

Statistical Modelling for Machine Learning (313307)

Model Question Paper — Set B · Model Answers

Q.1

a) Combined mean.

2 M

$$\bar{x}_{12} = (n_1\bar{x}_1 + n_2\bar{x}_2)/(n_1 + n_2) = (60 \times 55 + 40 \times 65)/(60 + 40) = (3300 + 2600)/100 = 59$$

(Formula 1 M, answer 1 M)

b) Sixth decile.

2 M

Ascending order (n = 12): 21, 26, 29, 30, 33, 34, 37, 38, 41, 45, 48, 52

$$D_6 = 6(n + 1)/10 \text{ th value} = 6 \times 13 / 10 = 7.8 \text{th value} = 7 \text{th value} + 0.8 \times (8 \text{th value} - 7 \text{th value}) = 37 + 0.8 \times (38 - 37) = 37.8$$

(Arrangement ½ M, position ½ M, D_6 1 M)

c) Covariance.

2 M

x	y	dx = x - $\bar{x}$	dy = y - $\bar{y}$	dx ²	dy ²	dx·dy
2	5	-4	-6	16	36	24
4	9	-2	-2	4	4	4
6	10	0	-1	0	1	0
8	14	2	3	4	9	6
10	17	4	6	16	36	24
$\Sigma x = 30$	$\Sigma y = 55$			40	86	58

$$\bar{x} = 6, \bar{y} = 11$$

$$\text{Cov}(x, y) = \Sigma(x - \bar{x})(y - \bar{y})/n = 58/5 = 11.6$$

(Table 1 M, answer 1 M)

d) Regression coefficient of x on y.

2 M

$$r^2 = b_{yx} \times b_{xy} \Rightarrow (0.8)^2 = 1.6 \times b_{xy} \Rightarrow b_{xy} = 0.64/1.6 = 0.4$$

(Property 1 M, answer 1 M)

e) Three coins — at least two heads.

2 M

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}, n(S) = 8.$$

$$A = \text{at least two heads} = \{HHH, HHT, HTH, THH\}, n(A) = 4 \Rightarrow P(A) = 4/8 = 1/2$$

(Sample space 1 M, answer 1 M)

f) Backward difference table.

2 M

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
1	4				
		8			
2	12		12		
		20		6	
3	32		18		0
		38		6	
4	70		24		
		62			
5	132				

(Bottom diagonal in bold.) $\nabla^2 y_5 = 24$, $\nabla^3 y_5 = 6$

(Table 1½ M, values ½ M)

g) Type I and Type II errors.

2 M

- **Type I error:** rejecting the null hypothesis H_0 when it is actually true. Its probability is the level of significance α (e.g. 5%).
- **Type II error:** accepting (not rejecting) H_0 when it is actually false. Its probability is denoted by β .

Decision	H_0 true	H_0 false
Reject H_0	Type I error (α)	Correct decision
Accept H_0	Correct decision	Type II error (β)

(Each definition 1 M)

Q.2

a) Less-than ogive and median.

4 M

Less-than cumulative frequencies are plotted against the upper class boundaries and joined by a smooth curve (ogive). Mark $N/2$ on the y-axis, draw a horizontal line to meet the ogive and drop a perpendicular to the x-axis — its foot is the median.

Marks less than	10	20	30	40	50	60
c.f.	4	14	32	54	66	70

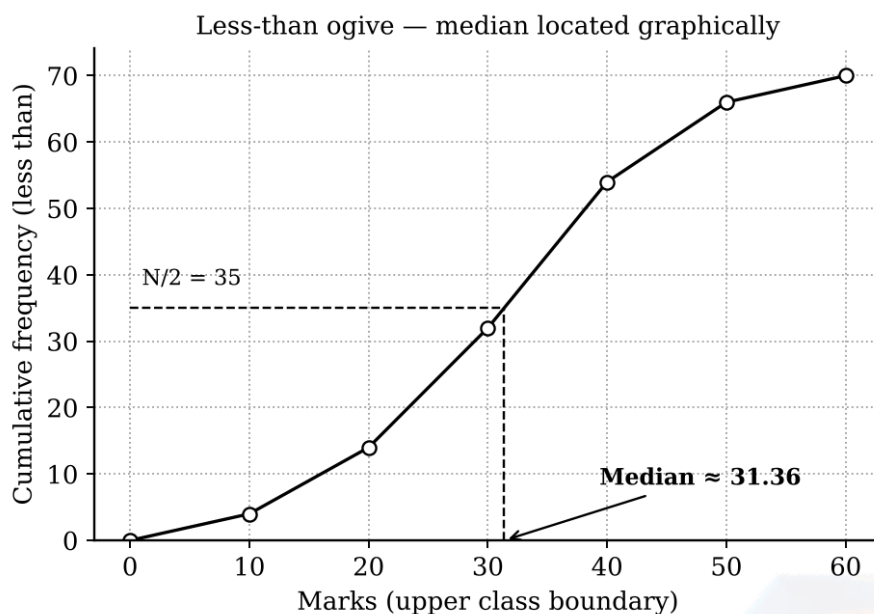


Fig.: Less-than ogive

Verification by formula:

$$N/2 = 70/2 = 35$$

Median class: the c.f. just greater than or equal to 35 is 54 $\Rightarrow$ Median class = **30–40**

$L = 30$, $cf = 32$ (c.f. of the class before), $f = 22$, $h = 10$

$$\text{Median} = L + ((N/2 - cf) / f) \times h = 30 + ((35 - 32) / 22) \times 10 = \mathbf{31.36}$$

(Ogive 2 M, median from graph 1 M, verification 1 M)

b) Histogram and mode.

