Q1) Weight (kg) of sample of 35 workers in a Company are found as follows.

56	78	65	49	63	58	70	61	53	69
57	69	90	78	64	71	65	49	56	59
50	57	62	70	68	54	49	87	68	71
55	78	80	73	85					

a) Is mean weight of workers 64 kg at 1% level of significance?
b) Find 99% confidence limit for mean weight of workers in Company.

⇒

Working Steps:

1. Calculate the sample mean ($\bar{x}$):
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

2. Calculate the sample standard deviation (s);
$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

3. Calculate the standard error of the mean
$$SE = S/\sqrt{n}$$

4. Calculate the z
$$z = \frac{\bar{x} - \mu_0}{SE(x)}$$

5. $P = \text{prob}(z > |z\ \text{cal}|)$

6. Compare p value with given level of significance then reject if $p < \alpha$ accept otherwise.

7. For Confidence limit you can use following formula as $\bar{x} \pm \left(z_\alpha \times \frac{S}{\sqrt{6}}\right)$.

(AR)
9nA1

a) Problem to test:

$H_0: \mu = 64$

$H_1: \mu \neq 64$

one sample t-test

data : Weight

$t = 0.71176$, $df = 34$, p-value $= 0.4815$

Alternative hypothesis: True mean is not equal to 64

Sample estimates:

mean of x is 65.34286

From above result,

$2p = 0.4815 > \alpha = 0.01$

Accept, H_0 at $\alpha = 0.01$

Therefore,

Weight of workers 64 kg at 1% level of significance.

b) 99% confidence limit for mean weight of workers in Company is 60.19525 to 70.49047.

AR

Q2) Marks secured in statistics by a sample of 33 and 36 student of Section A and section B are found as follows

Section A

37	29	50	58	24	41	25	49	56	49	17
56	48	35	29	43	28	7	21	33	40	38
50	17	32	50	18	44	49	47	38	51	51

Section B

38	57	37	43	55	53	48	40	50	34	24	15
46	58	55	49	53	58	50	37	44	52	40	56
51	15	10	37	51	38	42	57	13	7	32	42

i) Is there any significant difference in mean marks in stat. of Section A and Section B at 1% level of significance.

ii) Find 99% confidience limit for difference of mean marks.

Working steps:

1. Calculate the sample means $(\overline{x})$:

$$\overline{x}_A = \frac{\sum_{i=1}^{n_A} x_i}{n_A}, \quad \overline{x}_B = \frac{\sum_{i=1}^{n_B} x_i}{n_B}$$

2. Calculate the Sample standard deviation (S) :

$$S_A = \sqrt{\frac{\sum_{i=1}^{n_A} (x_i - \overline{x}_A)^2}{n_A - 1}}, \quad S_B = \sqrt{\frac{\sum_{i=1}^{n_B} (x_i - \overline{x}_B)^2}{n_B - 1}}$$

3. Calculate the standard error

$$S.E. = \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}$$

4. Calculate the z

$$z = \frac{\overline{x}_A - \overline{x}_B}{S.E.}$$

5. $P = prob(z > |z_{cal}|)$

6. Compare P value with given level of significance then reject if $P < \alpha$ accept otherwise

7. For Confidence limit

$$\overline{x}_A - \overline{x}_B \pm S.E.\, z_\alpha$$

QNA 2.

Problem to Test:
$H_0: \mu_A = \mu_B$
$H_1: \mu_A \neq \mu_B$

i) Welch Two Sample t-test

data: Section A and Section B

t = -0.9329, df = 67, p-value = 0.3542.

Alternative Hypothesis is : True difference in means is not equal to 0.

Sample estimates:

mean of x	mean of y
38.18182	41.30556

ii) Confidence limit for difference of mean marks
99 percent C.I:
-12.001101 to 5.753625.

Conclusion: There is no significant difference in the mean marks in statistics between section A and section B at 1 % level of significance. Also we are 99 % Confident that the true difference in mean marks falls within the calculated confidence interval.

