Statistics Year 1 (AS) exam questions

NOTE:

Please be aware that in the Year 12 collections, you will find questions from the Year 13 papers. However, these questions are intentionally included because they align with Year 12 content and topics.

Arancha Ruiz

Ch. 1: Data Collection

June 2024 Question 2

2. Keith is studying the variable Daily Mean Wind Direction, in degrees, from the large data set.

Keith summarised the data for Camborne from 1987 into 4 directions *A*, *B*, *C* and *D* representing North, South, East and West in some order.

Direction	*A*	*B*	*C*	*D*
Frequency	22	48	56	58

(a) Using your knowledge of the large data set state, giving a reason, which direction *A* represents.

(1)

The entry for Hurn on 27th September 1987 was 999

(b) State, giving a reason, what Keith should do with this value.

(2)

ANSWER

Qu 2	**Scheme**	**Marks**	**AO**
(a)	East		
	Since : The prevailing winds in Camborne/UK are from the South or West or North or East since lowest frequency or Camborne is in South/West (and so wind is less likely to come from East)	B1	2.2b
(b)		(1)	
	e.g. 'wind direction should be in the range 0 to 360'	M1	2.4/1.2
	e.g. 'so he should ignore/remove the value'	A1	1.1b
		(2)	
		(3 marks)	

June 2022 Question 3 (A-Level)

A bit of the large data set in this question, but it has binomial, so you will find it here.

November 2021 question 3

3. Helen is studying one of the qualitative variables from the large data set for Heathrow from 2015.

She started with the data from 3rd May and then took every 10th reading.

There were only 3 different outcomes with the following frequencies

Outcome	*A*	*B*	*C*
Frequency	16	2	1

(a) State the sampling technique Helen used.

(1)

(b) From your knowledge of the large data set

(i) suggest which variable was being studied,

(ii) state the name of outcome *A*.

(2)

George is also studying the same variable from the large data set for Heathrow from 2015. He started with the data from 5th May and then took every 10th reading and obtained the following

Outcome	*A*	*B*	*C*
Frequency	16	1	1

Helen and George decided they should examine all of the data for this variable for Heathrow from 2015 and obtained the following

Outcome	*A*	*B*	*C*
Frequency	155	26	3

(c) State what inference Helen and George could reliably make from their original samples about the outcomes of this variable at Heathrow, for the period covered by the large data set in 2015.

(1)

ANSWER

Qu	Scheme	Marks	AO
3. (a)	Systematic (sampling)	B1 **(1)**	1.2
(b)(i)	[Daily Mean] Wind Speed	B1	2.2a
(ii)	Light	B1 **(2)**	1.2
(c)	Variable *A* occurs most (around 80~90%) of the time	B1 **(1)** **(4 marks)**	2.2b

October 2020 question 2

It has correlation, so you will find it in chapter 4.

October 2020 question 4

It has questions about measures of location and spread, so you will find it in chapter 2.

June 2019 question 1

It has sampling and correlation, you will find it in chapter 4.

June 2019 question 4

It has 'large data set' and measures of location and spread, you will find it in chapter 2.

June 2018 Question 1 Paper 3 (A-Level)

1. Helen believes that the random variable C, representing cloud cover from the large data set, can be modelled by a discrete uniform distribution.

(a) Write down the probability distribution for C.

(2)

(b) Using this model, find the probability that cloud cover is less than 50%

(1)

Helen used all the data from the large data set for Hurn in 2015 and found that the proportion of days with cloud cover of less than 50% was 0.315

(c) Comment on the suitability of Helen's model in the light of this information.

(1)

(d) Suggest an appropriate refinement to Helen's model.

(1)

ANSWER

Qu 1	Scheme	Marks	AO
(a)	c: 0, 1, 2, 3, 4, 5, 6, 7, 8 $P(C=c)$: $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$	B1 B1ft (2)	1.2 1.2
(b)	$P(C<4)=\frac{4}{9}$ (accept 0.444 or better)	B1 (1)	3.4
(c)	Probability lower than expected suggests model is <u>not</u> good	B1ft (1)	3.5a
(d)	e.g. Cloud cover will vary from month to month and place to place So e.g. use a non-uniform distribution	B1 (1)	3.5c
		(5 marks)	

Ch. 2: Measures of Location and Spread

June 2024 Question 3 (A-level)

3. Ming is studying the large data set for Perth in 2015

He intended to use all the data available to find summary statistics for the Daily Mean Air Temperature, x °C.
Unfortunately, Ming selected an incorrect variable on the spreadsheet.
This incorrect variable gave a mean of 5.3 and a standard deviation of 12.4

(a) Using your knowledge of the large data set, suggest which variable Ming selected.

(1)

The correct values for the Daily Mean Air Temperature are summarised as

$$n = 184 \qquad \sum x = 2801.2 \qquad \sum x^2 = 44\,695.4$$

(b) Calculate the mean and standard deviation for these data.

(3)

One of the months from the large data set for Perth in 2015 has

- mean $\bar{x} = 19.4$
- standard deviation $\sigma_x = 2.83$

for Daily Mean Air Temperature.

(c) Suggest, giving a reason, a month these data may have come from.

(2)

ANSWER

Qu 3	Scheme	Marks	AO
(a)	Rain[fall] (allow [Mean] Windspeed)	B1 **(1)**	1.2
(b)	$[\bar{x} =]$ 15.2239… = awrt **15.2**	B1	1.1b
	$\sigma_x = \sqrt{\frac{44\,695.4}{184} - "15.22.."^2}$ or $\sqrt{11.1(422...)}$	M1	1.1b
	$= 3.33800\ldots$ awrt **3.34**	A1 **(3)**	1.1b
(c)	Mean is higher than average **OR** a summer/spring month If they say winter/autumn they must explain that these are hotter months for Perth.	M1	2.4
	[Perth is southern hemisphere or Australia so latest available] month is **Oct**	A1cso **(2)**	2.2b
		(6 marks)	

June 2023 Question 3 (A-Level)

3. Ben is studying the Daily Total Rainfall, x mm, in Leeming for 1987

He used all the data from the large data set and summarised the information in the following table.

x	0	0.1–0.5	0.6–1.0	1.1–1.9	2.0–4.0	4.1–6.9	7.0–12.0	12.1–20.9	21.0–32.0	tr
Frequency	55	18	18	21	17	9	9	6	2	29

(a) Explain how the data will need to be cleaned before Ben can start to calculate statistics such as the mean and standard deviation.

(2)

Using all 184 of these values, Ben estimates $\sum x = 390$ and $\sum x^2 = 4336$

(b) Calculate estimates for

(i) the mean Daily Total Rainfall,

(ii) the standard deviation of the Daily Total Rainfall.

(3)

Ben suggests using the statistic calculated in part (b)(i) to estimate the annual mean Daily Total Rainfall in Leeming for 1987

(c) Using your knowledge of the large data set,

(i) give a reason why these data would not be suitable,

(ii) state, giving a reason, how you would expect the estimate in part (b)(i) to differ from the actual annual mean Daily Total Rainfall in Leeming for 1987

(2)

ANSWER

Qu 3	Scheme	Marks	AO
(a)	Need to replace tr with a numerical value Value of tr is between 0 and 0.05 suggest using e.g 0.025 , 0 or value „ 0.05	M1 A1 **(2)**	1.2 1.1b
(b)(i)	$\left[\bar{x} = \frac{389.3 \sim 390.8}{184}\right] = 2.119\ldots$ awrt **2.12** allow $\frac{195}{92}$ or $2\frac{11}{92}$	B1	1.1b
(ii)	$[\sigma =]\sqrt{\frac{\text{(awrt)}4336}{184} - "\bar{x}^2"}$ or allow $\left[\sigma^2 =\right]\frac{\text{(awrt)}4336}{184} - "\bar{x}^2"$ or awrt 19.1 $= 4.367\ldots$ awrt **4.37**	M1 A1 **(3)**	1.1b 1.1b
(c)(i)	Only covers May~Oct (so not a suitable sample)	B1	1.1b
(ii)	e.g. Winter months are missing when we'd expect more rain so expect estimate in (b)(i) to be an underestimate (oe)	B1 **(2)**	2.4
		(7 marks)	

June 2022 Question 3 (A-Level)

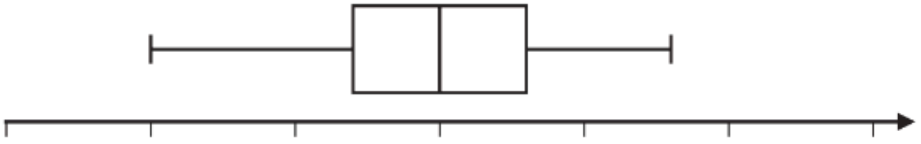
Finding mean and variance in this question, but it has binomial as well, so you will find it here.

June 2022 Question 4

4. Jiang is studying the variable Daily Mean Pressure from the large data set.

He drew the following box and whisker plot for these data for one of the months for one location using a linear scale but

- he failed to label all the values on the scale
- he gave an incorrect value for the median



Daily Mean Pressure (hPa)

Using your knowledge of the large data set, suggest a suitable value for

(a) the median,

(1)

(b) the range.

(1)

(*You are not expected to have memorised values from the large data set. The question is simply looking for sensible answers.*)

ANSWER

4. (a)	Accept 990 to 1030 inclusive	B1 **(1)**	1.1b
(b)	Any range between 10 and 50 inclusive	B1 **(1)**	1.1b
		(2 marks)	

November 2021 question 2

It has a histogram as well, so you will find it in chapter 3 instead.

October 2020 question 4

4. A lake contains three different types of carp.

There are an estimated 450 mirror carp, 300 leather carp and 850 common carp.

Tim wishes to investigate the health of the fish in the lake.

He decides to take a sample of 160 fish.

(a) Give a reason why stratified random sampling cannot be used.

(1)

(b) Explain how a sample of size 160 could be taken to ensure that the estimated populations of each type of carp are fairly represented.

You should state the name of the sampling method used.

(2)

As part of the health check, Tim weighed the fish.

IIis results are given in the table below.

Weight (w kg)	Frequency (f)	Midpoint (m kg)
$2 \leqslant w < 3.5$	8	2.75
$3.5 \leqslant w < 4$	32	3.75
$4 \leqslant w < 4.5$	64	4.25
$4.5 \leqslant w < 5$	40	4.75
$5 \leqslant w < 6$	16	5.5

(You may use $\sum fm = 692$ and $\sum fm^2 = 3053$)

(c) Calculate an estimate for the standard deviation of the weight of the carp.

(2)

Tim realised that he had transposed the figures for 2 of the weights of the fish.

He had recorded in the table 2.3 instead of 3.2 and 4.6 instead of 6.4

(d) Without calculating a new estimate for the standard deviation, state what effect

(i) using the correct figure of 3.2 instead of 2.3

(ii) using the correct figure of 6.4 instead of 4.6

would have on your estimated standard deviation.

Give a reason for each of your answers.