4 M

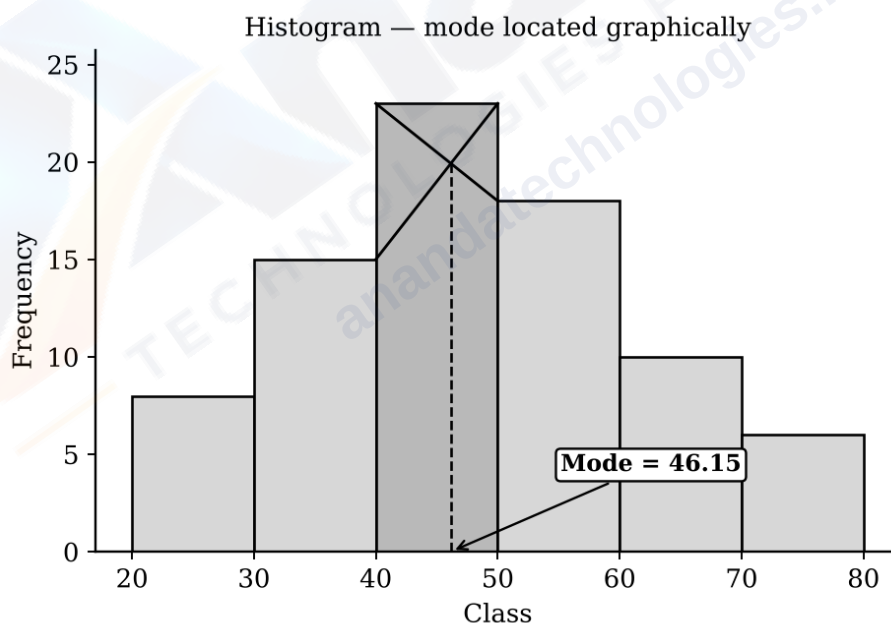


Fig.: Histogram — the crossing lines in the modal bar locate the mode

Lines are drawn from the top corners of the highest bar to the adjacent corners of the neighbouring bars; the perpendicular from their intersection gives the mode.

By formula:

$$\text{Maximum frequency} = 23 \Rightarrow \text{Modal class} = \mathbf{40-50}$$

$L = 40$, $f_1 = 23$ (modal class), $f_0 = 15$ (class before), $f_2 = 18$ (class after), $h = 10$

$$\text{Mode} = L + ((f_1 - f_0) / (2f_1 - f_0 - f_2)) \times h = 40 + ((23 - 15) / (2 \times 23 - 15 - 18)) \times 10 = 40 + (8 / 13) \times 10 = \mathbf{46.15}$$

(Histogram 2 M, graphical mode 1 M, formula 1 M)

c) Coefficient of variation — consistency.

4 M

Batsman A

x	$x - \bar{x}$	$(x - \bar{x})^2$
42	-4.67	21.7778
38	-8.67	75.1111
55	8.33	69.4444
61	14.33	205.4444
35	-11.67	136.1111
49	2.33	5.4444
$\Sigma x = 280$		$\Sigma(x - \bar{x})^2 = 513.3333$

Batsman B

x	$x - \bar{x}$	$(x - \bar{x})^2$
45	-2.67	7.1111
50	2.33	5.4444
47	-0.67	0.4444
52	4.33	18.7778
43	-4.67	21.7778
49	1.33	1.7778
$\Sigma x = 286$		$\Sigma(x - \bar{x})^2 = 55.3333$

A: $\bar{x} = 46.67$, $\sigma = \sqrt{(513.33/6)} = 9.2496$, C.V. = $(\sigma/\bar{x}) \times 100 = \mathbf{19.82\%}$

B: $\bar{x} = 47.67$, $\sigma = \sqrt{(55.33/6)} = 3.0368$, C.V. = $(\sigma/\bar{x}) \times 100 = \mathbf{6.37\%}$

Since C.V. of B is less, **batsman B is more consistent.**

(Mean and σ of each 1 M, C.V. 1 M, conclusion 1 M)

d) S.D. and C.V. from Sk (median form).

4 M

Karl Pearson's coefficient (median form): $Sk_p = 3(\text{Mean} - \text{Median})/\sigma$

$$0.42 = 3(36 - 34.6)/\sigma \Rightarrow \sigma = 3 \times 1.4 / 0.42 = 4.2/0.42 = \mathbf{10}$$

$$\text{C.V.} = (\sigma / \text{Mean}) \times 100 = (10/36) \times 100 = \mathbf{27.78\%}$$

(Formula 1 M, σ 2 M, C.V. 1 M)

Q.3

a) Bowley's coefficient of skewness.