$\boxed{\begin{array}{l} \because \text{P-value} > \alpha \text{ then accept } H_0 \\ 2 \times 0.3542 > 0.01 \end{array}}$

(AR)

Q3) Following information represents result of a sample of 32 students of B.Sc CSit ii semester

P P P F P F P P F F P P P P
F P P P F P P P F P P F P P
P F P P

i) Is pass percentage of B.SC csit ii semester 80% ? use 5% level of significance.

ii) On basis of sample pass percentage what is sample size required to study result of B.SC CSIT ii semester students at 95% confidence limit with 5% margin of error

→ Working Steps:

1. Calculate the sample pass percentage (P) using

$$\frac{\text{no. of pass}}{\text{Total sample size}} \times 100.$$

2. Calculate the Standard Error

$$S.E = \sqrt{\frac{p \times (1-P)}{n}}$$

3. Calculate, $Z = \frac{p - P}{SE}$

4. Calculate, $p = \text{prob}(Z > |Z_{cal}|)$

5. Compare p with given level of significance then reject if $p < \alpha$. Accept otherwise.

6. Calculate sample size,

$$n = \frac{Z_\alpha^2 \times p \times (1-p)}{e^2}$$

Where, e = margin of error.

Problem to test:

$H_0: P = 0.8$

$H_1: P \neq 0.8$

QN.3

i). 1-Sample proportion test with continuity correction

data: 23 out of 32, null probability 0.8

X-squared = 0.86133, df = 1, p-value = 0.3534

alternative hypothessis: true p is not equal to 0.8.

Sample estimates:

p is 0.71875

Here, $p > \alpha = 0.05$

So, H_0 is true. Therefore, pass % of BSCCSIT ii Semester 80 %.

ii). Sample size required to study result of B.SC CSIT ii semester student is 310.6294 ≈ 311.

95 % Confidence limit with 5% margin of error is 0.5302034 to 0.8560142.

(AR)

Q4) Following information represent result of a sample of 32 BIT i semester student and 36 BSC.CSIT i semester Students.

BIT i semester

P	F	P	P	P	P	F	F
F	F	P	P	P	P	P	P
P	P	F	F	P	P	F	P
P	F	F	P	P	F	P	P

BSC CSIT i semester

P	P	F	P	P	P	P	F	P
P	P	F	P	F	P	P	P	P
F	P	P	P	F	F	P	P	P
P	F	P	P	P	F	P	P	P

Is there any significant difference in pass percentage of BIT i semester and BSC CSIT i semester students ? use 5% level of significance.

⇒ Working steps:

0. Calculate the sample pass proportion ~~percentage~~ (P) for each group using

$$\frac{\text{no. of Pass}}{\text{Total Sample size}} \cancel{\times 100\%}$$

1. Calcalate the overall proportion of pass across both groups

$$p = \frac{\text{total no. of passed Students}}{\text{total no. of students}}$$

2. Calculate the standard error of the difference between proportions

$$S.E = \sqrt{p \times (1-p) \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where,
n_1 = total students in BIT i sem
n_2 = total students in CSIT i sem

3. Calculate, $Z = \frac{P_1 - P_2}{S.E.}$

4. Calculate, $p = \text{prob}(z > |z\,\text{cal}|)$

(Q.N.4)

5. Compare p value with given level of significance then reject if $p < \alpha$ accept otherwise.

Problem to Test:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$.

Let, Variables: CSITresult and BITresult

BITresult

F	P
11	21

CSITresult

F	P
9	27

2-Sample test for equality of proportions with Continuity Correction.

data: c(21,27) out of c(32,36)

X-squared = 0.3367, df=1, p-value = 0.5617

Alternative hypothesis is: two-sided

Sample estimates:

prop 1	prop 2
0.65625	0.75000

95% Confidence Interval: −0.340262 to 0.152762

From above result,

$2P = 2 \times 0.5617 > \alpha = 0.05$

Accept H_0 at $\alpha = 0.05$

Conclusion: There is no any significant difference in pass percentage of BIT I sem and BSC. CSIT I sem students.

(AR)

Q5. Marks secured by a sample of 22 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, 49, 27, 35, 41.

i) Is average marks in statistics I 30 at 5% level of significance using parametric test.

ii) Obtain 95% confidence limit for average marks of statistics I for all students appeared in examination.

iii) On basis of sample standard deviation obtained from marks of students in statistics I. What is sample size required for the study of marks distribution of students at 5% level of significance with 10% margin of error.

