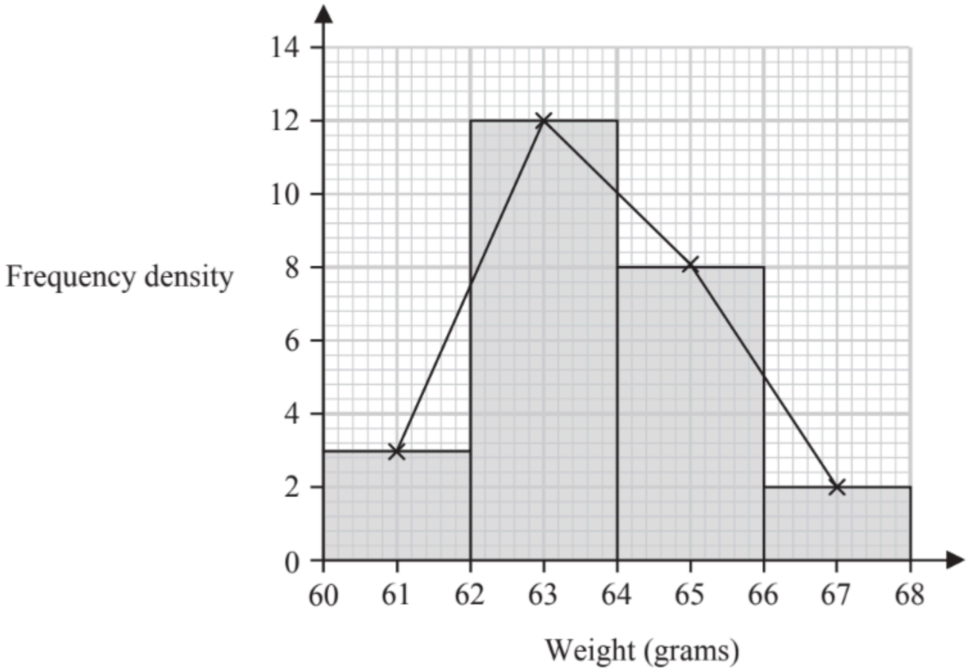
Statistics Year 1 (AS) exam questions

Arancha Ruiz

Ch. 3: Representations of Data

June 2023 Question 1

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



(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.

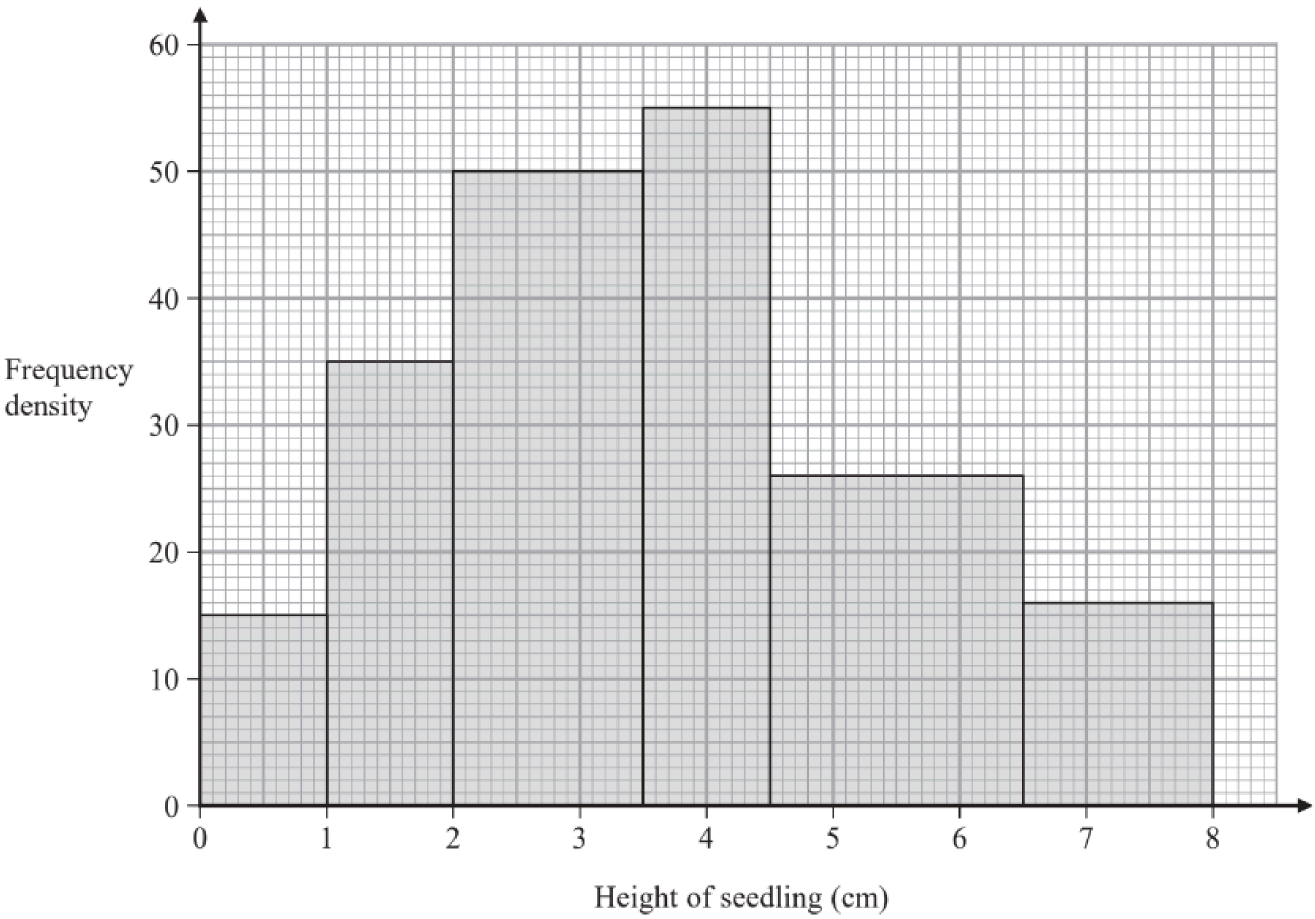
(1)