(2)

ANSWER

Question	Scheme	Marks	AOs
4(a)	It is not possible to have a sampling frame	B1	2.3
		(1)	
(b)	Quota sampling **and** (catch 85 common carp, 45 mirror carp and 30 leather carp) **or** (ignore any fish caught of a type where the quota is full)	M1	1.1a
	Quota sampling **and** catch 85 common carp, 45 mirror carp and 30 leather carp **and** ignore any fish caught of a type where the quota is full	A1	1.1b
		(2)	
(c)	$\sigma = \sqrt{\frac{3053}{160} - \left(\frac{692}{160}\right)^2}$	M1	1.1b
	$= 0.6129\ldots$ awrt 0.613	A1	1.1b
		(2)	
(d)(i)	This would have no effect as the piece of data would remain in the same class	B1	2.2a
(ii)	This would increase the standard deviation as change in mean is small and $6.4 - 4.6 \approx 3\sigma$ therefore estimate of standard deviation will increase	B1	2.2a
		(2)	
			(7 marks)

June 2019 question 4

4. Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

(a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

(1)

Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall (r mm)	Frequency	Midpoint (x mm)
$0 \leqslant r < 0.5$	121	0.25
$0.5 \leqslant r < 1.0$	10	0.75
$1.0 \leqslant r < 5.0$	24	3.0
$5.0 \leqslant r < 10.0$	12	7.5
$10.0 \leqslant r < 30.0$	17	20.0

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

(b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

(2)

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

(2)

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

(3)

ANSWER

Question	Scheme	Marks	AOs
4 (a)	Tr(ace) (data needs to be converted to numbers before the calculation can be carried out)	B1	2.4
		(1)	
(b)	$[1+]\frac{138-131}{24}\times 4$	M1	2.1
	$= 2.1666\ldots$ awrt **2.17**	A1	1.1b
		(2)	
(c)	$\sigma = \sqrt{\frac{7704.1875}{184} - \left(\frac{539.75}{184}\right)^2} = 5.7676\ldots$ $\sigma =$ awrt **5.77**	M1 A1	1.1b 1.1b
		(2)	
(d)(i)	Using class midpoints to estimate the mean assumes that the values are uniformly distributed **within the class(es)**.	B1	2.4
(ii)& (iii)	This is not the case here as the majority of the data (in the first class) are 0.	B1	2.3
	The actual mean is likely to be smaller than the estimate (since the first group has more values at 0 and close to 0)	dB1	2.2b
		(3)	
			(8 marks)

June 2018 Question 4

4. Helen is studying the daily mean wind speed for Camborne using the large data set from 1987. The data for one month are summarised in Table 1 below.

Windspeed	n/a	6	7	8	9	11	12	13	14	16
Frequency	13	2	3	2	2	3	1	2	1	2

Table 1

(a) Calculate the mean for these data.

(1)

(b) Calculate the standard deviation for these data and state the units.

(2)

The means and standard deviations of the daily mean wind speed for the other months from the large data set for Camborne in 1987 are given in Table 2 below. The data are not in month order.

Month	*A*	*B*	*C*	*D*	*E*
Mean	7.58	8.26	8.57	8.57	11.57
Standard Deviation	2.93	3.89	3.46	3.87	4.64

Table 2

(c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

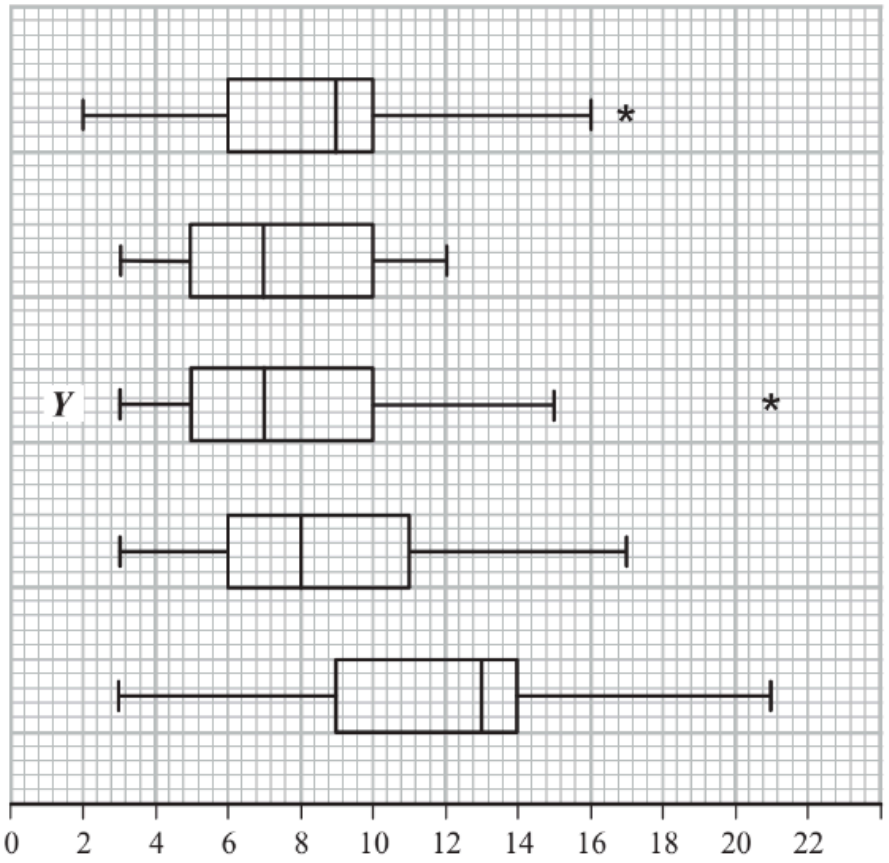
(2)

The data for these months are summarised in the box plots on the opposite page.
They are not in month order or the same order as in Table 2.

(d) (i) State the meaning of the * symbol on some of the box plots.

(ii) Suggest, giving your reasons, which of the months in Table 2 is most likely to be summarised in the box plot marked *Y*.

(3)



ANSWER

Qu	Scheme	Marks	AO
4 (a)	$\bar{x} = 10.2\,(2222\ldots)$ awrt **10.2**	B1 (1)	1.1b
(b)	$\sigma_x = 3.17\,(20227\ldots)$ awrt **3.17**	B1ft	1.1b
	Sight of "knots" or "kn" (condone knots/s etc)	B1 (2)	1.2
(c)	October ….. since it is windier in the autumn or month of the hurricane or latest month in the year	B1 B1 (2)	2.2b 2.4
(d)(i)	They represent outliers	B1	1.2
(ii)	*Y* has low median so expect lowish mean (but outlier so > 7) and *Y* has big range/IQR or spread so expect larger st.dev Suggests *B*	M1 A1 (3)	2.4 2.2b
		(8 marks)	

Ch. 3: Representations of Data

June 2024 Question 3

3. Customers in a shop have to queue to pay.

The partially completed table below and partially completed histogram opposite, give information about the time, x minutes, spent in the queue by each of 112 customers one day.

Time in queue (x minutes)	Frequency
1–2	64
2–3	
3–4	13
4–6	
6–8	3

No customer spent less than 1 minute or longer than 8 minutes in the queue.

(a) Complete the table.

(2)

(b) Complete the histogram.

(2)

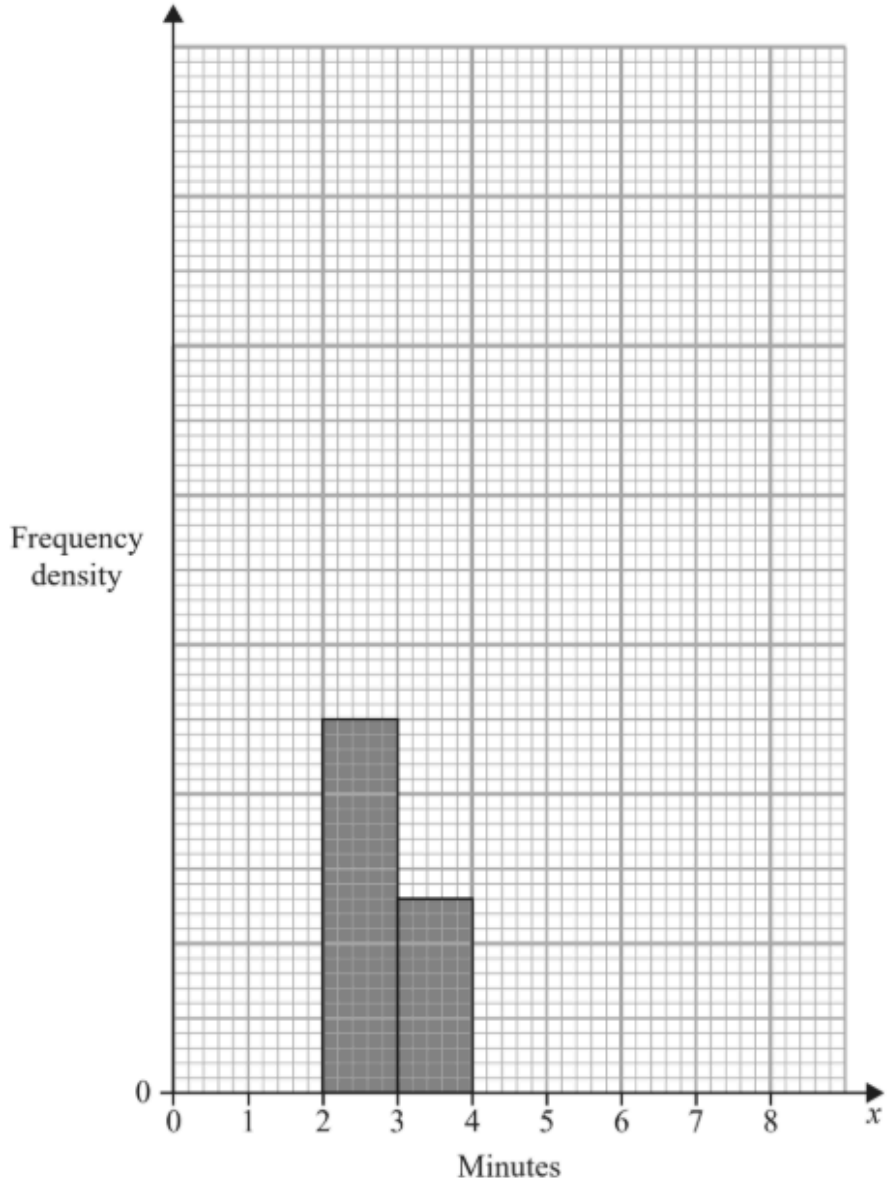
Ting decides to model the **frequency density** for these 112 customers by a curve with equation

$$y = \frac{k}{x^2} \qquad 1 \leqslant x \leqslant 8$$

where k is a constant.

(c) Find the value of k

(3)



ANSWER

Question	Scheme	Mark	AO
3(a)	[From 3~4 bar and freq in table deduce] fd scale of 1cm = 5	M1	2.2a
	2~3 has freq = **25** **and** 4~6 has freq = 112 – (64+13+3 + "25") = **7**	A1	1.1b
		(2)	
(b)	For a bar between 4~6 of height $\frac{"7"}{2}$ = 3.5 small squares or	M1	2.2a
	For a bar between 6~8 of height 1.5 small squares		
	For a fully correct histogram with all 3 bars plotted correctly	A1ft	1.1b
		(2)	
(c)	Require $\int_{(1)}^{(8)} \frac{k}{x^2}\,dx = 112$	M1	3.4
	$= \left[\frac{-k}{x}\right]_1^8 = \left(-\frac{k}{8}\right) - \left(-\frac{k}{1}\right) [= 112]$	M1	1.1b
	i.e. $\left[\frac{7}{8}k = 112 \Rightarrow\right]$ $k = \mathbf{128}$	A1	1.1b
		(3)	
		(7 marks)	

June 2024 Question 1

1. A coach recorded the heights of some adult rugby players and found the following summary statistics.

Median = 1.85 m

Range = 0.28 m

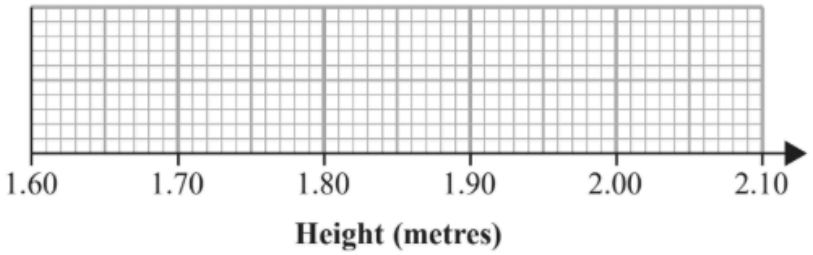
Interquartile range = 0.11 m

The coach also noticed that

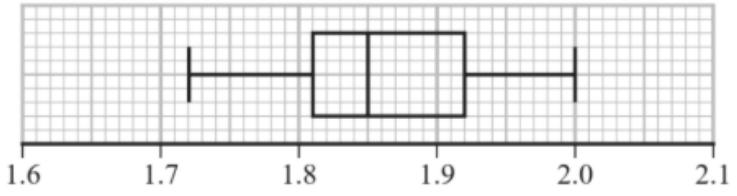
- the height of the shortest player is 1.72 m
- 25% of the players' heights are below the height of a player whose height is 1.81 m

Draw a box and whisker plot to represent this information on the grid below.

