision

Since $Z_{calc} < Z_{tabulated}$, do not reject H_0.

Step 5: Conclusion

The Medians of sample A and sample B are significantly identical.

 # For Kolmogorov Smirnov Test

Step 1: Setting up Hypothesis

Null Hypothesis; H_0: $\mathrm{F(x) = F(y)}$

Alternative Hypothesis; H_1: $\mathrm{F(x) \neq F(y)}$

Step 2: Test statistics

$D_0 = maximum|F(x) - F(y)|, \, where F(x) = \frac{cf_x}{N_x}, \, F(y) = \frac{cf_y}{N_y}$

Sample x	Sample y	Combined	Unique	f(x)	f(y)	F(x)	F(y)	|F(x) - F(y)|
34	11	34	10	0	1	0.00	0.07	0.07
48	53	48	11	0	1	0.00	0.07	0.07
21	27	21	21	1	0	0.07	0.00	0.07
52	38	52	24	1	0	0.07	0.00	0.07
31	47	31	26	0	1	0.00	0.07	0.07
43	50	43	27	0	2	0.00	0.13	0.13

29	26	29	28	0	1	0.00	0.07	0.07
37	38	37	29	1	0	0.07	0.00	0.07
24	44	24	31	1	0	0.07	0.00	0.07
52	33	52	33	0	2	0.00	0.13	0.13
49	27	49	34	2	0	0.14	0.00	0.14
34	33	34	37	1	0	0.07	0.00	0.07
40	41	40	38	0	2	0.00	0.13	0.13
48	10	48	40	1	0	0.07	0.00	0.07
	28	11	41	0	1	0.00	0.07	0.07
		53	43	1	0	0.07	0.00	0.07
		27	44	0	1	0.00	0.07	0.07
		38	47	0	1	0.00	0.07	0.07
		47	48	2	0	0.14	0.00	0.14
		50	49	1	0	0.07	0.00	0.07
		26	50	0	1	0.00	0.07	0.07
		38	52	2	0	0.14	0.00	0.14
		44	53	0	1	0.00	0.07	0.07
		33						
		27	Total	14	15			
		…						

Measurements	Formula	Values
D0	MAX(I2:I24)	0.142857
α	5/100	0.05
Dn1,n2,α	1.36*SQRT((E26+F26)/(E26*F26))	0.505392
Decision	IF(F27<F29, "Do not Reject H_0", "Reject H_0")	Do not Reject H_0

Step 3: Level of Significance and Critical Value
level of significance α= 0.05
critical value for n_1 and n_2 is $D_{n1,\, n2,\, \alpha}$ = 0.505

Step 4: Decision
Since, $D_0 < D_{n1,\, n2,\, \alpha}$, do not reject H_0.

Step 5: Conclusion
There is no significant difference between the two samples thus their medians.

Experiment 11: Chi Square for Independence of Attributes.
The following information is obtained from locality related to gender and eye color:

Person	Gender	Eye Color
A	Male	Black
B	Female	Black
C	Male	Brown
D	Male	Black
E	Female	Blue
F	Male	Brown
G	Female	Black
H	Male	Black
J	Female	Black
K	Female	Brown
L	Female	Black
N	Male	Black
O	Female	Blue
P	Female	Brown
Q	Male	Black
R	Female	Black
S	Male	Brown
T	Female	Black
U	Female	Black
V	Male	Brown

Is there any association between gender and eye color? Use 5% level of significance.

Step 1: Setting up Hypothesis.

Null Hypothesis H0: Attributes are independent.
Alternative Hypothesis H1: Attributes are dependent.

Step 2: Test statistic

$$\chi^2 = \sum\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

Person	Gender	Eye Color	Ai\Bj	Black	Brown	Blue	Total	Group	O_{ij}	E_{ij}= ($O_{i.}$ X $O_{.j}$)/N	O_{ij} - E_{ij}	(O_{ij} - E_{ij})^2/E_{ij}
A	Male	Black	Male	5	4	0	9	A_1B_1	5	5.4	-0.4	0.03
B	Female	Black	Female	7	2	2	11	A_1B_2	4	2.7	1.3	0.63
C	Male	Brown	Total	12	6	2	20	A_1B_3	0	0.9	-0.9	0.90
D	Male	Black						A_2B_1	7	6.6	0.4	0.02
E	Female	Blue						A_2B_2	2	3.3	-1.3	0.51
F	Male	Brown						A_2B_3	2	1.1	0.9	0.74
G	Female	Black						Total				2.83
H	Male	Black										

