9	a)	In two large populations, there are 30%, and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.	L4	CO5	5 M
	b)	The mean and standard deviation of a population are 11795 and 14054 respectively. If n=50 find a 95% confidence interval for the mean.	L3	CO3	5 M
		UNIT-V			
10	a)	Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins, show the sample standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test the hypothesis that the true variance are equal.	L3	CO3	5 M
	b)	In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have the same variance.	L3	CO3	5 M
		OR			
11		Given the following contingency table for hair colour and eye colour. Find the value of chi square. Is there good association between the two?	L4	CO5	10 M

		Fair	Brown	Black
Eye Colour	Blue	15	5	20
	Grey	20	10	20
	Brown	25	15	20

PVP 23

Code: 23BS1402

II B.Tech - II Semester – Regular Examinations - MAY 2025

PROBABILITY AND STATISTICS

(Common for ME, CSE, IT, AIML, DS)

Duration: 3 hours Max. Marks: 70

Note: 1. This question paper contains two Parts A and B.
2. Part-A contains 10 short answer questions. Each Question carries 2 Marks.
3. Part-B contains 5 essay questions with an internal choice from each unit. Each Question carries 10 marks.
4. All parts of Question paper must be answered in one place.

BL – Blooms Level CO – Course Outcome

PART – A

		BL	CO
1.a)	Define conditional probability.	L2	CO1
1.b)	What is the probability of getting an even prime number on throwing a die?	L2	CO1
1.c)	A binomial distribution has mean 3 and variance 2 then find number of trials.	L3	CO2
1.d)	Define probability density function.	L2	CO2
1.e)	Define coefficient of correlation.	L2	CO1
1.f)	Write the normal equations of parabola.	L2	CO2
1.g)	Define Type-I and Type-II error.	L2	CO1
1.h)	Write the formula of Test for single proportion.	L2	CO3
1.i)	Write the application of Chi-square test.	L2	CO3
1.j)	Write the properties of t - distribution.	L2	CO2

PART – B

			BL	CO	Max. Marks
		UNIT-I			
2		The following table gives the daily income of 150 workers of a factory. Find mean and mode.	L3	CO2	10 M

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	8	12	20	30	15	10	5

			BL	CO	Max. Marks
		OR			
3	a)	In a bolt factory machines A_1, A_2, A_3 manufacture respectively 25%, 35% and 40% of the total output. Of these 5, 4, and 2 percent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machine A_2 and A_3.	L3	CO2	5 M
	b)	Find the probability of drawing 2 red balls in succession from a bag containing 4 red and 5 blue balls when the ball that is drawn first is not replaced.	L3	CO2	5 M
		UNIT-II			
4		The probability density f(x) of a continuous random variable is given by $$f(x) = ce^{-\|x\|}, -\infty < x < \infty.$$ Show that $c = \frac{1}{2}$ and find the mean and variance of the distribution.	L3	CO2	10 M
		OR			
5	a)	If the masses of 300 students are normally distributed with mean 68 kgs and standard deviation 3kgs. How many students have masses.	L4	CO4	5 M

		i) greater than 72kgs. ii) less than or equal to 64 kgs iii) between 65 and 71 kgs inclusive.			
	b)	Out of 800 families with 5 children each, how many would you expect to have (i) 3 boys (ii) Either 2 or 3 boys (iii) at least one boy? Assume equal probability for boys and girls.	L4	CO4	5 M
		UNIT-III			
6	a)	Find the coefficient of correlation for the following data.	L3	CO2	5 M

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

	b)	Find the mean values of the variables X and Y and correlation coefficient from the following regression equations 2Y-X-50=0 and 3Y-2X-10=0.	L4	CO4	5 M
		OR			
7		Fit an exponential curve $Y=ae^{bx}$ for the given data	L3	CO2	10 M

X	40	65	90	5	30	10	80	85	70	25
Y	30	20	10	80	40	65	15	15	20	50

		UNIT-IV			
8	a)	A Sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample from a population with mean weight 56 kgs and standard deviation 25 kgs.	L3	CO3	5 M
	b)	Write the steps involved in test of significance for single sample proportion and population proportion.	L3	CO3	5 M
		OR			

