CBCS SCHEME

USN | | | | | | | | | |

SV

BCS301

Third Semester B.E./B.Tech. Degree Examination, June/July 2024
Mathematics for Computer Science

Time: 3 hrs. Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. VTU Formula Hand Book is permitted.
3. M : Marks , L: Bloom's level , C: Course outcomes.

		Module – 1	M	L	C
Q.1	a.	Obtain the mean and variance of Poisson distribution.	06	L2	CO2
	b.	Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) atleast one boy (iii) at most 2 girls. Assume equal probabilities for boys and girls.	07	L3	CO2
	c.	The length of telephone conversation in a booth has been an exponential distribution and found on an average to be 5 minutes. Find the probability that a random call made from this booth (i) ends less than 5 minutes (ii) between 5 and 10 minutes.	07	L2	CO2
		OR			
Q.2	a.	The probability distribution of a finite random variable X is given by (table below) (i) Find the value of k (ii) Variance (iii) $P(x \le 2)$	06	L2	CO1
	b.	The number of accidents in a year to taxi drivers in a city follows a Poisson distribution with mean 3. Out of 1000 taxi drivers find approximately number of drivers with (i) more than 3 accidents in a year (ii) at most 2 accidents in a year.	07	L3	CO2
	c.	The marks of 1000 students in an exam follows normal distribution with mean 70 and standard deviation 5. Find the students whose marks will be (i) less than 65 (ii) between 65 and 75. A(1) = 0.3413.	07	L3	CO2
		Module – 2			
Q.3	a.	Given the following joint distribution of the random variables X and Y. Find the corresponding marginal distribution. Also compute the covariance. (table below)	06	L3	CO2
	b.	A salesmen's territory consists of 3 cities A, B and C. He never sells in the same city for 2 consecutive days. If he sells in city A, then the next day he sells in city B. However if he sells in either B or C then the next day he is twice as likely to sell in city A as in the other city. In the long run how often does he sell in each of the cities.	07	L3	CO3
	c.	Show that $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix}$ is a regular stochastic matrix. Also find the associated unique fixed probability vector.	07	L2	CO2

Q.2 a. table:

X	-2	-1	0	1	2	3
P(X)	0.1	k	0.2	2k	0.3	k

Q.3 a. table:

X \ Y	1	3	9
2	1/8	1/24	1/12
4	1/4	1/4	0
6	1/8	1/24	1/12

		OR			
Q.4	a.	Define probability vector, regular stochastic matrix, fixed prob vector.	06	L1	CO3
	b.	The joint probability distribution of two discrete random variables X and Y is $f(x, y) = k(2x + y)$, where x and y are integers such that $0 \le x \le 2$, $0 \le y \le 3$. i) Find the value of the constant k. ii) Show that the random variables X and Y are dependent iii) Find $P(X \ge 1, Y \le 2)$.	07	L3	CO2
	c.	A fair coin is tossed thrice. The random variables X and Y are defined as X = 0 or 1 according as head or tail occurs on the first toss, y-number of heads. Compute c(X, Y)	07	L3	CO2
		Module – 3			
Q.5	a.	Define statistical hypothesis, null hypothesis, Type-I error and Type-II error.	06	L1	CO4
	b.	In 324 throws of a six faced die an even number turned up 181 times. Is it reasonable to think that the 'die' is an unbiased one at 99% level?	07	L2	CO4
	c.	Before an increase in excise duty on tea, 800 people out of sample of 1000 were consumers of tea. After the increase in duty, 800 people were consumers of tea in a sample of 1200 persons. Find whether there is significant decrease in the consumption of tea after the increase in duty at 1%. (One tailed test at 1% is 2.33).	07	L3	CO4
		OR			
Q.6	a.	A coin is tossed 1000 times and head turns up 540 times. Decide on the hypothesis that the coin is unbiased.	06	L2	CO4
	b.	In an exit poll enquiry it was revealed that 600 voters in one locality and 400 voters from another locality favoured 55% and 48% respectively a particular party to come to power. Test the hypothesis that there is a difference in the locality in respect of the opinion.	07	L3	CO4
	c.	A random sample for 1000 workers in company has mean wage of Rs.50 per day and standard deviation of Rs.15. Another sample of 1500 workers from another company has mean wage of Rs.45 per day and standard deviation of Rs.20. Does the mean rate of wages varies between the two companies at 95% confidence limit?	07	L3	CO4
		Module – 4			
Q.7	a.	The mean life time of a sample of 25 bulbs is found as 1550 hrs with standard deviation of 120 hours. The company manufacturing the bulbs claims that the average life of their bulbs is 1600 hrs. Is the claim acceptable at 5% level of significance?	06	L3	CO4
	b.	The two independent samples of eight and seven items respectively had the following values of the variable: Sample 1: 9, 11, 13, 11, 15, 9, 12, 14 Sample 2: 10, 12, 10, 14, 9, 8, 10 Do the two estimates of population variance differ significantly at 5% level of significance? F at 5% $(V_1 = 7, V_2 = 6) = 4.21$.	07	L3	CO4
	c.	Table gives the number of aircraft accidents that occurred during the various days of a week. Test whether the accidents are uniformly distributed over the week. $\chi^2_{5\%}(\gamma = 5) = 11.07$. Day: Mon, Tue, Wed, Thur, Fri, Sat Number of accidents: 15, 19, 13, 12, 16, 15	07	L3	CO4

OR

Q.8 a. Two random samples gave the following data: | 06 | L2 | CO4

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

Can we conclude that the two samples have been drawn from the same normal population? $F_{5\%}(10, 7) = 3.64$.

b. The following data relate to the marks obtained by 11 students in two tests. Second test is after intense coaching. Do the data indicate that the students have benefited by coaching? | 07 | L3 | CO4

Test 1	19	23	16	24	17	18	20	18	21	19	20
Test 2	17	24	20	24	20	22	20	20	18	22	19

($t_{5\%}(\gamma = 10)$ is 1.81)

c. The mean value of a random sample of 60 items was found to be 145 and standard deviation is 40. Find the 95% confidence limits for the population mean. | 07 | L2 | CO5

Module – 5

Q.9 a. The following figures relate to production in kgs of three variables A, B, C of wheat sown on 12 plots. | 10 | L3 | CO6

A	14	16	18		
B	14	13	15	22	
C	18	16	19	19	22

Apply one-way Annova using a 0.05 significance level in the production of the varieties. F_c at 5% (2, 9) d.f is 4.26.

b. Analyze and interpret the following statistics concerning output of wheat per field obtained as a result of experiment conducted to test four varieties of wheat viz., A, B, C and D under a Latin - Square design. | 10 | L3 | CO6

C 25	B 23	A 20	D 20
A 19	D 19	C 21	B 18
B 19	A 14	D 17	C 20
D 17	C 20	B 21	A 15

OR

Q.10 a. Four doctors each test four treatments for a certain disease and observe the number of days each patient takes to recover. The results are as follows: | 10 | L3 | CO6

Doctor/Treatment	T_1	T_2	T_3	T_4
D_1	10	14	19	20
D_2	11	15	17	21
D_3	9	12	16	19
D_4	8	13	17	20

Discuss the difference between doctors and treatments F_{at} 5% level (3, 9) is 3.86.

	b.	A study of the effect of different types of anesthesia on the length of post-operative hospital stay yielded for the following for cesarean patients. Group 'A' was given an epidural MS providing additional safety. Group 'B' was given an epidural and Group 'C' was given a spinal is considered to be less dangerous and Group 'D' was given general anesthesia is considered to be the most dangerous. Note that the data are in the form of distribution for each group.	10	L3	CO6

	Length of Stay	Number of patients
Group A	3	6
	4	14
Group B	4	18
	5	2
Group C	4	10
	5	9
	6	1
Group D	4	8
	5	12

Test for the existence of an effect due to anesthesia type at 0.01. $F_{0.01} = 4.13$

*****

Module 1

Q1. a. Question 1 a)

· Find mean and Standard Deviation of the Poisson Distribution

Proof :- $\text{Mean}(\mu) = \sum_{x=0}^{\infty} x \cdot P(x)$ —— (1)

$$= \sum_{x=0}^{\infty} x \cdot \frac{m^x e^{-m}}{x!} = \sum_{x=0}^{\infty} x \cdot \frac{m^x e^{-m}}{x(x-1)!}$$

$$\mu = \sum_{x=1}^{\infty} \frac{m^{x-1+1} e^{-m}}{(x-1)!} = m e^{-m} \sum_{x=1}^{\infty} \frac{m^{x-1}}{(x-1)!}$$

$$\mu = m e^{-m} \left[1 + \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!} + \cdots \right]$$

$$\mu = m e^{-m} \cdot e^{m} = m \cdot e^{0} = m$$

$\therefore \boxed{\text{Mean}(\mu) = m}$ or $\boxed{\sum_{x=0}^{\infty} x\, p(x) = m}$ —— (2)

· variance, $(V) = \sum_{x=0}^{\infty} x^2 p(x) - \left[\sum_{x=0}^{\infty} x\, p(x) \right]^2$ —— (3)

Consider, $\sum_{x=0}^{\infty} x^2 p(x) = \sum_{x=0}^{\infty} \left[x^2 - x + x \right] p(x)$

$$= \sum_{x=0}^{\infty} \left[x(x-1) + x \right] p(x)$$

$$= \sum_{x=0}^{\infty} x(x-1) p(x) + \sum_{x=0}^{\infty} x\, p(x)$$

$$= \sum_{x=0}^{\infty} x(x-1) \cdot \frac{m^x e^{-m}}{x!} + m$$

$$\sum_{x=0}^{\infty} x^2 p(x) = \sum_{x=0}^{\infty} \frac{\cancel{x(x-1)}.m^x e^{-m}}{\cancel{x(x-1)}(x-2)!} + m$$

$$= \sum_{x=2}^{\infty} \frac{m^{x-2+2} e^{-m}}{(x-2)!} + m$$

$$= m^2 e^{-m} \sum_{x=2}^{\infty} \left[\frac{m^{x-2}}{(x-2)!} \right] + m$$

$$= m^2 e^{-m} \left[1 + \frac{m}{1!} + \frac{m^2}{2!} + \cdots \right] + m$$

$$= m^2 e^{-m} . e^{m} + m$$

i.e., $\sum_{x=0}^{\infty} x^2 p(x) = m^2 + m$ —— (4)

using (4) and (2) in (3) we get

$$\text{variance } (V) = m^2 + m - (m)^2$$

$$= m^2 + m - m^2$$

$$\boxed{V = m}$$

$\therefore$ S.D $(\sigma) = \sqrt{V} = \sqrt{m}$

Thus we have for the poisson distribution

Mean$(\mu) = m$ and S.D$(\sigma) = \sqrt{m}$

<u>Note</u>:- Mean and variance are equal for the poisson distribution.

