Pearson

Topics 2 and 3

Measures of Location and Spread and Representation of Data

Bronze, Silver, Gold

Worksheets for

AS Level Mathematics

Teacher Notes

These Bronze, Silver and Gold worksheets are designed to be used either straight after the content has been taught or as part of a skills gap analysis, especially as students move into year 13.

They are drawn from the latest specification questions and legacy questions. The papers are between 25 and 35 marks.

The topic number on this worksheet relates to the corresponding chapter number in the 'Pearson Edexcel AS and A Level Mathematics: Statistics and Mechanics Year 1/AS' textbook.

Quick Links

(Press Ctrl, as you click with your mouse to follow these links)

- Bronze Questions
- Bronze Mark Scheme
- Silver Questions
- Silver Mark Scheme
- Gold Questions
- Gold Mark Scheme

Extension and Enrichment

If you have students that have enjoyed the challenge of the Gold questions, then they should have a go at the more challenging question from our Advanced Extension Award (AEA) papers. The Mathematics AEA is a single, 3 hour non-calculator paper, taken at the end of year 13. It helps students to develop high level problem solving and proof skills. It is entirely based on the content of the A Level Mathematics Course. No extra material needs to be covered to take the AEA in Mathematics. A second important difference is that marks are awarded for the clarity and quality of their solution. Developing this key skill, alongside the extra problem-solving experience, can pay dividends in the way they approach A Level Mathematics and Further Mathematics problems.

More information about the Advanced Extension Award can be found here on the Pearson Edexcel Website, or here on the Maths Emporium

Bronze Questions

Calculator

The total mark for this section is 35

Q1

Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8 5.6 2.3 9.4 0.0 0.5 1.8

Jenny then records the amount of rainfall, x mm, at the school each day for the following 21 days. The results for the 21 days are summarised below.

$$\sum x = 84.6$$

(a) Calculate the mean amount of rainfall during the whole 28 days.

(2)

Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and the number 0.5 should have been 5.0
Keith corrects these figures.

(b) State, giving your reason, the effect this will have on the mean.

(2)

(Total for Question 1 is 4 marks)

Q2

Sara is investigating the variation in daily maximum gust, t kn, for Camborne in June and July 1987.

She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

(a) State the sampling technique Sara used.

(1)

(b) From your knowledge of the large data set, explain why this process may not generate a sample of size 20.

(1)

The data Sara collected are summarised as follows

$$n = 20 \qquad \sum t = 374 \qquad \sum t^2 = 7600$$

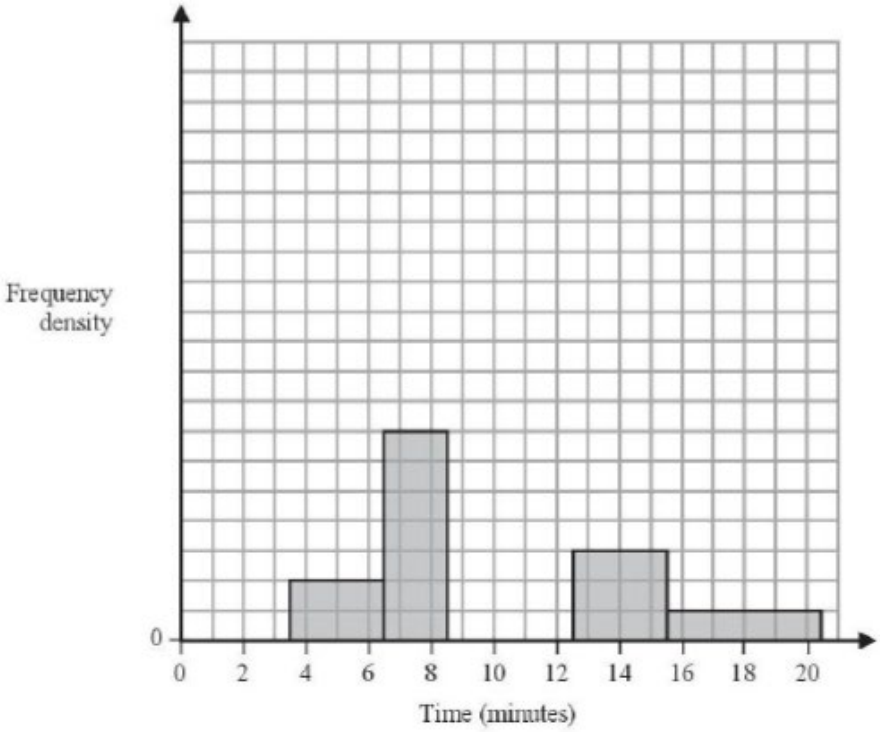
(c) Calculate the standard deviation.

(2)

(Total for Question 2 is 4 marks)

Q3

The partially completed histogram and the partially completed table show the time, to the nearest minute, that a random sample of motorists were delayed by roadworks on a stretch of motorway.



Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	17
10 – 12	45
13 – 15	9
16 – 20	

Estimate the percentage of these motorists who were delayed by the roadworks for between 8.5 and 13.5 minutes.

(Total for Question 3 is 5 marks)

Q4

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

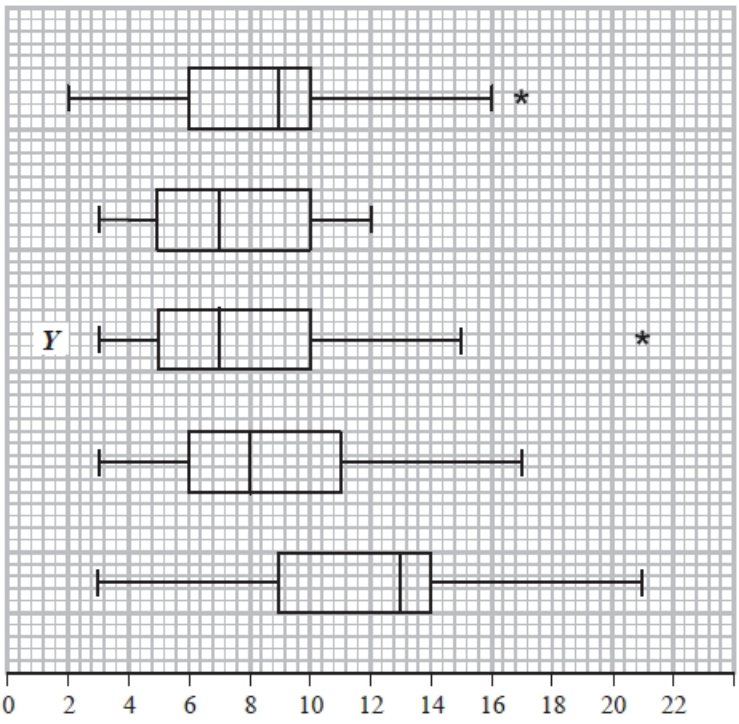
(2)

The data for these months are summarised in the box plots on the next 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*.

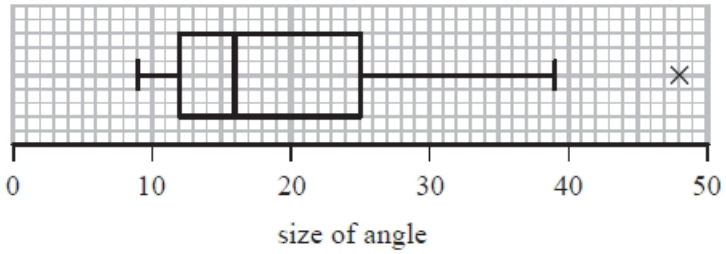
(3)



(Total for Question 4 is 8 marks)

Q5

Each of 60 students was asked to draw a 20° angle without using a protractor. The size of each angle drawn was measured. The results are summarised in the box plot below.



(a) Find the range for these data.