Subject Code: 23BS1402 PVP23

II B.Tech. II Semester Regular Examinations - MAY-2025

PROBABILITY AND STATISTICS

Key & Scheme of Evaluation

1.a) Let A and B two events in the sample space S and $P(A) \neq 0$. The probability of B after the event A has occurred is called the conditional probability of the event B given A and is denoted by $P(B/A) = \frac{P(A\cap B)}{P(A)}$ — **2**

1.b) No. of favourable cases (even prime number) = 1, No. of outcomes = 6 — **1**

Required probability $= \frac{1}{6}$ — **1**

1.c) $\text{mean} = np = 3, \ \text{variance} = npq = 2$ — **1**

Then, $q = \frac{2}{3}$ and $p = 1 - q = \frac{1}{3}$

Now $np = 3 \Rightarrow n\left(\frac{1}{3}\right) = 3 \Rightarrow n = 9$ — **1**

1.d) For a continuous random variable, X the probability distribution is called probability distribution function. i.e. $P\left(x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2}\right) = f(x)dx$ then $f(x)$ is called the probability distribution function of X. — **2**

1.e) A mathematical method for measuring the magnitude of linear relationship between two variables. — **2**

1.f) The normal equations to fit the parabola $y = a + bx + cx^2$ are given by

$$na + b\sum x + c\sum x^2 = \sum y$$

$$a\sum x + b\sum x^2 + c\sum x^3 = \sum xy$$

$$a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum x^2 y$$

— **1**, **1**

1.g) *Type I error:* It involves rejection of H_0 when it is true(T). — **1**

Type II error: It involves acceptance of H_0 when it is false(F). — **1**

1.h $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ or $Z = \frac{P - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ — **2**

1.i)
1. To test the goodness of fit — **1**
2. To test the independence of attributes — **1**
3. To test the homogeneity (population variance or population correlation coefficient)

Note: Award 1 mark for writing any(other) ONE application.

1.j)
1. The shape of t-distribution is bell shaped — **1**
2. Symmetrical about the mean — **1**
3. It is unimodal with mean=mode=median
4. Symmetrical about the line $t = 0$

Note: Award 1 mark for writing any(other) ONE property.

2. Note: In the question it is given no. of workers $N = 150$ instead of 100.

Class interval	Frequency, f	Mid x	$f.x$	Cumulative frequency
0-10	8	5	40	8
10-20	12	15	180	20
20-30	20	25	500	40
30-40	30	35	1050	70
40-50	15	45	675	85
50-60	10	55	550	95
60-70	5	65	325	**100**
	$\mathbf{\sum f = N = 100}$		$\mathbf{\sum f.x = 3320}$	

— **4**

$\text{Mean} = \frac{\sum f.x}{N} = \frac{3320}{100} = 33.2$ — **3**

Mode $= l + h \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2} = 30 + 10\left(\frac{30-20}{(30-20)+(30-15)}\right) = 30 + 10\left(\frac{10}{25}\right) = 34$ **3**

Note:-Marks can be awarded for alternate procedure(s) and/or solution(s)

3.a) Let $P(A_1), P(A_2), P(A_3)$ be the probabilities of the events that the bolts are manufactured by the machines A_1, A_2, A_3 respectively. Then

$$P(A_1) = \frac{25}{100}, P(A_2) = \frac{35}{100}, P(A_3) = \frac{40}{100}$$ **1**

Let D be the event that the bolt is defective.