⇒ working steps:

1. Calculate sample mean

$$\bar{x} = \frac{\Sigma x}{n}$$

2. Calculate standard error of mean

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

3. Compute, $Z = \dfrac{\bar{x} - \mu}{SE(\bar{x})}$

4. $p = prob(Z > |Z_{cal}|)$

5. If $p < \alpha$ (one tail) or $2p < \alpha$ (two tail) then Reject H_0 at (α) level of significance. Accept otherwise.

6. Sample size $= \dfrac{Z_\alpha^2 s^2}{e^2}$

7. For confidence interval, use

$$\bar{x} \pm SE(\bar{x}) Z_\alpha.$$

(QN-5)

Problem to Test:
$H_0: \mu = 30$
$H_1: \mu \neq 30$

(AR)

i) one Sample t-test

data: marks
$t = 3.2129$, $df = 21$, p-value $= 0.004177$
Alternative hypothesis is: true mean is not equal to 30.
Sample estimates:
mean of x is 38.31818

From above result,
$2P = 0.008354 < \alpha = 0.05$
Reject H_0 at $\alpha = 0.05$
So, Average marks in statistics is not 30 at 5% level of significance.

ii) 95 % Confidence limit is 32.93404 to 43.70232.

iii) Sample standard deviation $(S) = 12.14353$
Sample size $(n) = 56650.28 \approx 56650$.

Q6) 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

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

i) Is mean marks of section A and section B identical at 1% level of significance?

ii) Obtain 99% confidence limit for difference of mean.

⇒ Working steps:

1. Calculate sample means for the given samples.

$$\bar{x}_i = \frac{\sum x_i}{n_i}, \text{ where, } i = A, B$$

2. Calculate Standard error of difference between the means

$$SE(\bar{x}_A - \bar{x}_B) = \sqrt{S^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

Where,

$$S^2 = \frac{n_A s_A^2 + n_B s_B^2}{n_A + n_B - 2} = \frac{\sum(x_A - \bar{x}_A)^2 + \sum(x_B - \bar{x}_B)^2}{n_A + n_B - 2}$$

$$= \frac{\sum x_A^2 - n_A \bar{x}_A^2 + \sum x_B^2 - n_B \bar{x}_B^2}{n_A + n_B - 2}$$

3. Compute Z

$$Z = \frac{\bar{x}_A - \bar{x}_B}{SE(\bar{x}_A - \bar{x}_B)}$$

4. $P = prob(Z > |Z_{cal}|)$

5. Reject H_0 at α level of significance if $P < \alpha$ (one tail) or $2P \cancel{<} \alpha$ (two tail). Accept otherwise.

6. For confidence limit

$$(\bar{x}_A - \bar{x}_B) \pm SE(\bar{x}_A - \bar{x}_B)\, Z_\alpha (n_A + n_B - 2)$$

Q.N.6

Problem To Test:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$.

i) Welch Two Sample t-test

data: sectionAmarks and sectionBmarks

t = 1.1608, df = 26.497, p-value = 0.2561

Alternative hypothesis: true difference in means is not equal to 0.

Sample estimates:

mean of x	mean of y
38.71429	33.73333

From above result,

$2P > \alpha = 0.01$

Accept H_0 at $\alpha = 0.01$

So, Mean marks of section A and section B is identical at 1% level of significance.

ii) 99% Confidence limit for difference of mean is −6.924827 to 16.886732.

AR

Q7) Marks secured by a sample of 15 students of 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 parametric test at 1% level of significance.

⇒ Sol^n

Working Steps

1. Calculate difference between observations in the i^{th} sample.
   $d_i = x_i - y_i$ Where, x & y are 2 related samples.

2. Find sample mean of the differences
   $$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

3. Find sample standard deviation of the differences
   $$Sd = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} d^2 - n\bar{d}^2}{n-1}}$$

4. Calculate test statistic
   $$t = \frac{\bar{d}}{Sd/\sqrt{n}}$$

5. $P = prob(z > |z_{cal}|)$

6. Reject H_0 at α level of significance if $P < \alpha$ (one tail) or $2P < \alpha$ (two tail). Accept otherwise.

Problem To Test: (QN 7)

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$.

Pair t-test

data: markstesti and markstestii

$t = -0.27678$, $df = 14$, p-value $= 0.786$

Alternative hypothesis: true mean difference is not equal to 0.

