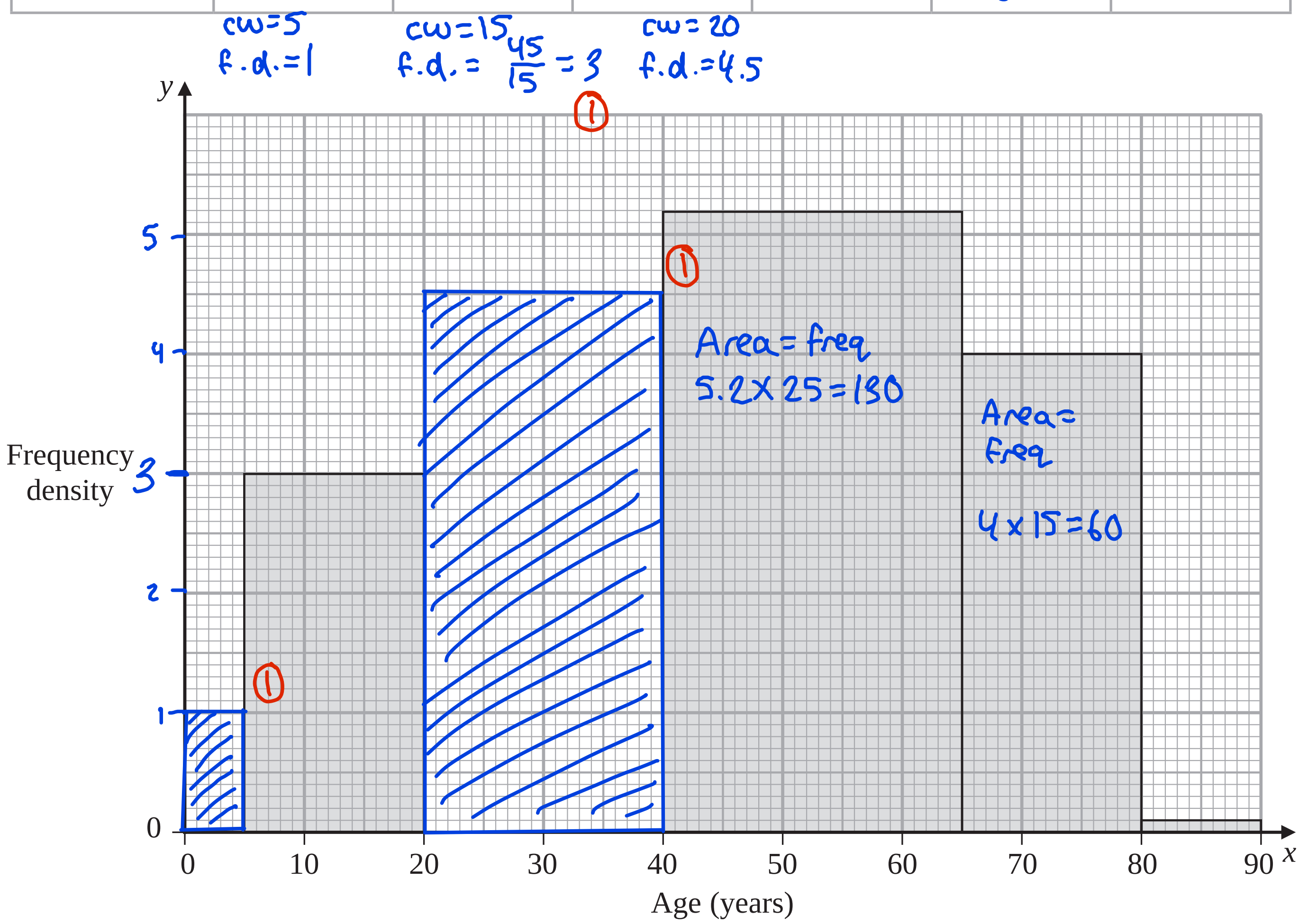
1. The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

Age (x years)	$0 \leqslant x < 5$	$5 \leqslant x < 20$	$20 \leqslant x < 40$	$40 \leqslant x < 65$	$65 \leqslant x < 80$	$80 \leqslant x < 90$
Frequency	5	45	90	130 (1)	60	1



(a) Complete the histogram.

(3)

(b) Use linear interpolation to estimate the median age.

(4)

An outlier is defined as a value greater than $Q_3 + 1.5 \times$ interquartile range.