Q1. b. Question 1 b)

Solution

Part (i): Families with 2 Boys and 2 Girls

The probability of having exactly 2 boys and 2 girls out of 4 children can be found using the binomial distribution:

$$P(X = 2) = \binom{4}{2} \times (0.5)^2 \times (0.5)^2 = 6 \times 0.0625 = 0.375$$

Expected number of families with 2 boys and 2 girls:

$$0.375 \times 800 = 300$$

Part (ii): Families with at Least One Boy

The probability of having at least one boy is complementary to the probability of having all girls:

$$P(\text{at least one boy}) = 1 - P(\text{all girls})$$

$$P(\text{all girls}) = (0.5)^4 = 0.0625$$

$$P(\text{at least one boy}) = 1 - 0.0625 = 0.9375$$

Expected number of families with at least one boy:

$$0.9375 \times 800 = 750$$

Part (iii): Families with at Most 2 Girls

"At most 2 girls" means either 0, 1, or 2 girls. We calculate the probabilities for each case and add them:

$$P(0 \text{ girls}) = (0.5)^4 = 0.0625$$

$$P(1 \text{ girl}) = \binom{4}{1} \times (0.5)^1 \times (0.5)^3 = 4 \times 0.5 \times 0.125 = 0.25$$

$$P(2 \text{ girls}) = \binom{4}{2} \times (0.5)^2 \times (0.5)^2 = 6 \times 0.0625 = 0.375$$

Total probability for at most 2 girls:

$$0.0625 + 0.25 + 0.375 = 0.6875$$

Expected number of families with at most 2 girls:

$$0.6875 \times 800 = 550$$

Summary of Results

- (i) Families with 2 boys and 2 girls: **300 families**
- (ii) Families with at least one boy: **750 families**
- (iii) Families with at most 2 girls: **550 families**

Q1. c. Question 1 c)

Sol:- We have, $f(x) = \alpha e^{-\alpha x}, \quad x > 0$

- Mean $= \frac{1}{\alpha}$

- By data $\frac{1}{\alpha} = 5 \Rightarrow \boxed{\alpha = \frac{1}{5}}$

- Hence, $f(x) = \frac{1}{5} e^{-x/5}$

(i) $p(x < 5) = \int_0^5 f(x)\, dx$

$$p(x<5) = \int_0^5 \frac{1}{5} e^{-x/5} dx = -\left[e^{-x/5}\right]_0^5$$

$$= 1 - e^{-1} = 0.6321$$

$\therefore \boxed{p(x<5) = 0.6321}$

(ii) $p(5 < x < 10) = \int_5^{10} f(x)\, dx = \int_5^{10} \frac{1}{5} e^{-x/5} dx$

$$p(5<x<10) = -\left[e^{-x/5}\right]_5^{10}$$

$$= \left(\frac{1}{e}\right) - \left(\frac{1}{e^2}\right) = 0.2325$$

$\therefore \boxed{p(5<x<10) = 0.2325}$

Q2. a. Question 2 a)

Sol^n: Let X be the random variable for the random values,
$x_1 = -2$, $x_2 = -1$, $x_3 = 0$, $x_4 = 1$, $x_5 = 2$, $x_6 = 3$,
and the given probabilities are,
$P(X = x_1) = P(-2) = 0.1$
$P(X = x_2) = P(-1) = k$
$P(X = x_3) = P(0) = 0.2$
$P(X = x_4) = P(1) = 2k$
$P(X = x_5) = P(2) = 0.3$
$P(X = x_6) = P(3) = k$

i) We know that,
$\sum_{i=1}^{6} P(X = x_i) = 1$

$$\Rightarrow 0.1 + k + 0.2 + 2k + 0.3 + k = 1$$
$$\Rightarrow 4k + 0.6 = 1$$
$$\Rightarrow 4k = 0.4$$
$$\Rightarrow k = 0.1$$

ii) Variance

The expected value $E(X)$ is calculated as follows:

$$E(X) = \sum x \cdot P(x) = (-2 \cdot 0.1) + (-1 \cdot 0.1) + (0 \cdot 0.2) + (1 \cdot 0.2) + (2 \cdot 0.3) + (3 \cdot 0.1)$$

Breaking it down:

$$E(X) = (-2 \cdot 0.1) + (-1 \cdot 0.1) + (0 \cdot 0.2) + (1 \cdot 0.2) + (2 \cdot 0.3) + (3 \cdot 0.1)$$
$$E(X) = -0.2 - 0.1 + 0 + 0.2 + 0.6 + 0.3 = 0.8$$

So, $E(X) = 0.8$.

Next, we calculate $E(X^2)$:

$$E(X^2) = \sum x^2 \cdot P(x) = ((-2)^2 \cdot 0.1) + ((-1)^2 \cdot 0.1) + (0^2 \cdot 0.2) + (1^2 \cdot 0.2) + (2^2 \cdot 0.3) + (3^2 \cdot 0.1$$

Breaking it down:

$$E(X^2) = (4 \cdot 0.1) + (1 \cdot 0.1) + (0 \cdot 0.2) + (1 \cdot 0.2) + (4 \cdot 0.3) + (9 \cdot 0.1)$$
$$E(X^2) = 0.4 + 0.1 + 0 + 0.2 + 1.2 + 0.9 = 2.8$$

So, $E(X^2) = 2.8$.

The variance $\mathrm{Var}(X)$ is given by:

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2$$

Substitute the values:

$$\mathrm{Var}(X) = 2.8 - (0.8)^2$$
$$\mathrm{Var}(X) = 2.8 - 0.64 = 2.16$$

So, the variance $\mathrm{Var}(X) = 2.16$.

iii) $P(x \leq 2)$

Finally, we calculate $P(X \leq 2)$:

$$P(X \leq 2) = P(-2) + P(-1) + P(0) + P(1) + P(2)$$

Using the probabilities:

$$P(X \leq 2) = 0.1 + 0.1 + 0.2 + 0.2 + 0.3 = 0.9$$

Thus, $P(X \leq 2) = 0.9$.

Q2. b. Question 2 b)

Given that the number of accidents in a year follows a Poisson distribution with mean $\lambda = 3$, we want to find the approximate number of taxi drivers with:

- (i) More than 3 accidents in a year
- (ii) At most 2 accidents in a year

The probability mass function of a Poisson distribution is:

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

(i) Probability of More Than 3 Accidents
We need to find $P(X > 3)$, which is given by:

$$P(X > 3) = 1 - P(X \leq 3)$$

where

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

Calculating each term individually with $\lambda = 3$:

$$P(X = 0) = \frac{3^0 e^{-3}}{0!} = e^{-3} \approx 0.0498$$
$$P(X = 1) = \frac{3^1 e^{-3}}{1!} = 3e^{-3} \approx 0.1494$$
$$P(X = 2) = \frac{3^2 e^{-3}}{2!} = \frac{9}{2}e^{-3} \approx 0.2240$$
$$P(X = 3) = \frac{3^3 e^{-3}}{3!} = \frac{27}{6}e^{-3} \approx 0.2240$$

Thus,

$$P(X \leq 3) = 0.0498 + 0.1494 + 0.2240 + 0.2240 = 0.6472$$

Then,

$$P(X > 3) = 1 - 0.6472 = 0.3528$$

The expected number of drivers with more than 3 accidents out of 1000 drivers is:

$$0.3528 \times 1000 = 352.8 \approx 353 \text{ drivers}$$

(ii) Probability of At Most 2 Accidents
For the probability of at most 2 accidents, $P(X \leq 2)$, we have:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

Calculating this sum:

$$P(X \leq 2) = 0.0498 + 0.1494 + 0.2240 = 0.4232$$

Thus, the expected number of drivers with at most 2 accidents out of 1000 drivers is:

$$0.4232 \times 1000 = 423.2 \approx 423 \text{ drivers}$$

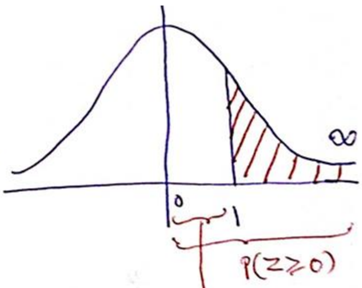
Summary

- Number of drivers with more than 3 accidents: Approximately 353 drivers
- Number of drivers with at most 2 accidents: Approximately 423 drivers