(1)

(b) Find the interquartile range for these data.

(1)

The students were then asked to draw a 70° angle.
The results are summarised in the table below.

Angle, a, (degrees)	Number of students
$55 \leqslant a < 60$	6
$60 \leqslant a < 65$	15
$65 \leqslant a < 70$	13
$70 \leqslant a < 75$	11
$75 \leqslant a < 80$	8
$80 \leqslant a < 85$	7

(c) Use linear interpolation to estimate the size of the median angle drawn. Give your answer to 1 decimal place.

(2)

(d) Show that the lower quartile is 63°

(2)

For these data, the upper quartile is 75°, the minimum is 55° and the maximum is 84°

An outlier is an observation that falls either
more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

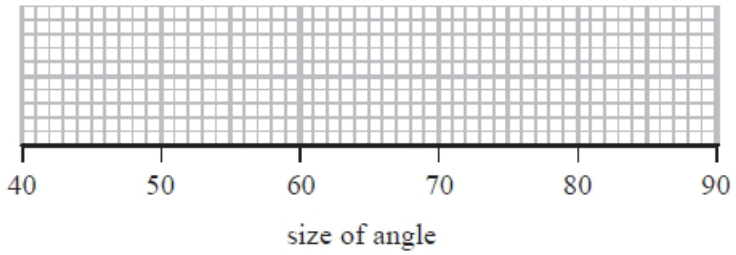
(e) (i) Show that there are no outliers for these data.

(ii) Draw a box plot for these data on the grid on the following page.

(5)

(f) State which angle the students were more accurate at drawing. Give reasons for your answer.

(3)



(Total for Question 5 is 14 marks)

Bronze Mark Scheme

Q1.

Question Number	Scheme	Marks
(a)	$2.8 + 5.6 + 2.3 + 9.4 + 0.5 + 1.8 + 84.6 = 107$ mean = 107 / 28 (= 3.821…) **(awrt 3.8)**	M1 A1 (2)
(b)	It will have no effect since one is 4.5 under what it should be and the other is 4.5 above what it should be.	B1 dB1 (2) **[4]**
	Notes	
(a)	M1 for a clear attempt to add the two sums. Accept a full expression or $2.8 + 5.6 + \ldots + 84.6 = x$ where $100<x<110$ i.e. seeing at least two correct terms of Keith's and the 84.6 with a slip. A1 for awrt 3.8 (Condone 1 dp/2sf here since data is given to 1 dp or 2 sf) Accept $\frac{107}{28}$ or $3\frac{23}{28}$ or any exact equivalent **Correct answer implies M1A1**	
(b)	1st B1 for clearly stating that it will have no effect. ("roughly the same" is B0 B0) 2nd dB1 for a supporting reason that mentions the fact that the increase and decrease are the same and gives some numerical value(s) to support this. e.g. Sum of Keith's observations is still 22.4 (or mean is still 3.2) or Sum is still 107 or $9.4 - 4.9 = 5 - 0.5$ (o.e.) This second B1 is dependent on their saying there is no effect so B0B1 is not possible.	

Q2.

Question	Scheme	Marks	AOs
(a)	Systematic (sample)	B1cao	1.2
(b)	In LDS some days have gaps because the data was not recorded	B1	2.4
(c)	$\left[\bar{t} = \frac{374}{20} = 18.7\right]$ $\sigma_t = \sqrt{\frac{7600}{20} - \bar{t}^2}$ $[=\sqrt{30.31}]$	M1	1.1a
	$= 5.5054\ldots$ awrt **5.51** (Accept use of $s_t = \sqrt{\frac{7600 - 20\bar{t}^2}{19}} = 5.6484\ldots$)	A1	1.1b
			(4 marks)

Part	Notes
(b)	B1 a correct explanation
(c)	M1 for a correct expression for $\bar{t}$ and σ_t or s_t. Ft an incorrect evaluation of $\bar{t}$
	A1 for σ_t = awrt 5.51 or s_t = awrt 5.65

Q3.

Question	Scheme	Marks	AOs
	$17 + 45 + \frac{1}{3}\times 9$ $[= 65]$	M1	2.2a
	$(7 - 8)$ **14** or $(16 - 20)$ **5** [Values may be seen in the table]	M1 A1	3.1a 1.1b
	Percentage of motorists is $\frac{"65"}{6+"14"+17+45+9+"5"} \times 100$	M1	3.1b
	$=$ **67.7%**	A1	1.1b
			(5 marks)

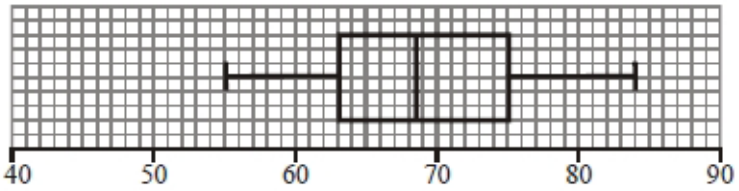
Part	Notes
	1st M1 for a fully correct expression for the number of motorists in the interval
	2nd M1 for clear use of frequency density in (4-6) or (13-15) cases to establish the fd scale. Then use of area to find frequency in one of the missing cases.
	1st A1 for both correct values seen
	3rd M1 for realising that total is required and attempting a correct expression for %
	2nd A1 for awrt 67.7%

Q4.

Qu	Scheme	Marks	AO
(a)	$\bar{x} = 10.2\ (2222\ldots)$ awrt **10.2**	B1 (1)	1.1b
(b)	$\sigma_x = 3.17\ (20227\ldots)$ awrt **3.17** Sight of "knots" or "kn" (condone knots/s etc)	B1ft B1 (2)	1.1b 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)	

	Notes
NB	$\bar{x} = \frac{184}{18}$ and $\sigma_x = \sqrt{\frac{2062}{18} - \bar{x}^2}$
(a)	B1 for $\bar{x} = 10.2$ (allow exact fraction)
(b)	1st B1ft allow 3.2 from a correct expr' accept $s = 3.26(3984\ldots)$ [ft use of n/a] Treating n/a as 0 May see $n = 31$ or $\bar{x} = 5.9354\ldots$ which is B0 in (a) but here in (b) it gives $\sigma_x = 5.59(34\ldots)$ or $s = 5.6858\ldots$(awrt 5.69) and scores 1st B1 2nd B1 accept kn accept in (a) or (b) (allow nautical miles/hour)
(c)	1st B1 choosing October but accept September. 2nd B1 for stating that (Camborne) is windier in autumn/winter months "because it is winter/autumn/windier/colder in "month" " Sep $\leqslant$ "month" $\leqslant$ Mar scores B1B1 for "month" = Sep or Oct and B0B1 for other months in range
(d)(i)	B1 for outlier or the idea of an extreme value allow "anomaly"
(ii)	M1 for a comment relating to location that mentions both median and mean and a comment relating to spread that mentions both range/IQR and standard deviation and leads to choosing B, C or D **Choosing A or E is M0** Incorrect/false statements score M0 e.g. Q_3 = (mean + σ) or identify Q_2 = mean or Y has small spread
ALT	**Use of outliers:** outlier is (mean + 3σ) ($B = 19.9$), ($C = 18.95$), ($D = 20.2$) Must see at least one of these values and compare to Y's outlier[leads to D or B]
	A1 for suitable inference i.e. B (accept D or B or D) M1 must be scored

Q5.