ANSWER

Question	Scheme	Marks	AOs
1(a)	$61\times(2\times3),\quad 63\times(2\times12),\quad 65\times(2\times8),\quad 67\times(2\times2)$	M1	2.1
	$\dfrac{61\times(2\times3)+63\times(2\times12)+65\times(2\times8)+67\times(2\times2)}{50}=63.72^*$	A1*cso	1.1b
		(2)	
(b)	$\sqrt{\dfrac{61^2\times6+63^2\times24+65^2\times16+67^2\times4}{50}-63.72^2}$	M1	1.1b
	$=\sqrt{2.5216}=1.58795...$ $\quad$ = awrt $\underline{\mathbf{1.59}}$	A1	1.1b
		(2)	
(c)	No effect (oe) since…e.g. • since addition/subtraction does not affect the standard deviation (only multiplication and division do) • the weights will have the same spread • the distance of each weight from the mean will not have changed • they all change by the same amount	B1	2.4
		(1)	
			(5 marks)

June 2022 Question 3

3. 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)

ANSWER

Qu	Scheme	Mark	AO
3. (a)			
	| Class | Frequency | Cum. Frequency |		
	| $0-1$ | 15 | 15 |	M1	2.1
	| $1-2$ | 35 | 50 |		
	| $2-3.5$ | 75 | 125 |	A1	1.1b
	| $3.5-4.5$ | 55 | 180 |		
	$[Q_2=](3.5)+\frac{\frac{256}{2}-\text{"125"}}{\text{"55"}}\times(4.5-3.5)$ <u>or</u> $(4.5)-\frac{\text{"180"}-\frac{256}{2}}{\text{"55"}}\times 1$	M1	2.1
	$=3.5545\ldots\ldots$ awrt <u>**3.55**</u>	A1	1.1b
		(4)	
(b)	Need area under curve to be 256 so $\int_{(0)}^{(8)} kx(8-x)\,\mathrm{d}x=256$	M1	3.1a
	$k\left[4x^2-\frac{x^3}{3}\right]_{(0)}^{(8)}=256$	M1	1.1b
	$\left\{k\left[4\times8^2-\frac{8}{3}\times8^2\right]=256\Rightarrow\right\}$ <u>$k=3$</u>	A1	1.1b
		(3)	
(c)	[By symmetry median =] <u>**4**</u>	B1	2.2a
		(1)	
		(8 marks)	