4 M

Class	Frequency f	c.f. (less than)
10-20	6	6
20-30	14	20
30-40	25	45
40-50	30	75
50-60	15	90
60-70	10	100
Total	N = 100	

$$Q_1: 1N/4 = 1 \times 100/4 = 25$$

Q1 class: the c.f. just greater than or equal to 25 is 45 $\Rightarrow$ Q1 class = **30–40**

$L = 30$, $cf = 20$ (c.f. of the class before), $f = 25$, $h = 10$

$$Q_1 = L + ((N/4 - cf) / f) \times h = 30 + ((25 - 20) / 25) \times 10 = \mathbf{32}$$

$$\text{Median: } N/2 = 100/2 = 50$$

Median class: the c.f. just greater than or equal to 50 is 75 $\Rightarrow$ Median class = **40–50**

$L = 40$, $cf = 45$ (c.f. of the class before), $f = 30$, $h = 10$

$$\text{Median} = L + ((N/2 - cf) / f) \times h = 40 + ((50 - 45) / 30) \times 10 = \mathbf{41.67}$$

$$Q_3: 3N/4 = 3 \times 100/4 = 75$$

Q3 class: the c.f. just greater than or equal to 75 is 75 $\Rightarrow$ Q3 class = **40–50**

$L = 40$, $cf = 45$ (c.f. of the class before), $f = 30$, $h = 10$

$$Q_3 = L + ((3N/4 - cf) / f) \times h = 40 + ((75 - 45) / 30) \times 10 = \mathbf{50}$$

$$Sk_B = (Q_3 + Q_1 - 2 \cdot \text{Median}) / (Q_3 - Q_1) = (50 + 32 - 2 \times 41.67) / (50 - 32) = \mathbf{-0.0741}$$

Skewness is negative, i.e. the distribution is negatively skewed (longer tail to the left; $\text{Mean} < \text{Median} < \text{Mode}$).

(Q_1 , Q_2 , Q_3 3 M , Sk_B 1 M)

b) Karl Pearson's coefficient of skewness (discrete series).

4 M

x	f	d = (x - 15)/5	fd	fd ²
5	4	-2	-8	16
10	9	-1	-9	9
15	16	0	0	0
20	12	1	12	12
25	6	2	12	24
30	3	3	9	27
Total	N = 50		$\Sigma fd = 16$	$\Sigma fd^2 = 88$

$$\text{Mean } \bar{x} = A + (\Sigma fd / N) \times h = 15 + (16/50) \times 5 = \mathbf{16.6}$$

$$\sigma = h \times \sqrt{(\Sigma fd^2 / N) - (\Sigma fd / N)^2} = 5 \times \sqrt{(88/50) - (16/50)^2} = 5 \times \sqrt{(1.7600 - 0.1024)} = \mathbf{6.4374}$$

Mode: the maximum frequency is 16 against $x = 15 \Rightarrow \text{Mode} = 15$ (by inspection)

$$Sk_p = (\text{Mean} - \text{Mode}) / \sigma = (16.6 - 15) / 6.4374 = \mathbf{0.2485}$$

Skewness is positive, i.e. the distribution is positively skewed (longer tail to the right; $\text{Mean} > \text{Median} > \text{Mode}$).

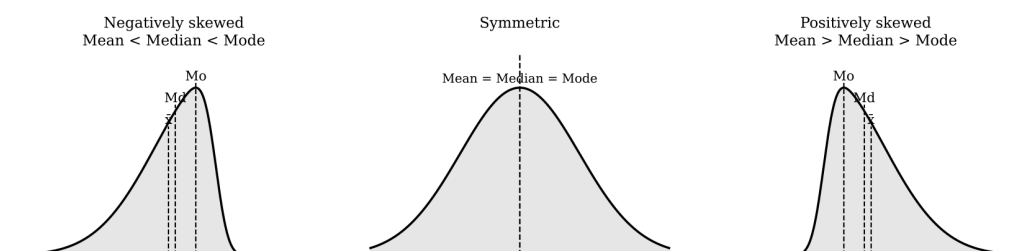


Fig.: Positions of mean, median and mode in skewed curves

(Mean 1 M, σ 1 M, mode $\frac{1}{2}$ M, Sk $1\frac{1}{2}$ M)

c) Spearman's rank correlation (tied ranks).

4 M

Rank 1 is given to the highest mark. Equal values get the average of the ranks they occupy: 65 occurs three times at positions 6, 7, 8 $\Rightarrow$ each gets rank $(6 + 7 + 8)/3 = 7$.

x	y	Rank R_1 (x)	Rank R_2 (y)	d = $R_1 - R_2$	d ²
70	60	5	7	-2	4
65	58	7	8	-1	1
80	72	3	3	0	0
65	70	7	4	3	9
90	85	1	1	0	0
75	66	4	5	-1	1
65	62	7	6	1	1
85	80	2	2	0	0
				$\Sigma d = 0$	$\Sigma d^2 = 16$

Ties: one value repeated 3 times ($m = 3$) $\Rightarrow$ correction factor C.F. = $\Sigma(m^3 - m)/12 = (27 - 3)/12 = 2$

$$\rho = 1 - 6[\Sigma d^2 + \text{C.F.}] / (n(n^2 - 1)) = 1 - 6(16 + 2) / (8 \times 63) = 1 - 108/504 = \mathbf{0.7857}$$

There is a high degree of positive correlation between the two examiners.