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$

(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier.

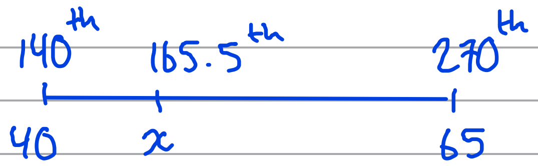
(2)

Question 1 continued

b) $\sum \text{frequency} = 5+45+90+130+60+1 = 331$ people 1

median $= 165.5^{\text{th}}$ person

$5+45+90 = 140 < 165.5$
$5+45+90+130 = 270 > 165.5$
so the median is in the $40 \le x < 65$ group



$$\frac{165.5-140}{270-140} = \frac{x-40}{65-40} \quad 1$$

$$\frac{51}{260} = \frac{x-40}{25}$$

$$x = \frac{51 \times 25}{260} + 40$$

$$x = 44.9038...$$
$$= 44.9 \text{ (3sf)} \quad 1$$

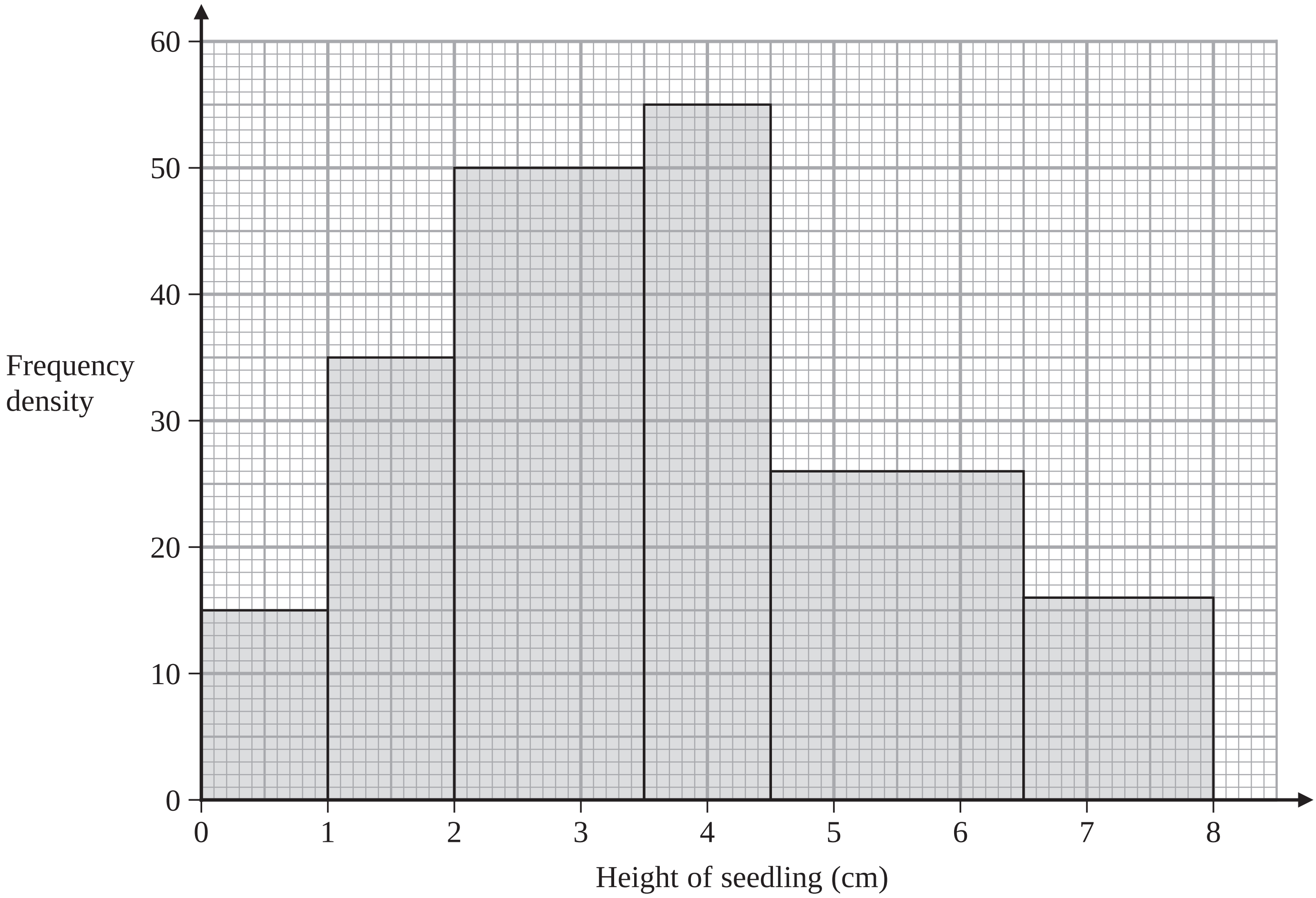
$\therefore$ median age $= 44.9$ years

c) $IQR = Q_3 - Q_1 = 58.9 - 27.3 = 31.6$

outlier $> 58.9 + 1.5(31.6)$ 1
outlier > 106.3

$90 < 106.3$ so oldest passenger is not an outlier. 1

2. The histogram summarises the heights of 256 seedlings two weeks after they were planted.



(a) Use linear interpolation to estimate the median height of the seedlings.

(4)

Chris decides to model the **frequency density** for these 256 seedlings by a curve with equation

$$y = kx(8 - x) \qquad 0 \leqslant x \leqslant 8$$

where k is a constant.

(b) Find the value of k

(3)

Using this model,

(c) write down the median height of the seedlings.

(1)

a)