Then, $P(D/A_1) = \frac{5}{100}$, $P(D/A_2) = \frac{4}{100}$, $P(D/A_3) = \frac{2}{100}$

By Baye's theorem, $P(A_2/D) = \frac{P(A_2)\,P(D/A_2)}{P(A_1)\,P(D/A_1) + P(A_2)\,P(D/A_2) + P(A_3)P(D/A_3)}$

$$= \frac{\left(\frac{35}{100}\right)\left(\frac{4}{100}\right)}{\left(\frac{25}{100}\right)\left(\frac{5}{100}\right) + \left(\frac{35}{100}\right)\left(\frac{4}{100}\right) + \left(\frac{40}{100}\right)\left(\frac{2}{100}\right)} = \frac{140}{345} (= 0.4058)$$ **2**

$$P(A_3/D) = \frac{P(A_3)\,P(D/A_3)}{P(A_1)\,P(D/A_1) + P(A_2)\,P(D/A_2) + P(A_3)P(D/A_3)}$$

$$= \frac{\left(\frac{40}{100}\right)\left(\frac{2}{100}\right)}{\left(\frac{25}{100}\right)\left(\frac{5}{100}\right) + \left(\frac{35}{100}\right)\left(\frac{4}{100}\right) + \left(\frac{40}{100}\right)\left(\frac{2}{100}\right)} = \frac{80}{345} (= 0.2319)$$ **2**

3.b) Total balls in the bag, 4 Red + 5 Blue = 9

Let A, B be the events that the two red balls are drawn in succession. **2**

Then $P(A) = \frac{4}{9}$, $P(B) = \frac{3}{8}$ (Drawn first ball is not replaced) **2**

$\therefore$ Required probability $= \frac{4}{9} \cdot \frac{3}{8} = \frac{1}{6} (= 0.1667)$ **1**

4) Since, $f(x)$ is probability density function, we have $\int_{-\infty}^{\infty} f(x)\,dx = 1$ **2**

$\Rightarrow \int_{-\infty}^{\infty} ce^{-|x|}\,dx = 1$

$\Rightarrow 2c\int_0^{\infty} e^{-x}\,dx = 1$ Intergand is even

$\Rightarrow 2c\left[\frac{e^{-x}}{-1}\right]_0^{\infty} = 1$

$\Rightarrow 2c(1) = 1 \Rightarrow c = \frac{1}{2}$ **2**

Mean of X is $\mu = \int_{-\infty}^{\infty} xf(x)\,dx$ **1**

$= \int_{-\infty}^{\infty} x \cdot ce^{-|x|}\,dx$

$= c\int_{-\infty}^{\infty} xe^{-|x|}\,dx = 0$ Intergand is odd **2**

Variance of X is $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

$= \int_{-\infty}^{\infty} x^2 \cdot ce^{-|x|}\,dx - \mu^2$ **1**

$= 2c\int_0^{\infty} x^2 e^{-x}\,dx - \mu^2$ Intergand is even

$= [x^2(-e^{-x}) - 2x(e^{-x}) + 2(-e^{-x})]_0^{\infty} - \mu^2 = 2 - 0^2 = 2$ **2**

Note:-Marks can be awarded for alternate procedure(s) and/or solution(s)

5.a) Given, $\mu = 68$, $\sigma = 3$.

In normal distribution, $z = \frac{x-\mu}{\sigma}$ **1**

i) When $X = 72$, $z = \frac{x-\mu}{\sigma} = \frac{72-68}{3} = 1.33$

$\therefore P(X > 72) = P(z > 1.33)$

$= 0.5 - A(1.33)$

$= 0.5 - 0.4082$

$= 0.0918$

Number of students have weight greater than 72 kgs are $0.0918 \text{x} 300 = 27.54 \cong 28$ **1**

$\therefore P(X \leq 64) = P(z \leq -1.33)$
$= 0.5 - A(-1.33)$
$= 0.5 - A(1.33)$
$= 0.5 - 0.4082$
$= 0.0918$

Number of students have weight less than or equal to 64 kgs are 0.0918 x 300 **1**
$= 27.54 \cong 28$

iii) When $X = 65,\ z = \frac{x-\mu}{\sigma} = \frac{65-68}{3} = -1$

When $X = 71,\ z = \frac{x-\mu}{\sigma} = \frac{71-68}{3} = 1$

$\therefore P(65 \leq X \leq 71) = P(-1 \leq z \leq 1)$
$= A(-1) + A(1)$
$= A(1) + A(1)$
$= 0.3413 + 0.3413 = 0.6826$

Number of students have weight between 65 and 71 kgs are 0.6826 x 300 **2**
$= 204.78 \cong 205$

5.b) Here number of families, $N = 800$.