Question	Scheme	Marks
(a)	[Range = 48 − 9] = **39**	B1 (1)
(b)	[IQR = 25 − 12]= **13**	B1 (1)
(c)	Median = $65+\frac{[9]}{13}\times 5 = \frac{890}{13}$ = awrt **68.5**° $\left[\text{Condone: } 65+\frac{[9.5]}{13}\times 5 = 68.7\right]$	M1 A1 (2)
(d)	Lower Quartile = $60+\frac{9}{15}\times 5 =$ **63** (*)	M1 A1cso (2)
(e)(i)	$63-1.5\times(75-63)=45$ $75+1.5\times(75-63)=93$ No data above 93 and no data below 45 or 55>45 etc or there are no outliers.	M1A1 A1
(ii)		M1 A1ft (5)
(f)	Median for the 70°angle is closer (to 70°)[than the 20° median is to 20°] The range/IQR for the 70° angle box plot is smaller/shorter Therefore, students were more accurate at drawing the 70° angle.	B1 B1 dB1 (3)
		(14 marks)

	Notes
(c)	M1 for an attempt (should have 65 or 70, 13 and 5)NB working down: $70-\frac{[4]}{13}\times 5$ Allow any correct method leading to $\frac{890}{13}$, the "5" may be implied by 65 and 70 seen A1 awrt 68.5 (condone 68.7 if (n+1) is used). Ans only of 68.5 is 2/2 but 68.7 needs M1
(d)	M1 for correct expression for the lower quartile (condone 9.25 if (n+1) used) Watch out for working down e.g. $65-\frac{6}{15}\times 5$ (M1) but e.g. $\frac{60+65}{2}=62.5=63$ is M0 A1 for correct solution with no incorrect working seen (condone (n+1) giving 63.08..)
(e)(i)	M1 for either correct calculation (may be implied by one correct limit) A1 for either 45 or 93 A1 for 45 and 93 and conclusion
(ii)	M1 for a box with 1 whisker drawn on each side (must see the line drawn) A1ft their median $63 < Q_2 < 75$ but quartiles (63 and 75), 55 and 84 must be correct.
Accuracy	Use 0.5 sq. accuracy so condone median on 68 or 69 if 68.5 seen
(f)	1st B1 for correct comparison of their **medians** (63 < (c) < 75) to true value 2nd B1 for correct comparison of their **range** or **IQR** ("spread" is B0) Allow saying IQRs of 12 and 13 are similar. Ignore mention of "skewness" or "outliers" 3rd dB1 dependent upon at least one previous B1 being scored for choosing 70°

Silver Questions

Calculator

The total mark for this section is 30

Q1

Sara was studying the relationship between rainfall, r mm, and humidity, h %, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara examined the rainfall figures and found

$$Q_1 = 0.1 \quad Q_2 = 0.9 \quad Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above Q_3 is called an outlier.

(a) Show that $r = 20.6$ is an outlier.

(1)

(b) Give a reason why Sara might

(i) include
(ii) exclude

this day's reading.

(2)

(Total for Question 1 is 3 marks)

Q2

The histogram in Figure 1 shows the time taken, to the nearest minute, for 140 runners to complete a fun run.

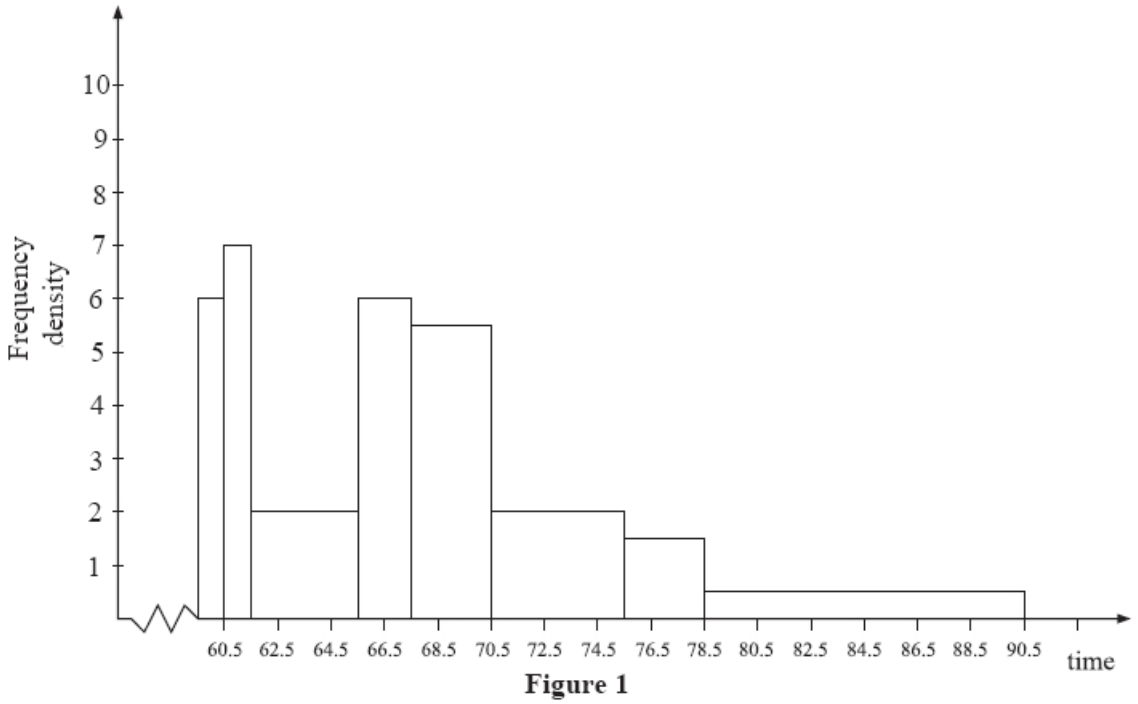


Figure 1

Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(Total for Question 2 is 5 marks)

Q3

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)

(Total for Question 3 is 8 marks)

Q4

The marks of a group of female students in a statistics test are summarised in Figure 1

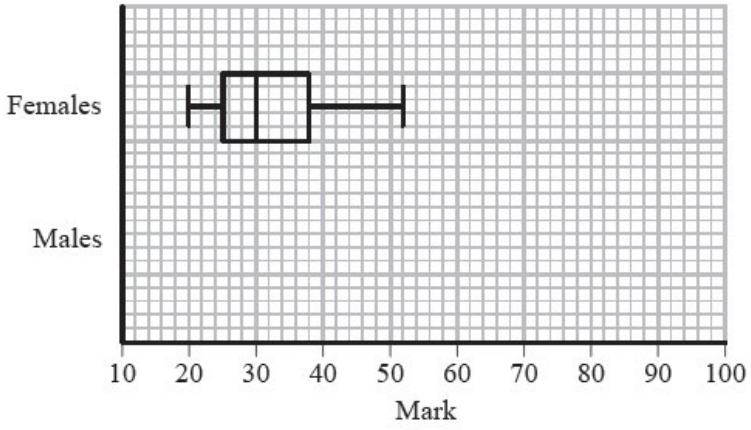


Figure 1

(a) Write down the mark which is exceeded by 75% of the female students.

(1)

The marks of a group of male students in the same statistics test are summarised by the stem and leaf diagram below.

Mark	(2|6 means 26)	Totals
1	4	(1)
2	6	(1)
3	4 4 7	(3)
4	0 6 6 7 7 8	(6)
5	0 0 1 1 1 3 6 7 7	(9)
6	2 2 3 3 3 8	(6)
7	0 0 8	(3)
8	5	(1)
9	0	(1)

(b) Find the median and interquartile range of the marks of the male students.

(3)

An outlier is a mark that is either more than 1.5 × interquartile range above the upper quartile or more than 1.5 × interquartile range below the lower quartile.

(c) In the space provided on Figure 1 draw a box plot to represent the marks of the male students, indicating clearly any outliers.

(5)

(d) Compare and contrast the marks of the male and the female students.

(2)

(Total for Question 4 is 11 marks)

Q5

The variable x was measured to the nearest whole number. Forty observations are given in the table below.

x	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 - 15 class has a width of 2 cm and a height of 5 cm. For the 16 - 18 class find

(a) the width,

(1)

(b) the height of the bar representing this class.