Measurements	Formula	Value
Chi-squared calc	M8	2.83
α	5/100	0.05
d.f.	(COUNTA(D2:D3)-1)*(COUNTA(E2:G2)-1)	2
Chi-squared tab	CHISQ.INV.RT(F12,F13)	5.991465
Decision	IF(F11<F14, "Do not Reject H_0", "Reject H_0")	Do not Reject H_0

Step 3: Level of Significance and Critical Value

level of significance α= 0.05

critical value for r =2 and c=3 is $\chi^2_{\alpha(r-1)(c-1)} = 5.99$

Step 4: Decision

Since, $\chi^2_{calc} < \chi^2_{tab}$, do not reject H_0.

Step 5: Conclusion

Thus, the attributes are independent of each other.

Experiment 12: Wilcoxon Matched Pair Signed Rank Test.

Marks secured by a sample of 15 students at a college in first test and second test of Statistics II are found as

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Test I	12	7	15	11	17	19	5	13	17	6	9	18	14	10	8
Test II	14	5	17	13	12	18	9	10	18	12	3	14	16	16	8

Is there improvement in marks in test II as compared to test I? Use nonparametric test at 5% level of significance.

Step 1: Setting up Hypothesis.

Null Hypothesis H_0: $Md_I = Md_{II}$

Alternate Hypothesis H_1: $Md_I < Md_{II}$

Find di = $y_i - x_i$ or $x_i - y_i$ for each pair of observations (x_i, y_i), i = 1,2,3…….n. Rank d_i irrespective of sign in ascending order but omit $d_i = 0$. If two or more di are equal, then assign the average rank and is called tied. In such case corrected sample size $n_c = n - t$, t is number of tied occurred. Assign sign to the ranks with respect to the sign of d_i. Sum the ranks of + sign and - sign separately to get S (+) and S (-) respectively. Finally get T = min {S (+), S (-)}

Step 2: Test statistic

$T = Min(S(+), S(-))$.

Student	Test I	Test II	$d_i = y_i - x_i$	$\|d_i\|$	R_i	Measurements	Formula	Values
1	12	14	2	2	5	S (+)	SUMIF(F2:F15,">0")	57
2	7	5	-2	2	-5	S (-)	ABS(SUMIF(F2:F15,"<0")	48
3	15	17	2	2	5	corrected n	COUNTA(A2:A16)-1	14
4	11	13	2	2	5	T_{calc}	MIN(J2:J3)	48
5	17	12	-5	5	-11	α	5/100	0.05
6	19	18	-1	1	-1.5	T_{tab}		26
7	5	9	4	4	9.5	Decision	IF(J5<J7,"Reject H0", "Do not Reject H0")	Do not Reject H_0
8	13	10	-3	3	-8			

Step 3: Level of Significance and Critical Value

Level of significance $\alpha = 0.05$

Critical Value $T_{\alpha,n} = 26$

Step 4: Decision

Since $T_{calc} > T_{tab}$, do not reject H_0.

Step 5: Conclusion

There is no significant improvement in test II compared to test I.

Experiment 13: Cochran Test

Four diets are fed to 9 cows, each diet for a month and the result of increase(I) and decrease(D) of milk given by different cows are found as follows:

Cow Diet	I	II	III	IV	V	VI	VII	VIII	IX
D1	I	I	D	I	D	I	I	D	I
D2	D	D	I	D	I	D	D	I	I
D3	I	D	I	D	D	I	I	D	D
D4	I	I	I	D	D	I	I	D	I

Test whether diets are equally effective or not at 1% level of significance.

Step 1: Setting up Hypothesis.

Null Hypothesis H_0: All the diets are equally effective.

Alternative Hypothesis H_1: All the diets are not equally effective.

Sum all the I(Positive) according to treatment to get Ri (Row wise) and according to objects to get Ci (Column wise), i = 1 , 2 , 3 , …… k and j = 1, 2, 3, …… n. Then get.