(4)

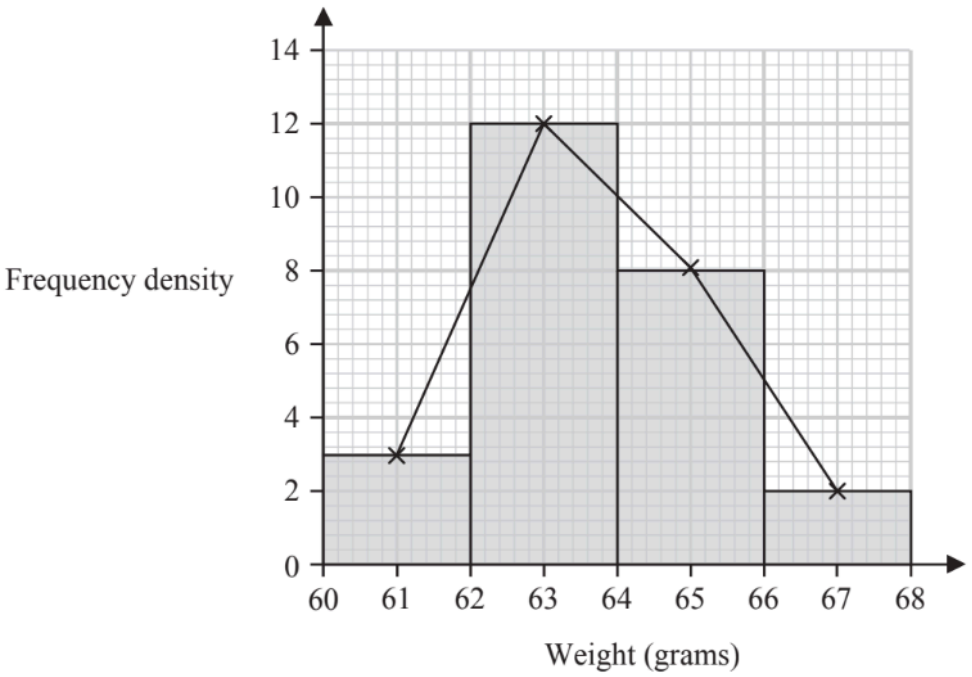


ANSWER

Question	Scheme	Marks	AO
1		B1 B1ft B1 B1ft **(4 marks)**	2.2a 1.1b 2.2a 1.1b
	Notes		

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.

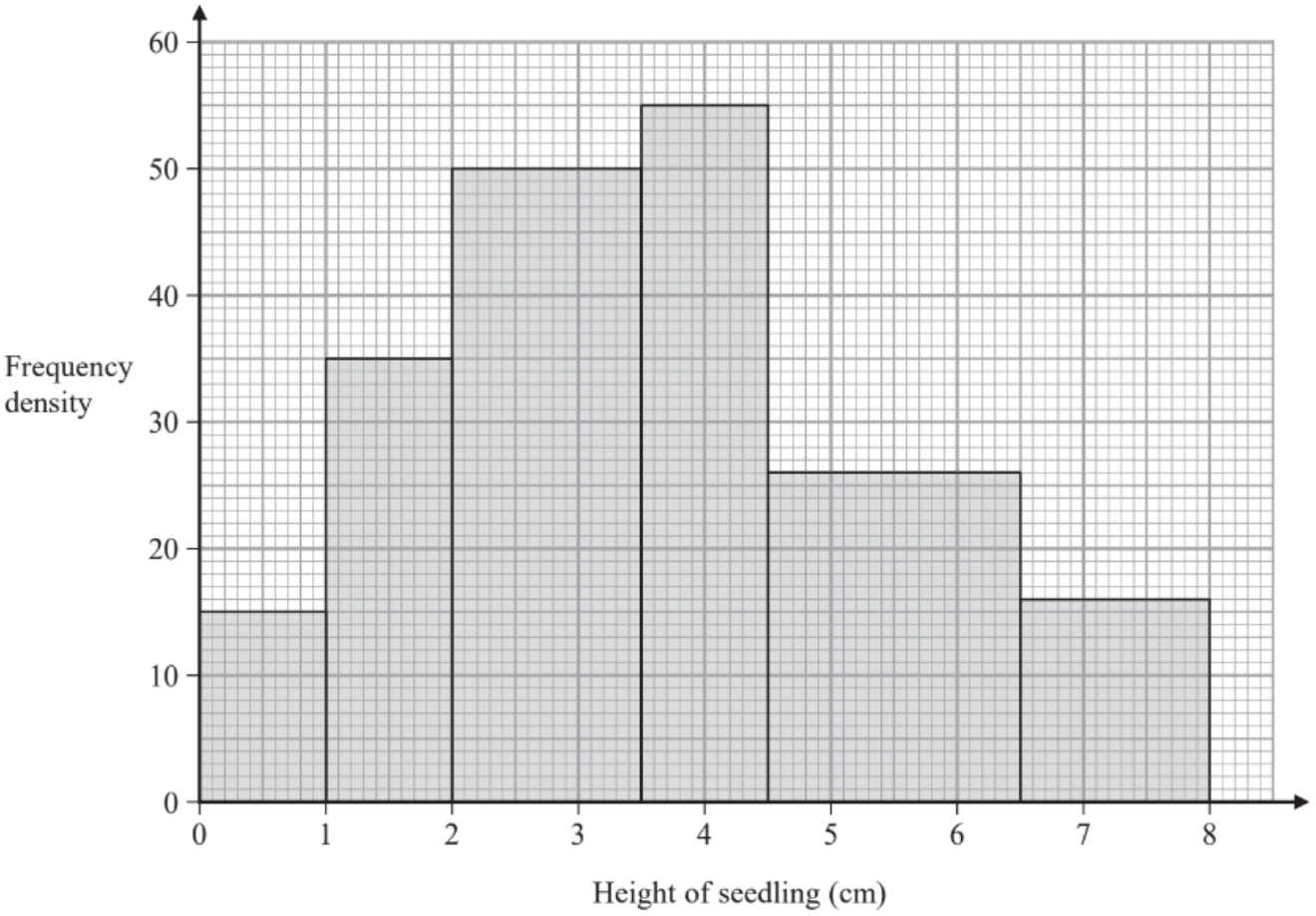
(1)

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
	$=\sqrt{2.5216}=1.58795...$ = awrt **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)	<table><tr><th>Class</th><th>Frequency</th><th>Cum. Frequency</th></tr><tr><td>0 – 1</td><td>15</td><td>15</td></tr><tr><td>1 – 2</td><td>35</td><td>50</td></tr><tr><td>2 – 3.5</td><td>75</td><td>125</td></tr><tr><td>3.5 – 4.5</td><td>55</td><td>180</td></tr></table>	M1 A1	2.1 1.1b
	$[Q_2 =](3.5)+\frac{\frac{256}{2}-"125"}{"55"}\times(4.5-3.5)$ or $(4.5)-\frac{"180"-\frac{256}{2}}{"55"}\times 1$	M1	2.1
	$= 3.5545\ldots\ldots$ awrt **3.55**	A1 **(4)**	1.1b
(b)	Need area under curve to be 256 so $\int_{(0)}^{(8)} kx(8-x)\,dx = 256$	M1	3.1a
	$k\left[4x^2-\frac{x^3}{3}\right]_{(0)}^{(8)} = 256$	M1	1.1b
	$\left\{k\left[4\times 8^2-\frac{8}{3}\times 8^2\right]=256\Rightarrow\right\}$ **$k = 3$**	A1 **(3)**	1.1b
(c)	[By symmetry median =] **4**	B1 **(1)**	2.2a
		(8 marks)	

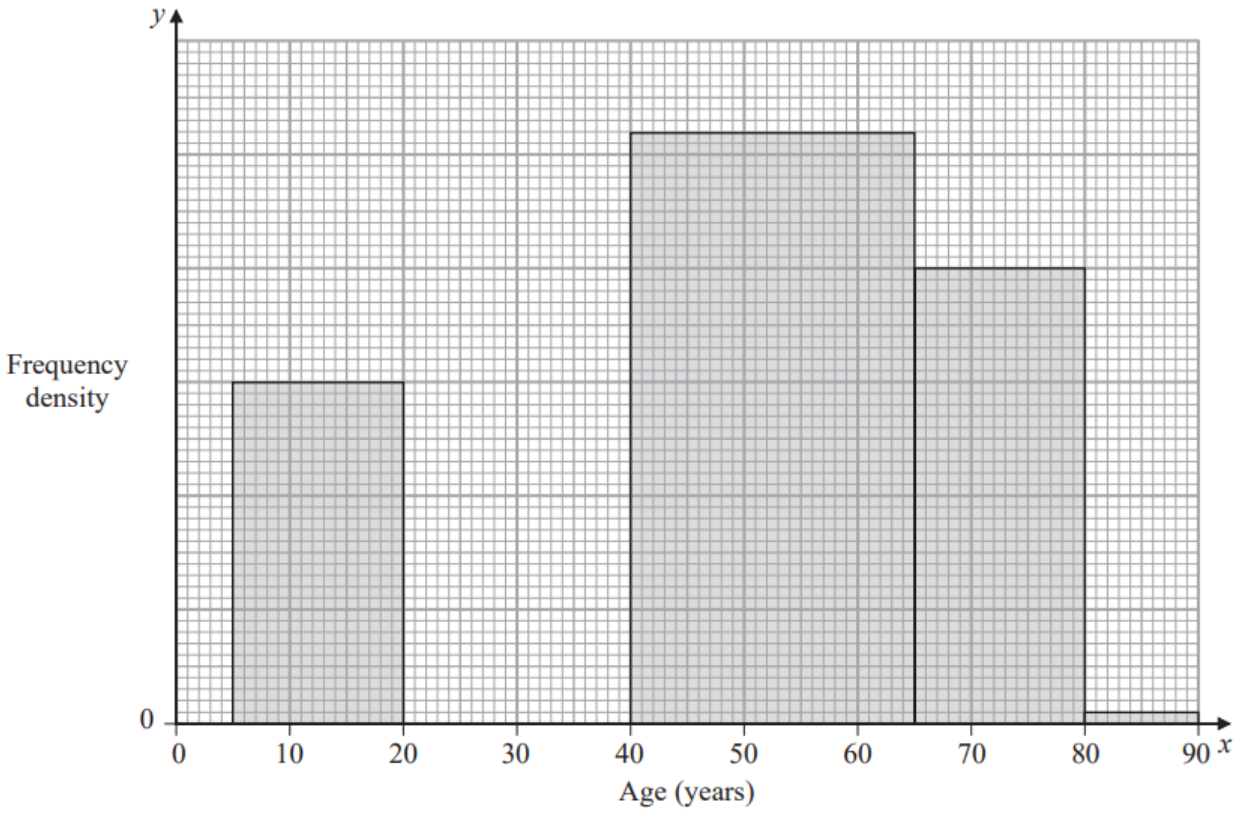
Video solution:
https://youtu.be/zf9owtfBqL0

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 or 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) **130** passengers or for [65, 80) **60** passengers	M1	2.1
	For attempt to find total number of passengers = **331**	A1ft	1.1b
	[Median =] $40+\frac{\frac{1}{2}("331")-140}{"130"}\times 25$ or $65-\frac{270-\frac{1}{2}("331")}{"130"}\times 25$ (o.e.)	M1	1.1b
	$= 44.9038\ldots$ = awrt **44.9**	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 not an outlier	A1	2.2a
		(2)	
		(9 marks)	

October 2020 question 1

1.