(2)

(Total for Question 5 is 3 marks)

Silver Mark Scheme

Q1

Question	Scheme	Marks	AOs
(a)	IQR = 2.3 and $20.6 \gg 2.4 + 1.5\times 2.3$ (= 5.85) (Compare correct values)	B1	1.1b
		(1)	
(b)(i)	e.g. it is a piece of data and we should consider all the data (o.e.)	B1	2.4
(ii)	e.g. it is an extreme value and could unduly influence the analysis or it could be a mistake	B1	2.4
		(2)	
		(3 marks)	

Part	Notes
(a)	B1 for sight of the correct calculation and suitable comparison with 20.6
(b)(i)	B1 for a suitable reason for including the data point
(ii)	B1 for a suitable reason for excluding the data point

Q2

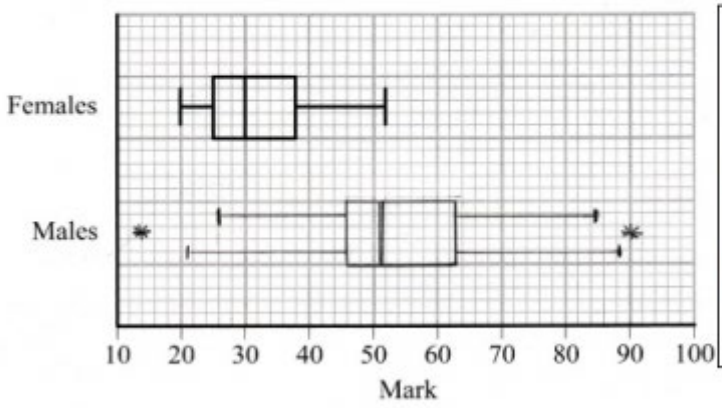
Question Number	Scheme	Marks
	<table><tr><td>Width</td><td>1</td><td>1</td><td>4</td><td>2</td><td>3</td><td>5</td><td>3</td><td>12</td></tr><tr><td>Freq. Density</td><td>6</td><td>7</td><td>2</td><td>6</td><td>5.5</td><td>2</td><td>1.5</td><td>0.5</td></tr></table> 0.5×12 or 6 Total area is $(1\times6)+(1\times7)+(4\times2)+....,=70$ $(90.5-78.5)\times\frac{1}{2}\times\frac{140}{\text{their } 70}$ "70 seen anywhere" Number of runners is 12	**M1** **A1** **M1** **B1** **A1** **(5)** **Total 5 marks**
	1st M1 for attempt at width of the correct bar (90.5 - 78.5) [Maybe on histogram or in table] 1st A1 for 0.5×12 or 6 (may be seen on the histogram. Must be related to the area of the bar above 78.5 - 90.5. 2nd M1 for attempting area of correct bar $\times\frac{140}{\text{their } 70}$ B1 for 70 seen anywhere in their working 2nd A1 for correct answer of 12. Minimum working required is $2\times0.5\times12$ where the 2 should come from $\frac{140}{70}$ Beware 90.5 - 78.5 = 12 (this scores M1A0M0B0A0) Common answer is $0.5\times12=6$ (this scores M1A1M0B0A0) If unsure send to review e.g. $2\times0.5\times12=12$ without 70 being seen	

Q3

Question	Scheme	Marks	AOs
(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)	$\sigma = \sqrt{\frac{7704.1875}{184} - \left(\frac{539.75}{184}\right)^2} = 5.7676\ldots$ σ = 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)

Notes	
(a)	**B1:** Identifying tr(ace) data Ignore comments about n/a, missing data, anomalies, etc.
(b)	**M1:** Correct fraction $\frac{7}{24}\times 4$ allow working down $[5]-\frac{155-138}{24}\times 4$ allow a correct equation leading to a correct fraction e.g. $\frac{x-1}{5-1}=\frac{138-131}{155-131}$ for M1 Use of $(n+1)$ with 138.75 allow $\frac{7.75}{24}\times 4$ **A1:** awrt 2.17 (condone $\frac{13}{6}$) awrt 2.29 from $(n+1)$ (condone $\frac{55}{24}$)
(c)	**M1:** Correct expression for standard deviation (allow mean = awrt 2.93) **A1:** awrt 5.77 correct answer only scores M1A1 (allow $s = 5.78$) **SC:** 5.76 with no working scores M1A0
(d)(i)	**B1:** Explaining that data assumed to be spread evenly across each class (o.e.) e.g. The midpoint of each class is the mean of each class or all the values in the class are located at the midpoint condone normally distributed within each class
Mark together (ii)&(iii)	**B1:** Demonstrating an understanding of the LDS that the majority of data values (in the first class) are at 0 or close to 0 (trace). **dB1:** (dependent upon 2nd B1) Correct inference based on knowledge of the LDS **SC:** If B1 is scored in (i) for 'The data are spread evenly across each class,' then in (ii) 'The data are not evenly distributed in the classes' scores B1 but in (iii) 'the actual mean is smaller' with no further justification scores B0

Q4

Question	Scheme	Marks
(a)	25 (allow any x where $24 < x < 26$)	B1 (1)
(b)	Q_2 (or median or m) = **51** **IQR** = 63 – 46 ,= **17** (or $Q_3 - Q_1 = 17$)	B1 M1, A1 (3)
(c)	Outliers given by $46 - 1.5 \times 17 = 20.5$ or $63 + 1.5 \times 17 = 88.5$ Outliers limits are **20.5** and **88.5**  Allow lower whisker to 20.5 and upper whisker to 88.5 Do **not** allow a mix of whiskers e.g 20.5 and 85 Do **not** allow both sets of whiskers	M1 A1 M1 A1ft B1 (5)
(d)	**Medians:** Median for females lower than males **IQR:** IQR for females smaller than males. Allow "lower/higher" but not "wider" **Range:** Range of females is less than males **Skewness:** Male and female marks are both positively skew Ignore other statements about average, spread, mean, st. Dev, variation, outliers etc	B1ft B1ft (2) **(11 marks)**

	Notes
	Mark (b) and (c) together BUT must see clear statement that median (or m or Q_2) = 51 and IQR = 17
(b)	M1 for 2 quartiles (at least one correct) and attempt to find the difference. Must see their 63 – their 46 A1 for 17 only. [Answer only of IQR= 17 scores M1A1]
(c)	**A fully correct box-plot (either version) with no supporting work scores 5/5. Otherwise:** 1st M1 for correct attempt to calc' at least one limit for outliers, ft their quartiles or IQR or award for sight of 20.5 or 88.5 1st A1 for identifying both limits of 20.5 and 88.5 2nd M1 for a box with an upper and a lower whisker(s) with at least 2 correct values (or correct ft) (condone no median marked) (condone 2 upper or 2 lower whiskers) 2nd A1ft for their 20.5 or 26 ,46, 51, 63 and 85 or their 88.5 in appropriate places and readable off their scale. Follow through their 20.5 and their 88.5 only, other values need to be correct If there are 2 upper or 2 lower whiskers A0 B1 for only 2 outliers appropriately marked at 14 and 90 Do not award if whiskers go beyond these values. **Apply ± 0.5 square accuracy for diagram** A box plot not on the graph paper can only score the 1st M1A1
(d)	**In (d) ft from their diagrams (if no diagram then use their values)** 1st B1ft for one correct comment comparing median, IQR , range or skewness 2nd B1ft for a second correct comment comparing median, IQR, range or skewness Do not allow contradictory statements

Q5

Question Number	Scheme	Marks
(a)	1(cm) cao	B1
(b)	10 cm^2 represents 15 10/15 cm^2 represents 1 or $1cm^2$ represents 1.5 Therefore frequency of 9 is $\frac{10}{15}\times 9$ or $\frac{9}{1.5}$ Require $x\frac{2}{3}$ or $\div 1.5$ height = 6(cm)	M1 A1 [3]
Notes	If 3(a) and 3(b) incorrect, but their (a) x their (b)=6 then award B0M1A0 3(b) Alternative method: f/cw=15/6=2.5 represented by 5 so factor x2 award M1 So f/cw=9/3=3 represented by 3x2=6. Award A1.	