(Ranks with ties 1 M, d² 1 M, correction 1 M, ρ 1 M)

d) Straight-line trend by least squares.

4 M

Let $u = x - 2021$ (middle year), so that $\Sigma u = 0$. Trend line: $y = a + bu$.

u	y	u ²	uy
-3	12	9	-36
-2	15	4	-30
-1	14	1	-14
0	18	0	0
1	21	1	21
2	20	4	40
3	25	9	75
$\Sigma u = 0$	$\Sigma y = 125$	$\Sigma u^2 = 28$	$\Sigma uy = 56$

Normal equations for $y = a + bu$:

$$\Sigma y = na + b\Sigma u \quad \dots(1)$$

$$\Sigma uy = a\Sigma u + b\Sigma u^2 \quad \dots(2)$$

$$\text{Substituting: } 125 = 7a + 0b \quad \dots(1) \quad 56 = 0a + 28b \quad \dots(2)$$

$$\text{Since } \Sigma u = 0: \quad a = \Sigma y/n = 125/7 = \mathbf{17.8571}, \quad b = \Sigma uy/\Sigma u^2 = 56/28 = \mathbf{2}$$

Trend line: $y = 17.8571 + 2u, u = x - 2021$

For 2026, $u = 5$: $y = 17.8571 + 2 \times 5 = \mathbf{27.86 \text{ lakh}}$

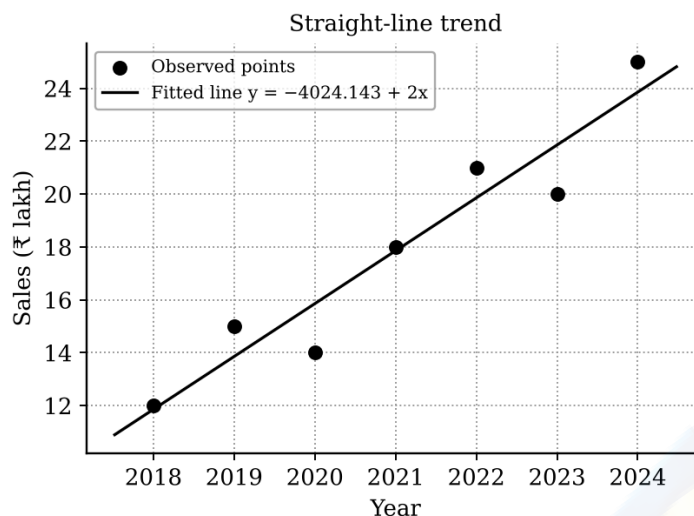


Fig.: Actual sales and the fitted trend line

(Table 1 M, normal equations 1 M, a and b 1 M, estimate 1 M)

Q.4

a) Regression line from given sums.

4 M

$$\bar{x} = \Sigma x/n = 40/8 = 5, \quad \bar{y} = \Sigma y/n = 56/8 = 7$$

$$b_{yx} = (n\Sigma xy - \Sigma x \Sigma y) / (n\Sigma x^2 - (\Sigma x)^2) = (8 \times 310 - 40 \times 56) / (8 \times 240 - 40^2) = (2480 - 2240) / (1920 - 1600) = 240/320 = \mathbf{0.75}$$

$$\text{Line of } y \text{ on } x: y - \bar{y} = b_{yx}(x - \bar{x}) \Rightarrow y - 7 = 0.75(x - 5) \Rightarrow \mathbf{y = 0.75x + 3.25}$$

$$\text{At } x = 7: y = 0.75 \times 7 + 3.25 = \mathbf{8.5}$$

$$r = (n\Sigma xy - \Sigma x \Sigma y) / \sqrt{[(n\Sigma x^2 - (\Sigma x)^2)(n\Sigma y^2 - (\Sigma y)^2)]}$$

$$= (8 \times 310 - 40 \times 56) / \sqrt{[(8 \times 240 - 40^2)(8 \times 436 - 56^2)]}$$

$$= 240 / \sqrt{(320 \times 352)} = 240 / 335.6188 = \mathbf{0.7151}$$

(Means $\frac{1}{2}$ M, b_{yx} 1 M, line 1 M, estimate $\frac{1}{2}$ M, r 1 M)

b) Karl Pearson's coefficient of correlation.