Q2. c. Question 2 c)

Sol:- Given, mean $(\mu) = 70$, S.D $(\sigma) = 5$

Hence, S.n.v $= \frac{x-\mu}{\sigma} = \frac{x-70}{5}$



(i) To find $P(x < 65)$

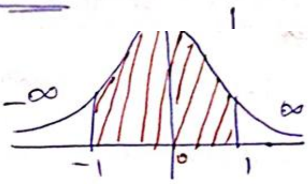
- If $x = 65$, $Z = \frac{65-70}{5} = \frac{-5}{5} = -1 \Rightarrow Z = -1$
- Hence we have to find $P(z < -1)$

$\{\phi(1) = 0.3413$ using table$\}$

$$P(Z<-1) = P(Z>1)$$
$$= P(Z \geq 0) - P(0 \leq Z \leq 1)$$
$$= 0.5 - \phi(1)$$
$$= 0.5 - 0.3413 = \underline{0.1587}$$

$\therefore$ number of Students scoring less than 65 marks

$$= 1000 \times 0.1587 = 158.7 \approx \underline{159}$$



(ii) To find $P(65 < x < 75)$

- We have to find $P(-1 < Z < 1)$

i.e, $P(-1 < Z < 1) = P(-1 < Z < 0) + P(0 < Z < 1)$
$$= P(0 < Z < 1) + P(0 < Z < 1)$$
$$= 2P(0 < Z < 1)$$
$$= 2\phi(1)$$
$$= 2(0.3413)$$
$$= 0.6826$$

$\therefore$ number of Student scoring marks between 65 and 75

$$= 1000 \times 0.6826 = 682.6 \approx \underline{683}$$

Module 2

Q3. a. Question 3 a)

Sol:- The joint probability distribution table is as follows

X \ Y	$y_1 = 1$	$y_2 = 3$	$y_3 = 9$	Sum
$x_1 = 2$	$P_{11} = \frac{1}{8}$	$P_{12} = \frac{1}{24}$	$P_{13} = \frac{1}{12}$	$f(x_1) = \frac{1}{4}$
$x_2 = 4$	$P_{21} = \frac{1}{4}$	$P_{22} = \frac{1}{4}$	$P_{23} = 0$	$f(x_2) = \frac{1}{2}$
$x_3 = 6$	$P_{31} = \frac{1}{8}$	$P_{32} = \frac{1}{24}$	$P_{33} = \frac{1}{12}$	$f(x_3) = \frac{1}{4}$
Sum	$g(y_1) = \frac{1}{2}$	$g(y_2) = \frac{1}{3}$	$g(y_3) = \frac{1}{6}$	1

Marginal distributions of X and Y is a follows.

x_i	$x_1 = 2$	$x_2 = 4$	$x_3 = 6$
$f(x_i)$	$f(x_1) = \frac{1}{4}$	$f(x_2) = \frac{1}{2}$	$f(x_3) = \frac{1}{4}$

y_j	$y_1 = 1$	$y_2 = 3$	$y_3 = 9$
$g(y_j)$	$g(y_1) = \frac{1}{2}$	$g(y_2) = \frac{1}{3}$	$g(y_3) = \frac{1}{6}$

To find $Cov(X, Y)$

$$Cov(X,Y) = E(XY) - E(X) \cdot E(Y) \quad \text{--- (1)}$$

Now,

$$\cdot E(XY) = \sum_{i=1}^{3} \sum_{j=1}^{3} x_i y_j P_{ij}$$

$$= (2)(1)\left(\frac{1}{8}\right) + (2)(3)\left(\frac{1}{24}\right) + (2)(9)\left(\frac{1}{12}\right)$$

$$+ (4)(1)\left(\frac{1}{4}\right) + (4)(3)\left(\frac{1}{4}\right) + (4)(9)(0)$$

$$+ (6)(1)\left(\frac{1}{8}\right) + (6)(3)\left(\frac{1}{24}\right) + (6)(9)\left(\frac{1}{12}\right)$$

$$\boxed{E(XY) = 12}$$

$$\boxed{E(XY) = 12}$$

- $E(X) = \sum_{i=1}^{3} x_i f(x_i) = (2)\left(\frac{1}{4}\right) + (4)\left(\frac{1}{2}\right) + (6)\left(\frac{1}{4}\right) = 4$

- $E(Y) = \sum_{j=1}^{3} y_j g(y_j) = (1)\left(\frac{1}{2}\right) + (3)\left(\frac{1}{3}\right) + (9)\left(\frac{1}{6}\right) = 3$

$\therefore \text{Cov}(X,Y) = 12 - (4)(3) = 0$

$$\Rightarrow \boxed{\text{Cov}(X,Y) = 0}$$

(14)

TO find ~~$\rho(X,Y)$~~ correlation of X and Y

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{0}{\sigma_x \sigma_y} = 0$$

$$\therefore \boxed{\sigma(X,Y) = 0}$$

Q3. b. Question 3 b)

Sol:- Since there are 3 cities A, B and C, the state space = {A, B, C}

- So, t.p.m is a 3×3 matrix

$$B \xrightarrow{x} C$$

$$C \xrightarrow{x} B$$

$$B \xrightarrow[2x]{} A \quad \text{or} \quad C \xrightarrow[2x]{} A$$

$$\text{PRESENT} \begin{array}{c|ccc} & A & B & C \;\text{FUTURE} \\ \hline A & 0 & 1 & 0 \\ B & 2x & 0 & x \\ C & 2x & x & 0 \end{array}$$

$$2x + x + 0 = 1 \Rightarrow 3x = 1$$

$$\Rightarrow \boxed{x = \frac{1}{3}}$$

Hence, the required t.p.m is

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix}$$

We have to find $vP = v$

where, v is the probability vector i.e., $v = (v_1, v_2, v_3)$

$\hookrightarrow v_1 + v_2 + v_3 = 1$

i.e., $$[v_1, v_2, v_3]\begin{bmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix} = [v_1, v_2, v_3]$$

$$\left[\frac{2}{3}v_2 + \frac{2}{3}v_3,\ v_1 + \frac{v_3}{3},\ \frac{v_2}{3}\right] = [v_1, v_2, v_3]$$

i.e., $\frac{2}{3}v_2 + \frac{2}{3}v_3 = v_1$, $v_1 + \frac{v_3}{3} = v_2$, $\frac{v_2}{3} = v_3$

i.e., $2v_2 + 2v_3 = 3v_1$ — (1)

$3v_1 + v_3 = 3v_2$ — (2)

$v_2 = 3v_3$ — (3)

using $v_3 = 1 - v_2 - v_1$ in (1), yields

(1) $\Rightarrow 2v_2 + 2(1 - v_2 - v_1) = 3v_1$

$\Rightarrow 2v_2 + 2 - 2v_2 - 2v_1 - 3v_1 = 0$

$\Rightarrow 2 - 5v_1 = 0$

$\Rightarrow 5v_1 = 2$

$\Rightarrow \boxed{v_1 = \frac{2}{5}}$, (or) $\boxed{v_1 = 40\%}$

$\frac{2}{5} \times 100 = 40\%$

(2) $\Rightarrow 3\left(\frac{2}{5}\right) + \left(1 - v_2 - \frac{2}{5}\right) = 3v_2$

$\Rightarrow \frac{6}{5} + 1 - v_2 - \frac{2}{5} = 3v_2$

(56)

$\Rightarrow \frac{6}{5} + 1 - \frac{2}{5} = 3V_2 + V_2$

$\Rightarrow \frac{9}{5} = 4V_2$

$\Rightarrow \boxed{V_2 = \frac{9}{20}}$ (or) $\boxed{V_2 = 45\%}$

$\frac{9}{20} \times 100 = 45\%$

$V_3 = 1 - V_2 - V_1 = 1 - \frac{9}{20} - \frac{2}{5} = \frac{3}{20}$

$\therefore \boxed{V_3 = \frac{3}{20}}$ (or) $\boxed{V_3 = 15\%}$

$\frac{3}{20} \times 100 = 15\%$

$\therefore V = (V_1, V_2, V_3) = \left(\frac{2}{5}, \frac{9}{20}, \frac{3}{20}\right) = (40\%, 45\%, 15\%)$

A B C

$\therefore$ The salesman sells in City A is 40%, City B is 45% and City C is 15%

Q3. c. Question 3 c)

Sol^n: Since the given matrix P is of order 3x3, the required fixed probability vector Q must be also order of 3x3.

Let $Q = [x \quad y \quad z]$, $\quad$ *For every* $x \geq 0, y \geq 0, z \geq 0 \& x + y + z = 1$

Also, QP=Q

$$\therefore [x \quad y \quad z]\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} \frac{z}{2} & x + \frac{z}{2} & y \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \frac{z}{2} & x + \frac{z}{2} & y \end{bmatrix} = [x \quad y \quad z]$$

$$\Rightarrow \frac{z}{2} = x, x + \frac{z}{2} = y, y = z$$

$$\Rightarrow \frac{1}{2}(1 - x - y) = x \ldots\ldots (1), x + \frac{1}{2}(1 - x - y) = y \ldots (2), y = 1 - x - y \ldots (3)$$

$$\Rightarrow 3x + y = 1, x - 3y = -1, x + 2y = 1$$

$$\Rightarrow x = \frac{1}{5}, y = \frac{2}{5} \Rightarrow z = 1 - \frac{1}{5} - \frac{2}{5} = \frac{2}{5}$$

Hence the required fixed probability vector is $Q = [x \quad y \quad z] = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{bmatrix}$

Q4. a. Question 4 a)

→ Probability vector

A vector $V = (V_1, V_2, ---, V_n)$ is called a probability vector if each one of its components are non negative and their sum is equal to unity.