class	frequency	cumulative frequency
0 - 1	15	15 ①
1 - 2	35	50
2 - 3.5	75	125
3.5 - 4.5	55	180
4.5 - 6.5	52	232
6.5 - 8	24	256 ①

128th term is within this class → 3.5 - 4.5

$\frac{256}{2} = 128^{th}$ → median is the 128th term

$$\text{Median } (Q_2) = 3.5 + \frac{\frac{256}{2} - 125}{55} \times (4.5 - 3.5) \quad ①$$

$$= 3.55 \ (3 \text{ s.f.}) \quad ①$$

b) $\int_0^8 kx(8-x)\,dx = 256$ ① ← which is the total frequency (area under the curve)

$$k\int_0^8 x(8-x)\,dx = 256$$

$$k\int_0^8 8x - x^2\,dx = 256$$

$$k\left[4x^2 - \frac{1}{3}x^3\right]_0^8 = 256 \quad ①$$

$$k\left[4(8)^2 - \frac{1}{3}(8)^3\right] = 256 \quad ①$$

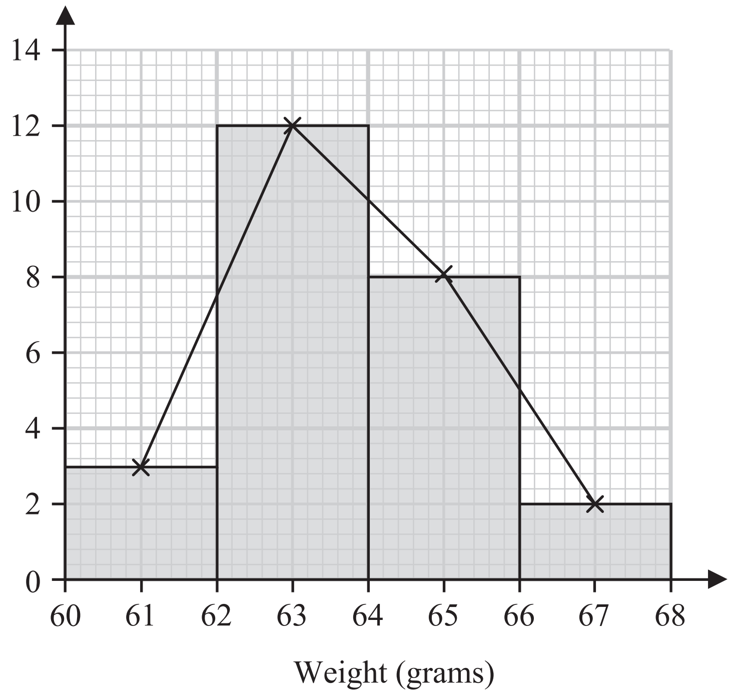
$$k\left(\frac{256}{3}\right) = 256$$

$$k = 3 \quad ①$$

c) By symmetry, median = 4 ①

since the quadratic $y = 3x(8-x)$ has roots 0 and 8, the median value must be between them since parabolas are symmetrical

3. The histogram and its frequency polygon below give information about the weights, in grams, of 50 plums.



frequency = class width × frequency density

(a) Show that an estimate for the mean weight of the 50 plums is 63.72 grams.
(2)

(b) Calculate an estimate for the standard deviation of the 50 plums.
(2)

Later it was discovered that the scales used to weigh the plums were broken.