Sample estimates:

mean difference is -0.2666667

From above result,

$P = 0.786 > \alpha = 0.01$

H_0 is accept at $\alpha = 0.01$

Conclusion: There is no improvement in marks in test II as compared to test I.

Q8) 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.

→

Working Steps (for i)

1. Count the no. of runs (r) of the symbols assigned to the sample items.

* For small sample size (n_1 or $n_2 \leq 20$)

Where, n_1 & n_2 : No. of observations

2. For α level of significance, obtain critical values $\underline{r}$ and $\bar{r}$ from run table.
3. Reject Ho at α level of significance if $r \notin (\underline{r}, \bar{r})$, Accept otherwise.

* For large sample size (n_1 or $n_2 > 20$)

a. Calculate μ_r and σ_r

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 \quad \text{and} \quad \sigma_r = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

b. Calculate Z

$$Z = \frac{r - \mu_r}{\sigma_r}$$

For α other than 5%, even if small size is small, use Z-test as $Z = \frac{|r-\mu_r| - 0.5}{\sigma_r}$

c. At α level of significance, find critical (tabulated) value Z_α

d. Compute $|Z|$ with Z_α

- Reject Ho at α level of significance if $|Z| > Z_\alpha$. Accept otherwise.

Working steps (for ii)

1. Calculate x_0

$x_0 = \min\{n_1, n_2\}$, where, n_1 and n_2 are no. of observations belonging to '2' groups from n sample s.t. $n_1 + n_2 = n$

2. calculate p for small sample size $[n \leq 25]$

$$P = prob(X \leq x_0) = \sum_{x=0}^{x_0} C(n,x)\, p^x q^{n-x}$$

where, p is defined is null hypothesis and $q = 1-p$.

3. Reject H_0 at α level of significance if $P < \alpha$ (one-tail) or $2P < \alpha$ (two tail). Accept otherwise.

* For large sample size $(n > 25)$

$x_0 \sim N(\mu_x, \sigma_x^2)$

a. Calculate μ_x and σ_x

$\mu_x = np$ and $\sigma_x = \sqrt{npq}$

b. Calculate Z using continuity correction

$$Z = \frac{(x_0 \pm 0.5) - np}{\sqrt{npq}}, \quad \begin{cases} \text{use } +0.5 \text{ if } x_0 < np \\ \text{use } -0.5 \text{ if } x_0 > np \end{cases}$$

c. Find critical value Z_α at α level of significance.

d. Reject H_0 at α level of significance if $Z > Z_\alpha$. Accept otherwise.

→ problem to test (for i)

H_0: Outcomes are in random order

H_1: Outcomes are not in random order

→ Output From PSPP

NPAR TEST
/RUNS (MEDIAN) = Outcomes.

Runs Test	outcomes
Test value (median)	1.00
Cases < Test value	14
Cases ≥ Test value	16
Total Cases	30
Number of Runs	19
Z	0.96
Asymp. Sig. (2-tailed)	0.338

// End

→ From above result
$2P = 2 \times 0.338 = 0.676 > \alpha = 0.05$
Accept H_0 at α level of significance

Conclusion: outcomes are in random order.

ii) problem to test
$H_0: p = 1/2$
$H_1: p \neq 1/2$

→ Output From R-studio

Exact binomial test

data: 16 and 30
number of successes = 16, number of trials = 30, p-value = 0.8555
alternative hypothesis: true probability of success is not equal to 0.5

95 % Confidence Interval:
0.3432552   0.7165812

Sample estimates:
Probability of success
0.5333333

From above result

$2P = 2 \times 0.8555 = 1.711 > \alpha\ 0.01$

Accept H_0 at α level of significance.

Conclusion :

Coin is not biased.

Q9) 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, 49, 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 Kolmugorov Smirnov Test
iii) Are marks uniformly distributed? use chi square test using 5% level of significance.