Examples :- $u = (0, 1)$; $V = (1/2, 1/2)$, $W = (1/4, 1/4, 1/2)$ are all probability vectors.

Note :- If 'v' is not a probability vector but each one of the V_i ($i = 1$ to n) are non negative then λV is a probability vector where $\boxed{\lambda = \frac{1}{\sum_{i=1}^{n} V_i}}$

Example :- If $V = (1, 2, 3)$ then $\lambda = \frac{1}{(1+2+3)} = \frac{1}{6}$

$\therefore \lambda V = \frac{1}{6}(1, 2, 3) = \left(\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right)$ is a probability vector.

→ Stochastic Matrix

A square matrix $P = (p_{ij})$ having every row is the form of a probability vector is called a stochastic matrix.

Examples (i) Identity matrix (I) of any order

$$I_{(2)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} ; \quad I_{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(ii) $\begin{bmatrix} 1/2 & 1/2 \\ 0 & 1 \end{bmatrix}$ (iii) $\begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$

→ Regular Stochastic Matrix

A stochastic matrix P is said to be a regular stochastic matrix if all the entries of some power P^n are positive.

Example :- $A = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$

Consider, $A^2 = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$

$\therefore A$ is a regular stochastic matrix ($n = 2$)

Q4. b. Question 4 b)

Given Data

The joint probability distribution of two discrete random variables X and Y is:

$$f(x,y) = k(2x+y)$$

where x and y are integers such that:

- $0 \le x \le 2$
- $0 \le y \le 3$

We are asked to:

1. Find the value of the constant k.
2. Show that the random variables X and Y are dependent.
3. Find $P(X \ge 1, Y \le 2)$.

Step 1: Finding the Value of k

Since $f(x,y)$ is a probability distribution, the sum of all probabilities must be 1. Thus,

$$\sum_{x=0}^{2}\sum_{y=0}^{3} f(x,y) = 1$$

Substituting $f(x,y) = k(2x+y)$, we get:

$$\sum_{x=0}^{2}\sum_{y=0}^{3} k(2x+y) = 1$$

Let's calculate this sum step-by-step.

The value of the constant k is $\frac{1}{42}$.

Step 2: Show that X and Y are Dependent

To determine if X and Y are dependent, we need to check if $f(x, y)$ can be factored into functions of x and y alone, i.e., if:

$$f(x, y) = g(x) \cdot h(y)$$

If this is not possible, then X and Y are dependent.

Substituting $f(x, y) = \frac{1}{42}(2x + y)$ and examining whether it can be factored will reveal if the variables are dependent.

Let's analyze this.

The expression $f(x, y) = \frac{1}{42}(2x + y)$ cannot be separated into a product of a function of x alone and a function of y alone, as it directly depends on both x and y through the term $2x + y$.

Thus, $f(x, y)$ cannot be expressed as $g(x) \cdot h(y)$, meaning that X and Y are dependent.

Step 3: Find $P(X \geq 1, Y \leq 2)$

We are asked to find the probability $P(X \geq 1, Y \leq 2)$. This probability is given by summing $f(x, y)$ for all values of x and y that satisfy $X \geq 1$ and $Y \leq 2$.

This corresponds to the pairs:

- $(x = 1, y = 0)$, $(x = 1, y = 1)$, $(x = 1, y = 2)$
- $(x = 2, y = 0)$, $(x = 2, y = 1)$, $(x = 2, y = 2)$

Let's calculate $P(X \geq 1, Y \leq 2)$ by summing $f(x, y)$ for these values.

The probability $P(X \geq 1, Y \leq 2)$ is $\frac{4}{7}$.

Summary of Solutions

1. The value of k is $\frac{1}{42}$.
2. The random variables X and Y are dependent.
3. $P(X \geq 1, Y \leq 2) = \frac{4}{7}$.

Q4. c. Question 4 c)

Solution Steps:

1. **Calculate $E(X)$**

   Since X depends only on the first toss, and there is a $\frac{1}{2}$ probability for each outcome:

$$E(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

2. **Calculate $E(Y)$**

   Y represents the number of heads in three tosses, which follows a Binomial distribution with $n = 3$ and $p = \frac{1}{2}$. The expected value of a Binomial random variable is $E(Y) = n \cdot p$:

$$E(Y) = 3 \cdot \frac{1}{2} = \frac{3}{2}$$

3. **Calculate $E(XY)$**

   To find $E(XY)$, we consider the joint distribution of X and Y. We'll examine the possible values of X and Y, and the associated probabilities for each outcome.

Possible Cases:

- **Case 1**: $X = 0$ (Head on the first toss)
  - $Y = 1$: One head (only the first toss is a head) — Probability $\frac{1}{8}$
  - $Y = 2$: Two heads (first and one other toss) — Probability $\frac{3}{8}$
  - $Y = 3$: Three heads (all tosses are heads) — Probability $\frac{1}{8}$
- **Case 2**: $X = 1$ (Tail on the first toss)
  - $Y = 0$: Zero heads — Probability $\frac{1}{8}$
  - $Y = 1$: One head (only one of the last two tosses is a head) — Probability $\frac{3}{8}$
  - $Y = 2$: Two heads (both of the last two tosses are heads) — Probability $\frac{1}{4}$

Calculating $E(XY)$

Using these probabilities:

$$E(XY) = \sum (X \cdot Y \cdot P(X, Y))$$

$$= (0 \cdot 1 \cdot \frac{1}{8}) + (0 \cdot 2 \cdot \frac{3}{8}) + (0 \cdot 3 \cdot \frac{1}{8}) + (1 \cdot 0 \cdot \frac{1}{8}) + (1 \cdot 1 \cdot \frac{3}{8}) + (1 \cdot 2 \cdot \frac{1}{4})$$

Simplifying, we get:

$$E(XY) = 0 + 0 + 0 + 0 + \frac{3}{8} + \frac{2}{4} = \frac{3}{8} + \frac{1}{2} = \frac{7}{8}$$

4. **Calculate $\text{Cov}(X, Y)$**

   The covariance is given by:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

   Substituting the values we have:

$$\text{Cov}(X, Y) = \frac{7}{8} - \frac{1}{2} \cdot \frac{3}{2} = \frac{7}{8} - \frac{3}{4} = \frac{7-6}{8} = \frac{1}{8}$$

Answer

The covariance $\text{Cov}(X, Y)$ is $\frac{1}{8}$.

Module – 3

Q5. a. Question 5 a)

(*) Test of Hypothesis

→ Hypothesis

In order to arrive at a decision regarding the population through a sample of the population, we have to make certain assumption referred to as hypothesis which may or may not be true.

This hypothesis is classified as

Null Hypothesis denoted by H_0	Alternative Hypothesis denoted by H_1

→ Null Hypothesis

Null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it

(*) Type-I and Type-II Errors

In statistics, Type-I error is a false positive conclusion while a Type-II error is a false negative conclusion.

(or)

- The error of rejecting the hypothesis when the hypothesis is true is called Type-I error.
- The error of accepting the hypothesis when the hypothesis is false is called Type-II error.

Decision	Accepting the hypothesis	Rejecting the hypothesis
Hypothesis true	Correct decision	Wrong decision (Type I error)
Hypothesis false	Wrong decision (Type II error)	Correct decision

Q5. b. Question 5 b)

Step 1: Define Hypotheses

- **Null Hypothesis** (H_0): The die is fair, so the probability of rolling an even number (2, 4, or 6) is $p = 0.5$.
- **Alternative Hypothesis** (H_1): The die is not fair, meaning $p \neq 0.5$.

Step 2: Set the Significance Level

We're testing at the 99% confidence level, so our significance level is:

$$\alpha = 1 - 0.99 = 0.01$$

Step 3: Calculate Expected Mean and Standard Deviation

Since we have a large number of trials (324), we can approximate the binomial distribution with a normal distribution:

$$X \sim N(np, np(1-p))$$

where:

- $n = 324$ (total throws),
- $p = 0.5$ (probability of rolling an even number under H_0).

Expected Mean (μ)

The expected mean, μ, is calculated as:

$$\mu = np = 324 \times 0.5 = 162$$

Standard Deviation (σ)

The standard deviation, σ, is:

$$\sigma = \sqrt{np(1-p)} = \sqrt{324 \times 0.5 \times 0.5} = \sqrt{81} = 9$$

Step 4: Calculate the Test Statistic (Z-Score)

We observed 181 even-numbered outcomes, so our observed value of X is 181.

The Z-score, which measures how many standard deviations the observed result is from the expected mean, is calculated as:

$$z = \frac{X - \mu}{\sigma} = \frac{181 - 162}{9} = \frac{19}{9} \approx 2.11$$

Step 5: Determine the Critical Value

Since this is a two-tailed test (we're checking if the die could be biased in either direction), we need the critical Z-value for a 99% confidence level, which is approximately:

$$z_{\text{critical}} = \pm 2.576$$

Step 6: Decision Rule

- **If** $|z| \leq 2.576$, we **fail to reject the null hypothesis**.
- **If** $|z| > 2.576$, we **reject the null hypothesis**.