Each plum actually weighs 5 grams less than originally thought.

(c) State the effect this will have on the estimate of the standard deviation in part (b). Give a reason for your answer.
(1)

a)

class	frequency (f)	midpoint (x)	fx
60 - 62	2×3 = 6	61	6×61 = 366
62 - 64	2×12 = 24	63	24×63 = 1512
64 - 66	2×8 = 16	65	16×65 = 1040 ①
66 - 68	2×2 = 4	67	4×67 = 268 +
			3186

$\frac{\Sigma fx}{n} = \frac{3186}{50} = 63.72$ ①

b) standard deviation $= \sqrt{\frac{\Sigma fx^2}{n} - \left(\frac{\Sigma fx}{n}\right)^2}$

1

$$\sigma = \sqrt{\frac{6\times61^2+24\times63^2+16\times65^2+4\times67^2}{50}-(63.72)^2} = 1.58795\ldots$$

$\sigma = 1.59$ (3sf) 1

c) the standard deviation will be unchanged, since addition and subtraction dont affect standard deviation.

1

σ_x is a measure of spread of data, so translation of all data points doesnt change σ_x.

4. A medical researcher is studying the number of hours, T, a patient stays in hospital following a particular operation.

The histogram on the page opposite summarises the results for a random sample of 90 patients.

(a) Use the histogram to estimate $P(10 < T < 30)$

(2)

For these 90 patients the time spent in hospital following the operation had

- a mean of 14.9 hours
- a standard deviation of 9.3 hours

Tomas suggests that T can be modelled by $N(14.9, 9.3^2)$

(b) With reference to the histogram, state, giving a reason, whether or not Tomas' model could be suitable.

(1)

Xiang suggests that the frequency polygon based on this histogram could be modelled by a curve with equation

$$y = kxe^{-x} \quad 0 \leqslant x \leqslant 4$$

where

- x is measured in **tens of hours**
- k is a constant

(c) Use algebraic integration to show that

$$\int_0^n xe^{-x}\,dx = 1 - (n+1)e^{-n}$$

(4)

(d) Show that, for Xiang's model, $k = 99$ to the nearest integer.

(3)

(e) Estimate $P(10 < T < 30)$ using

(i) Tomas' model of $T \sim N(14.9, 9.3^2)$

(1)