$\sum_{i=1}^{k} R_i, \sum_{i=1}^{k} R_i^2, \sum_{i=1}^{k} C_j, \sum_{i=1}^{k} C_j^2.$

Step 2: Test statistics

$$Q = \frac{(k-1)\left[K\sum R_i^2 - \left(\sum R_i\right)^2\right]}{K\sum C_j - \sum C_j^2} \sim \chi^2.$$

Cow Diet	I	II	III	IV	V	VI	VII	VIII	IX	R_i	R_i^2
D1	I	I	D	I	D	I	I	D	I	6	36
D2	D	D	I	D	I	D	D	I	I	4	16
D3	I	D	I	D	D	I	I	D	D	4	16
D4	I	I	I	D	D	I	I	D	I	6	36
C_i	3	2	3	1	1	3	3	1	3	20	104
C_i^2	9	4	9	1	1	9	9	1	9	52	

Measurements	Formula	Values
$\Sigma R_i = \Sigma C_i$	K6	20

ΣR_i^2	L6	104
ΣC_i^2	K7	52
k	COUNTA(A2:A5)	4
Q	((P5-1)*(P5*P3-POWER(P2,2)))/(P5*P2-P4)	1.714285714
α	1/100	0.01
$\chi_{\alpha(k-1)}$	CHISQ.INV.RT(P7,P5-1)	11.34486673
Decision	IF(P6<P8, "Do not reject H_0", "Reject H_0")	Do not reject H_0

Step 3: Level of Significance and Critical value

Level of Significance $\alpha = 0.01$

Critical Value$\chi_{\alpha\,(k-1)}^2$ = 11.34

Step 4: Decision

Since $Q < \chi^2_{\alpha,(k-1)}$, do not reject H_0.

Step 5: Conclusion

All the Diets are equally Effective.

Experiment 14: Kruskal Wallis H test

The following data represents the operating times in hours for four types of laptops before a charge is required.

Dell	5.3	4.8	6.1	3.5			
Acer	5.2	5.8	3.9	4.6	4.9	5.1	5.6
HP	4.5	5.2	3.8	4.8	5.3		
Lenovo	4.7	6.2	5.7	5.5	3.9	4.8	

Is operating time for all laptops equal at 5% level of significance use nonparametric test?

Step 1: Setting up Hypothesis.

Null Hypothesis H_0: $Md_1 = Md_2 = Md_3 = \ldots\ldots\ldots\ldots\ldots = Md_k$

Alternative Hypothesis H_1: At least one Md_i is different i = 1, 2, 3, ………., k.

Combine n_1, n_2, n_3 ……… and n_k such that $n_1 + n_2 + n_3 + \ldots\ldots\ldots + n_k = n$ and rank these n observations in ascending order. If two or more observations are equal, then assign average rank and is called tied. Sum the ranks of sample of sizes n_1, n_2, n_3, ……… and n_k separately to get R_1, R_2, R_3, ……, R_k.

Step 2: Test statistic : $H = \frac{12}{n(n+1)}\sum\frac{R_i^2}{n_i} - 3(n + 1) \sim \chi^2_{(k-1)}$.

If tied occurs, then corrected test statistic is,

$$H = \frac{\frac{12}{n(n+1)}\sum\frac{R_i^2}{n_i} - 3(n+1)}{1-\sum\frac{\left(t_i^3 - t_i\right)}{n^3-n}}$$

, t_i=number of times i^{th} rank is repeated. Critical value for $n_i < 5$ & k = 4,

Computers	Operating Time							R_i	R_i^2/n_i
Dell	5.3	4.8	6.1	3.5					
R1	15.5	9	21	1				46.5	540.5625
Acer	5.2	5.8	3.9	4.6	4.9	5.1	5.6		
R2	13.5	20	3.5	6	11	12	18	84	1008
HP	4.5	5.2	3.8	4.8	5.3				
R3	5	13.5	2	9	15.5			45	405
Lenovo	4.7	6.2	5.7	5.5	3.9	4.8			
R4	7	22	19	17	3.5	9		77.5	1001.042
Total									2954.604

Measurements	Symbols	Formulae	Values	$_F$Values
Sample Sizes	n_1	COUNTA(B2:H2)	4.00	
	n_2	COUNTA(B4:H4)	7.00	
	n_3	COUNTA(B6:H6)	5.00	
	n_4	COUNTA(B9:H9)	6.00	
Total Sample Size	n	SUM(O2:O6)	22.00	
Repetition	3.5	(POWER(O9,3)-O9)/(POWER(O8,3)-O8)	2.00	0.0006
	9	(POWER(O10,3)-O10)/(POWER(O8,3)-O8)	3.00	0.0023
	13.5	(POWER(O11,3)-O11)/(POWER(O8,3)-O8)	2.00	0.0006
	15.5	(POWER(O12,3)-O12)/(POWER(O8,3)-O8)	2.00	0.0006
H		(((12/(O8*(O8+1)))*J10)-(3*(O8+1)))/(1-SUM(P9:P12))	1.07	

Level of Sig.	α	5/100	0.05
Chi Square Tab	χ^2_{tab}	CHISQ.INV.RT(O14, 4-1)	7.81
Decision		IF(O13<O15,"Do not reject H_0", "Reject H_0")	Do not reject H_0

Step 3: Level of Significance and Critical Value

Level of Significance α = 0.05

Critical Value $\chi^2_{\alpha,(k-a)} = 7.81$

Step 4: Decision

Since $H > \chi^2_{tab}$, Do not reject H_0.