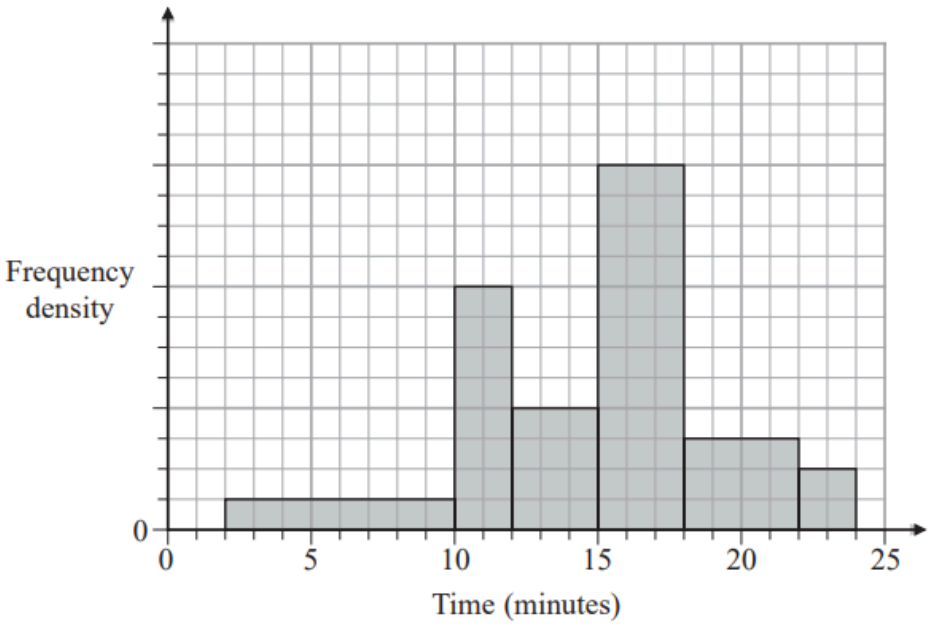


Figure 1

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

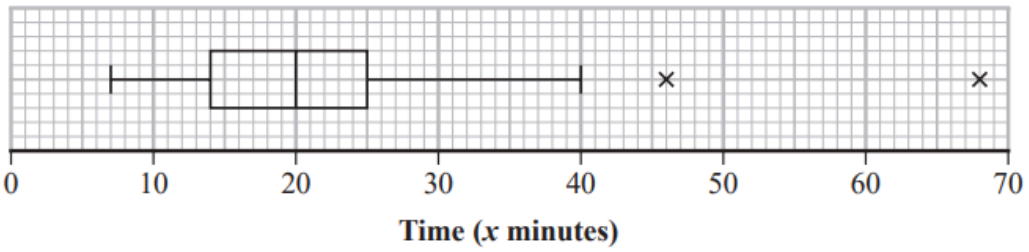
You will find this question in chapter 4 because it has correlation.

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.



Time (x minutes)

(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 =] **61** (only)	B1 **(1)**	1.1b
(b)	[25 – 14] = **11**	B1 **(1)**	1.1b
(c)	$\left[\mu \text{ or } \bar{x} = \frac{607.5}{27} = \right] = \mathbf{22.5}$	B1 **(1)**	1.1b
(d)	$\sigma = \sqrt{\frac{17\,623.25}{27} - "22.5"^2}$ or $\sqrt{146.4629...}$	M1	1.1b
	= 12.10218… awrt **12.1**	A1 **(2)**	1.1b
(e)	$\mu + 3\sigma = "22.5" + 3 \times "12.1..." =$ awrt 59 so only **one** outlier	B1ft **(1)**	1.1b
(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 **(3)**	2.2b
(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 **(1)**	2.4
		(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.

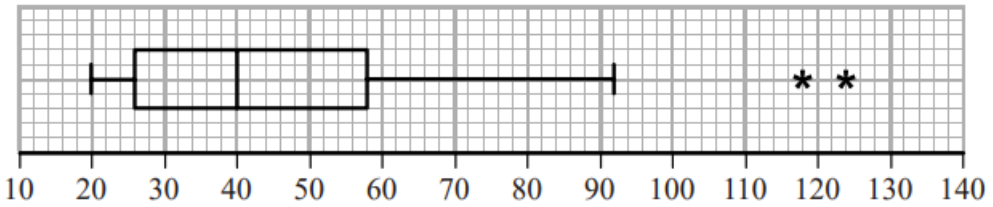
(2)

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.



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

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

(1)

ANSWER

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

Ch. 4: Correlation

June 2023 Question 2

2. Fred and Nadine are investigating whether there is a linear relationship between Daily Mean Pressure, p hPa, and Daily Mean Air Temperature, t °C, in Beijing using the 2015 data from the large data set.

Fred randomly selects one month from the data set and draws the scatter diagram in Figure 1 using the data from that month.

The scale has been left off the horizontal axis.

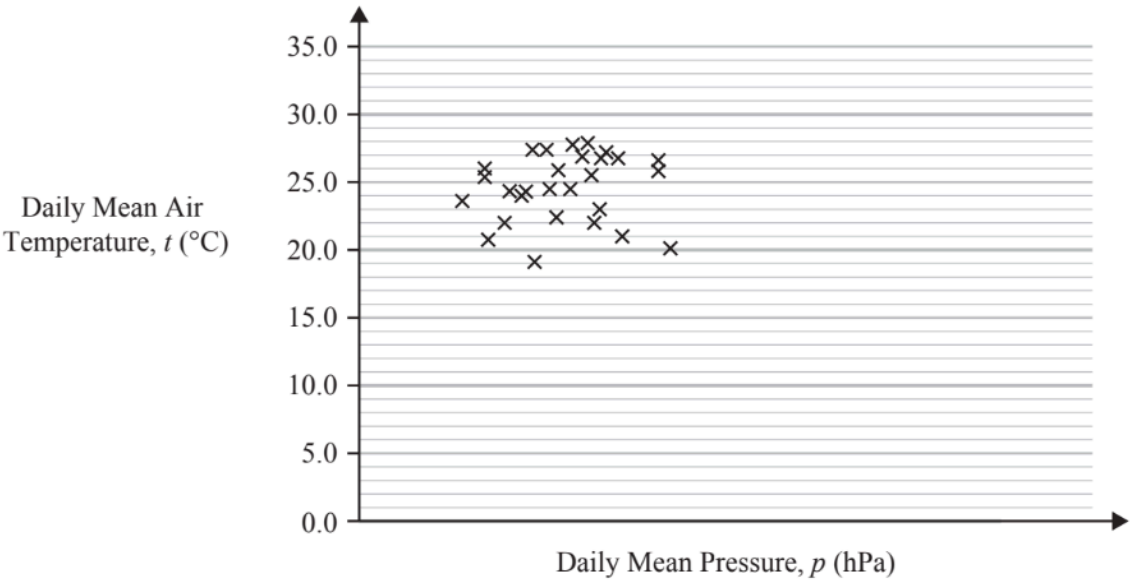


Figure 1

(a) Describe the correlation shown in Figure 1.

(1)

Nadine chooses to use all of the data for Beijing from 2015 and draws the scatter diagram in Figure 2.

She uses the same scales as Fred.

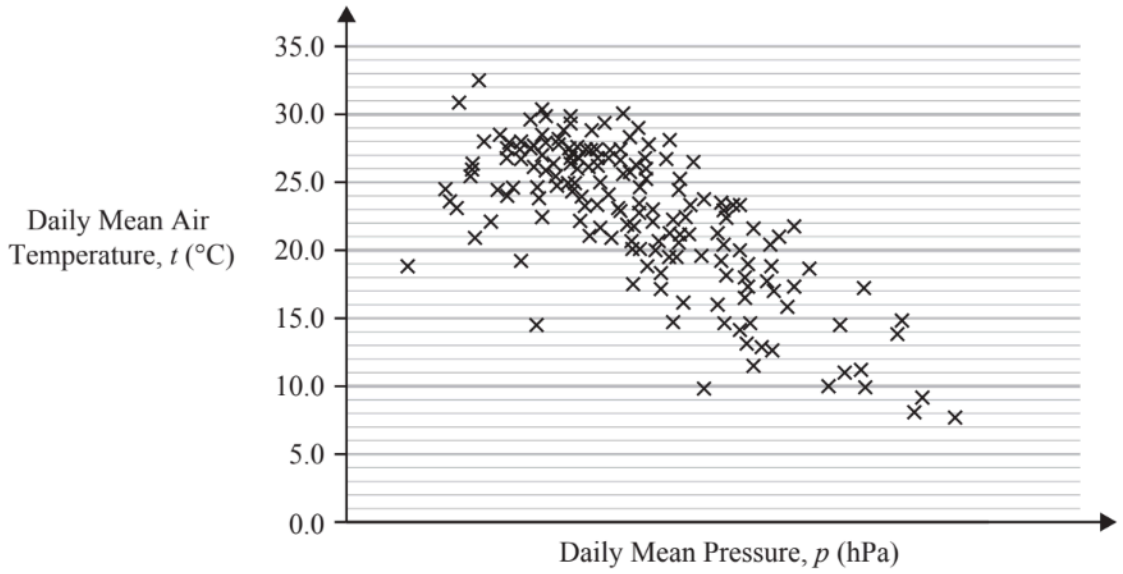


Figure 2

(b) Explain, in context, what Nadine can infer about the relationship between p and t using the information shown in Figure 2.

(1)

(c) Using your knowledge of the large data set, state a value of p for which interpolation can be used with Figure 2 to predict a value of t.

(1)

(d) Using your knowledge of the large data set, explain why it is not meaningful to look for a linear relationship between Daily Mean Wind Speed (Beaufort Conversion) and Daily Mean Air Temperature in Beijing in 2015.

(1)

ANSWER

Question	Scheme	Marks	AOs
2(a)	No (correlation)/weak (correlation)	B1	1.1b
		(1)	
(b)	(Negative correlation…) As p(ressure) increases, t(emperature) decreases.	B1	2.2b
		(1)	
(c)	990 to 1040 (hPa)	B1	3.4 LDS
		(1)	
(d)	Daily mean wind speed (Beaufort) is a qualitative variable.	B1	2.4 LDS
		(1)	
			(4 marks)

June 2022 Question 1

1. The relationship between two variables p and t is modelled by the regression line with equation

$$p = 22 - 1.1\,t$$

The model is based on observations of the independent variable, t, between 1 and 10

(a) Describe the correlation between p and t implied by this model.

(1)

Given that p is measured in centimetres and t is measured in days,

(b) state the units of the gradient of the regression line.

(1)

Using the model,

(c) calculate the change in p over a 3-day period.

(2)

Tisam uses this model to estimate the value of p when $t = 19$

(d) Comment, giving a reason, on the reliability of this estimate.

(1)

ANSWER

Qu	Scheme	Marks	AO
1. (a)	Negative (since gradient of regression line is negative)	B1 **(1)**	1.2
(b)	cm/day (o.e. e.g. cm day^{-1})	B1 **(1)**	2.2a
(c)	$3\times[\pm]1.1$ = decrease of 3.3 [cm]	M1 A1 **(2)**	3.4 1.1b
(d)	19 is (well) outside the range [1, 10] or involves extrapolation (o.e.) so (possibly) unreliable/ inaccurate (o.e.)	B1 **(1)**	2.4
		(5 marks)	

Video solution:

https://youtu.be/ZtQSYJYy51A

October 2020 question 2

2. Jerry is studying visibility for Camborne using the large data set June 1987.

The table below contains two extracts from the large data set.

It shows the daily maximum relative humidity and the daily mean visibility.