4 M

x	y	dx = x - $\bar{x}$	dy = y - $\bar{y}$	dx ²	dy ²	dx·dy
15	52	-7.5	9.33	56.25	87.11	-70
18	48	-4.5	5.33	20.25	28.44	-24
21	45	-1.5	2.33	2.25	5.44	-3.5
24	40	1.5	-2.67	2.25	7.11	-4
27	38	4.5	-4.67	20.25	21.78	-21
30	33	7.5	-9.67	56.25	93.44	-72.5
$\Sigma x = 135$	$\Sigma y = 256$			157.5	243.33	-195

$$\bar{x} = \Sigma x/n = 22.5, \quad \bar{y} = \Sigma y/n = 42.667$$

$$r = \Sigma dx \cdot dy / \sqrt{(\Sigma dx^2 \cdot \Sigma dy^2)} = -195 / \sqrt{(157.5 \times 243.33)} = -195 / 195.7677 = -0.9961$$

Interpretation: $r = -0.9961 \Rightarrow$ high degree of negative correlation: demand falls as price rises (law of demand).

(Table 2 M, r 1 M, interpretation 1 M)

c) Addition theorem and conditional probability.

4 M

(i) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.4 - 0.2 = 0.7$

(ii) $P(A' \cap B') = P((A \cup B)') = 1 - P(A \cup B) = 1 - 0.7 = 0.3$ (De Morgan's law)

(iii) $P(A | B) = P(A \cap B)/P(B) = 0.2/0.4 = 0.5$

(iv) $P(A) \times P(B) = 0.5 \times 0.4 = 0.2 = P(A \cap B) \Rightarrow$ **A and B are independent.** (Also $P(A | B) = P(A) = 0.5$.)

(Each part 1 M)

d) Probability using combinations.

4 M

Total balls = 6 + 4 + 5 = 15. Three balls can be drawn in ${}^{15}C_3 = (15 \times 14 \times 13)/(3 \times 2 \times 1) = 455$ ways.

(i) One of each colour: ${}^6C_1 \times {}^4C_1 \times {}^5C_1 = 6 \times 4 \times 5 = 120$ ways $\Rightarrow P = 120/455 = 24/91 = 0.2637$

(ii) All three red: ${}^6C_3 = 20$ ways $\Rightarrow P = 20/455 = 4/91 = 0.0440$

(n(S) 1 M, each part 1½ M)

e) Bayes' theorem — two bags.

4 M

Let E_1 = Bag I is chosen, E_2 = Bag II is chosen, W = ball drawn is white. $P(E_1) = P(E_2) = 1/2$ (coin).

$$P(W | E_1) = 3/8, \quad P(W | E_2) = 6/8$$

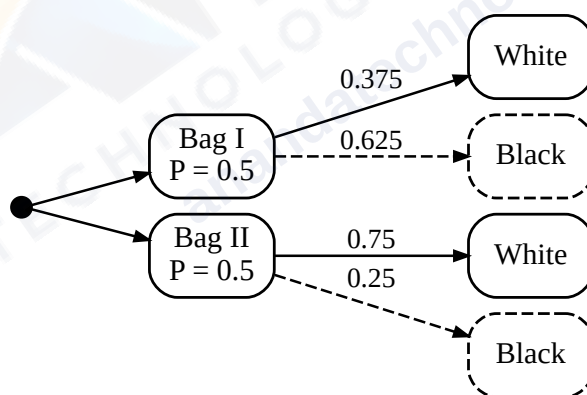


Fig.: Probability tree

$$P(W) = (1/2)(3/8) + (1/2)(6/8) = 3/16 + 6/16 = 9/16$$

$$P(E_2 | W) = P(E_2)P(W | E_2) / P(W) = (6/16)/(9/16) = 2/3 = 0.6667$$

(Notation 1 M, P(W) 1 M, Bayes 2 M)

Q.5

a) Lagrange's formula; Newton's forward formula for a polynomial.

6 M

(i)

Lagrange's interpolation formula:

$$y = \frac{(x - x_1)(x - x_2)(x - x_3)}{[(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)]} \cdot y_0 + \frac{(x - x_0)(x - x_2)(x - x_3)}{[(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)]} \cdot y_1 + \frac{(x - x_0)(x - x_1)(x - x_3)}{[(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)]} \cdot y_2 + \frac{(x - x_0)(x - x_1)(x - x_2)}{[(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)]} \cdot y_3$$

Here $x_0, x_1, \dots = 0, 2, 3, 6$ and $y_0, y_1, \dots = -4, 2, 14, 158$. Putting $x = 4$:

$$\begin{aligned} y(4) &= \frac{(4 - 2)(4 - 3)(4 - 6)}{[(0 - 2)(0 - 3)(0 - 6)]} \times -4 \\ &+ \frac{(4 - 0)(4 - 3)(4 - 6)}{[(2 - 0)(2 - 3)(2 - 6)]} \times 2 \\ &+ \frac{(4 - 0)(4 - 2)(4 - 6)}{[(3 - 0)(3 - 2)(3 - 6)]} \times 14 \\ &+ \frac{(4 - 0)(4 - 2)(4 - 3)}{[(6 - 0)(6 - 2)(6 - 3)]} \times 158 \\ &= (-4/-36) \times -4 + (-8/8) \times 2 + (-16/-9) \times 14 + (8/72) \times 158 \\ &= -0.4444 - 2 + 24.8889 + 17.5556 \end{aligned}$$

$$\Rightarrow f(4) = 40$$

(ii) Forward difference table:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0	1				
		3			
1	4		4		
		7		0	
2	11		4		0
		11		0	
3	22		4		
		15			
4	37				