Step 5: Conclusion

Thus at least one Md_i is different, so not all operating times for laptops are equal.

Experiment 15: Friedman F test

The scores of 7 students in Statistics II in three tests are found as

Student Test	A	B	C	D	E	F	G
I	15	13	8	12	9	16	13
II	14	16	12	10	14	11	6
III	10	12	5	16	8	14	16

(i) Is there any significant difference in marks in the three tests?

(ii) Is there any significant difference in the marks of seven students?

Use nonparametric test at 1% level of significance.

Step 1: Setting up Hypothesis.

Null Hypothesis H_0: $Md_1 = Md_2 = Md_3 = \ldots\ldots\ldots\ldots = Md_k$

Alternative Hypothesis H_1: At least one Md_i is different i = 1, 2, 3, ………. k.

Rank k sample observations for each block separately from 1 to k. in ascending order. If two or more observations are the same, then assign average rank which is also called tied. Obtain sum of ranks for each sample to get Ri, i = 1, 2, 3 , ………k

Step 2: Test statistic

$$F_r = \frac{12}{nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3n(k+1).$$

If tied occurs, then corrected test statistic is,

$$F_r = \frac{\frac{12}{nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3n(k+1)}{1-\sum\frac{t_i^3-t_i}{n(k^3-k)}}$$, t_i = number of times i^{th} rank is repeated.

Test Student	I	R_1	II	R_2	III	R_3	R_i	R_i^2
A	15	2	14	2.5	10	5	9.5	90.25
B	13	3.5	16	1	12	4	8.5	72.25
C	8	7	12	4	5	7	18	324
D	12	5	10	6	16	1.5	12.5	156.25
E	9	6	14	2.5	8	6	14.5	210.25
F	16	1	11	5	14	3	9	81
G	13	3.5	6	7	16	1.5	12	144
Total								1078

Measurements	Formulae	Values	$_T$Value	$_F$Value
n	COUNTA(A2:A8)	7		
k	COUNTA(B1,D1,F1)	3		
Rank Repeat	(POWER(N4,3) - N4)/(M2*(POWER(M3,3)-M3))	3.5	2	0.035714
	(POWER(N5,3) - N5)/(M2*(POWER(M3,3)-M3))	2.5	2	0.035714
	(POWER(N6,3) - N6)/(M2*(POWER(M3,3)-M3))	1.5	2	0.035714
F_r	((12/(M2*M3*(M3+1)))*I9-(3*M2*(M3+1)))/(1-SUM(O4:O6))	78.4		
P		0		
α	1/100	0.01		
Decision	IF(M8<M9, "Reject H0", "Do not Reject H0")	Reject H0		

Step 3: Level of Significance and Critical value

For n=7, k=3 critical value p is obtained from Friedman table, p = P (F_r > 73.96) = 0.

Step 4: Decision
Since $p < \alpha$, reject H_0.

Step 5: Conclusion
There are significant Differences.

Experiment 16: Completely Randomized Design
Let A, H, D and L represent Acer, HP, Dell and Lenovo laptop and following information represents their operating time in hours before charge is required.

A 5.2	H 3.8	D 4.6	H 5.2	D 3.6	L 4.5
L 5.6	A 3.9	H 4.6	L 6.2	L 4.8	A 3.5
H 4.4	D 3.6	L 5.2	D 4.8	A 4.2	D 5.4
A 6.1	L 4.7	A 3.2	H 5.3	D 4.8	H 3.9

Carryout analysis of the design at 1% level of significance. (CRD)
Working formula
Mathematical model: $y_{ij} = \mu + \tau_i + e_{ij}$; $i = 1, 2, 3, ..., t; j = 1, 2, 3, ..., r.$

Step 1: Setting up Hypothesis.
Null Hypothesis H_{0T}: $\mu_{1.} = \mu_{2.} = \mu_{3.} = ... = \mu_{t.}$ (There is no significant difference between the treatments)
Alternative Hypothesis H_{1T}: At least one $\mu_{i.}$ is different. i = 1, 2, 3……. t (There is at least one significant difference between treatments)

Step 2: Test statistic:
F_T = MST/MSE, where MST = SST/ (t – 1), MSE = SSE / t(r – 1)

$$TSS = \sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}^2 - C.F.$$

$$SST = \frac{\sum_{i=1}^{t} T_{i.}^2}{r} - C.F. \quad where C.F. = \frac{G^2}{N}$$

SSE = TSS - SST

Using EXCEL, Insert/Data/Data Analysis/ANOVA: Single factor.