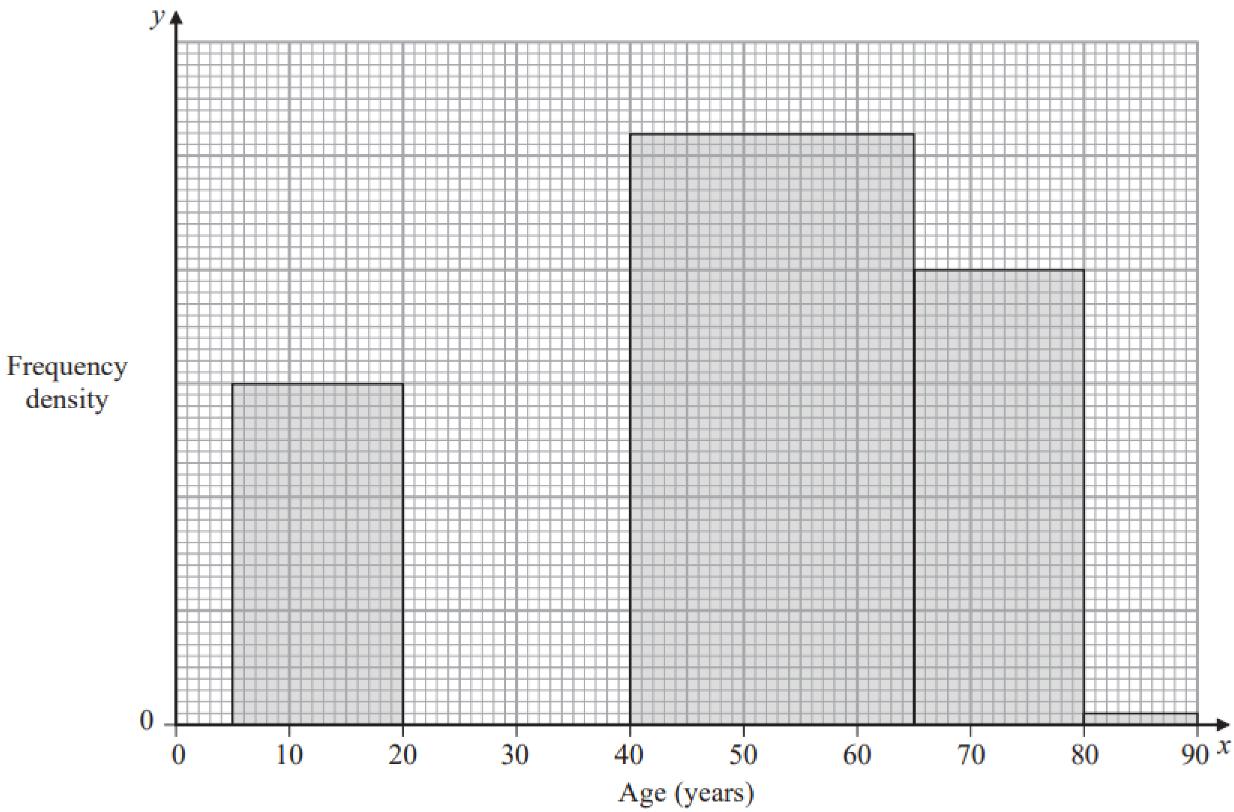
Video solution:
https://youtu.be/zf9owtfBqL0

November 2021 question 2

2. 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			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)

ANSWER

Qu	Scheme	Marks	AO
2. (a)	From [5,20) fd = 3 <u>or</u> 1 large square = 2.5 passengers o.e.	M1	2.2a
	Correct bar above [0, 5)	A1	1.1b
	Correct bar above [20, 40)	A1	1.1b
		(3)	
(b)	For [40, 65) **<u>130</u>** passengers <u>or</u> for [65, 80) **<u>60</u>** passengers	M1	2.1
	For attempt to find total number of passengers = **<u>331</u>**	A1ft	1.1b
	[Median =] $40+\frac{\frac{1}{2}(\text{"331"})-140}{\text{"130"}}\times 25$ <u>or</u> $65-\frac{270-\frac{1}{2}(\text{"331"})}{\text{"130"}}\times 25$ (o.e.)	M1	1.1b
	$= 44.9038\ldots$ = awrt **<u>44.9</u>**	A1	1.1b
		(4)	
(c)	Upper outlier limit $= 58.9 + 1.5\times(58.9 - 27.3) = 106\ (.3) > 90$	M1	2.4
	So oldest passenger is <u>not</u> an outlier	A1	2.2a
		(2)	
		(9 marks)	

October 2020 question 1

1.