Date	Daily Maximum Relative Humidity	Daily Mean Visibility
Units	**%**	
10/06/1987	90	5300
28/06/1987	100	0

(The units for Daily Mean Visibility are deliberately omitted.)

Given that daily mean visibility is given to the nearest 100,

(a) write down the range of distances in metres that corresponds to the recorded value 0 for the daily mean visibility.

(1)

Jerry drew the following scatter diagram, Figure 2, and calculated some statistics using the June 1987 data for Camborne from the large data set.

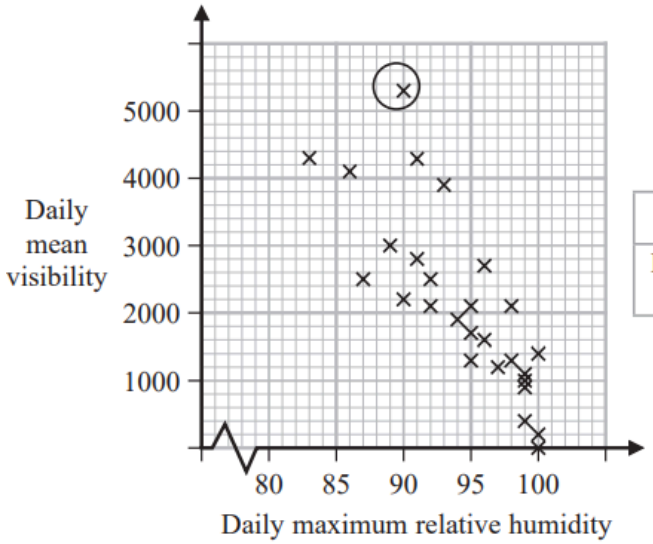


Figure 2

	Q_1	IQR
Daily mean visibility	1100	1600
Daily maximum relative humidity (%)	92	8

Jerry defines an outlier as a value that is more than 1.5 times the interquartile range above Q_3 or more than 1.5 times the interquartile range below Q_1.

(b) Show that the point circled on the scatter diagram is an outlier for visibility.

(2)

(c) Interpret the correlation between the daily mean visibility and the daily maximum relative humidity.

(1)

Jerry drew the following scatter diagram, Figure 3, using the June 1987 data for Camborne from the large data set, but forgot to label the x–axis.

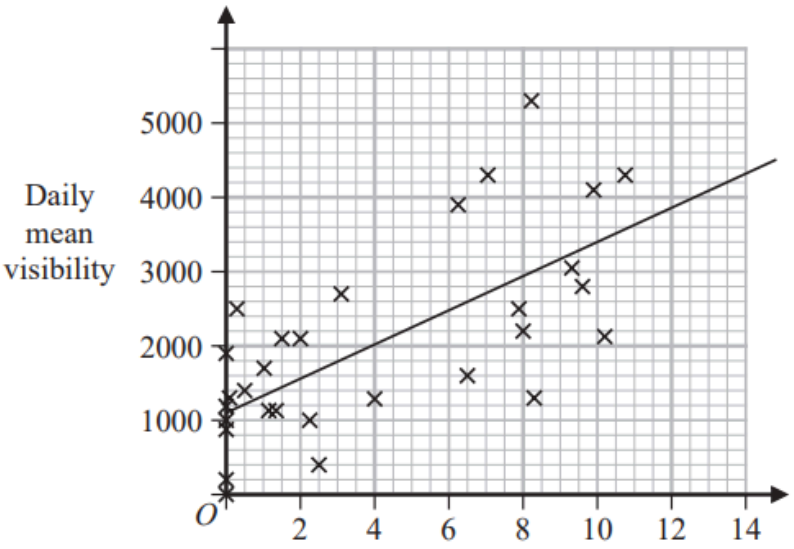


Figure 3

(d) Using your knowledge of the large data set, suggest which variable the x-axis on this scatter diagram represents.

(1)

ANSWER

Question	Scheme	Marks	AOs
2(a)	0 to 500 m	B1	1.2
		(1)	
(b)	$1100+1600+1.5\times1600$ [= 5100]	M1	2.1
	5300 > 5100 therefore outlier	A1	1.1b
		(2)	
(c)	As the humidity increases the mean visibility decreases	B1	2.4
		(1)	
(d)	(Hours of) sunshine	B1	2.2b
		(1)	
			(5 marks)

June 2019 question 1

1. A sixth form college has 84 students in Year 12 and 56 students in Year 13

The head teacher selects a stratified sample of 40 students, stratified by year group.

(a) Describe how this sample could be taken.

(3)

The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.
For the sample of 40 students, he finds the equation of the regression line of p on s to be

$$p = 26.1 + 5.60s$$

(b) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

(1)

(c) Describe one limitation of this regression model.

(1)

ANSWER

Question	Scheme	Marks	AOs
1(a)	Label **each** year group	B1	1.1b
	Use random numbers to select a ...	B1	1.1b
	Simple random sample of 24 Year 12s and 16 Year 13s.	B1	1.1b
		(3)	
(b)	Increase by 2.8 marks	B1	3.4
		(1)	
(c)	e.g. 'the best performance is predicted for the students who never wake up'	B1	3.5b
		(1)	
			(5 marks)

June 2018 Question 1

1. A company is introducing a job evaluation scheme. Points (x) will be awarded to each job based on the qualifications and skills needed and the level of responsibility. Pay (£y) will then be allocated to each job according to the number of points awarded.

Before the scheme is introduced, a random sample of 8 employees was taken and the linear regression equation of pay on points was $y = 4.5x - 47$

(a) Describe the correlation between points and pay.

(1)

(b) Give an interpretation of the gradient of this regression line.

(1)

(c) Explain why this model might not be appropriate for all jobs in the company.

(1)

ANSWER

Qu	Scheme	Marks	AO
1 (a)	Positive (correlation)	B1 (1)	1.2
(b)	Every extra point gives £4.5(0) more on pay (o.e.)	B1 (1)	3.4
(c)	e.g. For points < 11 it would give pay < 0 which is ridiculous	B1 (1)	2.4
		(3 marks)	

Ch. 5: Probability

June 2023 Question 3

3. In an after-school club, students can choose to take part in Art, Music, both or neither.

There are 45 students that attend the after-school club. Of these

- 25 students take part in Art
- 12 students take part in both Art and Music
- the number of students that take part in Music is x

(a) Find the range of possible values of x

(2)

One of the 45 students is selected at random.

Event A is the event that the student selected takes part in Art.

Event M is the event that the student selected takes part in Music.

(b) Determine whether or not it is possible for the events A and M to be independent.

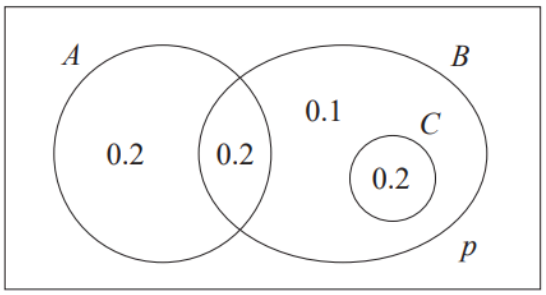
(4)

ANSWER

Que.	Scheme	Marks	AOs
3(a)	$45 - 25 = 20$ or e.g. '$25 \leqslant 13 + 12 + y \leqslant 45$'	M1	2.1
	$12 \leqslant x \leqslant 32$	A1	1.1b
		(2)	
(b)	To be independent $\mathrm{P}(A) \times \mathrm{P}(M) = \mathrm{P}(A \text{ and } M)$	M1	1.1a
	$\mathrm{P}(M) = \dfrac{\mathrm{P}(A \text{ and } M)}{\mathrm{P}(A)} = \dfrac{\frac{12}{45}}{\frac{25}{45}} = \dfrac{12}{25}$ or $\dfrac{25}{45} \times \mathrm{P}(M) = \dfrac{12}{45}$ or $\dfrac{25}{45} \times \dfrac{x}{45} = \dfrac{12}{45}$ or $\dfrac{25}{45} \times \dfrac{12+y}{45} = \dfrac{12}{45}$	A1	2.1
	The number of students taking part in music would be $\dfrac{12}{25} \times 45 = 21.6$ The number of students taking part in music but not art would be $y = 9.6$	A1	1.1b
	…so it is not possible for A and M to be independent (since it must be a whole number).	A1	2.2a
		(4)	
			(6 marks)

November 2021 question 1

1.



The Venn diagram, where p is a probability, shows the 3 events A, B and C with their associated probabilities.

(a) Find the value of p.

(1)

(b) Write down a pair of mutually exclusive events from A, B and C.

(1)

ANSWER

Qu	Scheme	Marks	AO
1 (a)	$[p = 1 - (0.2 + 0.2 + 0.1 + 0.2)] =$ **0.3**	B1 **(1)**	1.1b
(b)	A and C are mutually exclusive. [NOT P(A) and P(C)]	B1 **(1)**	1.2
		(2 marks)	

November 2021 question 5

5. Two bags, **A** and **B**, each contain balls which are either red or yellow or green.

Bag **A** contains 4 red, 3 yellow and n green balls.
Bag **B** contains 5 red, 3 yellow and 1 green ball.

A ball is selected at random from bag **A** and placed into bag **B**.
A ball is then selected at random from bag **B** and placed into bag **A**.

The probability that bag **A** now contains an equal number of red, yellow and green balls is p.

Given that $p > 0$, find the possible values of n and p.

(5)

ANSWER

Qu	Scheme	Marks	AO
5	Must end up with 3 of each colour or 4 of each colour	M1	3.1b
	$\underline{n = 2}$ requires 1st red and 2nd green or red from **A** and green from **B**	M1	2.2a
	$P(1^{st} \text{ red and } 2^{nd} \text{ green}) = \frac{4}{9} \times \frac{1}{10} = \frac{4}{90}$ or $\frac{2}{45}$ $\quad p = \underline{\underline{\frac{2}{45}}}$	A1	1.1b
	$\underline{n = 5}$ requires 1st green and 2nd yellow or green from **A** and yellow from **B**	M1	2.2a
	$P(1^{st} \text{ green and } 2^{nd} \text{ yellow}) = \frac{5}{12} \times \frac{3}{10} = \frac{15}{120}$ or $\frac{1}{8}$ $\quad p = \underline{\underline{\frac{1}{8}}}$	A1	1.1b
		(5)	
		(5 marks)	

October 2020 question 3

3. In a game, a player can score 0, 1, 2, 3 or 4 points each time the game is played.

The random variable S, representing the player's score, has the following probability distribution where a, b and c are constants.

s	0	1	2	3	4
$P(S=s)$	a	b	c	0.1	0.15

The probability of scoring less than 2 points is twice the probability of scoring at least 2 points.

Each game played is independent of previous games played.

John plays the game twice and adds the two scores together to get a total.

Calculate the probability that the total is 6 points.

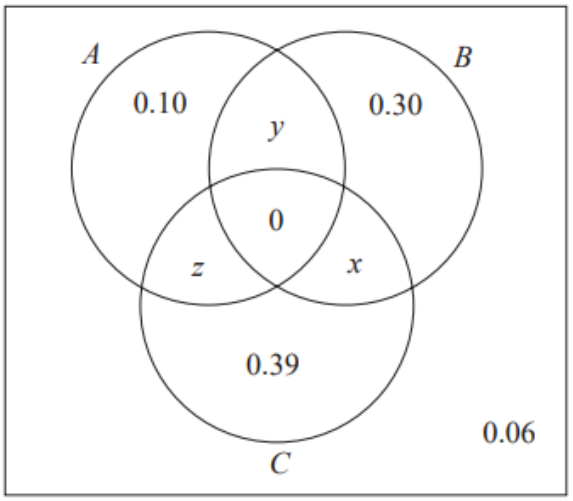
(6)

ANSWER

Question	Scheme	Marks	AOs
3	Overall method	M1	2.1
	$a+b=2c+0.5$ oe **or** $a+b=2(1-a-b)$	B1	2.2a
	$a+b+c=0.75$ oe	B1	1.1b
	$3c=0.25$ $\left[c=0.0833...\text{ or }\frac{1}{12}\right]$	M1	1.1b
	P(scoring 2,4 or 4,2 or 3,3) = $2\times"\frac{1}{12}"\times0.15+0.1^2$	M1	3.1b
	$=0.035$ oe	A1cso	1.1b
		(6)	
			(6 marks)

June 2019 question 2

2. The Venn diagram shows three events, A, B and C, and their associated probabilities.



Events B and C are mutually exclusive.
Events A and C are independent.