ANOVA

values

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.308	3	.769	1.197	.336
Within Groups	12.857	20	.643		
Total	15.165	23			

Treatments							$T_{i.}$
A	5.2	3.9	3.5	4.2	6.1	3.2	26.1
D	4.6	3.6	3.6	4.8	5.4	4.8	26.8
H	3.8	5.2	4.6	4.4	5.3	3.9	27.2
L	4.4	5.6	6.2	4.8	5.2	4.7	30.9
Total							111

Measurements	Formulae	Values
G	H6	111
N	COUNTA(B2:G5)	24
C.F.	POWER(K2,2)/K3	513.375
Σy_{ij}^2	SUM(POWER(B2:G5,2)	528.54
r	COUNTA(B2:G2)	6
t	COUNTA(A2:A5)	4
α	5/100	0.05
TSS	K5-K4	15.165
SST	(SUM(POWER(H2:H5,2))/K6)-K4	2.308333333
SSE	K9-K10	12.85666667
d.f.	M2 & "," & M3	3,20
Decision	IF(P2<Q2,"Do not reject H_0", "Reject H_0")	Do not reject H_0

S.V.	d.f.	S.S.	M.S.	F_{cal}	F_{tab}
Treatment	3	2.308333	0.769444	1.196958	3.098
Error	20	12.85667	0.642833		
Total	23	15.165			

Step 3: Level of significance and Critical Value

Level of significance $\alpha = 0.05$

Critical Value $F_{\alpha,(t-1),t(r-1)} = 3.098$

Step 4: Decision

Since $F_{calc} < F_{tab}$, do not reject H_0.

Step 5: Conclusion

Hence there is no significant difference between the laptops(treatments).

Experiment 17: Randomized Block Design

Let A, H, D and L represent Acer, HP, Dell and Lenovo laptop and following information represents their operating time in hours before charge is required.

A 5.0	H 3.6	D 4.8	A 4.2	D 3.8	L 4.6
L 5.4	A 4.9	H 4.3	L 5.2	L 5.8	A 5.5
H 4.8	D 4.6	L 5.5	D 4.6	A 5.2	D 5.0
D 6.0	L 4.5	A 3.9	H 5.1	H 4.9	H 4.9

Carryout analysis of the design at 1% level of significance. (RBD)

Working formula

Mathematical model:

$$y_{ij} = \mu + \tau_i + \beta_j + e_{ij}$$

Where, y_{ij} = j^{th} block receiving i^{th} treatment; i = 1, 2, ..., t, j = 1, 2, ..., b

Step 1: Setting up Hypothesis.

Null Hypothesis H_{0T}: $\mu_{1.} = \mu_{2.} = \mu_{3.} = \ldots = \mu_{t}$. (There is no significant difference between the treatments)

Alternative Hypothesis H_{1T}: At least one $\mu_{i.}$ is different. i = 1, 2, 3……. t (There is at least one significant difference between treatments)

Null Hypothesis H_{0B}: $\mu_{.1} = \mu_{.2} = \mu_{3.} = \ldots = \mu_{.b}$ (There is no significant difference between blocks)

Alternative Hypothesis H_{1T}: At least one $\mu_{.j}$ is different. j = 1, 2, 3……. b (There is at least one significant difference between blocks)

Step 2: Test statistic

F_T = MST/MSE, F_B = MSB/MSE,

where MST = SST/(t – 1) , MSB = SSB/(b – 1), MSE = SSE / (t-1)(b – 1)

$$TSS = \sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}^2 - C.F.$$

$$SST = \frac{\sum_{i=1}^{t} T_{i.}^2}{b} - C.F. \quad where C.F. = \frac{G^2}{N}$$

$$SSB = \frac{\sum_{j=1}^{r} T_{.j}^2}{t} - C.F$$

SSE = TSS – SST-SSB

Using EXCEL, Insert/Data/Data Analysis/ANOVA: Two factor without replication.