→

working steps (for i)

1. Calculate median from given data then assign a symbol
   say A if observation > Md
   B if observation < md
   omit if observation = md.
2. Calculate the number of run (r).
3. Calculate no. of A and B and denoted by n_1 and n_2.
4. Calculate popl^n mean using
$$\mu = \frac{2n_1n_2}{n_1+n_2}+1$$
5. Calculate popl^n SD(σ) by using
$$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$
6. Calculate Z statistics
$$Z_{calc} = \frac{r-\mu}{\sigma}$$
7. Calculate P using
$$p = prob(Z > |Z_{cal}|)$$
8. Compare p value within given level of significance then reject if $p < \alpha$: Accept otherwise.

working steps (for ii)

1. Calculate Cfo and relative Cfo (Cfo/n) for given frequency.
   Where, fo = given frequency
2. Calculate probability a/c to pop/n distribution then obtain frequency (fe) then Cfe and relative Cfe (Cfe/n).
3. Calculate absolute value of difference of Cfo/n and Cfe/n
   It is test statistic
   $$D_{calc} = \max\left|\frac{Cfo}{n} - \frac{Cfe}{n}\right|$$
4. Calculate P using
   $$P = P(D_0 > |D_{cal}|)$$
5. Compare P value with given level of significance then reject if $P < \alpha$. Accept otherwise.

working steps (for iii)

1. Calculate expected frequency (E_i) using
   $$E_i = NP_i$$
   Where, P_i is probability of happening, N is no. of data given.
2. Calculate test statistic
   $$\chi = \sum_i \frac{(O_i - E_i)^2}{E_i}$$ where, O_i = Observed frequency Which will be given.
3. Calculated P using,
   $$P = Prob(\chi > |\chi_{cal}|)$$
4. Compare P with given level of significance then reject if $P < \alpha$. Accept otherwise.

For (I)

problem to test:

H_0: Samples are random

H_1: Samples are not random

Runs Test

	marks
Test value (median)	38.00
Cases < Test Value	14
Cases ≥ Test Value	18
Total Cases	32
Number of Runs	16
Z	-0.09
Asymp. Sig. (2-tailed)	0.927

From above result

$P = 0.927 > \alpha = 0.05$

Accept H_0 at $\alpha = 5\%$.

Conclusion: Samples are random

For (II)

Problem to test:

H_0: Marks are uniformly distributed

H_1: Marks are not uniformly distributed

one Sample kolmogorov-smirnov Test

		marks
N		32
uniform Parameters	Minimum	10.00
	Maximum	56.00
Most Extreme Differences	Absolute	0.27
	Positive	0.06
	Negative	-0.27
Kolmogorov-smirnov Z		1.54
Asymp. Sig. (2-tailed)		0.012

From above result

$P = 0.012 < 0.05$

Reject H_0 at $\alpha = 5\%$.

Conclusion: Marks are not uniformly distributed.

For (iii)

Problem to test:

H_0: Marks are uniformly distributed.

H_1: Marks are not uniformly distributed

Test Statistics

	Chi-square	df	Asymp. Sig.
marks	9.25	21	0.987

Conclusion: Marks are uniformly distributed.

10) 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

Section A 34 48 21 52 31 43 29 37 24 52 49 34 40 48

Section B 11 53 27 38 47 50 26 38 44 33 27 33 41 10 2

Is 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

→ <u>working steps</u> for (i and ii)

1. Calculate difference d of related samples.
2. Rank d irrespective of sign omit if $d = 0$. If two or more d are same then assign average rank.
3. Calculate rank of d according to sign to get $S(+)$ & $S(-)$
4. Calculate test statistic

$$T = \min\{S(+) \text{ and } S(-)\}$$

5. Calculate P

$$P = prob(T > |T_{cal}|)$$

6. Compare P with given level of significance then reject if $P < \alpha$. Accept otherwise.

<u>working steps</u> (for iii)

1. Calculate cf_A and relative cf_A (cf_A/n_A) for given frequency.
2. Calculate probability according to pop^n distribution then obtain frequency (fe) the cfe and relative cfe (cfe/n).
3. Calculate test statistic

$$D = \max\left|\frac{cf_A}{n_A} - \frac{cf_B}{n_B}\right|$$

CamScanner

4. ~~Comp~~ Calculate P using

$$P = prob(D > |D_{cal}|)$$

5. Compare P with given level of significance then reject if $P < \alpha$ accept otherwise.

→ For (i)

Problem to Test

$H_0 : Md_A = Md_B$
$H_1 : Md_A \neq Md_B$

Wilcoxon rank sum test with continuity correction

data: sectionAmarks and sectionBmarks

W = 128, p-value = 0.3258

Alternative hypothesis : true location shift is not equal to zero (0).