Step 7: Compare and Conclude

We found:

$$|z| = 2.11$$

Since $2.11 < 2.576$, we fail to reject the null hypothesis.

Conclusion

At the 99% confidence level, we do not have sufficient evidence to conclude that the die is biased. Thus, it is reasonable to think that the die might be fair.

Q5. c. Question 5 c)

1. **Define Hypotheses:**
   - **Null Hypothesis** H_0: There is no decrease in the proportion of tea consumers after the increase in duty. Thus, $p_1 = p_2$.
   - **Alternative Hypothesis** H_1: There is a decrease in the proportion of tea consumers after the increase in duty, i.e., $p_1 > p_2$.

   Here:

   - p_1 is the proportion of tea consumers before the increase.
   - p_2 is the proportion of tea consumers after the increase.
2. **Set the Significance Level:**
   - We are testing at the 1% level of significance, so $\alpha = 0.01$.
   - For a one-tailed test at the 1% significance level, the critical z-value is given as $z = 2.33$.

3. **Calculate Sample Proportions:**

   Before the increase in duty:

$$p_1 = \frac{800}{1000} = 0.8$$

   After the increase in duty:

$$p_2 = \frac{800}{1200} = 0.6667$$

4. **Calculate the Pooled Proportion p:**

   Since we're comparing two proportions, we use the pooled proportion:

$$p = \frac{800 + 800}{1000 + 1200} = \frac{1600}{2200} = 0.7273$$

5. **Calculate the Standard Error SE:**

   The standard error SE for the difference in proportions is:

$$SE = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

   where:

   - $n_1 = 1000$ (sample size before),
   - $n_2 = 1200$ (sample size after).

   Substituting in the values:

$$SE = \sqrt{0.7273 \times (1 - 0.7273) \times \left(\frac{1}{1000} + \frac{1}{1200}\right)}$$

$$SE = \sqrt{0.7273 \times 0.2727 \times \left(\frac{1}{1000} + \frac{1}{1200}\right)}$$

$$SE = \sqrt{0.7273 \times 0.2727 \times 0.001833}$$

$$SE \approx 0.0212$$

6. **Calculate the Z-Score:**

   Now, we calculate the Z-score to determine if the difference in proportions is statistically significant:

$$z = \frac{p_1 - p_2}{SE} = \frac{0.8 - 0.6667}{0.0212}$$

$$z = \frac{0.1333}{0.0212} \approx 6.29$$

7. **Decision Rule:**

   - If $z \geq 2.33$, we reject the null hypothesis at the 1% significance level.
   - If $z < 2.33$, we fail to reject the null hypothesis.

Since $z = 6.29$ is much greater than 2.33, we **reject the null hypothesis**.

Conclusion

There is significant evidence at the 1% level to conclude that there has been a decrease in tea consumption after the increase in excise duty.

Q6.a. Question 6 a)

Sol:- Given, $n = 1000$

$x = 540$ (i.e., probability of success out of 1000 trials)

- Suppose the coin is unbiased
- The probability of getting a head in one toss, $p = \frac{1}{2}$
- Since $p + q = 1 \Rightarrow \boxed{q = \frac{1}{2}}$
- Expected number of heads in 1000 tosses, $\mu = np$

i.e., $\boxed{\mu = 1000 \times \frac{1}{2} = 500}$

- S.D of Sampling $(\sigma) = \sqrt{npq} = \sqrt{\left(1000 \times \frac{1}{2} \times \frac{1}{2}\right)} = 250$
- Consider,

$$Z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}} = \frac{540 - 500}{250} = \frac{40}{250} = 2.53 < 2.58$$

As $Z < 2.58$, the hypothesis is accepted at 1% level of significance i.e., we conclude that the coin is unbiased at 5% level of significance.

Q6.b. Question 6 b)

1. **Define Hypotheses:**
   - **Null Hypothesis** H_0: There is no difference in the proportion of support for the party between the two localities. Thus, $p_1 = p_2$.
   - **Alternative Hypothesis** H_1: There is a difference in the proportion of support for the party between the two localities, i.e., $p_1 \neq p_2$.

   Here:

   - p_1 is the proportion of support in the first locality.
   - p_2 is the proportion of support in the second locality.

2. **Calculate Sample Proportions:**

   For the first locality:

   $$p_1 = 0.55$$

   Sample size, $n_1 = 600$.

   For the second locality:

   $$p_2 = 0.48$$

   Sample size, $n_2 = 400$.

3. **Calculate the Pooled Proportion** p**:**

   Since we're comparing two proportions, we use the pooled proportion p, calculated as:

   $$p = \frac{p_1 \times n_1 + p_2 \times n_2}{n_1 + n_2} = \frac{(0.55 \times 600) + (0.48 \times 400)}{600 + 400}$$

   $$p = \frac{330 + 192}{1000} = \frac{522}{1000} = 0.522$$

4. **Calculate the Standard Error** SE**:**

   The standard error SE for the difference in proportions is given by:

   $$SE = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

   Substituting in the values:

   $$SE = \sqrt{0.522 \times (1 - 0.522) \times \left(\frac{1}{600} + \frac{1}{400}\right)}$$

   $$SE = \sqrt{0.522 \times 0.478 \times \left(\frac{1}{600} + \frac{1}{400}\right)}$$

   $$SE = \sqrt{0.522 \times 0.478 \times 0.004167}$$

   $$SE \approx 0.0222$$

5. **Calculate the Z-Score:**

   Now, we calculate the Z-score to determine if the difference in proportions is statistically significant:

$$z = \frac{p_1 - p_2}{SE} = \frac{0.55 - 0.48}{0.0222}$$

$$z = \frac{0.07}{0.0222} \approx 3.15$$

6. **Decision Rule:**

   For a two-tailed test, if the Z-score is greater than the critical value at a standard significance level (e.g., 1.96 for a 5% significance level), we reject the null hypothesis.

Since $z = 3.15$ is greater than 1.96, we **reject the null hypothesis** at the 5% significance level.

Conclusion

There is significant evidence to conclude that there is a difference in opinion between the two localities regarding support for the party.

Q6.c. Question 6 c)

1. **Define Hypotheses:**
   - **Null Hypothesis H_0:** There is no difference in the mean wage rates between the two companies. Thus, $\mu_1 = \mu_2$.
   - **Alternative Hypothesis H_1:** There is a difference in the mean wage rates between the two companies, i.e., $\mu_1 \neq \mu_2$.

   Here:

   - μ_1 is the mean wage in the first company.
   - μ_2 is the mean wage in the second company.
2. **Given Data:**
   - Sample 1 (Company 1):
     - Sample size, $n_1 = 1000$
     - Sample mean, $\bar{X}_1 = 50$
     - Sample standard deviation, $s_1 = 15$
   - Sample 2 (Company 2):
     - Sample size, $n_2 = 1500$
     - Sample mean, $\bar{X}_2 = 45$
     - Sample standard deviation, $s_2 = 20$

3. **Set the Significance Level:**
   - We are testing at a 95% confidence level, so $\alpha = 0.05$.
   - For a two-tailed test, the critical z-value at the 95% confidence level is approximately ± 1.96.
4. **Calculate the Standard Error (SE):**

   Since we are comparing the means of two independent samples, the standard error of the difference in means is calculated as:

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

   Substituting in the values:

$$SE = \sqrt{\frac{15^2}{1000} + \frac{20^2}{1500}}$$

$$SE = \sqrt{\frac{225}{1000} + \frac{400}{1500}}$$

$$SE = \sqrt{0.225 + 0.2667} = \sqrt{0.4917} \approx 0.701$$

5. **Calculate the Z-Score:**

   The Z-score for the difference in sample means is given by:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{SE} = \frac{50 - 45}{0.701}$$

$$z = \frac{5}{0.701} \approx 7.13$$

6. **Decision Rule:**
   - If $|z| > 1.96$, we reject the null hypothesis at the 5% significance level.
   - If $|z| \leq 1.96$, we fail to reject the null hypothesis.

Since $|z| = 7.13$ is greater than 1.96, we **reject the null hypothesis**.

Conclusion

At the 95% confidence level, there is significant evidence to conclude that the mean wage rates vary between the two companies.

Module – 4

Q7. a. Question 7 a)

1. **Define Hypotheses:**
   - **Null Hypothesis H_0**: The mean lifetime of the bulbs is 1600 hours, so $\mu = 1600$.
   - **Alternative Hypothesis H_1**: The mean lifetime of the bulbs is not 1600 hours, so $\mu \neq 1600$.
2. **Given Data:**
   - Sample mean, $\bar{X} = 1550$ hours
   - Sample standard deviation, $s = 120$ hours
   - Sample size, $n = 25$
   - Claimed mean by the company, $\mu = 1600$
3. **Set the Significance Level:**
   - We are testing at a 5% significance level, so $\alpha = 0.05$.
   - For a two-tailed test at the 5% significance level with $n - 1 = 24$ degrees of freedom, the critical t-value is approximately ± 2.064.
4. **Calculate the Standard Error (SE):**

   The standard error SE is calculated as:

$$SE = \frac{s}{\sqrt{n}} = \frac{120}{\sqrt{25}} = \frac{120}{5} = 24$$

5. **Calculate the t-Score:**

   The t-score for the sample mean is calculated as:

$$t = \frac{\bar{X} - \mu}{SE} = \frac{1550 - 1600}{24} = \frac{-50}{24} \approx -2.083$$

6. **Decision Rule:**
   - If $|t| > 2.064$, we reject the null hypothesis at the 5% significance level.
   - If $|t| \leq 2.064$, we fail to reject the null hypothesis.