Treatments							$T_{i.}$
A	5.0	4.9	3.9	4.2	5.2	5.5	28.7
D	6.0	4.6	4.8	4.6	3.8	5.0	28.8
H	4.8	3.6	4.3	5.1	4.9	4.9	27.6
L	5.4	4.5	5.5	5.2	5.8	5.5	31.9
$T_{.j}$	21.2	17.6	18.5	19.1	19.7	20.9	117

Measurements	Formulae	Values
G	H6	117
N	COUNTA(B2:G5)	24
C.F.	POWER(K2,2)/K3	570.375
Σy_{ij}^2	SUM(POWER(B2:G5,2)	579.06
b	COUNTA(B2:G2)	6
t	COUNTA(A2:A5)	4
α	5/100	0.01
TSS	K5-K4	8.685
SST	(SUM(POWER(H2:H5,2))/K6)-K4	1.708333333
SSB	(SUM(POWER(B6:G6,2))/K7)-K4	2.415
SSE	K9-K10	4.561666667
df	M2 & “,” & M3	3,15
	M3 & “,” & M4	5,15
Decision T	IF(P2<Q2,”Do not reject H_0”, “Reject H_0”)	Do not reject H_0
Decision B	IF(P3<Q3,”Do not reject H_0”, “Reject H_0”)	Do not reject H_0

S.V.	d.f.	S.S.	M.S.	F_{cal}	F_{tab}
Treatment	3	1.708333	0.569444	1.872488	5.417
Blocks	5	2.415	0.483	1.588235	4.556
Error	15	4.561667	0.304111		
Total	23	8.685			

Step 3: Level of Significance and Critical Value

Level of significance $\alpha = 0.01$

Critical Value $F_{\alpha,(t-1),(t-1)(r-1)} = 5.417,\ F_{\alpha,(r-1),(t-1)(r-1)} = 4.556$

Step 4: Decision

Since $F_{cal} < F_{tab}$ for both block and treatments, do not reject both H_0.

Step 5: Conclusion

There is not any significant difference between the blocks and treatments.

Experiment 18: Latin Square Design

Let A, H, D and L represent Acer, HP, Dell and Lenovo laptops and the following information represents their operating time in hours before charge is required.

A 4.2	H 4.8	D 4.2	L 6.2
L 4.6	A 5.9	H 4.8	D 5.2
H 5.4	D 5.6	L 5.6	A 4.8
D 4.1	L 5.7	A 4.2	H 4.3

Carryout analysis of the design at 5% level of significance. (LSD)

Working formula

Mathematical model: $y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + e_{ijk}$

Where y_{ijk} = i^{th} row and j^{th} column receiving k^{th} treatment; I = 1, 2, ..., m; j = 1, 2, ..., m; k = 1, 2...m

Step 1: Setting up Hypothesis.

Null Hypothesis H_{0R}: $\mu_{1..} = \mu_{2..} = \mu_{3..} = \ldots = \mu_{m..}$ (There is no significant difference between rows)

Alternative Hypothesis H_{1R}: At least one $\mu_{i..}$ is different. I = 1, 2, 3……. M (There is at least one significant difference between rows)

Null Hypothesis H_{0C}: $\mu_{.1.} = \mu_{.2.} = \mu_{.3.} = \ldots = \mu_{.m.}$ (There is no significant difference between columns)

Alternative Hypothesis H_{1C}: At least one $\mu_{.j.}$ is different. J = 1, 2, 3……. M (There is at least one significant difference between columns)

Null Hypothesis H_{0T}: $\mu_{..1} = \mu_{..2} = \mu_{..3} = \ldots = \mu_{..m}$ (There is no significant difference between treatments)

Alternative Hypothesis H_{1T}: At least one $\mu_{..k}$ is different. K = 1, 2, 3……. M (There is at least one significant difference between treatments)

Step 2: Test statistic

F_R = MSR/MSE, F_C = MSC/MSE, F_T = MST/MSE, where MSR = SSR/ (m – 1), MSC = SSC/ (m – 1), MST = SST/ (m – 1), MSE = SSE / (m-1) (m-2)

$$TSS = \sum_{i,j,k}^{m} y_{ij}^2 - C.F.$$

$$SSR = \frac{\sum_{i=1}^{m} T_{i..}^2}{m} - C.F. \quad where C.F. = \frac{G^2}{N}$$

$$SSC = \frac{\sum_{j=1}^{m} T_{.j.}^2}{m} - C.F$$

$$SST = \frac{\sum_{k=1}^{m} T_{..k}^2}{m} - C.F$$

SSE = TSS – SSR-SSC-SST

Rows	Columns				$T_{i..}$
	I	II	III	IV	
I	4.2	4.8	4.2	6.2	19.4
II	4.6	5.9	4.8	5.2	20.5
III	5.4	5.6	5.6	4.8	21.4
IV	4.1	5.7	4.2	4.3	18.3
$T_{.j.}$	18.3	22	18.8	20.5	79.6