Gold Questions

Calculator

The total mark for this section is 33

Q1

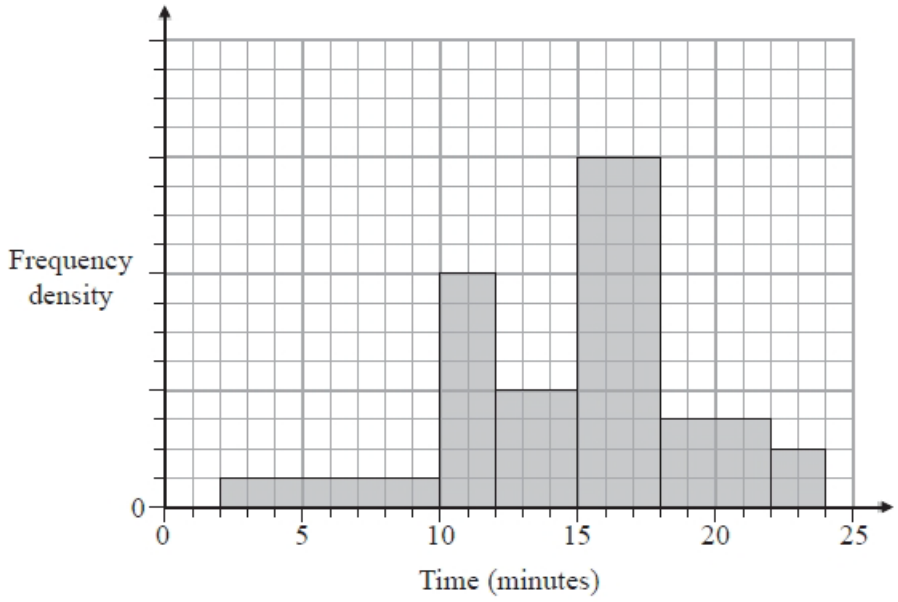


Figure 1

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.

(Total for Question 1 is 4 marks)

Q2

The mark, x, scored by each student who sat a statistics examination is coded using

$$y = 1.4x - 20$$

The coded marks have mean 60.8 and standard deviation 6.60

Find the mean and the standard deviation of x.

(4)

(Total for Question 2 is 4 marks)

Q3

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(a) Estimate the mean and standard deviation of these data.

(5)

(b) Use linear interpolation to estimate the value of the median.

(2)

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures.

(1)

(d) Estimate the interquartile range of this distribution.

(2)

(e) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data.

(1)

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(f) Without further calculations, explain the effect this would have on each of the estimates found in parts (a), (b), (c) and (d).

(3)

(Total for Question 3 is 14 marks)

Q4

A midwife records the weights, in kg, of a sample of 50 babies born at a hospital. Her results are given in the table below.

Weight (w kg)	Frequency (f)	Weight midpoint (x)
$0 \leqslant w < 2$	1	1
$2 \leqslant w < 3$	8	2.5
$3 \leqslant w < 3.5$	17	3.25
$3.5 \leqslant w < 4$	17	3.75
$4 \leqslant w < 5$	7	4.5

[You may use $\sum fx^2 = 611.375$]

A histogram has been drawn to represent these data.

The bar representing the weight $2 \leq w < 3$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing a weight of $3 \leq w < 3.5$

(3)

(b) Use linear interpolation to estimate the median weight of these babies.

(2)

(c) (i) Show that an estimate of the mean weight of these babies is 3.43 kg.

(ii) Find an estimate of the standard deviation of the weights of these babies.

(3)

A newborn baby weighing 3.43 kg is born at the hospital.

(f) Without carrying out any further calculations, state, giving a reason, what effect the addition of this newborn baby to the sample would have on your estimate of the

(i) mean,
(ii) standard deviation.

(3)

(Total for Question 4 is 11 marks)

Gold Mark Scheme

Q1.

Question	Scheme	Marks	AOs
	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

Notes		
	M1:	For clear use of frequency density to establish the fd scale and then use the area to find frequency of <11 minutes. Allow maximum of 3 errors in either the heights or widths in total if working shown. They may calculate the area using other size squares. Allow for realising they need to find the total number of squares (88) maximum of 4 errors in either the heights or widths and number < 11 minutes(16) - must have a maximum of 1 error in either the heights or widths (and not use the 78 as part of calulation)
	A1:	For correct values seen. Allow for 88 and 16
	M1:	For realising the need to find the total and calculating a percentage. (with "their 24" as the numerator). Allow $(8\times1+2\times8)\times"1.5"$ instead of $"24"+1\times8\times"1.5"$ If working shown can allow maximum of 2 errors in either the heights or widths in the calculation of the total. Allow "their 24" / 132 oe
	A1:	awrt 18

Q2.

Question Number	Scheme	Marks
	mean $= \frac{60.8+20}{1.4}$ or $60.8 = 1.4x - 20$ (o.e.)	M1
	$= 57.7142\ldots$ awrt 57.7	A1
	standard deviation $= \frac{6.60}{1.4}$ or $6.60 = 1.4x$	M1
	$= 4.7142\ldots$ awrt **4.71**	A1
		(4)
		Total 4
	Notes	
	1st M1 sub. 60.8 for y into a correct equation. Allow use of x or any other letter or expression for mean 1st A1 for awrt 57.7 or $\frac{404}{7}$ (o.e.). Correct answer only is 2/2 2nd M1 sub. 6.60 or 6.6 for y and ignoring the 20 Allow use of x or any other letter or expression for st. dev. $6.60^2 = 1.4^2x^2$ is M0 until we see them take a square root. 2nd A1 for awrt 4.71 or $\frac{33}{7}$ (o.e.). Correct answer only is 2/2	

Q3.

Question	Scheme	Marks
(a)	$\sum ft = 4837.5$ (allow 4838 or 4840)	B1
	$\text{Mean} = \frac{"4837.5"}{200} = 24.1875$ awrt **24.2** or $\frac{387}{16}$	M1 A1
	$\sigma = \sqrt{\frac{134281.25}{200} - \left(\frac{4837.5}{200}\right)^2}$	M1
	$= 9.293$ ..... (accept s =9.32) awrt **9.29**	A1
		(5)
(b)	$Q_2 = [20.5] + \frac{(100/100.5 - 62)}{88} \times 5 = 22.659...$ awrt **22.7**	M1 A1
		(2)
(c)	$Q_1 = 10.5 + \frac{(50/50.25)}{62} \times 10 [= 18.56]$ (*) (n + 1 gives 18.604…)	B1 cso
		(1)
(d)	Q_3= 25.5 (Use of n + 1 gives 25.734…)	B1
	IQR = 6.9 (Use of n + 1 gives 7.1)	B1 ft
		(2)
(e)	The data is skewed (condone "negative skew")	B1
		(1)
(f)	Mean decreases and st. dev. remains the same. [Must mention mean and st. dev.] (from(a))	B1
	The median and quartiles would decrease. [Must refer to median and at least Q_1.] ((b)(c))	B1
	The IQR would remain unchanged (from (d))	B1 **(3)**
		(14 marks)