Showing your working, find the value of x, the value of y and the value of z.

(5)

ANSWER

Question	Scheme	Marks	AOs
2	$x = 0$	B1	2.2a
	$P(A) = 0.1 + z + y \quad P(C) = 0.39 + z[+x] \quad P(A \text{ and } C) = z$	M1	2.1
	$P(A \text{ and } C) = P(A) \times P(C) \rightarrow z = (0.1 + z + y) \times (0.39 + z[+x])$	M1	1.1b
	$[\sum p = 1]$ $0.06 + 0.3 + 0.39 + 0.1 + z + y[+x] = 1 \rightarrow \quad [z + y[+x] = 0.15]$	M1	1.1b
	Solving (simultaneously) leading to $\underline{z = 0.13} \quad \underline{y = 0.02}$	A1	1.1b
			(5 marks)

June 2018 Question 2

2. A factory buys 10% of its components from supplier A, 30% from supplier B and the rest from supplier C. It is known that 6% of the components it buys are faulty.

Of the components bought from supplier A, 9% are faulty and of the components bought from supplier B, 3% are faulty.

(a) Find the percentage of components bought from supplier C that are faulty.

(3)

A component is selected at random.

(b) Explain why the event "the component was bought from supplier B" is not statistically independent from the event "the component is faulty".

(1)

ANSWER

Qu	Scheme	Marks	AO
2 (a)	[Let $p = \text{P}(F \mid C)$]		
	Tree diagram or some other method to find an equation for p	M1	2.1
	$0.1\times0.09+0.3\times0.03+0.6\times p=0.06$	A1	1.1b
	$p=0.07$ i.e. **7%**	A1	1.1b
		(3)	
(b)	e.g. $\text{P}(B \text{ and } F) = 0.3\times0.03=0.009$ but		
	$\text{P}(B)\times\text{P}(F)=0.3\times0.06=0.018$	B1	2.4
	These are not equal so not independent		
		(1)	
		(4 marks)	

June 2018 Question 3 Paper 3 (A-Level)

3. In an experiment a group of children each repeatedly throw a dart at a target.
For each child, the random variable H represents the number of times the dart hits the target in the first 10 throws.

Peta models H as B(10, 0.1)

(a) State two assumptions Peta needs to make to use her model.

(2)

(b) Using Peta's model, find $P(H \geqslant 4)$

(1)

For each child the random variable F represents the number of the throw on which the dart first hits the target.

Using Peta's assumptions about this experiment,

(c) find $P(F = 5)$

(2)

Thomas assumes that in this experiment no child will need more than 10 throws for the dart to hit the target for the first time. He models $P(F = n)$ as

$$P(F = n) = 0.01 + (n - 1) \times \alpha$$

where α is a constant.

(d) Find the value of α

(4)

(e) Using Thomas' model, find $P(F = 5)$

(1)

(f) Explain how Peta's and Thomas' models differ in describing the probability that a dart hits the target in this experiment.

(1)

ANSWER

Qu 3	Scheme	Marks	AO
(a)	The probability of a dart hitting the target is constant (from child to child and for each throw by each child) (o.e.)	B1	1.2
	The throws of each of the darts are independent (o.e.)	B1 (2)	1.2
(b)	$[P(H \geqslant 4) = 1 - P(H \leqslant 3) = 1 - 0.9872 = 0.012795.. =]$ awrt **0.0128**	B1 (1)	1.1b
(c)	$P(F = 5) = 0.9^4 \times 0.1, = 0.06561$ = awrt **0.0656**	M1, A1 (2)	3.4 1.1b
(d)	n | 1 | 2 | … | 10 $P(F = n)$ | 0.01 | $0.01 + \alpha$ | … | $0.01+9\alpha$	M1	3.1b
	Sum of probs = 1 $\Rightarrow \frac{10}{2}[2\times0.01+9\alpha]=1$	M1A1	3.1a 1.1b
	[i.e. $5(0.02 + 9\alpha) = 1$ or $0.1 + 45\alpha = 1$] so α = **0.02**	A1 (4)	1.1b
(e)	$P(F = 5|$ Thomas' model) = **0.09**	B1ft (1)	3.4
(f)	Peta's model assumes the probability of hitting target is constant (o.e.) **and** Thomas' model assumes this probability increases with each attempt(o.e.)	B1 (1)	3.5a
		(11 marks)	

Ch. 6: Statistical Distributions

June 2024 Question 5

5. A biased 4-sided spinner has the numbers 6, 7, 8 and 10 on it.

The discrete random variable X represents the score when the spinner is spun once and has the following probability distribution,

x	6	7	8	10
$P(X = x)$	0.5	0.2	q	q

where q is a probability.

(a) Find the value of q

(1)

Karen spins the spinner repeatedly until she **either** gets a 7 **or** she has taken 4 spins.

(b) Show that the probability that Karen stops after taking her 3rd spin is 0.128

(2)

The random variable S represents the number of spins Karen takes.

(c) Find the probability distribution for S

(4)

The random variable N represents the number of times Karen gets a 7

(d) Find $P(S > N)$

(1)

ANSWER

<table>
<tr><th>Qu 5</th><th>Scheme</th><th>Marks</th><th>AO</th></tr>
<tr><td>(a)</td><td>$[2q = 0.3]$ $\left[q = \frac{1-(0.5+0.2)}{2}\right]$ $[q =]$ <u>0.15</u></td><td>B1
(1)</td><td>1.1b</td></tr>
<tr><td>(b)</td><td>Realising require sequence: $\overline{7},\overline{7},7$ may see $0.8\times0.8\times0.2$ o.e.
$=$ <u>0.128</u> *</td><td>M1
A1*
(2)</td><td>3.4
1.1b</td></tr>
<tr><td>(c)</td><td>Possible values for S are: 1, 2, 3 or 4 only
$[P(S=1)] = 0.2$ <u>and</u> $[P(S=2) = 0.8\times0.2 =] 0.16$
$P(S=4) = 0.8^3\times0.2+0.8^4$ $[= 0.512]$
<u>or</u> $1-['P(S=1)'+'P(S=2)'+0.128]$
<table><tr><td>s</td><td>1</td><td>2</td><td>3</td><td>4</td></tr><tr><td>$P(S=s)$</td><td>0.2 $=\frac{1}{5}$</td><td>0.16 $=\frac{4}{25}$</td><td>0.128 $=\frac{16}{125}$</td><td>0.512 $=\frac{64}{125}$</td></tr></table></td><td>B1
M1

M1

A1
(4)</td><td>3.3
3.4

3.4

1.1b</td></tr>
<tr><td>(d)</td><td>$[= 1 - P(S = 1 = N) = 1 - 0.2\ ] =$ <u>0.8</u></td><td>B1
(1)</td><td>1.1b</td></tr>
<tr><td></td><td></td><td>(8 marks)</td><td></td></tr>
</table>

June 2023 Question 5

5. Julia selects 3 letters at random, one at a time without replacement, from the word

VARIANCE

The discrete random variable X represents the number of times she selects a letter A.

(a) Find the complete probability distribution of X.

(5)

Yuki selects 10 letters at random, one at a time **with** replacement, from the word

DEVIATION

(b) Find the probability that he selects the letter E at least 4 times.

(3)

ANSWER

Que.	Scheme	Marks	AOs
5(a)	$X = 0, 1, 2$ only	B1	3.1b
	$[\mathrm{P}(X=0)=]\frac{6}{8}\times\frac{5}{7}\times\frac{4}{6}$	M1	1.1b
	$[\mathrm{P}(X=1)=]3\times\frac{2}{8}\times\frac{6}{7}\times\frac{5}{6}$ or $[\mathrm{P}(X=2)=]3\times\frac{2}{8}\times\frac{1}{7}\times\frac{6}{6}$	M1	2.1
	<table><tr><td>x</td><td>0</td><td>1</td><td>2</td></tr><tr><td>$\mathrm{P}(X=x)$</td><td>$\frac{5}{14}$</td><td>$\frac{15}{28}$</td><td>$\frac{3}{28}$</td></tr></table>	A1 A1	1.1b 1.1b
		(5)	
(b)	$J \sim \mathrm{B}(10, \frac{1}{9})$	M1	3.1b
	$\mathrm{P}(J\geqslant 4)=1-\mathrm{P}(J\leqslant 3)$ or $\mathrm{P}(J\geqslant 4)=\mathrm{P}(J=4)+\mathrm{P}(J=5)+\ldots+\mathrm{P}(J=10)$ or $1-0.981(57\ldots)$	M1	3.4
	$=$ awrt 0.0184	A1	1.1b
		(3)	
			(8 marks)

June 2022 Question 3 (A-Level)

3. Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set.

(1)

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \qquad \sum r = 174.9 \qquad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

(3)

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.
The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2)

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

ANSWER

Question	Scheme	Marks	AOs
3(a)	tr	B1	1.2
		(1)	
(b)(i)	$\mu = \frac{174.9}{31} = 5.6419\ldots$ awrt 5.64	B1	1.1b
(ii)	$\sigma_r = \sqrt{\frac{3523.283}{31} - \mu^2}$	M1	1.1b
	$= 9.04559\ldots$ awrt 9.05	A1	1.1b
		(3)	
(c)	Leuchars is in the North and Camborne is in the South	M1	2.4
	The mean is smaller for Leuchars than Camborne therefore there is no evidence that Dian's belief is true	A1ft	2.2b
		(2)	
(d)	eg $p = 0.27$ is unlikely to be constant.	B1	2.4
		(1)	
			(7 marks)

June 2022 Question 5

5. Manon has two biased spinners, one red and one green.

The random variable R represents the score when the red spinner is spun.
The random variable G represents the score when the green spinner is spun.

The probability distributions for R and G are given below.

r	2	3
$P(R = r)$	$\frac{1}{4}$	$\frac{3}{4}$

g	1	4
$P(G = g)$	$\frac{2}{3}$	$\frac{1}{3}$

Manon spins each spinner once and adds the two scores.

(a) Find the probability that

(i) the sum of the two scores is 7

(ii) the sum of the two scores is less than 4

(3)

The random variable $X = mR + nG$ where m and n are integers.

$$P(X = 20) = \frac{1}{6} \quad \text{and} \quad P(X = 50) = \frac{1}{4}$$

(b) Find the value of m and the value of n

(5)

ANSWER

Qu	Scheme	Mark	AO
5. (a)(i)	Require $R = 3$ and $G = 4$ so probability is $\frac{3}{4} \times \frac{1}{3}$	M1	2.1
	$= \underline{\frac{1}{4}}$ or $\underline{\mathbf{0.25}}$	A1	1.1b
(ii)	[R must be 2 and $G = 1$ so $\frac{1}{4} \times \frac{2}{3}$] $= \underline{\frac{1}{6}}$	A1	1.1b
		(3)	
(b)	P($X = 50$) = 0.25 must mean $R = 3$ and $G = 4$	M1	3.1a
	so $3m + 4n = 50$	A1	1.1b
	P($X = 20$) = $\frac{1}{6} \Rightarrow R = 2, G = 1$ so $2m + n = 20$	A1	2.1
	Solving: $3m + 4(20 - 2m) = 50$ (o.e.)	M1	1.1b
	$\underline{\boldsymbol{m = 6}}$ and $\underline{\boldsymbol{n = 8}}$	A1	3.2a
		(5)	
		(8 marks)	

November 2021 question 4

Included in chapter 7 as it has hypothesis testing.