From above result:

$P = 0.325 > \alpha = 0.05$

Accept H_0 at $\alpha = 5\%$.

Conclusion: Yes, the median marks of section A and section B is identical.

For (ii)

problem to test:

$H_0 : Md_A = Md_B$
$H_1 : Md_A \neq Md_B$

Exact two-sample Kolmogorov-Smirov Test

data: sectionAmarks and sectionBmarks

D = 0.25714, P-value = 0.5416

Alternative hypothesis : two sided

From above result,

$P = 0.5416 > \alpha = 0.05$

Accept H_0 at $\alpha = 0.05$.

Conclusion: The median of section A and Section B are identical.

(AR)

Q4) Following information are 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
I	Female	Black
J	Female	Brown
K	Female	Black
L	male	Black
N	Female	Blue
O	Female	Brown
P	Male	Black
Q	Female	Black
R	male	Brown
S	Female	Black
T	Female	Black
U	Male	Brown
V		

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

→ Working Steps

1. Calculate Expected frequency (E_i) using
$E_i = Np_i$
Where, P_i = Probability of happening
N = No. of data given

2. Calculate test statistic
$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$
Where O_i = observed frequency which will be given

3. Calculate p using
$p = \text{Prob}(\chi^2 > |\chi_{cal}|)$

4. Compare P with given level of significance then reject if $P < \alpha$. Accept otherwise.

→ Problem to test

$H_0: O_i = E_i$

$H_1: O_i \neq E_i$

Eye color / Gender	Bk	Bl	Br
F	7	2	2
M	5	0	4

Pearson's Chi-Squared Test

data: table (Gender, EyeColor)

X squared = 2.8263, df = 2, p-value = 0.2431

From above result

$P = 0.24 > 0.05$

Accept H_0 at $\alpha = 5\%$

Conclusion:

There is no any association between gender and eye color.

Q12) Marks secured by a sample of 15 students of 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	05	13	17	6	9	18	14	10	8
Test II	14	5	17	13	12	18	09	10	18	12	3	14	16	16	8

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

→ Working step

1. Calculate difference d of related samples.
2. Rank d irrespective of sign omit if $d = 0$. If two or more d are same then assign average rank.
3. Calculate rank of d according to sign to get $S\oplus$ and $S\ominus$
4. Calculate test statistic using
   $T = \min\{S\oplus \text{ and } S\ominus\}$
5. Calculate p using
   $p = prob(T \geq |T_{cal}|)$
6. Compare p with given level of significance than reject if $p < \alpha$; Accept otherwise.

Problem to test

$H_0 : Md_I = Md_{II}$

$H_1 : Md_I < Md_{II}$

wilxcon signal rank test

data: Test I marks and Test II marks

$V = 48$, p-value $= 0.4002$

Alternative hypothesis = true location shift is less than 0.

CamScanner

From above result

$$p = 0.40 > 0.05$$

Accept H_0 at $\alpha = 5\%$

Conclusion:

There is no improvement in marks in test II as Compared to test I.

Q 13) 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.

Diet \ Cow	I	II	III	IV	V	VI	VII	VIII	IX
D_1	I	I	D	I	D	I	I	D	I
D_2	D	D	I	D	I	D	D	I	I
D_3	I	D	I	D	D	I	I	D	D
D_4	I	I	I	D	D	I	I	D	I

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

→ Working steps

For k related samples each of size 'n' selected from dichotomized related population

1. Sum of all the y (positives) according to treatment to get R_i (row wise) and according to objects to get C_j (column wise) where $i = 1, 2, 3, ---, k$ and $j = 1, 2, 3, ---, n$.

2. Calculate, $\sum_{i=1}^{k} R_i$, $\sum_{i=1}^{k} R_i^2$, $\sum_{j=1}^{n} C_j$, and $\sum_{j=1}^{n} C_j^2$

3. Compute the test statistic

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

4. Reject H_0 at α level of significance if $Q > \chi^2_{\alpha}(k-1)$. Accept otherwise.

- For p-value, reject H_0 at α level of significance if $p < \alpha$ (one tail) or $2p < \alpha$ (two tail). Accept otherwise.