	Notes
	Correct answers only score full marks in each part except (c)
(a)	B1 for 4837.5 or 4838 or 4840 seen.
	If no $\sum ft$ seen (or attempt at $\sum ft$ seen), B1 can be implied by a correct mean of awrt 24.2
	1st M1 for attempt at their $\frac{\sum ft}{\sum f}$ allow 1sf so $\sum f$ = awrt 200 and $\sum ft$ = awrt 5000.
	Or award M1 for a clear attempt at mean where at least 4 correct products of $\sum ft$ are seen
	2nd M1 for correct expression including square root seen. Follow through their mean. Allow a transcription error in 134281.25 but not an incorrect re-calculation.
(b)	M1 for a correct fraction ×5. Ignore end point but must be +. Allow use of (n + 1) giving 100.5…
(c)	B1cso for a fully correct expression including end point. NB Answer is given. Allow use of (n + 1) giving 50.25…but use of 50.5 scores B0
(d)	1st B1 for 25.5 (or awrt 25.7 using n + 1)
	2nd B1ft for their Q_3 − their Q_1 (or 18.6) (provided > 0) Accept awrt 2sf. Correct ans. only scores 2/2
(e)	B1 Must mention that the data is skewed or not symmetrical. Do not award for "outliers"
(f)	1st B1 for one correct comment from the above. May refer to parts (a), (b), (c) or (d)
	2nd B1 for two correct comments from the above
	3rd B1 for all 3 correct comments from the above

Q4.

Question Number	Scheme	Marks
(a)	Width = 0.5 (cm) e.g 4 [cm^2] represents 8 babies or frequency densities are 8 and 34 Height = 17 (cm)	B1 M1 A1 (3)
(b)	$[Q_2 =]$ $\{3\} + \frac{(25-9)}{(26-9)} \times 0.5$, or $\{3.5\} - \frac{(25-24)}{(41-24)} \times 0.5$ = awrt **3.47** (allow $\frac{59}{17}$)	M1, A1 (2)
(c)(i)	$\sum fx = 1\times1 + 2.5\times8 + 3.25\times17 + 3.75\times17 + 4.5\times7 = 171.5$, $\bar{x} = \frac{171.5}{50} = (3.43)$ (*)	B1cso
(ii)	$\sqrt{\frac{611.375}{50} - 3.43^2}$ $= 0.680147\ldots$ = awrt **0.680** (Accept 0.68)	M1, A1 (3)
(d)(i) (ii)	No change in mean (since weight is the same) s.d. will decrease (Extra value is at "centre" so data more concentrated) Both statements correct and correct reasons for each	B1 B1 dB1 (3) [11 marks]

	Notes
(a)	M1 for clear representation of area with frequency or height × width = 8.5 A1 for 17 (cm) [Must be clear it is height not frequency] (Ans only must satisfy $h \times w = 8.5$)
(b)	M1 for $\frac{16}{17} \times 0.5$ or if using $n+1$ for $\frac{16.5}{17} \times 0.5$ May see $-\frac{1}{17} \times 0.5$ if working down A1 for awrt 3.47 (or $\frac{59}{17}$) [check from correct working] or (if using $(n+1)$ for 3.485 or awrt 3.49)
(c)(i)	B1cso for Σfx (at least 3 correct & no incorrect products seen) and correct $\frac{\sum fx}{50}$ or $\frac{171.5}{50}$
(ii)	M1 for a correct expression including square root. Must use 3.43 no ft A1 for awrt 0.680 (accept 0.68). Allow use of s = awrt 0.687 (Ans only 2/2)
(d)(i) (ii)	1st B1 for no change in mean {send a correct argument for decrease to review} 2nd B1 for s.d. decreases 3rd dB1 dep on 1st and 2nd Bs for a correct reason for both mean and sd e.g. "new mean the same so within 1 s.d. of old mean"

Pearson

Topic 4

Correlation

Bronze, Silver, Gold
Worksheets for
AS Level Mathematics

Teacher Notes

These Bronze, Silver and Gold worksheets are designed to be used either straight after the content has been taught or as part of a skills gap analysis, especially as students move into year 13.

They are drawn from the latest specification questions and legacy questions. The papers are between 25 and 35 marks.

The topic number on this worksheet relates to the corresponding chapter number in the 'Pearson Edexcel AS and A Level Mathematics: Statistics and Mechanics Year 1/AS' textbook.

Quick Links

(Press Ctrl, as you click with your mouse to follow these links)

- Bronze Questions
- Bronze Mark Scheme
- Silver Questions
- Silver Mark Scheme
- Gold Questions
- Gold Mark Scheme

Extension and Enrichment

If you have students that have enjoyed the challenge of the Gold questions, then they should have a go at the more challenging question from our Advanced Extension Award (AEA) papers. The Mathematics AEA is a single, 3 hour non-calculator paper, taken at the end of year 13. It helps students to develop high level problem solving and proof skills. It is entirely based on the content of the A Level Mathematics Course. No extra material needs to be covered to take the AEA in Mathematics. A second important difference is that marks are awarded for the clarity and quality of their solution. Developing this key skill, alongside the extra problem-solving experience, can pay dividends in the way they approach A Level Mathematics and Further Mathematics problems.

More information about the Advanced Extension Award can be found here on the Pearson Edexcel Website, or here on the Maths Emporium.

Bronze Questions

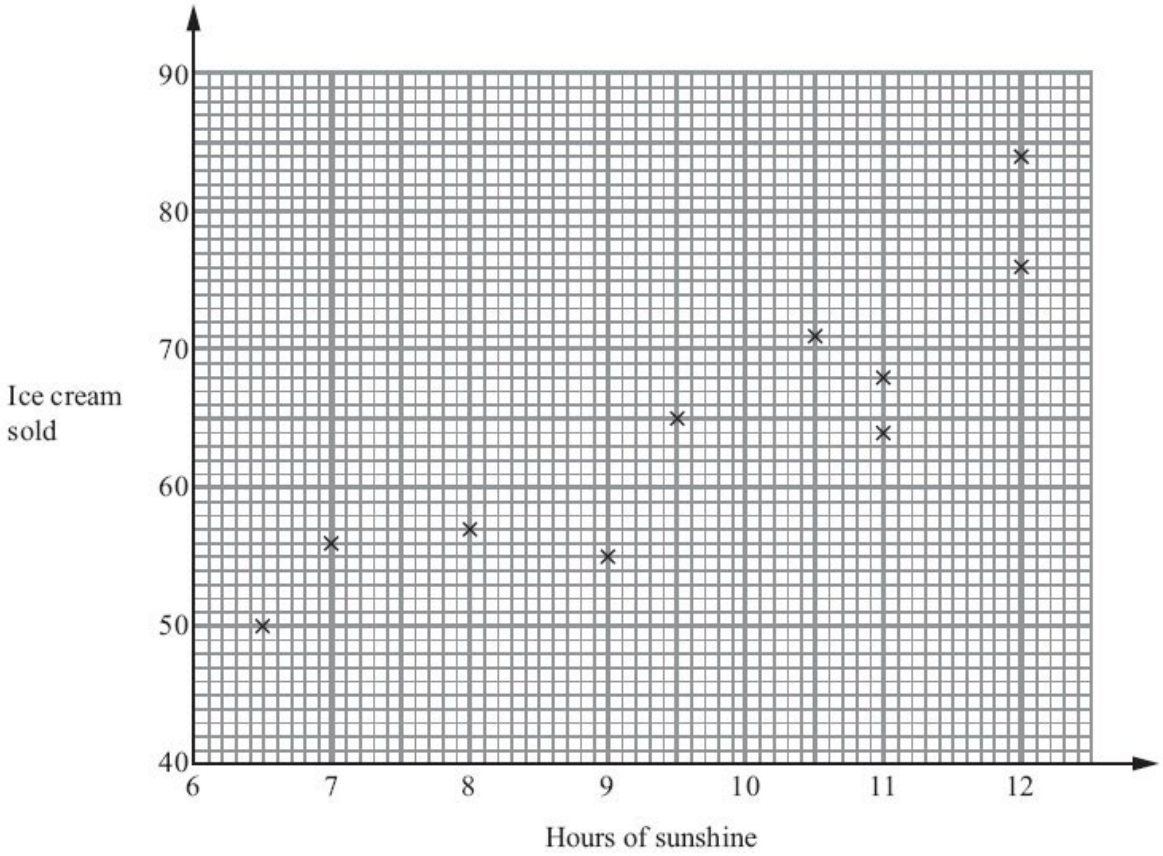
Calculator

The total mark for this section is 28

Q1

A beach cafe sells ice creams.
Each day the manager records the number of hours of sunshine and the number of ice creams sold.