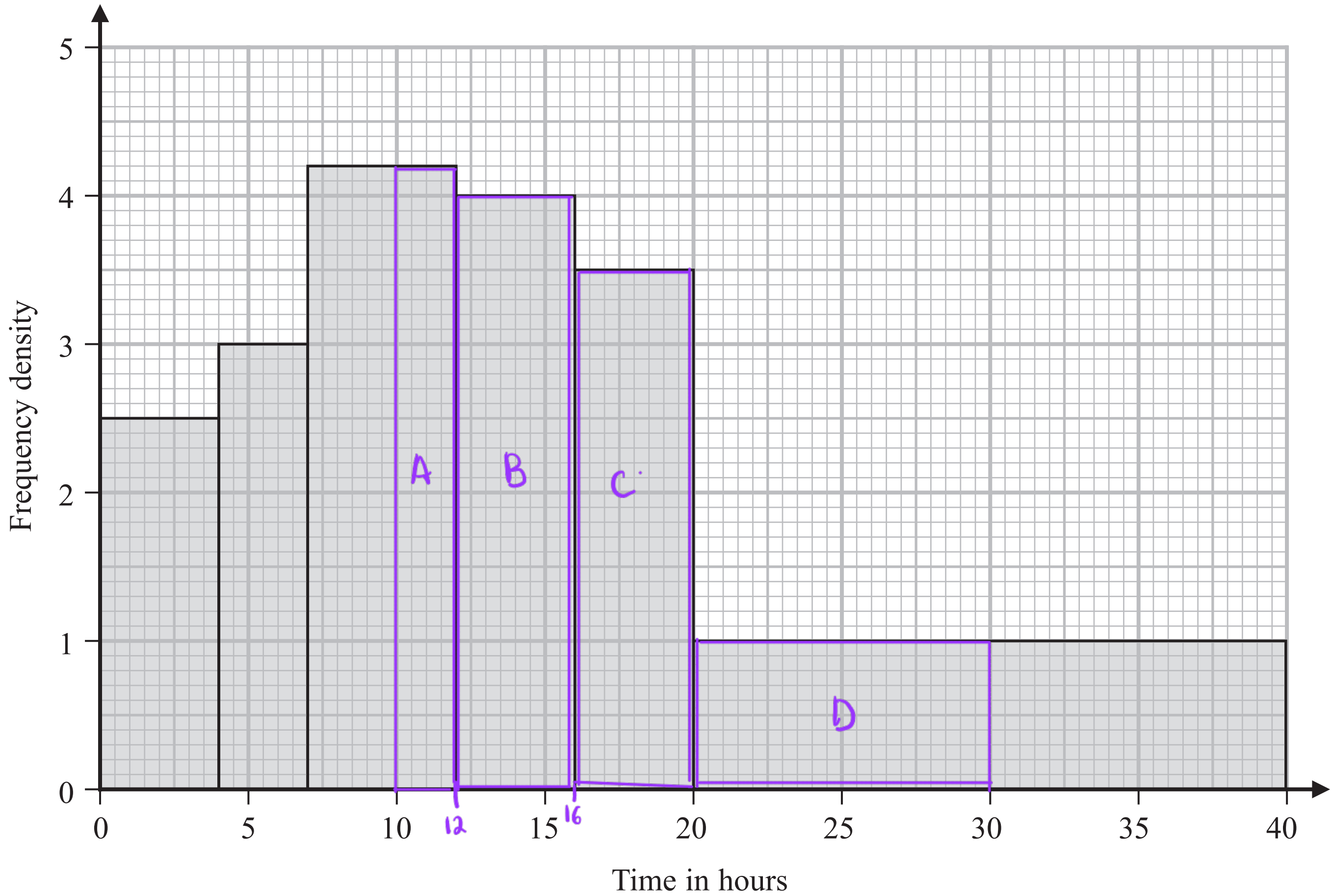
(ii) Xiang's curve with equation $y = 99xe^{-x}$ and the answer to part (c)

(2)

The researcher decides to use Xiang's curve to model $P(a < T < b)$

(f) State one limitation of Xiang's model.

(1)



a) $P(10 < T < 30) = P(A) + P(B) + P(C) + P(D)$

$= \dfrac{(2 \times 4.2) + (4 \times 4) + (4 \times 3.5) + (16 \times 1)}{90}$ (1)

$= \dfrac{8.4 + 16 + 14 + 10}{90}$

$= \dfrac{48.4}{90}$

$= 0.5377... = 0.54$ (2 s.f.) (1)

(b) It does not look suitable because a normal distribution is symmetrical and the histogram is not. (1)

c) $u = x \qquad v = -e^{-x}$

$u' = 1 \qquad v' = e^{-x}$ (1)

$u \times v - \int vu'$

$\therefore \int xe^{-x}\,dx = -xe^{-x} - \int (-e^{-x})\,dx$ (1)

$= -xe^{-x} + \int e^{-x}\,dx$

$= -xe^{-x} - e^{-x}$

$\therefore \int_0^n xe^{-x}\,dx = \left[-xe^{-x} - e^{-x}\right]_0^n$

$= (-ne^{-n} - e^{-n}) - (-e^{0})$ (1)

$= -ne^{-n} - e^{-n} + 1$

$= 1 - (n+1)e^{-n}$ (1) (shown)

d) Area under frequency polygon = 90 when $n = 4$

$\therefore k \int xe^{-x}\,dx = 90$ (1)

$\therefore k\{1 - (n+1)e^{-n}\} = 90$

when $n = 4$: $k(1 - 5e^{-4}) = 90$ (1)

$k = \dfrac{90}{1 - 5e^{-4}}$

$k = 99.07\ldots$ (1)

$\therefore k \approx 99$

e) (i) $T \sim N(14.9, 9.3^2)$

$P(10 < T < 30) = 0.6486.... \approx 0.649$ (3s.f.) 1

(ii) No° of patients

$\approx \int_1^3 99xe^{-x}\,dx$

$= 99\{(1-4e^{-3}) - (1-2e^{-1})\}$

$\approx 53.1....$ 1

$\therefore$ Probability $= \dfrac{53.1....}{90}$

$= 0.590...$ 1

(f) Some patients might stay longer than 40 hours. 1