October 2020 question 5

Same with this question, so you will find it in chapter 7.

June 2019 question 3

3. A fair 5-sided spinner has sides numbered 1, 2, 3, 4 and 5

The spinner is spun once and the score of the side it lands on is recorded.

(a) Write down the name of the distribution that can be used to model the score of the side it lands on.

(1)

The spinner is spun 28 times.

The random variable X represents the number of times the spinner lands on 2

(b) (i) Find the probability that the spinner lands on 2 at least 7 times.

(ii) Find $P(4 \leqslant X < 8)$

(5)

ANSWER

Question	Scheme	Marks	AOs
3(a)	(Discrete) uniform (distribution)	B1	1.2
		(1)	
(b)	B(28, 0.2)	B1	3.3
(i)	$P(X \geq 7) = 1 - P(X \leq 6)$ [= 1 – 0.6784…]	M1	3.4
	awrt **0.322**	A1	1.1b
(ii)	$P(4 \leq X < 8) = P(X \leq 7) - P(X \leq 3)$ [= 0.818… – 0.160…]	M1	3.1b
	awrt **0.658**	A1	1.1b
		(5)	
			(6 marks)

June 2019 question 4 (A level, not AS)

4. Magali is studying the mean total cloud cover, in oktas, for Leuchars in 1987 using data from the large data set. The daily mean total cloud cover for all 184 days from the large data set is summarised in the table below.

Daily mean total cloud cover (oktas)	0	1	2	3	4	5	6	7	8
Frequency (number of days)	0	1	4	7	10	30	52	52	28

One of the 184 days is selected at random.

(a) Find the probability that it has a daily mean total cloud cover of 6 or greater.

(1)

Magali is investigating whether the daily mean total cloud cover can be modelled using a binomial distribution.

She uses the random variable X to denote the daily mean total cloud cover and believes that $X \sim B(8, 0.76)$

Using Magali's model,

(b) (i) find $P(X \geqslant 6)$

(2)

(ii) find, to 1 decimal place, the expected number of days in a sample of 184 days with a daily mean total cloud cover of 7

(2)

(c) Explain whether or not your answers to part (b) support the use of Magali's model.

(1)

There were 28 days that had a daily mean total cloud cover of 8
For these 28 days the daily mean total cloud cover for the **following** day is shown in the table below.

Daily mean total cloud cover (oktas)	0	1	2	3	4	5	6	7	8
Frequency (number of days)	0	0	1	1	2	1	5	9	9

(d) Find the proportion of these days when the daily mean total cloud cover was 6 or greater.

(1)

(e) Comment on Magali's model in light of your answer to part (d).

(2)

ANSWER

Question	Scheme	Marks	AOs
4 (a)	$\frac{132}{184} = 0.71739...$ awrt **0.717**	B1	1.1b
		(1)	
(b)(i)	$P(X \geq 6) = 1 - P(X \leq 5)$ or $P([X =]6) + P([X =]7) + P([X =]8)$	M1	3.4
	$= 1 - 0.296722...$ awrt **0.703**	A1	1.1b
		(2)	
(b)(ii)	$184 \times P(X = 7)$ $[= 184 \times 0.2811\ldots]$	M1	1.1b
	$= 51.7385...$ awrt **51.7**	A1	1.1b
		(2)	
(c)	Part (a) and part (b)(i) are similar **and** the expected number of 7s (51.7 or 0.281) matches with the number of 7s found in the data set (52 or 0.283) so Magali's model is supported.	B1ft	3.5a
		(1)	
(d)	$\frac{23}{28} = 0.82142...$ awrt **0.821**	B1	1.1b
		(1)	
(e)	Any one of… • Part (d)/'0.821' differs from part (a)/(b)(i)/(0.7…) • there is a **greater/different probability of high cloud cover/more likely to have high cloud cover** if the previous day had high cloud cover • **independence(o.e.)** does not hold	B1	2.4
	…therefore Magali's (binomial) model may not be suitable.	dB1	3.5a
		(2)	
			(9 marks)

June 2018 Question 5

5. A biased spinner can only land on one of the numbers 1, 2, 3 or 4. The random variable X represents the number that the spinner lands on after a single spin and $P(X = r) = P(X = r + 2)$ for $r = 1, 2$

Given that $P(X = 2) = 0.35$

(a) find the complete probability distribution of X.

(2)

Ambroh spins the spinner 60 times.

(b) Find the probability that more than half of the spins land on the number 4
Give your answer to 3 significant figures.

(3)

The random variable $Y = \frac{12}{X}$

(c) Find $P(Y - X \leqslant 4)$

(3)

ANSWER

Qu	Scheme	Marks	AO
5(a)	$P(X=4) = P(X=2)$ so $P(X=4) = 0.35$ $P(X=1) = P(X=3)$ and $P(X=1) + P(X=3) = 1 - 0.7$ So x: 1, 2, 3, 4 $P(X=x)$: 0.15, 0.35, 0.15, [0.35]	M1 A1 (2)	2.1 1.1b
(b)	Let A = number of spins that land on 4 $A \sim B(60, \text{“}0.35\text{”})$	B1ft	3.3
	$[P(A > 30) =]$ $1 - P(A \leqslant 30)$	M1	3.4
	$= 1 - 0.99411\ldots$ = **awrt 0.00589**	A1 (3)	1.1b
(c)	$Y - X \leqslant 4 \Rightarrow \frac{12}{X} - X \leqslant 4$ or $12 - X^2 \leqslant 4X$ (since $X > 0$) o.e.	M1	3.1a
	i.e. $0 \leqslant X^2 + 4X - 12 \Rightarrow 0 \leqslant (X+6)(X-2)$ so $X \geqslant 2$	M1	1.1b
	$P(Y - X \leqslant 4) = P(X \geqslant 2) = 0.35 + 0.15 + 0.35$ = **0.85**	A1 (3)	3.2a
		(8 marks)	

Ch. 7: Hypothesis Testing

June 2024 Question 4 (A-level)

4. The proportion of left-handed adults in a country is 10%
Freya believes that the proportion of left-handed adults under the age of 25 in this country is different from 10%
She takes a random sample of 40 adults under the age of 25 from this country to investigate her belief.

(a) Find the critical region for a suitable test to assess Freya's belief.
You should
- state your hypotheses clearly
- use a 5% level of significance
- state the probability of rejection in each tail

(4)

(b) Write down the actual significance level of your test in part (a)

(1)

In Freya's sample 7 adults were left-handed.

(c) With reference to your answer in part (a) comment on Freya's belief.

(1)

ANSWER

Qu 4	Scheme	Marks	AO
(a)	$H_0 : p = 0.1$ $H_1 : p \neq 0.1$ [Allow 10% for 0.1]	B1	2.5
	$[X \sim B(40, 0.1)] \Rightarrow P(X = 0) = 0.0148$ [Allow any letter for X]	M1	3.3
	$P(X \geqslant 9) = 1 - P(X \leqslant 8) = 1 - 0.9845 = 0.0155$	A1	1.1b
	Critical region is $\{X = 0\} \cup \{X \geqslant 9\}$ (o.e.)	A1	1.1b
		(4)	
(b)	["0.0148" + "0.0155"] = 0.0303	B1ft	1.1b
		(1)	
(c)	[Provided 7 is not in their CR] insufficient evidence to support Freya's belief	B1	2.2b
		(1)	
		(6 marks)	

June 2024 Question 4

4. The random variable $X \sim B(27, 0.35)$

(a) Find (i) $P(X = 10)$

(ii) $P(12 \leqslant X < 15)$

(3)

Historical records show that the proportion of defective items produced by a machine is 0.12

Following a maintenance service of the machine, a random sample of 60 items is taken and 3 defective items are found.

(b) Carry out a suitable test to determine whether the proportion of defective items produced by the machine has decreased following the maintenance service. You should state your hypotheses clearly and use a 5% level of significance.

(4)

(c) Write down the p-value for your test in part (b)

(1)

ANSWER

Qu 4	Scheme	Marks	AO
(a)	(i) 0.153588… awrt **0.154**	B1	3.4
	(ii) $P(X \leqslant 14) - P(X \leqslant 11)$ with at least 1 from $P(X \leqslant 14) = 0.97707\ldots$; $P(X \leqslant 11) = 0.797603\ldots$	M1	2.1
	$= 0.17947\ldots$ awrt **0.179**	A1	1.1b
		(3)	
(b)	$H_0 : p = 0.12$ $H_1 : p < 0.12$	B1	2.5
	[D = no. of defective items in sample] $D \sim B(60, 0.12)$	M1	2.1
	$[P(D \leqslant 3)] = 0.06013\ldots$ awrt **0.060**	A1	1.1b
	or $[P(D \leqslant 2)]$ = awrt 0.0196 with reference to CR [so CR: $D \leqslant 2$] [0.06… > 5% not significant, do not reject H_0]		
	Insufficient evidence that proportion of defective items has decreased	A1	2.2b
		(4)	
(c)	"**0.06…**"	B1ft	1.2
		(1)	
		(8 marks)	

June 2023 Question 4

4. Past information shows that 25% of adults in a large population have a particular allergy.

Rylan believes that the proportion that has the allergy differs from 25%

He takes a random sample of 50 adults from the population.

Rylan carries out a test of the null hypothesis H_0: $p = 0.25$ using a 5% level of significance.

(a) Write down the alternative hypothesis for Rylan's test.

(1)

(b) Find the critical region for this test.
You should state the probability associated with each tail, which should be as close to 2.5% as possible.

(4)

(c) State the actual probability of incorrectly rejecting H_0 for this test.

(1)

Rylan finds that 10 of the adults in his sample have the allergy.

(d) State the conclusion of Rylan's hypothesis test.

(1)

ANSWER

Que.	Scheme	Marks	AOs
4(a)	$[H_1 :] p \neq 0.25$	B1	2.5
		(1)	
(b)	$X \sim B(50, 0.25)$	B1	3.3
	$[P(X \leqslant 6) =]0.0194$ <u>or</u> $[P(X \leqslant 18) =]0.9713$ <u>or</u> $[P(X \geqslant 19) =]0.0287$ <u>or</u> $X \leqslant 6$ <u>or</u> $X \geqslant 19$	M1	3.4
	$[P(X \leqslant 6) =]$awrt 0.0194 and $[P(X \geqslant 19) =]$awrt 0.0287	A1	1.1b
	CR: $X \leqslant 6$ or $X \geqslant 19$	A1	1.1b
		(4)	
(c)	[0.0194 + 0.0287 =] awrt 0.048	B1ft	1.1b
		(1)	
(d)	(Do not reject H_0,) there is insufficient evidence to suggest that the **proportion** of those with the **allergy** differs from 25%/**Rylan's belief** not supported	B1	2.2b
		(1)	
			(7 marks)

June 2023 Question 2 (A-Level)

2. A machine fills packets with sweets and $\frac{1}{7}$ of the packets also contain a prize.

The packets of sweets are placed in boxes before being delivered to shops.
There are 40 packets of sweets in each box.

The random variable T represents the number of packets of sweets that contain a prize in each box.

(a) State a condition needed for T to be modelled by $\mathrm{B}(40, \frac{1}{7})$

(1)

A box is selected at random.

(b) Using $T \sim \mathrm{B}(40, \frac{1}{7})$ find

(i) the probability that the box has exactly 6 packets containing a prize,

(ii) the probability that the box has fewer than 3 packets containing a prize.

(2)

Kamil's sweet shop buys 5 boxes of these sweets.

(c) Find the probability that exactly 2 of these 5 boxes have fewer than 3 packets containing a prize.

(2)

Kamil claims that the proportion of packets containing a prize is less than $\frac{1}{7}$

A random sample of 110 packets is taken and 9 packets contain a prize.