The scatter graph shows this information.



On another day there were 11.5 hours of sunshine and 73 ice creams sold.

(a) Show this information on the scatter graph.

(1)

(b) Describe the relationship between the number of hours of sunshine and the number of ice creams sold.

(1)

One day had 10 hours of sunshine.

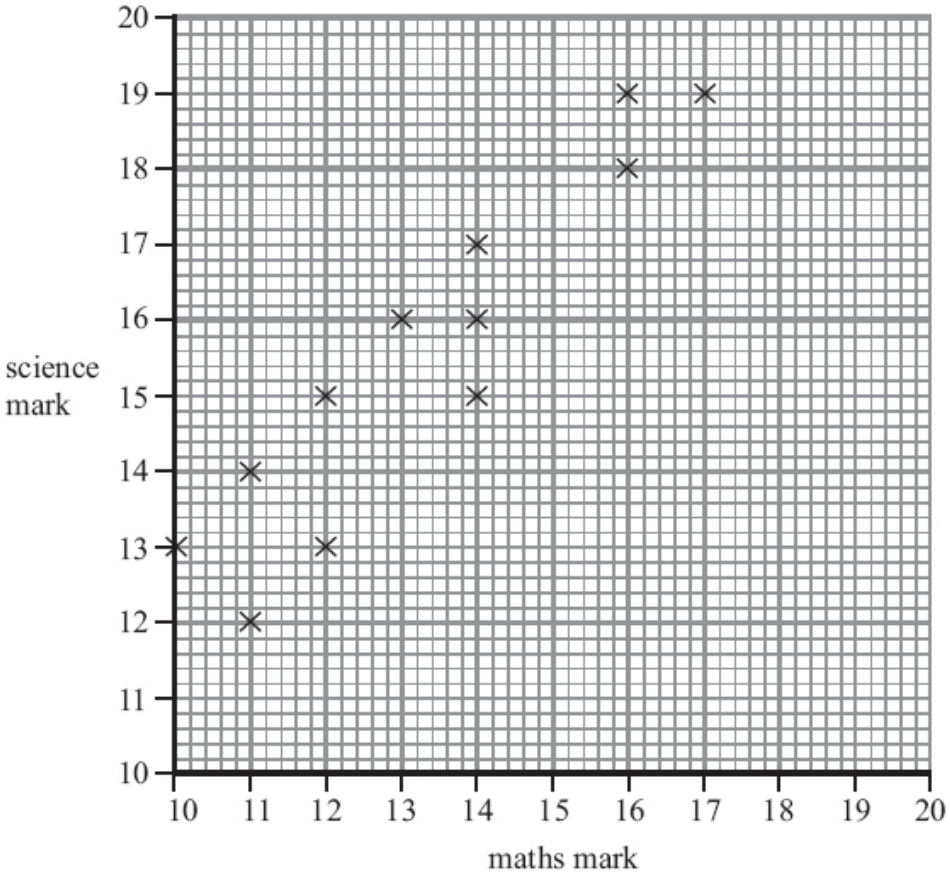
(c) Estimate how many ice creams were sold.

(2)

(Total for Question 1 is 4 marks)

Q2

Mr Kent's students did a maths test and a science test.
The scatter graph shows the marks of 12 of these students.



The table shows the marks of two more students.

Name	maths	science
Masood	12	14
Nimer	17	20

(a) Show this information on the scatter graph.

(1)

(b) What type of correlation does this scatter graph show?

.

(1)

David did the maths test.
He was absent for the science test.

David's mark in the maths test was 15

(c) Estimate a science mark for David.

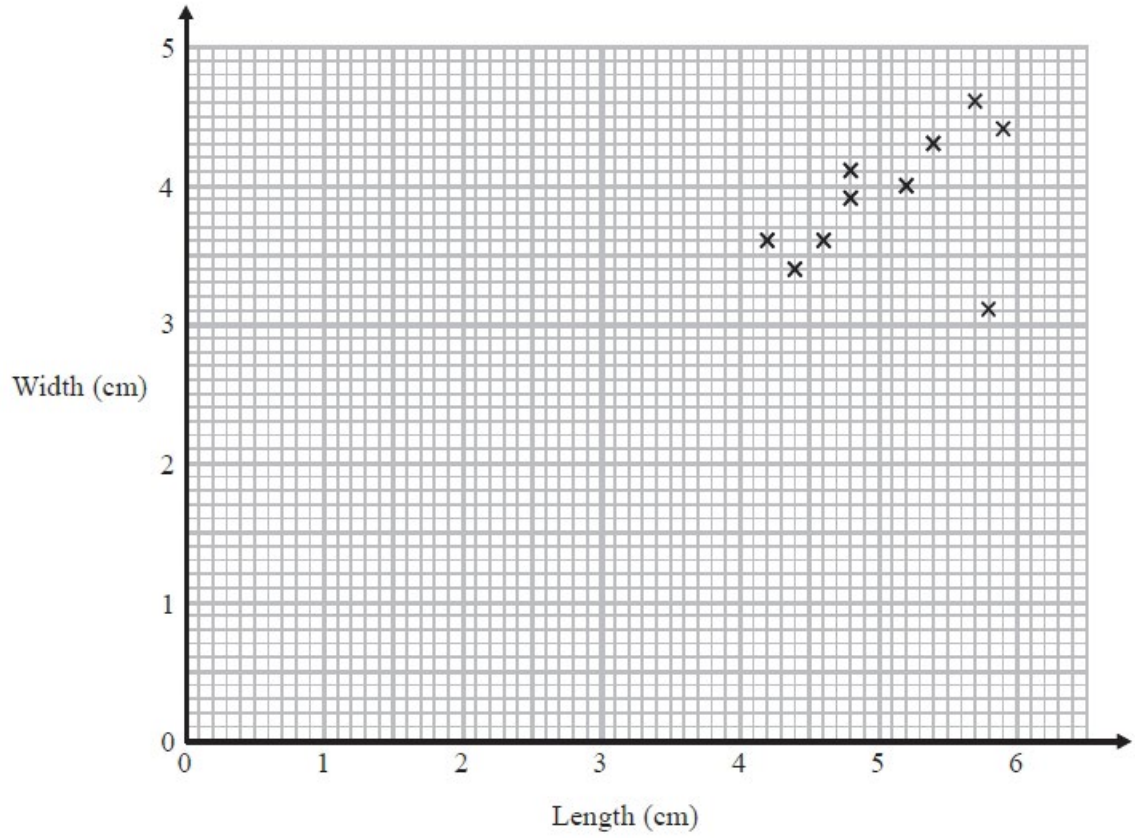
(2)

(Total for Question 2 is 4 marks)

Q3

Katie measured the length and the width of each of 10 pine cones from the same tree.

She used her results to draw this scatter graph.



(a) Describe one improvement Katie can make to her scatter graph.

(1)

The point representing the results for one of the pine cones is an outlier.

(b) Explain how the results for this pine cone differ from the results for the other pine cones.