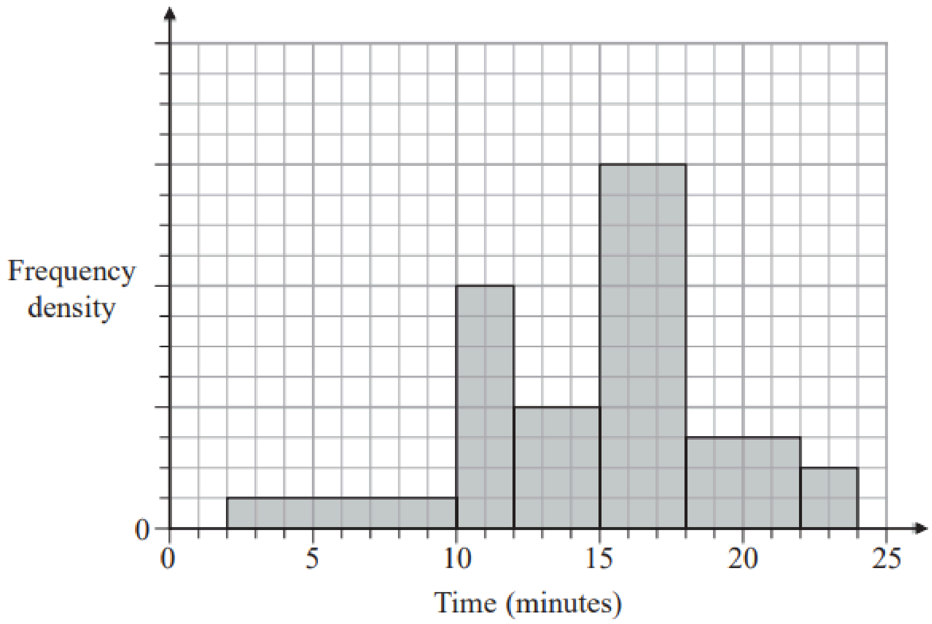


Figure 1

The histogram in Figure 1 shows the times taken to complete a crossword by a random sample of students.

The number of students who completed the crossword in more than 15 minutes is 78

Estimate the percentage of students who took less than 11 minutes to complete the crossword.

(4)

ANSWER

Question	Scheme	Marks	AOs
1	1 square is $\frac{78}{12\times3+3\times4+2\times2} = \left[\frac{78}{52}=1.5\right]$ **and** $(8\times1+1\times8)\times"1.5"$	M1	3.1a
	24 students took less than 11 minutes	A1	1.1b
	Percentage of students = $\frac{"24"}{78+"24"+1\times8\times"1.5"+3\times4\times"1.5"}\times100$	M1	3.1b
	= 18.18… awrt 18%	A1	1.1b
		(4)	
			Total 4

October 2020 question 2

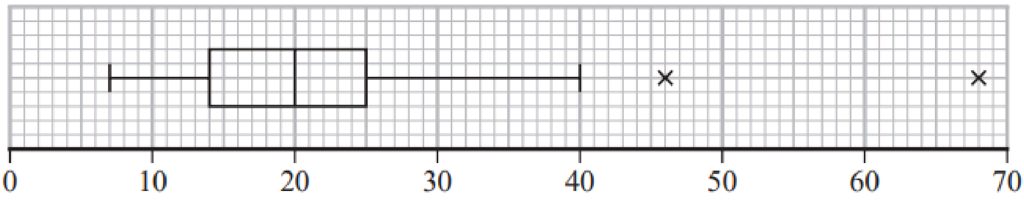
You will find this question in chapter 4 because it has correlation.

October 2020 question 3 (A-level, not AS)

3. Each member of a group of 27 people was timed when completing a puzzle.

The time taken, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



(a) Find the range of the times.

(1)

(b) Find the interquartile range of the times.

(1)

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17\,623.25$

(c) calculate the mean time taken to complete the puzzle,

(1)

(d) calculate the standard deviation of the times taken to complete the puzzle.

(2)

Taruni defines an outlier as a value more than 3 standard deviations above the mean.

(e) State how many outliers Taruni would say there are in these data, giving a reason for your answer.

(1)

Adam and Beth also completed the puzzle in a minutes and b minutes respectively, where $a > b$.
When their times are included with the data of the other 27 people

- the median time increases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b, explaining how your values satisfy the above conditions.

(3)

(g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d).