Second differences are constant (4), so y is a polynomial of degree 2. Here $x_0 = 0, h = 1 \Rightarrow p = (x - 0)/1 = x$.

$$y = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!} \Delta^2 y_0 = 1 + x(3) + \frac{x(x-1)(4)}{2} = 1 + 3x + 2x^2 - 2x$$

$$\Rightarrow \boxed{y = f(x) = 2x^2 + x + 1} \quad (\text{check: } f(4) = 32 + 4 + 1 = 37 \checkmark)$$

((i) 3 M, (ii) table 1 M, polynomial 2 M)

b) Newton's backward formula; operator Δ .

6 M

(i) Backward difference table (bottom diagonal in bold):

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
20	4.4721				
		0.5279			
25	5		-0.0507		
		0.4772		0.0124	
30	5.4772		-0.0383		-0.0045
		0.4389		0.0079	
35	5.9161		-0.0304		
		0.4085			
40	6.3246				

Here $x_n = 40$, $h = 5$, $x = 38 \Rightarrow p = (x - x_n)/h = (38 - 40)/5 = -0.4$

Newton's backward interpolation formula:

$$y = y_n + p\nabla y_n + p(p+1)/2! \nabla^2 y_n + p(p+1)(p+2)/3! \nabla^3 y_n + p(p+1)(p+2)(p+3)/4! \nabla^4 y_n$$

$$y(38) = 6.3246 + (-0.4)(0.4085) + (-0.4)(0.6)(-0.0304)/2 + (-0.4)(0.6)(1.6)(0.0079)/6 + (-0.4)(0.6)(1.6)(2.6)(-0.0045)/24$$

$$= 6.3246 - 0.1634 + 0.00365 - 0.00051 + 0.00019$$

$$\Rightarrow \sqrt{38} \approx 6.1645 \quad (\text{actual } \sqrt{38} = 6.1644)$$

(ii) $\Delta f(x) = f(x+h) - f(x)$, $h = 1$.

$$\Delta(x^3) = (x+1)^3 - x^3 = x^3 + 3x^2 + 3x + 1 - x^3 = 3x^2 + 3x + 1$$

$$\Delta^2(x^3) = \Delta(3x^2 + 3x + 1) = [3(x+1)^2 + 3(x+1) + 1] - [3x^2 + 3x + 1] = 3(2x+1) + 3 = 6x + 6$$

((i) table 1 M, formula and value 2 M; (ii) 3 M)

c) Extrapolation — cubes.

6 M

Difference table:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
3	27			
		37		
4	64		24	
		61		6
5	125		30	
		91		
6	216			

Third difference is 6 (constant for $y = x^3$), so the formula can be used outside the table (extrapolation).

(a) $(8)^3$:

$$\text{Here } x_0 = 3, h = 1, x = 8 \Rightarrow p = (x - x_0)/h = (8 - 3)/1 = 5$$

Newton's forward interpolation formula:

$$y = y_0 + p\Delta y_0 + p(p-1)/2! \Delta^2 y_0 + p(p-1)(p-2)/3! \Delta^3 y_0$$

$$y(8) = 27 + (5)(37) + (5)(4)(24)/2 + (5)(4)(3)(6)/6$$

$$= 27 + 185 + 240 + 60$$

$$\Rightarrow (8)^3 = 512$$

(b) $(2)^3$:

Here $x_0 = 3$, $h = 1$, $x = 2 \Rightarrow p = (x - x_0)/h = (2 - 3)/1 = -1$

Newton's forward interpolation formula:

$$y = y_0 + p\Delta y_0 + p(p-1)/2! \Delta^2 y_0 + p(p-1)(p-2)/3! \Delta^3 y_0$$

$$y(2) = 27 + (-1)(37) + (-1)(-2)(24)/2 + (-1)(-2)(-3)(6)/6$$

$$= 27 - 37 + 24 - 6$$

$$\Rightarrow (2)^3 = 8$$

(Table 2 M, each value 2 M)

Q.6

a) Chi-square — fitting a binomial distribution.

6 M

H₀: the coins are unbiased, i.e. the data follow a binomial distribution with $n = 4$, $p = 1/2$.

H₁: the coins are not unbiased.

Expected frequency $E = N \cdot {}^4C_x (1/2)^4 = 160 \times {}^4C_x/16 = 10 \times {}^4C_x \Rightarrow$ for $x = 0, 1, 2, 3, 4$: $E = 10, 40, 60, 40, 10$

Heads	O	E	O - E	(O - E) ²	(O - E) ² /E
0	8	10	-2	4	0.4000
1	34	40	-6	36	0.9000
2	69	60	9	81	1.3500
3	43	40	3	9	0.2250
4	6	10	-4	16	1.6000
Total	160	160			$\chi^2 = 4.4750$

d.f. = $n - 1 = 5 - 1 = 4$ (p is given, not estimated); table value = 9.49.