Since $|t| = 2.083$ is slightly greater than 2.064, we **reject the null hypothesis**.

Conclusion

At the 5% significance level, there is sufficient evidence to suggest that the mean lifetime of the bulbs differs from the claimed 1600 hours. Therefore, the company's claim is not acceptable based on this sample.

Q7. b. Question 7 b)

1. **Define Hypotheses:**
   - **Null Hypothesis** H_0: The population variances are equal, so $\sigma_1^2 = \sigma_2^2$.
   - **Alternative Hypothesis** H_1: The population variances are not equal, so $\sigma_1^2 \neq \sigma_2^2$.
2. **Given Data:**
   - **Sample 1:** $n_1 = 8$
     - Data: $9, 11, 13, 11, 15, 9, 12, 14$
   - **Sample 2:** $n_2 = 7$
     - Data: $10, 12, 10, 14, 9, 8, 10$

   We need to calculate the sample variances s_1^2 and s_2^2 for these two samples.
3. **Calculate Sample Variances:**

   Let's calculate the mean and variance for each sample.

Sample 1:

- Mean ($\bar{X}_1$):

$$\bar{X}_1 = \frac{9+11+13+11+15+9+12+14}{8} = \frac{94}{8} = 11.75$$

- Variance (s_1^2):

$$s_1^2 = \frac{\sum(X_i - \bar{X}_1)^2}{n_1 - 1}$$

$$s_1^2 = \frac{(9-11.75)^2+(11-11.75)^2+(13-11.75)^2+(11-11.75)^2+(15-11.75)^2+(9-11.75)^2+(12-11.75)^2+(14-11.75)^2}{7}$$

Calculating each term:

$$= \frac{(2.75)^2+(0.75)^2+(1.25)^2+(0.75)^2+(3.25)^2+(2.75)^2+(0.25)^2+(2.25)^2}{7}$$

$$= \frac{7.5625+0.5625+1.5625+0.5625+10.5625+7.5625+0.0625+5.0625}{7} = \frac{33.5}{7} \approx 4.79$$

So, $s_1^2 \approx 4.79$.

Sample 2:

- Mean ($\bar{X}_2$):

$$\bar{X}_2 = \frac{10+12+10+14+9+8+10}{7} = \frac{73}{7} \approx 10.43$$

- Variance (s_2^2):

$$s_2^2 = \frac{\sum(X_i - \bar{X}_2)^2}{n_2 - 1}$$

$$s_2^2 = \frac{(10-10.43)^2+(12-10.43)^2+(10-10.43)^2+(14-10.43)^2+(9-10.43)^2+(8-10.43)^2+(10-10.43)^2}{6}$$

Calculating each term:

$$= \frac{(0.43)^2+(1.57)^2+(0.43)^2+(3.57)^2+(1.43)^2+(2.43)^2+(0.43)^2}{6}$$

$$= \frac{0.1849+2.4649+0.1849+12.7449+2.0449+5.9049+0.1849}{6} = \frac{23.7184}{6} \approx 3.95$$

So, $s_2^2 \approx 3.95$.

4. **Calculate the F-Statistic:**

   The F-statistic for testing the equality of variances is calculated as:

$$F = \frac{s_1^2}{s_2^2} = \frac{4.79}{3.95} \approx 1.21$$

5. **Determine the Critical Value:**
   - Given that $\alpha = 0.05$ and $v_1 = 7$ (for Sample 1) and $v_2 = 6$ (for Sample 2), the critical F-value at the 5% significance level is $F_{0.05}(7, 6) = 4.21$.
6. **Decision Rule:**
   - If $F > 4.21$, we reject the null hypothesis.
   - If $F \leq 4.21$, we fail to reject the null hypothesis.

Since $F = 1.21$ is less than 4.21, we **fail to reject the null hypothesis**.

Conclusion

At the 5% significance level, there is not enough evidence to conclude that the variances of the two samples are significantly different. Therefore, we accept the hypothesis that the population variances are equal.

Q7. c. Question 7 c)

1. **Define Hypotheses:**
   - **Null Hypothesis H_0**: The accidents are uniformly distributed across the days of the week.
   - **Alternative Hypothesis H_1**: The accidents are not uniformly distributed across the days of the week.
2. **Given Data:**

   Days: $\text{Mon, Tue, Wed, Thur, Fri, Sat}$

   Observed number of accidents: $O = [15, 19, 13, 12, 16, 15]$

   Total number of accidents over the week:

$$\text{Total} = 15 + 19 + 13 + 12 + 16 + 15 = 90$$

3. **Determine the Expected Frequency:**

   If accidents are uniformly distributed across 6 days, each day would have the same expected number of accidents. Thus, the expected frequency for each day is:

$$E = \frac{\text{Total Accidents}}{\text{Number of Days}} = \frac{90}{6} = 15$$

   So, the expected frequency E for each day is 15.

4. **Calculate the Chi-Square Test Statistic:**

   The Chi-Square statistic χ^2 is calculated using the formula:

   $$\chi^2 = \sum \frac{(O - E)^2}{E}$$

   where O is the observed frequency, and E is the expected frequency for each day.

   Let's calculate χ^2 for each day:

   - For Monday: $\frac{(15-15)^2}{15} = \frac{0^2}{15} = 0$
   - For Tuesday: $\frac{(19-15)^2}{15} = \frac{4^2}{15} = \frac{16}{15} \approx 1.067$
   - For Wednesday: $\frac{(13-15)^2}{15} = \frac{(-2)^2}{15} = \frac{4}{15} \approx 0.267$
   - For Thursday: $\frac{(12-15)^2}{15} = \frac{(-3)^2}{15} = \frac{9}{15} = 0.6$
   - For Friday: $\frac{(16-15)^2}{15} = \frac{1^2}{15} = \frac{1}{15} \approx 0.067$
   - For Saturday: $\frac{(15-15)^2}{15} = \frac{0^2}{15} = 0$

   Summing these values gives:

   $$\chi^2 = 0 + 1.067 + 0.267 + 0.6 + 0.067 + 0 = 2.001$$

5. **Determine the Critical Value:**
   - The degrees of freedom $\text{df} = \text{Number of Days} - 1 = 6 - 1 = 5$.
   - Given $\alpha = 0.05$, the critical value from the Chi-Square table for χ^2 with 5 degrees of freedom is $\chi^2_{0.05,5} = 11.07$.

6. **Decision Rule:**
   - If χ^2 calculated > χ^2 critical, we reject the null hypothesis.
   - If χ^2 calculated $\leq \chi^2$ critical, we fail to reject the null hypothesis.

Since $\chi^2 = 2.001$ is less than 11.07, we **fail to reject the null hypothesis**.

Conclusion

At the 5% significance level, there is no evidence to suggest that the accidents are not uniformly distributed over the days of the week. Therefore, we conclude that the accidents appear to be uniformly distributed across the week.

Q8. a. Question 8 a)

1. **Define Hypotheses:**
   - **Null Hypothesis H_0**: The variances of the two populations are equal ($\sigma_1^2 = \sigma_2^2$).
   - **Alternative Hypothesis H_1**: The variances of the two populations are not equal ($\sigma_1^2 \neq \sigma_2^2$).
2. **Given Data:**
   - Sample 1:
     - Size, $n_1 = 8$
     - Mean, $\bar{X}_1 = 9.6$ (not used in variance comparison)
     - Variance, $s_1^2 = 1.2$
   - Sample 2:
     - Size, $n_2 = 11$
     - Mean, $\bar{X}_2 = 16.5$ (not used in variance comparison)
     - Variance, $s_2^2 = 2.5$

   The larger variance is $s_2^2 = 2.5$, and the smaller variance is $s_1^2 = 1.2$.
3. **Calculate the F-Statistic:**

   The F-statistic for comparing two variances is given by:

   $$F = \frac{\text{larger variance}}{\text{smaller variance}} = \frac{s_2^2}{s_1^2} = \frac{2.5}{1.2} \approx 2.083$$
4. **Determine the Critical Value:**
   - The degrees of freedom for Sample 1 (smaller sample) is $v_1 = n_1 - 1 = 8 - 1 = 7$.
   - The degrees of freedom for Sample 2 (larger sample) is $v_2 = n_2 - 1 = 11 - 1 = 10$.
   - At a significance level of 5%, the critical value for F with $v_1 = 7$ and $v_2 = 10$ is given as $F_{0.05,(10,7)} = 3.64$.
5. **Decision Rule:**
   - If F calculated > F critical, we reject the null hypothesis.
   - If F calculated $\leq F$ critical, we fail to reject the null hypothesis.

Since $F = 2.083$ is less than 3.64, we **fail to reject the null hypothesis.**

Q8. b. Question 8 b)

Step 1: Define Hypotheses

1. **Null Hypothesis (H_0)**: Coaching did not significantly improve the students' scores. ($\mu_d = 0$), where μ_d is the mean of the differences.
2. **Alternative Hypothesis (H_1)**: Coaching significantly improved the students' scores. ($\mu_d > 0$)

This is a one-tailed test, as we're specifically testing for improvement.

Step 3: Calculate the Mean and Standard Deviation of the Differences

1. **Mean of Differences ($\bar{d}$)**:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-2+1+4+0+3+4+0+2-3+3-1}{11} \approx 1.00$$

2. **Standard Deviation of Differences (s_d)**:

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

Step 4: Calculate the t-statistic

The formula for the t-statistic in a paired t-test is:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where:

- $\bar{d}$ is the mean of the differences,
- s_d is the standard deviation of the differences,
- $n = 11$ is the number of pairs.