(d) Use a suitable test to assess Kamil's claim.
You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

ANSWER

Qu 2	Scheme	Marks	AO
(a)	Comment in context about either **independence** or **random** packing e.g. "prizes must be placed in packets at random/independently of each other" **or** about **constant probability** e.g. "the probability of a packet containing a prize is constant/ the same/fixed"	B1 **(1)**	3.5b
(b)(i)	[P($T = 6$) =] 0.17273… awrt **0.173**	B1	1.1b
(ii)	[P($T < 3$) = P(T ,, 2) =] 0.061587… awrt **0.0616**	B1 **(2)**	1.1b
(c)	[K= no. of boxes with fewer than 3 packets containing a prize] $K \sim$ B(5, "0.0616") P($K = 2$) = 0.031344… in the range **[0.0313~0.0314]**	M1 A1 **(2)**	1.1b 1.1b
(d)	$H_0 : p = \frac{1}{7}$ $H_1 : p < \frac{1}{7}$ [X = no of packets containing a prize] X~B(110, $\frac{1}{7}$) [P(X ,, 9)] = 0.038292… [Significant result or reject H_0] E.g. there is evidence to support Kamil's claim	B1 M1 A1 A1 **(4)**	2.5 3.3 3.4 2.2b
		(9 marks)	

June 2022 Question 4 (A-Level)

4. A dentist knows from past records that 10% of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

A random sample of 50 of the dentist's customers is taken.

(a) Write down

- a null hypothesis corresponding to no change in the proportion of customers who arrive late
- an alternative hypothesis corresponding to the manager's belief

(1)

(b) Using a 5% level of significance, find the critical region for a two-tailed test of the null hypothesis in (a)
You should state the probability of rejection in each tail, which should be less than 0.025

(3)

(c) Find the actual level of significance of the test based on your critical region from part (b)

(1)

The manager observes that 15 of the 50 customers arrived late for their appointment.

(d) With reference to part (b), comment on the manager's belief.

(1)

ANSWER

Question	Scheme	Marks	AOs
4(a)	$H_0: p = 0.1$ $H_1: p \neq 0.1$	B1	2.5
		(1)	
(b)	Use of $X \sim B(50, 0.1)$ implied by sight of one of awrt 0.0052 or awrt 0.9755 or awrt 0.0245	M1	3.4
	Critical regions $X = 0$ or $X \geqslant 10$	A1	1.1b
	$X = 0$ and $X \geqslant 10$ plus $P(X = 0) =$ awrt 0.0052 and $P(X \geqslant 10) =$ awrt 0.0245	A1	1.1b
	SC: Both CR correct with no probabilities and no distribution seen scores M0A1A0		
		(3)	
(c)	0.0297	B1ft	1.1b
		(1)	
(d)	15 is in the critical region therefore there is evidence to support the **manager**'s belief	B1ft	2.2b
		(1)	
			(6 marks)

Video solution:

https://youtu.be/ImAuJLYpWdQ

June 2022 Question 2

2. A manufacturer of sweets knows that 8% of the bags of sugar delivered from supplier A will be damp.
A random sample of 35 bags of sugar is taken from supplier A.

(a) Using a suitable model, find the probability that the number of bags of sugar that are damp is

(i) exactly 2

(ii) more than 3

(3)

Supplier B claims that when it supplies bags of sugar, the proportion of bags that are damp is less than 8%

The manufacturer takes a random sample of 70 bags of sugar from supplier B and finds that only 2 of the bags are damp.

(b) Carry out a suitable test to assess supplier B's claim.
You should state your hypotheses clearly and use a 10% level of significance.

(4)

ANSWER

Qu	Scheme	Mark	AO
2. (a)	[D = number of bags that are damp] $D \sim B(35, 0.08)$ NB $0.08 = \frac{2}{25}$	M1	3.3
(i)	$P(D = 2) = 0.2430497\ldots$ awrt **0.243**	A1	3.4
(ii)	$P(D > 3) = [1 - P(D \leq 3) = 1 - 0.69397...] = 0.30602\ldots$ awrt **0.306**	A1	1.1b
		(3)	
(b)	$H_0 : p = 0.08$ $H_1 : p < 0.08$	B1	2.5
	[$X \sim$] B(70, 0.08)	M1	2.1
	$[P(X \leq 2)] = 0.0739756\ldots$ awrt **0.074**	A1	1.1b
	[0.074 < 0.10 so significant, reject H_0 so…]		
	there is evidence to support supplier B's claim (o.e.)	A1	2.2b
		(4)	
		(7 marks)	

Video solution:
https://youtu.be/0UCgR4qAd9k

November 2021 question 4

4. A nursery has a sack containing a large number of coloured beads of which 14% are coloured red.

Aliya takes a random sample of 18 beads from the sack to make a bracelet.

(a) State a suitable binomial distribution to model the number of red beads in Aliya's bracelet.

(1)

(b) Use this binomial distribution to find the probability that

(i) Aliya has just 1 red bead in her bracelet,

(ii) there are at least 4 red beads in Aliya's bracelet.

(3)

(c) Comment on the suitability of a binomial distribution to model this situation.

(1)

After several children have used beads from the sack, the nursery teacher decides to test whether or not the proportion of red beads in the sack has changed.
She takes a random sample of 75 beads and finds 4 red beads.

(d) Stating your hypotheses clearly, use a 5% significance level to carry out a suitable test for the teacher.

(4)

(e) Find the p-value in this case.

(1)

ANSWER

Qu	Scheme	Marks	AO
4. (a)	[R = no. of red beads in Aliya's bracelet] $R \sim B(18, 0.14)$	B1 **(1)**	3.3
(b)(i)	$P(R = 1) = 0.19403\ldots$ awrt **0.194**	B1	1.1b
(ii)	$P(R \geqslant 4) = 1 - P(R \leqslant 3) = 1 - [0.76184\ldots]$	M1	3.4
	$= 0.2381588\ldots$ awrt **0.238**	A1 **(3)**	1.1b
(c)	Requires $p = 0.14$ to be constant so need a large number of beads in the sack to ensure that removing 18 beads does not appreciably affect this probability, then it could be suitable.	B1 **(1)**	3.5b
(d)	$H_0 : p = 0.14$ $H_1 : p \neq 0.14$	B1	2.5
	[X = number of red beads in the sample] $X \sim B(75, 0.14)$	M1	3.3
	$P(X \leqslant 4) = 0.01506\ldots$ or if B(75, 0.14) seen awrt 0.02	A1	3.4
	{0.02 < 0.025 so significant or reject H_0 } There is evidence that the proportion of red beads has changed	A1 **(4)**	2.2b
(e)	p-value is $2 \times$"0.01506..." = 0.030123... = awrt 0.03	B1ft **(1)**	1.1b
		(10 marks)	

October 2020 question 5

5. Afrika works in a call centre.

She assumes that calls are independent and knows, from past experience, that on each sales call that she makes there is a probability of $\frac{1}{6}$ that it is successful.

Afrika makes 9 sales calls.

(a) Calculate the probability that at least 3 of these sales calls will be successful.

(2)

The probability of Afrika making a successful sales call is the same each day.

Afrika makes 9 sales calls on each of 5 different days.

(b) Calculate the probability that at least 3 of the sales calls will be successful on exactly 1 of these days.

(2)

Rowan works in the same call centre as Afrika and believes he is a more successful salesperson.

To check Rowan's belief, Afrika monitors the next 35 sales calls Rowan makes and finds that 11 of the sales calls are successful.

(c) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence to support Rowan's belief.

(4)

ANSWER

Question	Scheme	Marks	AOs
5(a)	Let C = the number of successful calls. $C \sim B\left(9, \frac{1}{6}\right)$	M1	3.3
	$P(C \geq 3) = 1 - P(C \leq 2) = 0.1782\ldots$ **awrt 0.178**	A1	1.1b
		(2)	
(b)	Let X = the number of occasions when at least 3 calls are successful. $P(X = 1) = 5 \times ("0.1782...") \times ("0.8217...")^4$	M1	1.1b
	$= 0.4061\ldots$ **awrt 0.406**	A1	1.1b
		(2)	
(c)	$H_0 : p = \frac{1}{6}$ $H_1 : p > \frac{1}{6}$	B1	2.5
	Let R = the number of successful calls $R \sim B\left(35, \frac{1}{6}\right)$	M1	3.3
	$P(R \geq 11) = 1 - P(R \leq 10) = 0.02...$	A1	3.4
	There is sufficient evidence to support that **Rowan** has more successful sales calls than Afrika.	A1	2.2b
		(4)	
			(8 marks)

Video solution:

https://youtu.be/D5PhpBhxxaQ

June 2019 question 5

5. Past records show that 15% of customers at a shop buy chocolate. The shopkeeper believes that moving the chocolate closer to the till will increase the proportion of customers buying chocolate.

After moving the chocolate closer to the till, a random sample of 30 customers is taken and 8 of them are found to have bought chocolate.

Julie carries out a hypothesis test, at the 5% level of significance, to test the shopkeeper's belief.

Julie's hypothesis test is shown below.

$H_0 : p = 0.15$

$H_1 : p \geqslant 0.15$

Let X = the number of customers who buy chocolate.

$X \sim B(30, 0.15)$

$P(X = 8) = 0.0420$

$0.0420 < 0.05$ so reject H_0

There is sufficient evidence to suggest that the proportion of customers buying chocolate has increased.

(a) Identify the first two errors that Julie has made in her hypothesis test.

(2)

(b) Explain whether or not these errors will affect the conclusion of her hypothesis test. Give a reason for your answer.

(1)

(c) Find, using a 5% level of significance, the critical region for a one-tailed test of the shopkeeper's belief. The probability in the tail should be less than 0.05

(2)

(d) Find the actual level of significance of this test.

(1)

ANSWER

Question	Scheme	Marks	AOs
5(a)	The alternative hypothesis should be $H_1 : p > 0.15$	B1	2.5
	The calculation of the test statistic should be $P(X \geq 8)$ [= 0.0698]	B1	2.3
		(2)	
(b)	These will affect the conclusion (as the null hypothesis should not be rejected) since $P(X \geq 8)$ [= 0.0698] is greater than 0.05	B1	2.4
		(1)	
(c)	$P(X \leq 8) = 0.9722\ldots > 0.95$ or $P(X \geq 9) = 0.0277\ldots < 0.05$	M1	2.1
	CR: $\{X \geq 9\}$	A1	1.1b
		(2)	
(d)	awrt **0.0278**	B1ft	1.1b
		(1)	
			(6 marks)

June 2018 Question 3

3. Naasir is playing a game with two friends. The game is designed to be a game of chance so that the probability of Naasir winning each game is $\frac{1}{3}$

Naasir and his friends play the game 15 times.

(a) Find the probability that Naasir wins

(i) exactly 2 games,

(ii) more than 5 games.

(3)

Naasir claims he has a method to help him win more than $\frac{1}{3}$ of the games. To test this claim, the three of them played the game again 32 times and Naasir won 16 of these games.

(b) Stating your hypotheses clearly, test Naasir's claim at the 5% level of significance.

(4)

ANSWER

Qu	Scheme	Marks	AO
3 (a)	Let N = the number of games Naasir wins $N \sim B(15, \frac{1}{3})$	M1	3.3
(i)	$P(N = 2) = 0.059946\ldots$ **awrt 0.0599**	A1	1.1b
(ii)	$P(N > 5) = 1 - P(N \leqslant 5) = 0.38162\ldots$ **awrt 0.382**	A1	1.1b
		(3)	
(b)	$H_0 : p = \frac{1}{3}$ $H_1 : p > \frac{1}{3}$	B1	2.5
	Let X = the number of games Naasir wins $X \sim B(32, \frac{1}{3})$	M1	3.3
	$P(X \geqslant 16) = 1 - P(X \leqslant 15) = 0.03765$ (< 0.05)	A1	3.4
	[Significant result so reject H_0 (the null model) and conclude:] There is evidence to support Naasir's claim (o.e.)	A1	3.5a
		(4)	
		(7 marks)	