$n = 5,\ p = q = \frac{1}{2}$

Let $p(r)$ denotes the probability that the family has r boys out of 5 children

where $p(r) = P(X = r) = \binom{n}{r} p^r q^{n-r}$ **2**

i) Probability of one family contains 3 boys, $p(3) = \binom{5}{3} p^3 q^{5-3}$

$$= 10\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5}{16} = 0.3125$$

Out of 800 families, $800 \cdot p(3) = 800 \cdot \frac{5}{16} = 250$ families **1**

ii) Probability of one family contains either 2 or 3 boys, $p(2) + p(3)$

$$= \binom{5}{2} p^2 q^{5-2} + \binom{5}{3} p^3 q^{5-3}$$

$$= 10\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 + 10\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2$$

$$= \frac{5}{16} + \frac{5}{16} = \frac{10}{16} = \frac{5}{8}$$

Out of 800 families, $800 \cdot [p(2) + p(3)] = 800 \cdot \frac{5}{8} = 500$ families. **1**

iii) Probability of one family contains atleast one boy, $P(X \geq 1) = 1 - P(X = 0)$

$$= 1 - \binom{5}{0} p^0 q^{5-0}$$

$$= 1 - \left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{31}{32}$$

Out of 800 families, $800 \cdot \frac{31}{32} = 775$ families. **1**

6.a) $\bar{x} = \frac{\Sigma x}{n} = \frac{544}{8} = 68 \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{552}{8} = 69$ **2**

	x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	X^2	Y^2	XY
	65	67	-3	-2	9	4	6
	66	68	-2	-1	4	1	2
	67	65	-1	-4	1	16	4
	67	68	-1	-1	1	1	1
	68	72	0	3	0	9	0
	69	72	1	3	1	9	3
	70	69	2	0	4	0	0
	72	71	4	2	16	4	8
Sum	**544**	**552**			**36**	**44**	**24**

2

$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} = \frac{24}{\sqrt{(36)(44)}} = 0.6030$ **1**

6.b) Let $\bar{x}$ and $\bar{y}$ be the mean values of variables x and y respectively.

Solving the given two lines of regressions, we get $\bar{x} = 130$, $\bar{y} = 90$ **2**

Let $2y - x = 50$ be the regression equation of y on x

$$\Rightarrow 2y = x + 50$$
$$\Rightarrow y = 0.5x + 25$$
$$\therefore b_{yx} = 0.5$$ **1**

Let $3y - 2x = 10$ be the regression equation of x on y

$$\Rightarrow -2x = -3y + 10$$
$$\Rightarrow x = 1.5y - 5$$
$$\therefore b_{xy} = 1.5$$ **1**

$$\therefore r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{(0.5)(1.5)} = 0.866$$ **1**

7) Let $y = ae^{bx}$

$\Rightarrow log\, y = log\, a + log\, e^{bx}$

$\Rightarrow Y = A + bx \longrightarrow \boxed{1}$ where $Y = log\, y$, $A = log\, a$

The normal equations to fit $\boxed{1}$ are given by $nA + b\sum x = \sum Y$ and $A\sum x + b\sum x^2 = \sum xY$ **3**

	x	y	x^2	$Y = log\, y$	xY
	40	30	1600	3.4012	136.048
	65	20	4225	2.9957	194.7205
	90	10	8100	2.3026	207.234
	5	80	25	4.3820	21.91
	30	40	900	3.6889	110.667
	10	65	100	4.1744	41.744
	80	15	6400	2.7081	216.648
	85	15	7225	2.7081	230.1885
	70	20	4900	2.9957	209.699
	25	50	625	3.9120	97.8
Sum	**500**		**34100**	**33.2687**	**1466.659**

5

$$\therefore 10A + 500b = 33.2687 \quad \text{and} \quad 500A + 34100b = 1466.659$$
$$\Rightarrow A = log\, a = 4.4081 \text{ and } b = -0.0216$$
$$\Rightarrow a = e^A = e^{4.4081} = 82.1133 \text{ and } b = -0.0216$$
$$\therefore y = ae^{bx} = 82.1133e^{-0.0216x}$$ **2**