Treatments	I	II	III	IV	T..k
A	4.2	5.9	4.2	4.8	19.1
D	4.1	5.6	4.2	5.2	19.1
H	5.4	4.8	4.8	4.3	19.3
L	4.6	5.7	5.6	6.2	22.1
Total					79.6

Measurements	Formulae	Values
G	L6	79.6
N	COUNTA(H2:K5)	16
C.F.	POWER(O2,2)/O3	396.01
Σy_{ijk}^2	SUM(POWER(H2:K5,2))	403.16
m	COUNTA(H2:K2)	4
α	5/100	0.05
TSS	O5-O4	7.15
SSR	SUM(POWER(F3:F6,2))/O6-O4	1.355
SSC	SUM(POWER(B7:E7,2))/O6-O4	2.135
SST	SUM(POWER(L2:L5,2))/O6-O4	1.62
SSE	O8-O9-O10-O11	2.04
Decision R	IF(T2<U2,"Do not Reject H_0", "Reject H_0")	Do not Reject H_0

Decision C	IF(T3<U3,"Do not Reject H_0", "Reject H_0")	Do not Reject H_0
Decision T	IF(T4<U4,"Do not Reject H_0", "Reject H_0")	Do not Reject H_0

S.V.	d.f.	S.S	M.S.	F_{cal}	F_{tab}
Row	3	1.355	0.451667	1.328431	4.757
Column	3	2.135	0.711667	2.093137	4.757
Treatment	3	1.62	0.54	1.588235	4.757
Error	6	2.04	0.34		
Total	15	7.15			

Step 3: Level of Significance and Critical Value

Level of Significance α = 0.05

Critical Value $F_{\alpha,(m-1),(m-1)(m-2)} = 4.757$

Step 4: Decision:

Since $F_{cal} < F_{tab}$ for rows, columns, and treatments, Reject H_0 for all three.

Step 5: Conclusion

There is no significant difference in rows, columns, and treatments.

Unit 4: Multiple Correlation and Regression

Experiment 19: Regression and Correlations.

A developer of food for pig wishes to determine what relationship exists among ‘age of a pig’ when it starts receiving a newly developed food supplement, the initial weight of the pig and the amount of weight it gains in a week period with the food supplement. The following information is the result of a study of eight piglets.

Piglet No.	Initial wt. (lbs.) (x_1)	Initial age (weeks) (x_2)	Wt. gain (y)
1	39	8	7
2	52	6	6
3	49	7	8

4	46	12	10
5	61	9	9
6	35	6	5
7	25	7	3
8	55	5	4

(i) Determine the least square equation that best describes these three variables.
(ii) Calculate the standard error.
(iii) How much gain in weight of a pig in a week can we expect with the food supplement if it were 9 weeks old and weighted 49 lbs.
(iv) Test the significance of regression coefficient and overall fit of the regression equation.
(v) Conduct residual analysis.
(vi) Determine partial correlations, multiple correlation and coefficient of multiple determination. Interpret the results.

Working Formula:

(i) The multiple regression equation of Y on X_1 and X_2 is

$Y= a + b_1X_1 + b_2X_2$ ………………………... (1), using LSM, we get the normal equations are,

$\Sigma Y= n a + b_1\Sigma X_1 + b_2\Sigma X_2$……………….. (2)

$\Sigma YX_1= a\Sigma X_1 + b_1\Sigma X_1^2 + b_2\Sigma X_2X_1$………. (3)

$\Sigma YX_2= a\Sigma X_2 + b_1\Sigma X_1X_2 + b_2\Sigma X_2^2$………. (4)

On solving equations 2,3,4 we get values of a, b1 and b2, after putting the values we get the fitted regression equation is $Y= a + b_1X_1 + b_2X_2$

(ii)Standard Error $(S.E) = \sqrt{\frac{SSE}{n-k-1}}$

$$SSE = \sum\left(Y - \hat{Y}\right)^2 = \sum Y^2 - a\sum Y - b_1\sum X_1Y - b_2\sum X_2Y.$$

(iii)$\hat{y} = a + b_1X_1 + b_2X_2$

(iv)To test the significance of the regression coefficient, we have the following procedure

Step 1: Setting up Hypothesis.