Step 5: Decision Rule

Given a one-tailed test with $\alpha = 0.05$ and $n - 1 = 10$ degrees of freedom, the critical t-value is 1.81.

Step 6: Conclusion

If the calculated t-statistic is greater than 1.81, we reject H_0 and conclude that coaching had a significant effect.

Let's calculate this.

The calculations are as follows:

- Mean of differences ($\bar{d}$) = 0.82
- Standard deviation of differences (s_d) = 2.23
- Calculated t-statistic = 1.22

Since the calculated t-statistic (1.22) is less than the critical t-value (1.81), we **fail to reject the null hypothesis**.

Q8. c. Question 8 c)

Given Values:

- Sample mean ($\bar{x}$) = 145
- Sample standard deviation (s) = 40
- Sample size (n) = 60
- Confidence level = 95% (so $Z = 1.96$)

Step 1: Calculate the Standard Error (SE)

$$\text{Standard Error (SE)} = \frac{s}{\sqrt{n}} = \frac{40}{\sqrt{60}} \approx 5.16$$

Step 2: Calculate the Margin of Error (ME)

$$\text{Margin of Error (ME)} = Z \cdot \text{SE} = 1.96 \times 5.16 \approx 10.11$$

Step 3: Calculate the Confidence Interval

$$\text{Lower Limit} = \bar{x} - \text{ME} = 145 - 10.11 \approx 134.89$$

$$\text{Upper Limit} = \bar{x} + \text{ME} = 145 + 10.11 \approx 155.11$$

The 95% confidence interval for the population mean is approximately **(134.89, 155.11)**.

Module – 5

Q9. a. Question 9 a)

Step 1: Gather and Organize the Data

The data for the production (in kg) is organized as follows:

$$\text{Variety A: } 14, 16, 18$$

$$\text{Variety B: } 14, 13, 15, 22$$

$$\text{Variety C: } 18, 16, 19, 22$$

Step 2: Calculate the Group Means and Overall Mean

1. **Calculate the mean for each variety:**
   - Mean of Variety A (mean_A):

$$\text{mean}_A = \frac{14 + 16 + 18}{3} = \frac{48}{3} = 16$$

   - Mean of Variety B (mean_B):

$$\text{mean}_B = \frac{14 + 13 + 15 + 22}{4} = \frac{64}{4} = 16$$

   - Mean of Variety C (mean_C):

$$\text{mean}_C = \frac{18 + 16 + 19 + 22}{4} = \frac{75}{4} = 18.75$$

2. **Calculate the overall mean ($\text{mean}_{\text{overall}}$):**
   - Total observations = 3 (A) + 4 (B) + 4 (C) = 11
   - Sum of all values = $14 + 16 + 18 + 14 + 13 + 15 + 22 + 18 + 16 + 19 + 22 = 187$

$$\text{mean}_{\text{overall}} = \frac{187}{11} \approx 17$$

Step 3: Calculate the Sum of Squares

1. **Sum of Squares Between (SSB):**

$$\text{SSB} = n_A(\text{mean}_A - \text{mean}_{\text{overall}})^2 + n_B(\text{mean}_B - \text{mean}_{\text{overall}})^2 + n_C(\text{mean}_C - \text{mean}_{\text{overall}})^2$$

where $n_A = 3$, $n_B = 4$, and $n_C = 4$.

Plugging in the values:

$$\text{SSB} = 3 \times (16 - 17)^2 + 4 \times (16 - 17)^2 + 4 \times (18.75 - 17)^2$$

$$= 3 \times (-1)^2 + 4 \times (-1)^2 + 4 \times (1.75)^2$$

$$= 3 \times 1 + 4 \times 1 + 4 \times 3.0625 = 3 + 4 + 12.25 = 19.25$$

2. **Sum of Squares Within (SSW):**
   - Calculate the squared differences between each observation and its group mean:
     - For Variety A: $(14-16)^2+(16-16)^2+(18-16)^2=4+0+4=8$
     - For Variety B: $(14-16)^2+(13-16)^2+(15-16)^2+(22-16)^2=4+9+1+36=50$
     - For Variety C: $(18-18.75)^2+(16-18.75)^2+(19-18.75)^2+(22-18.75)^2=0.5625+7.5625+0.0625+10.5625=18.75$

$$\text{SSW}=8+50+18.75=76.75$$

Step 4: Calculate the Degrees of Freedom

- **Degrees of Freedom Between (df$_{\text{between}}$):**

$$df_{\text{between}}=k-1=3-1=2$$

- **Degrees of Freedom Within (df$_{\text{within}}$):**

$$df_{\text{within}}=N-k=11-3=8$$

Step 5: Calculate the Mean Squares

1. **Mean Square Between (MSB):**

$$\text{MSB}=\frac{\text{SSB}}{df_{\text{between}}}=\frac{19.25}{2}=9.625$$

2. **Mean Square Within (MSW):**

$$\text{MSW}=\frac{\text{SSW}}{df_{\text{within}}}=\frac{76.75}{8}=9.59375$$

Step 6: Calculate the F-Statistic

$$F=\frac{\text{MSB}}{\text{MSW}}=\frac{9.625}{9.59375}\approx 1.003$$

Step 7: Compare with Critical Value

The critical F-value at a 5% significance level with $df_{\text{between}}=2$ and $df_{\text{within}}=8$ is **4.26**.

Since our calculated F-statistic (1.003) is less than the critical F-value (4.26), we **fail to reject the null hypothesis**.

Q9. b. Question 9 b)

Using coding Method, subtract by 20. $n = 4+4+4+4 = 16$

C 5	B 3	A 0	D 0	8
A −1	D −1	C 1	B −2	−3
B −1	A −6	D −3	C 0	−10
D −3	C 0	B 1	A −5	−7
0	−4	−1	−7	−12

Square the values

25	9	0	0	
1	1	1	4	
1	36	9	0	
9	0	1	25	
36	46	11	29	122

i) Correction factor $= \frac{T^2}{n} = \frac{(12)^2}{16} = 9$

ii) $SST = \Sigma X^2 - \frac{T^2}{n} = 122 - 9 = 113$

iii) $SSR = \frac{8^2 + 3^2 + 10^2 + 7^2}{4} - 9 = 46.5$

iv) $SSC = \frac{0^2 + 4^2 + 1^2 + 7^2}{4} - 9 = 7.5$

From letters [Take values with respective letters]

$\Sigma A = -1 + 6 + 0 - 5 = -12$ $\quad \Sigma C = 5 + 0 + 1 + 0 = 6$

$\Sigma B = -1 + 3 + 1 - 2 = 1$ $\quad \Sigma D = -3 - 1 - 3 + 0 = -7$

v) $SSL = \dfrac{12^2+1^2+6^2+7^2}{4} - 9 = 48.5$.

vi) $SSE = SST - SSR - SSC - SSL$

$= 113 - 46.5 - 7.5 - 48.5 = 10.5$.

vii) ANOVA Table.

Source of Variation	Sum of Squares	degrees of freedom	Mean Sum of Squares	F-ratio.
Rows	SSR 46.5	4-1 3	MSR: 46.5/3 = 15.5	F_R = 15.5/1.75 = 8.85 > $4.76_{F_{3,6}}$
columns	SSC 7.5	4-1 3	MSC = 7.5/3 = 2.5	F_C = 2.5/1.75 = 1.42 < $4.76_{F_{3,6}}$
letters	SSL 48.5	4-1 3	MSL = 48.5/3 = 16.1	F_L = 16.1/1.75 = 9.2 > $4.76_{F_{3,6}}$
Error.	SSE 10.5	(4-1)(4-2) 6	MSE = 10.5/6 = 1.75	

∴ Hypothesis accepted for columns.
Hypothesis rejected for rows and letters.

We can conclude that variance between rows and variance between varieties (letters) are significant but variance between columns is insignificant.

Q10. a. Question 10 a)

Step 1: Set Up Hypotheses

1. **For Doctors (Row Factor):**
   - **Null Hypothesis ($H0_{\text{doctor}}$)**: There is no significant difference in recovery times across doctors.
   - **Alternative Hypothesis ($H1_{\text{doctor}}$)**: There is a significant difference in recovery times across doctors.
2. **For Treatments (Column Factor):**
   - **Null Hypothesis ($H0_{\text{treatment}}$)**: There is no significant difference in recovery times across treatments.
   - **Alternative Hypothesis ($H1_{\text{treatment}}$)**: There is a significant difference in recovery times across treatments.