Note:- Marks can be awarded for alternate procedure(s) or method(s) or solution(s)

8.a) Given, population mean, $\mu = 56$ population standard deviation, $\sigma = 25$

Sample size, $n = 64$ Sample mean, $\bar{x} = 70$ **1**

Null Hypothesis, H_0*:* The sample can be regarded as came from a population i.e. $\mu = 56$

Alternative Hypothesis, H_1*:* The sample cannot be regarded as came from a population i.e. $\mu \neq 56$ **2**

Level of significance, α*:* $\alpha = 5\,\% = 0.05$

For 5 % level of significance, $z_{\alpha/2} = 1.96$

Test Statistic: $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{70-56}{25/\sqrt{64}} = \frac{14}{25/8} = 4.48$ **1**

	Since, $\|z\| > z_{\alpha/2}$ null hypothesis, H_0 is rejected. ∴ The sample is cannot be regarded as came from a population with mean 56.	1
8.b)	Suppose a large sample of size n has a sample proportion p. To test the population proportion P, has a specified value P_0 *Null Hypothesis, H_0:* This is a statement of no effect or no difference. $H_0: P = P_0$ *Alternative Hypothesis, H_1:* The alternative hypothesis is • Right-tailed $H_1: P > P_0$ • Left-tailed $H_1: P < P_0$ • Two-tailed $H_1: P \neq P_0$ *Level of significance, α:* The values for significance level, α are 0.05 (5%) and 0.01 (1%) *Test Statistic:* The test statistic calculated as $Z = \frac{P - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ or $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ *Conclusion:* Compare the computed value z with the critical value z_α or $z_{\alpha/2}$ If $\|z\| < z_\alpha$ we conclude that it is not significant and accept the null hypothesis. If $\|z\| > z_\alpha$ we conclude that it is significant and reject the null hypothesis.	2 2 1
9.a)	Given, $n_1 = 1200$ $P_1 = \frac{x_1}{n_1} = \frac{30}{100} = 0.3$ $Q_1 = 1 - P_1 = 0.7$ $n_2 = 900$ $P_2 = \frac{x_2}{n_2} = \frac{25}{100} = 0.25$ $Q_2 = 1 - P_2 = 0.75$ *Null Hypothesis, H_0:* $P_1 = P_2$ *Alternative Hypothesis, H_1:* $P_1 \neq P_2$ *Level of significance, α:* $\alpha = 5\ \% = 0.05$ For 5 % level of significance, $z_{\alpha/2} = 1.96$ *Test Statistic:* $z = \frac{P_1 - P_2}{\sqrt{\frac{P_!Q_!}{n_1} + \frac{P_2Q_2}{n_2}}} = \frac{0.3 - 0.25}{\sqrt{\left(\frac{0.3\text{x}0.7}{1200} + \frac{0.25\text{x}0.75}{900}\right)}} = 2.56$ *Conclusion:* Here $\|z\| = 2.56 > 1.96$ i.e. $\|z\| > z_{\alpha/2}$ Since, $\|z\| >< z_{\alpha/2}$ null hypothesis, H_0 is rejected. ∴ The difference in proportions is to be hidden unlikely.	2 1 1 1
9.b)	Here $n = 50,\ \overline{x} = 11795,\ \sigma = 14054$ 95 % confidence interval is $\left(\overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(11795 \pm 1.96 \text{ x } \frac{14054}{\sqrt{50}}\right)$ $= (11795 \pm 3896)$ $= (11795 - 3896,\ 11795 + 3896)$ $= (7899, 15691)$	2 1 2
10.a)	Given, First sample size, $n_1 = 11$ Second sample size, $n_2 = 9$ First sample standard deviation, $s_1 = 0.8$ Second sample standard deviation, $s_2 = 0.5$ Then, $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11(0.8)^2}{10} = 0.704$ $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{9(0.5)^2}{8} = 0.28125 \cong 0.2813$ *Null Hypothesis, H_0:* There is no significant difference between the variances. i.e. $\sigma_1^2 = \sigma_2^2$ *Alternative Hypothesis, H_1:* There is significant difference between the variances. i.e. $\sigma_1^2 \neq \sigma_2^2$ *Level of significance, α:* $\alpha = 5\ \% = 0.05$ For 5 % level of significance with $F_\alpha(\nu_1, \nu_2) = F_\alpha(10, 8) = 3.35$ *Test Statistic:* $F = \frac{S_1^2}{S_2^2} = \frac{0.704}{0.2813} = 2.5027$ *Conclusion:* Here $F = 2.5027 < 3.35$ i.e. $F < F_\alpha$ Since, $F < F_\alpha$ null hypothesis, H_0 is accepted. i.e. There is no significant difference between the variances.	2 1 1 1