(1)

ANSWER

Qu 3	Scheme	Marks	AO
(a)	$[68-7=]$ $\underline{\mathbf{61}}$ (only)	B1	1.1b
		(1)	
(b)	$[25-14]=\underline{\mathbf{11}}$	B1	1.1b
		(1)	
(c)	$\left[\mu \text{ or } \bar{x}=\frac{607.5}{27}=\right]=\underline{\mathbf{22.5}}$	B1	1.1b
		(1)	
(d)	$\sigma=\sqrt{\frac{17\,623.25}{27}-\text{"}22.5\text{"}^2}$ $\underline{\text{or}}$ $\sqrt{146.4629...}$	M1	1.1b
	$=12.10218\ldots$ awrt $\underline{\mathbf{12.1}}$	A1	1.1b
		(2)	
(e)	$\mu+3\sigma=\text{"}22.5\text{"}+3\times\text{"}12.1...\text{"}=$ awrt 59 so only **<u>one</u>** outlier	B1ft	1.1b
		(1)	
(f)	Median increases implies that both values must be > 20	M1	3.1b
	Mean is the same means that $a+b=45$	M1	1.1b
	So possible values are: e.g. $b=21$ and $a=24$ (o.e.)	A1	2.2b
		(3)	
(g)	Both values will be less than 1 standard deviation from the mean and so the standard deviation of all 29 values will be smaller	B1	2.4
		(1)	
		(10 marks)	

June 2018 Question 4 Paper 3 (A-Level)

4. Charlie is studying the time it takes members of his company to travel to the office. He stands by the door to the office from 08 40 to 08 50 one morning and asks workers, as they arrive, how long their journey was.

(a) State the sampling method Charlie used.

(1)

(b) State and briefly describe an alternative method of non-random sampling Charlie could have used to obtain a sample of 40 workers.

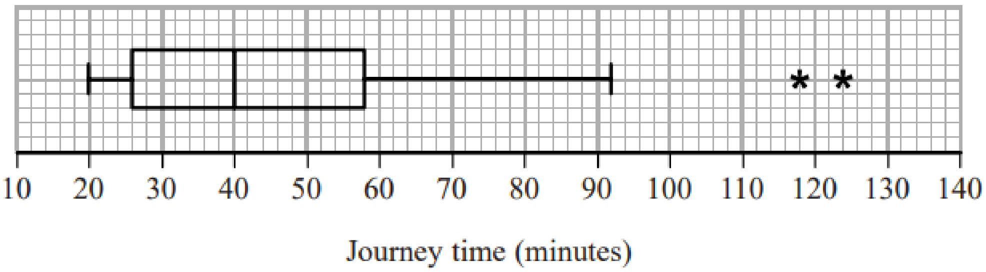
(2)

Taruni decided to ask every member of the company the time, x minutes, it takes them to travel to the office.

(c) State the data selection process Taruni used.

(1)

Taruni's results are summarised by the box plot and summary statistics below.



$$n = 95 \qquad \sum x = 4133 \qquad \sum x^2 = 202\,294$$

(d) Write down the interquartile range for these data.

(1)

ANSWER

Qu 4	Scheme	Marks	AO
(a)	Convenience <u>or</u> opportunity [sampling]	B1	1.2
		(1)	
(b)	Quota [sampling]	B1	1.1a
	e.g. Take 4 people every 10 minutes	B1	1.1b
		(2)	
(c)	Census	B1	1.2
		(1)	
(d)	[58 – 26 =] **<u>32</u>** (min)	B1	1.1b
		(1)	
(e)	$\mu = \frac{4133}{95} = 43.505263\ldots$ **awrt <u>43.5</u> (min)**	B1	1.1b
	$\sigma_x = \sqrt{\frac{202\,294}{95} - \mu^2} = \sqrt{236.7026\ldots}$	M1	1.1b
	= 15.385… **awrt <u>15.4</u> (min)**	A1	1.1b
		(3)	
(f)	There are outliers in the data (or data is skew) which will affect mean and sd	B1	2.4
	Therefore use median and IQR	dB1	2.4
		(2)	
(g)	Value of 20, LQ at 26 and outliers will not change <u>or</u> state that median and upper quartile are the values that <u>do</u> change	B1	1.1b
	<u>More values now below 40 than above</u> so Q_2 <u>or</u> Q_3 will change and be lower	M1	2.1
	Both Q_2 <u>and</u> Q_3 will be lower	A1	2.4
		(3)	
		(13 marks)	