Since $\chi^2 = 4.475 < 9.49$, **H₀ is accepted**: the coins may be regarded as unbiased.

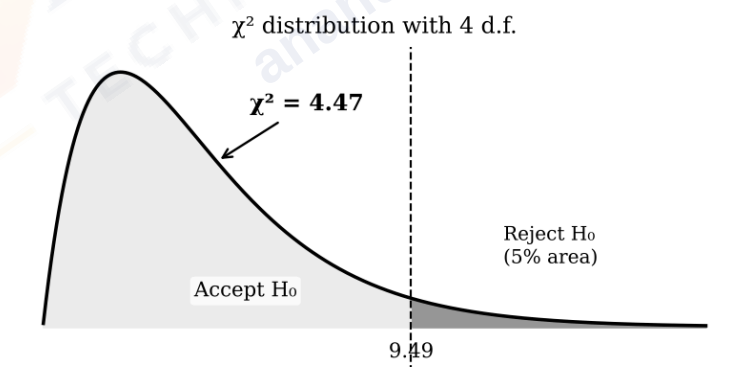


Fig.: Calculated χ^2 compared with the table value

(Hypotheses 1 M, expected 2 M, χ^2 2 M, conclusion 1 M)

b) Large-sample tests: two proportions; one mean.

6 M

(i)

$H_0: P_1 = P_2$ (no difference between the two cities).

$H_1: P_1 \neq P_2$.

$p_1 = 480/800 = 0.60$, $p_2 = 420/600 = 0.70$; pooled $\hat{p} = (480 + 420)/(800 + 600) = 900/1400 = 0.6429$, $\hat{q} = 0.3571$

S.E. = $\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)} = \sqrt{(0.6429 \times 0.3571 \times (1/800 + 1/600))} = 0.02588$

$z = (p_1 - p_2)/S.E. = (0.60 - 0.70)/0.02588 = -3.864$

Since $|z| = 3.864 > 1.96$ (table value), the difference is **significant**. H_0 is **rejected**. The proportions differ significantly.

(ii)

$H_0: \mu = 1200$ hours (claim is true).

$H_1: \mu \neq 1200$ hours.

$n = 64$ (large), $\bar{x} = 1180$, $s = 96$. S.E. = $s/\sqrt{n} = 96/8 = 12$

$z = (\bar{x} - \mu)/(s/\sqrt{n}) = (1180 - 1200)/12 = -1.667$

Since $|z| = 1.667 < 1.96$ (table value), the difference is **not significant**. H_0 is **accepted**. The claim of mean life 1200 hours **can be accepted**.

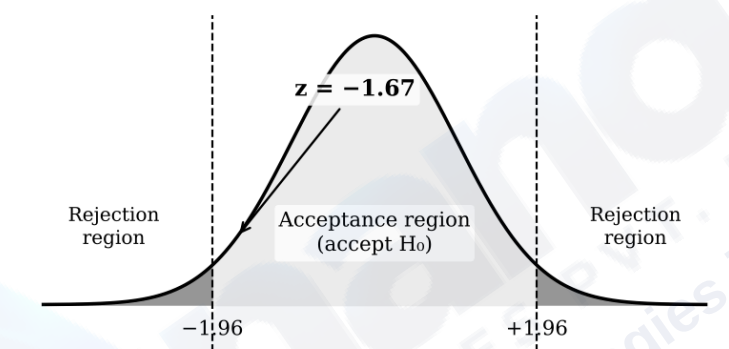


Fig.: $z = -1.67$ lies in the acceptance region

((i) 3 M, (ii) 3 M)

c) Paired t-test.

6 M

$H_0: \mu_d = 0$ (the diet has no effect on weight).

$H_1: \mu_d \neq 0$ (the diet changes the weight).

Let $d = \text{Before} - \text{After}$ for each person:

d	d - $\bar{d}$	(d - $\bar{d}$) ²
2	-0.5	0.25
3	0.5	0.25
2	-0.5	0.25
1	-1.5	2.25
5	2.5	6.25
0	-2.5	6.25
4	1.5	2.25
3	0.5	0.25
$\Sigma d = 20$		$\Sigma(d - \bar{d})^2 = 18$

$\bar{d} = \Sigma d/n = 20/8 = 2.5$, $s = \sqrt{(\Sigma(d - \bar{d})^2/(n - 1))} = \sqrt{(18/7)} = 1.6036$

$t = \bar{d}/(s/\sqrt{n}) = 2.5/(1.6036/\sqrt{8}) = 2.5/0.5669 = \mathbf{4.410}$

d.f. = $n - 1 = 7$; $t_{0.05} = 2.365$. Since $t = 4.410 > 2.365$, **H_0 is rejected**: the diet programme is effective in reducing weight.