Null hypothesis: H_0: $\beta_1=\beta_2$, there is no linear relationship between dependent variable Y and independent variables X_1 and X_2.

Alternative Hypothesis: H_1: At least one β_i is different from zero.

Step 2: Test Statistics:

$F = \frac{MSR}{MSE} \sim F$ distribution with (k, n-k-1), where k is the no. of independent variables.

MSR = mean sum of square due to regression =SSR/k

MSE = mean sum square due to error=SSE/(n-k-1)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.916[a]	.840	.776	1.161	.840	13.092	2	5	.010

a. Predictors: (Constant), Initial Age, InitialWT

b. Dependent Variable: WtGain

Correlations

		WtGain	InitialWT	InitialAge
Pearson Correlation	WtGain	1.000	.514	.794
	InitialWT	.514	1.000	.072
	InitialAge	.794	.072	1.000
Sig. (1-tailed)	WtGain	.	.096	.009
	InitialWT	.096	.	.433
	InitialAge	.009	.433	.
N	WtGain	8	8	8
	InitialWT	8	8	8
	InitialAge	8	8	8

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35.266	2	17.633	13.092	.010[b]
	Residual	6.734	5	1.347		
	Total	42.000	7			

a. Dependent Variable: WtGain

b. Predictors: (Constant), InitialAge, InitialWT

Coefficients

Model		Unstandardized Coefficients: B	Std. Error	Standardized Coefficients: Beta	t	Sig.	95.0% Confidence Interval for B: Lower Bound	Upper Bound	Correlations: Zero-order	Partial	Part
1	(Constant)	-4.190	2.222		-1.886	.118	-9.902	1.521			
	InitialWT	.096	.038	.459	2.555	.051	-.001	.193	.514	.753	.458
	InitialAge	.846	.200	.761	4.238	.008	.333	1.359	.794	.884	.759

a. Dependent Variable: WtGain

From the tables,

1. The least square equation that best describes the 3 variables is: $y = (-4.190) + (0.096) X_1 + (0.846) X_2$

2. The standard Error is 1.161.
3. Given, X_1 = 49 lbs. and X_2 = 9 weeks.
   Y = (-4.190) +(0.096) * 49+(0.846) *9 = 8.128lbs.
   Thus, we can expect 8.128 lbs. gain in weight in a week.
4. For testing null hypothesis $B_0 = 0$: since p value = 0.118, it is insignificant.
   For testing null hypothesis $B_1 = 0$: since p value = 0.051, it is insignificant.
   For testing null hypothesis $B_2 = 0$: since p value = 0.008, it is significant.
   For testing null hypothesis: overall fit of the regression coefficients = 0, since here the p value = 0.010 < 0.05 for F test, which indicates overall fit significant.
5. Residual statistics is as follows:

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4.13	10.38	6.50	2.245	8
Residual	-1.323	1.562	.000	.981	8
Std. Predicted Value	-1.055	1.727	.000	1.000	8
Std. Residual	-1.140	1.346	.000	.845	8

a. Dependent Variable: WtGain

6. For Correlations
   a. Pearson correlation between Wt. Gain and Initial Weight (r_{12}) = 0.514
   b. Pearson correlation between Wt. Gain and Initial Age (r_{13}) = 0.794
   For multiple correlations
   a. Multiple correlation coefficient (R) = 0.916
   For coefficient of multiple determination
   a. R^2 = 0.840
   b. Adjusted R^2 = 0.776
   For Partial Correlations:
   a. Partial correlation between Wt. Gain and Initial Weight, controlling for Initial Age ($r_{12.3}$)= 0.753
   b. Partial correlation between Wt. Gain and Initial Age, controlling for Initial Weight ($r_{13.2}$)= 0.884
   For Interpretation:
   a. The multiple correlation coefficient (R) of 0.916 indicates a strong positive relationship between the predictors (InitialWT and InitialAge) and the response variable (WtGain).
   b. The coefficient of multiple determination (R^2) of 0.840 means that 84% of the variance in WtGain is explained by the two predictors.
   c. The high partial correlations indicate that both InitialWT and InitialAge individually have strong relationships with WtGain when controlling for the other variable.

d. The regression model is statistically significant, as indicated by the significance tests of the coefficients and the F-statistic.