Q.No.	Solution	Marks
10.b)	Given, First sample size, $n_1 = 8$ Second sample size, $n_2 = 10$ Also, $\sum(x_i - \bar{x})^2 = 84.4$ $\sum(y_j - \bar{y})^2 = 102.6$	2
	$S_1^2 = \frac{\sum(x_i-\bar{x})^2}{n_1 - 1} = \frac{84.4}{7} = 12.0571$ $S_2^2 = \frac{\sum(y_j-\bar{y})^2}{n_2 - 1} = \frac{102.6}{9} = 11.4$	1
	Null Hypothesis, H_0: There is no significant difference between the variances. i.e. $\sigma_1^2 = \sigma_2^2$ *Alternative Hypothesis, H_1:* There is significant difference between the variances. i.e. $\sigma_1^2 \neq \sigma_2^2$ *Level of significance, α:* $\alpha = 5\ \% = 0.05$ For 5 % level of significance with $F_\alpha(\nu_1, \nu_2) = F_\alpha(7, 9) = 3.29$	1
	Test Statistic: $F = \frac{S_1^2}{S_2^2} = \frac{12.0571}{11.4} = 1.0576$	1
	Conclusion: Here $F = 1.0576 < 3.29$ i.e. $F < F_\alpha$ Since, $F < F_\alpha$ null hypothesis, H_0 is accepted. i.e. There is no significant difference between the variances.	
11)		

Eye colour	Hair colour			
	Fair	Brown	Black	Total
Blue	15	5	20	40
Grey	20	10	20	50
Brown	25	15	20	60
Total	60	30	60	150

Expected frequency table:

$E(15) = \frac{(60)(40)}{150} = 16$	$E(5) = \frac{(30)(40)}{150} = 8$	$E(20) = \frac{(60)(40)}{150} = 16$
$E(20) = \frac{(60)(50)}{150} = 20$	$E(10) = \frac{(30)(50)}{150} = 10$	$E(20) = \frac{(60)(50)}{150} = 20$
$E(25) = \frac{(60)(60)}{150} = 24$	$E(15) = \frac{(30)(60)}{150} = 12$	$E(20) = \frac{(60)(60)}{150} = 24$

3

Observed frequency, O_i	Expected frequency, E_i	$O_i - E_i$	$\frac{(O_i - E_i)^2}{E_i}$
15	16	1	0.0625
5	8	-3	1.125
20	16	4	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.0417
15	12	3	0.75
20	24	-4	0.6667
		Sum	3.6459

3

Null Hypothesis, H_0: The hair and eye colours are not associated
Alternative Hypothesis, H_1: The hair and eye colours are associated. 1
Level of significance, α: $\alpha = 5\ \% = 0.05$
For 5 % level of significance with $\nu = (m-1)(n-1) = (2)(2) = 4,\ \chi_\alpha^2 = 9.488$ 2

Test Statistic: $\chi^2 = \sum_{i=0}^{n} \frac{(O_i - E_i)^2}{E_i} = 3.6459$

Conclusion: Here $\chi^2 = 3.6459$ i.e. $\chi^2 < \chi_\alpha^2$
Since, $\chi^2 < {\chi^2}_\alpha$ null hypothesis, H_0 is accepted.
i.e. The hair and eye colours are not associated. 1

Note:- Marks can be awarded for alternate procedure(s) or method(s) and/or solution(s).