(Hypotheses 1 M, differences 1 M, s 2 M, t and conclusion 2 M)

Check your answers with Python (optional)

The numerical answers above were computed by program and cross-checked with the standard **numpy** / **scipy** / **sympy** functions. Students can run the same check (this is not part of the exam answer):

```
# Set B: verify key numerical answers with numpy / scipy / sympy
import numpy as np
from scipy import stats
import sympy as sp

print("Q1 b) D6 =", round(float(np.percentile([34, 21, 45, 29, 38, 52, 26, 41,
33, 48, 30, 37], 60, method='weibull')), 4))
print("Q1 c) cov =", round(float(np.cov([2, 4, 6, 8, 10], [5, 9, 10, 14, 17],
bias=True)[0, 1]), 4))

A = [42, 38, 55, 61, 35, 49]; B = [45, 50, 47, 52, 43, 49]
print("Q2 c) CV(A) = %.2f %% , CV(B) = %.2f %%" % (stats.variation(A) * 100,
stats.variation(B) * 100))

x = np.array([5, 10, 15, 20, 25, 30]); f = np.array([4, 9, 16, 12, 6, 3])
m = np.average(x, weights=f); s = np.sqrt(np.average((x - m) ** 2, weights=f))
print("Q3 b) mean = %.4f, sd = %.4f, Sk = %.4f" % (m, s, (m - 15) / s))

x = [70, 65, 80, 65, 90, 75, 65, 85]; y = [60, 58, 72, 70, 85, 66, 62, 80]
R1 = stats.rankdata([-v for v in x]); R2 = stats.rankdata([-v for v in y]) #
rank 1 = highest, ties averaged
d2 = sum((R1 - R2) ** 2); n = len(x); CF = (3 ** 3 - 3) / 12 #
65 repeated 3 times
print("Q3 c) sum d^2 =", d2, "; textbook rho with C.F. = %.4f" % (1 - 6 * (d2 +
CF) / (n * (n * n - 1))))
# (scipy's spearmanr = Pearson r of the averaged ranks; with ties it differs
```

```

slightly from the textbook formula)
print("      scipy spearmanr =", round(stats.spearmanr(x, y).statistic, 4))
b, a = np.polyfit(np.arange(-3, 4), [12, 15, 14, 18, 21, 20, 25], 1)
print("Q3 d) trend: y = %.4f + %.4f u ; 2026 -> %.2f" % (a, b, a + b * 5))

print("Q4 b) r =", round(stats.pearsonr([15, 18, 21, 24, 27, 30], [52, 48, 45,
40, 38, 33]).statistic, 4))

X = sp.symbols('x')
print("Q5 a) f(4) =", sp.interpolate([(0, -4), (2, 2), (3, 14), (6, 158)],
X).subs(X, 4),
      "; curve:", sp.expand(sp.interpolate([(0, 1), (1, 4), (2, 11), (3, 22),
(4, 37)], X)))
xs = [20, 25, 30, 35, 40]; ys = [4.4721, 5.0000, 5.4772, 5.9161, 6.3246]
print("Q5 b) sqrt(38) by interpolating polynomial = %.4f" %
np.polyval(np.polyfit(xs, ys, 4), 38))
print("Q5 c) 8^3, 2^3 =", [round(float(np.polyval(np.polyfit([3, 4, 5, 6], [27,
64, 125, 216], 3), v)), 4) for v in (8, 2)])

exp = 160 * stats.binom.pmf(range(5), 4, 0.5)
r = stats.chisquare([8, 34, 69, 43, 6], exp)
print("Q6 a) expected =", exp.round(2).tolist(), "; chi2 = %.4f, p-value = %.4f"
% (r.statistic, r.pvalue))
r = stats.ttest_rel([72, 68, 80, 75, 90, 66, 84, 78], [70, 65, 78, 74, 85, 66,
80, 75])
print("Q6 c) paired t = %.4f, p-value = %.4f" % (r.statistic, r.pvalue))

```

Output:

```

Q1 b) D6 = 37.8
Q1 c) cov = 11.6
Q2 c) CV(A) = 19.82 %, CV(B) = 6.37 %
Q3 b) mean = 16.6000, sd = 6.4374, Sk = 0.2485
Q3 c) sum d^2 = 16.0 ; textbook rho with C.F. = 0.7857
      scipy spearmanr = 0.8051
Q3 d) trend: y = 17.8571 + 2.0000 u ; 2026 -> 27.86
Q4 b) r = -0.9961
Q5 a) f(4) = 40 ; curve: 2*x**2 + x + 1
Q5 b) sqrt(38) by interpolating polynomial = 6.1645
Q5 c) 8^3, 2^3 = [512.0, 8.0]
Q6 a) expected = [10.0, 40.0, 60.0, 40.0, 10.0] ; chi2 = 4.4750, p-value =
0.3455
Q6 c) paired t = 4.4096, p-value = 0.0031

```

Note: This is a **model question paper prepared by Ananda Technologies** only for practice and exam preparation, based on the pattern of previous MSBTE question papers and the K-scheme syllabus. It is **not an official MSBTE paper**, and it does **not** mean that these questions will come in the exam. Answers follow standard textbook points; marks shown are indicative. Free for students — anandatechnologies.in