Step 2: Organize the Data

The data provided is:

Doctor/Treatment	T_1	T_2	T_3	T_4
D_1	10	14	19	20
D_2	11	15	17	21
D_3	9	12	16	19
D_4	8	13	17	20

Step 3: Calculate the Means

1. **Calculate the Mean for Each Row (Doctors)**
   - Mean of D_1: $\frac{10+14+19+20}{4} = 15.75$
   - Mean of D_2: $\frac{11+15+17+21}{4} = 16$
   - Mean of D_3: $\frac{9+12+16+19}{4} = 14$
   - Mean of D_4: $\frac{8+13+17+20}{4} = 14.5$
2. **Calculate the Mean for Each Column (Treatments)**
   - Mean of T_1: $\frac{10+11+9+8}{4} = 9.5$
   - Mean of T_2: $\frac{14+15+12+13}{4} = 13.5$
   - Mean of T_3: $\frac{19+17+16+17}{4} = 17.25$
   - Mean of T_4: $\frac{20+21+19+20}{4} = 20$
3. **Calculate the Grand Mean (Overall Mean)**
   - Sum of all values = $10 + 14 + 19 + 20 + 11 + 15 + 17 + 21 + 9 + 12 + 16 + 19 + 8 + 13 + 17 + 20 = 231$
   - Total observations = 16

$$\text{Grand Mean} = \frac{231}{16} = 14.4375$$

Step 4: Calculate the Sum of Squares

1. **Total Sum of Squares (SST)**
   - Calculate the squared differences between each observation and the grand mean.

$$\mathrm{SST} = (10 - 14.4375)^2 + (14 - 14.4375)^2 + \cdots + (20 - 14.4375)^2$$

2. **Sum of Squares for Doctors ($\mathrm{SS}_{\mathrm{doctors}}$)**
   - Multiply each doctor mean's squared deviation from the grand mean by the number of treatments per doctor.

$$\mathrm{SS}_{\mathrm{doctors}} = 4 \times (15.75 - 14.4375)^2 + 4 \times (16 - 14.4375)^2 + \cdots + 4 \times (14.5 - 14.4375)^2$$

3. **Sum of Squares for Treatments ($\mathrm{SS}_{\mathrm{treatments}}$)**
   - Multiply each treatment mean's squared deviation from the grand mean by the number of doctors per treatment.

$$\mathrm{SS}_{\mathrm{treatments}} = 4 \times (9.5 - 14.4375)^2 + 4 \times (13.5 - 14.4375)^2 + \cdots + 4 \times (20 - 14.4375)^2$$

4. **Sum of Squares Within (SSW)**
   - This is the sum of the squared differences within each cell (each observation from its corresponding row and column mean).

$$\mathrm{SSW} = \mathrm{SST} - \mathrm{SS}_{\mathrm{doctors}} - \mathrm{SS}_{\mathrm{treatments}}$$

Step 5: Calculate the Degrees of Freedom

- **Degrees of Freedom for Doctors**: $df_{\mathrm{doctors}} = r - 1 = 4 - 1 = 3$
- **Degrees of Freedom for Treatments**: $df_{\mathrm{treatments}} = c - 1 = 4 - 1 = 3$
- **Degrees of Freedom for Within**: $df_{\mathrm{within}} = (r - 1)(c - 1) = 3 \times 3 = 9$

Step 6: Calculate the Mean Squares

1. **Mean Square for Doctors ($\mathrm{MS}_{\mathrm{doctors}}$)**

$$\mathrm{MS}_{\mathrm{doctors}} = \frac{\mathrm{SS}_{\mathrm{doctors}}}{df_{\mathrm{doctors}}}$$

2. **Mean Square for Treatments ($\mathrm{MS}_{\mathrm{treatments}}$)**

$$\mathrm{MS}_{\mathrm{treatments}} = \frac{\mathrm{SS}_{\mathrm{treatments}}}{df_{\mathrm{treatments}}}$$

3. **Mean Square for Within ($\mathrm{MS}_{\mathrm{within}}$)**

$$\mathrm{MS}_{\mathrm{within}} = \frac{\mathrm{SSW}}{df_{\mathrm{within}}}$$

Step 7: Calculate the F-Statistics

1. **F for Doctors**

$$F_{\text{doctors}} = \frac{\text{MS}_{\text{doctors}}}{\text{MS}_{\text{within}}}$$

2. **F for Treatments**

$$F_{\text{treatments}} = \frac{\text{MS}_{\text{treatments}}}{\text{MS}_{\text{within}}}$$

Step 8: Compare with Critical Value

The critical F-value at a 5% significance level with $df_{\text{between}} = 3$ and $df_{\text{within}} = 9$ is **3.86**.

- If $F_{\text{doctors}} > 3.86$, we reject the null hypothesis for doctors.
- If $F_{\text{treatments}} > 3.86$, we reject the null hypothesis for treatments.

Q10. b. Question 10 b)

Step 1: Extract Data and Organize

Based on the table:

- **Group A**: Length of Stay = 3 (6 patients), 4 (14 patients)
- **Group B**: Length of Stay = 4 (18 patients), 5 (2 patients)
- **Group C**: Length of Stay = 4 (10 patients), 5 (9 patients), 6 (1 patient)
- **Group D**: Length of Stay = 4 (8 patients), 5 (12 patients)

We'll expand these frequencies into data points for each group to find individual means and sums.

Step 2: Calculate the Sum and Mean for Each Group

We calculate the total length of stay and mean for each group.

Group A:

- Data: $3 \times 6 + 4 \times 14$
- Sum = $3 \cdot 6 + 4 \cdot 14 = 18 + 56 = 74$
- Number of Patients = $6 + 14 = 20$
- Mean for Group A, $\bar{X}_A = \frac{74}{20} = 3.7$

Group B:

- Data: $4 \times 18 + 5 \times 2$
- Sum = $4 \cdot 18 + 5 \cdot 2 = 72 + 10 = 82$
- Number of Patients = $18 + 2 = 20$
- Mean for Group B, $\bar{X}_B = \frac{82}{20} = 4.1$

Group C:

- Data: $4 \times 10 + 5 \times 9 + 6 \times 1$
- Sum = $4 \cdot 10 + 5 \cdot 9 + 6 \cdot 1 = 40 + 45 + 6 = 91$
- Number of Patients = $10 + 9 + 1 = 20$
- Mean for Group C, $\bar{X}_C = \frac{91}{20} = 4.55$

Group D:

- Data: $4 \times 8 + 5 \times 12$
- Sum = $4 \cdot 8 + 5 \cdot 12 = 32 + 60 = 92$
- Number of Patients = $8 + 12 = 20$
- Mean for Group D, $\bar{X}_D = \frac{92}{20} = 4.6$

Step 3: Calculate the Overall Mean

The overall mean is calculated by summing all observations and dividing by the total number of patients.

Total Sum = $74 + 82 + 91 + 92 = 339$
Total Number of Patients = $20 + 20 + 20 + 20 = 80$

Overall Mean, $\bar{X} = \frac{339}{80} = 4.2375$

Step 4: Calculate the Sum of Squares

4.1 Between-Groups Sum of Squares (SSB)

SSB measures the variation due to the differences between group means.

$$SSB = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2$$

where n_i is the number of patients in each group, X_i is the mean of each group, and X is the overall mean.

Using the values:

- For Group A: $20 \cdot (3.7 - 4.2375)^2$
- For Group B: $20 \cdot (4.1 - 4.2375)^2$
- For Group C: $20 \cdot (4.55 - 4.2375)^2$
- For Group D: $20 \cdot (4.6 - 4.2375)^2$

Calculating each term:

$$SSB_A = 20 \cdot (3.7 - 4.2375)^2 = 20 \cdot 0.2885^2 = 1.6653$$

$$SSB_B = 20 \cdot (4.1 - 4.2375)^2 = 20 \cdot 0.0185^2 = 0.00685$$

$$SSB_C = 20 \cdot (4.55 - 4.2375)^2 = 20 \cdot 0.3125^2 = 1.9531$$

$$SSB_D = 20 \cdot (4.6 - 4.2375)^2 = 20 \cdot 0.3625^2 = 2.6281$$

Adding these together:

$$SSB = 1.6653 + 0.00685 + 1.9531 + 2.6281 = 6.25335$$

4.2 Within-Groups Sum of Squares (SSW)

SSW measures the variation within each group.

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

I'll continue the calculations for the within-groups sum of squares next, and then we'll calculate the F-ratio.

The Between-Groups Sum of Squares (SSB) is approximately 10.7375.

Step 5: Calculate Within-Groups Sum of Squares (SSW)

To calculate SSW, we need to measure the variance within each group based on the individual data points and their respective group means.

Group A:

- $(3 - 3.7)^2 \times 6 + (4 - 3.7)^2 \times 14$

Group B:

- $(4 - 4.1)^2 \times 18 + (5 - 4.1)^2 \times 2$

Group C:

- $(4 - 4.55)^2 \times 10 + (5 - 4.55)^2 \times 9 + (6 - 4.55)^2 \times 1$

Group D:

- $(4 - 4.6)^2 \times 8 + (5 - 4.6)^2 \times 12$

Step 6: Calculate Degrees of Freedom

- **Between-Groups Degrees of Freedom ($\mathrm{df}_{\mathrm{between}}$):** $k - 1 = 4 - 1 = 3$
- **Within-Groups Degrees of Freedom ($\mathrm{df}_{\mathrm{within}}$):** $N - k = 80 - 4 = 76$

Step 7: Calculate Mean Squares

- **Mean Square Between (MSB):**

$$MSB = \frac{SSB}{df_{\mathrm{between}}} = \frac{10.7375}{3} = 3.5792$$

- **Mean Square Within (MSW):**

$$MSW = \frac{SSW}{df_{\mathrm{within}}} = \frac{17.75}{76} = 0.2336$$

Step 8: Calculate the F-Ratio

$$F = \frac{MSB}{MSW} = \frac{3.5792}{0.2336} \approx 15.32$$

Step 9: Compare with the Critical Value

With a significance level of 0.01 and degrees of freedom $df_{\mathrm{between}} = 3$ and $df_{\mathrm{within}} = 76$, we look up the critical F-value, $F_{0.01}$. If the calculated F-ratio (15.32) exceeds this critical value, we reject the null hypothesis.