Problem to test

$H_0: P_1 = P_2 = P_3 = P_4$

H_1: At least one P_i is different, $i = 1, 2, 3, 4$

Output from RStudio

D1		D2		D3		D4	
D	I	D	I	D	I	D	I
3	6	5	4	5	4	3	6

4 - Sample test for equality of proportions
Without Continuity Correction

data: c(6,4,4,6) out of (9,9,9,9)

X-squared = 1.8, df = 3, p-value = 0.6149

alternative hypothesis: two-sided

Sample estimates:

prop 1	prop 2	prop 3	prop 4
0.6666667	0.4444444	0.4444444	0.6666667

// End.

From above result

$2P = 2 \times 0.6149 > \alpha = 1\% = 0.01$

Accept H_0 at $\alpha = 0.01$ level of significance.

Conclusion:

Diets are not effective at 1% level of significance.

AR

Q14) Following data represent the operating times in hours for four types of laptop 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	

Are operating time for all laptops equal at 5% level of significance use non parametric test?

→

Working steps

1. Calculate total no. of sample size (n)
$$n = n_1 + n_2 + n_3 + n_4$$

2. Rank these n observations in ascending order.
If two or more observations are equal then assign average rank.

3. Calculate the number of times rank (t)

4. Calculate test statistic, H
$$H = \frac{\frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)}{1 - \sum \frac{(t_i^3 - t_i)}{n^3 - n}}$$

5. Calculate p using
$$p = prob(H > |H_{cal}|)$$

6. Compare p with given level of significance than reject H_0 if $p < \alpha$. Accept otherwise.

CS CamScanner

(QN 14)

→ Problem to Test

H_0 : $Md_1 = Md_2 = Md_3 = Md_4$

H_1 : At least one Md_i is different, $i = 1, 2, 3, 4$.

Kruskal-Wallis rank sum test

data: OperatingTime by laptop

Kruskal-Wallis chi-square = 1.0739, df = 3, p-value = 0.7834

From above result,

$P = 0.7834 > 0.05$

Accept H_0 at $\alpha = 0.05$.

Conclusion:

Yes, operating time for all laptops are equal.

Q.No.15) The scores of 7 students in Statistics II in three test are found as

Test \ Student	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 differente in marks in three test?

ii) Is there any significant difference in marks of seven students? Use non parametric test at 1% level of significance.

→ Working Steps

1. Rank k sample observations for each block separately from 1 to k in ascending order.
   - If two or more observations are same then assign average rank which is called tied.

2. Calculate the number of times rank is repeated (t_i).

3. Calculate 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

$$F_r = \frac{\frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1)}{1 - \sum \frac{(t_i^3 - t_i)}{n(k^3-k)}}$$

Where, n = no. of sample size
k = no. of independent sample

4. Calculate p using,

$$p = Prob(F_r > |F_{r(calc)}|)$$

5. Compare p with given level of significance then reject Ho if $p < \alpha$. Accept otherwise.

(QN15)

i)

Problem to test

H_0: $Md_1 = Md_2 = Md_3$

H_1: At least one Md_j, $j = 1, 2, 3$. is different

// output

/FRIEDMAN = TestI TestII TestIII

Ranks

	Mean Rank
Test I	2.29
Test II	2.00
Tes III	1.71

Test statistics

N	7
Chi-Square	1.12
df	2
Asymp. Sig.	0.565

from above result,

$2P = 2 \times 0.565 > \alpha = 0.01$

Accept H_0 at $\alpha = 0.01$.

Conclusion: No, there is not any significant difference in marks in three test ?

ii) problem to test

H_0: $Md_1 = Md_2 = Md_3 = Md_4 = Md_5 = Md_6 = Md_7$

H_1: At least one Md_j is different, $j = 1, 2, 3, 4, 5, 6, 7$

// output

/FRIEDMAN = A B C D E F G

Ranks

	mean Rank
A	4.83
B	5.17
C	2.00
D	3.83
E	3.17
F	5.00
G	4.00

Test Statistics

N	3
Chi-square	5.09
df	6
Asymp. Sig.	0.532

// End.

From aboved result,

$2P = 2 \times 0.532 > \alpha = 0.01$

Accept H_0 at $\alpha = 0.01$

Conclusion: No, there is not any significant difference in marks of seven students.

CamScanner