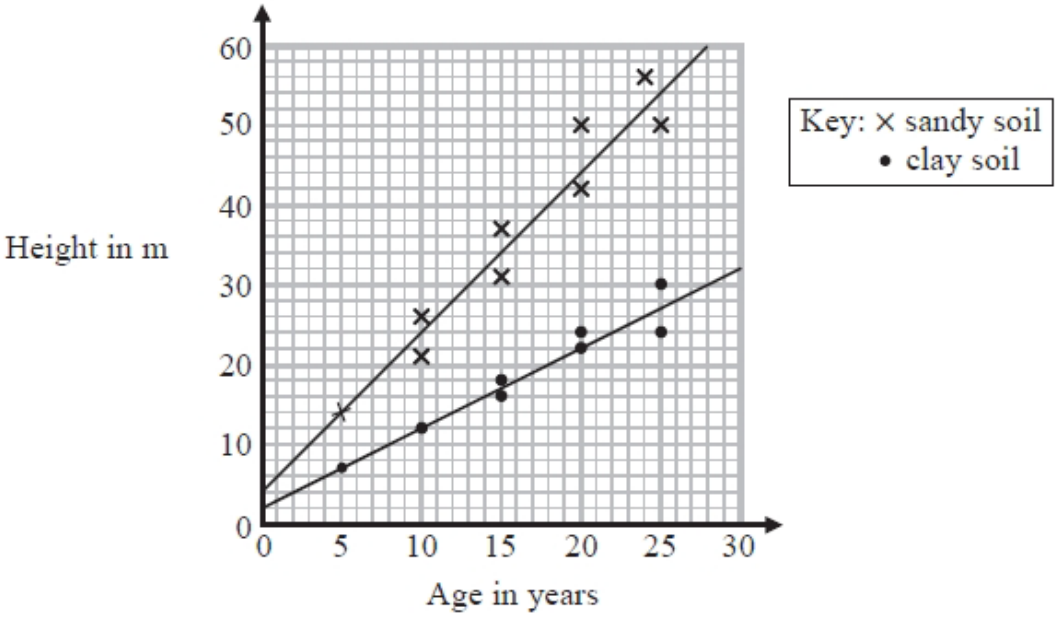
(1)

(Total for Question 3 is 2 marks)

Q4

Bill wants to compare the heights of pine trees growing in sandy soil with the heights of pine trees growing in clay soil.

The scatter diagram gives some information about the heights and the ages of some pine trees.



(a) Describe the relationship between the height of pine trees and the age of pine trees growing in sandy soil.

(1)

A pine tree growing in clay soil is 18 years old.

(b) Find an estimate for the height of this tree.

(1)

A pine tree is growing in sandy soil.

(c) Work out an estimate for how much the height of this tree increases in a year.

(2)

(d) Compare the rate of increase of the height of trees growing in clay soil with the rate of increase of the height of trees growing in sandy soil.

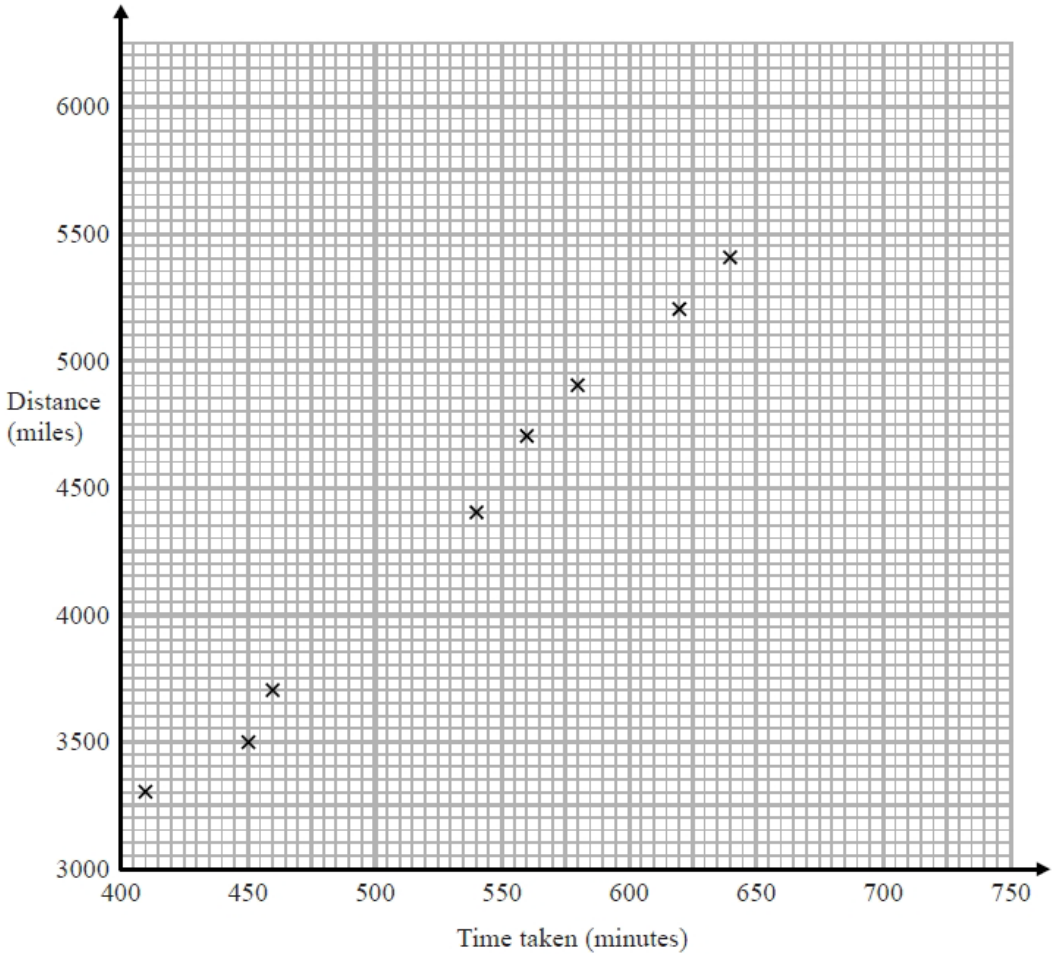
(2)

(Total for Question 4 is 6 marks)

Q5

Oliver records the distance from London to each of eight cities in the USA.
He also records the time taken to fly from London to each of these cities.

The scatter graph shows this information.



Chicago is a city in the USA.
Chicago is 4000 miles from London.

(a) (i) By drawing a line of best fit, find an estimate for the time taken to fly from London to Chicago.

(2)

(ii) Why is your answer to part (i) only an estimate?

(1)

(b) (i) Calculate the gradient of your line of best fit.

(2)

(ii) Give an interpretation of the gradient of your line of best fit.

(1)

(Total for Question 5 is 6 marks)

Q6

The age, t years, and weight, w grams, of each of 10 coins were recorded.

Given that the equation of the regression line of w on t is $w = 11.6 - 0.0263t$

(a) State, with a reason, which variable is the explanatory variable.

(2)

(b) Using this model, estimate

(i) the weight of a coin which is 5 years old,
(ii) the effect of an increase of 4 years in age on the weight of a coin.

(2)

It was discovered that a coin in the original sample, which was 5 years old and weighed 20 grams, was a fake.

(c) State, without any further calculations, whether the exclusion of this coin would increase or decrease the value of the
correlation. Give a reason for your answer.

(2)

(Total for Question 6 is 6 marks)

Bronze Mark Scheme

Q1.

Question	Working	Answer	Mark	Notes
(a)		Point at (11.5, 73)	1	B1 Point plotted ±½ small square
(b)			1	B1 for description of dynamic relationship eg "the more hours of sunshine, the more ice creams sold" or positive correlation]Note: 'sunnier' implies 'more hours of sunshine']
(c)		62 - 70	2	B2 for answer in the range 62-70 OR M1 for a single straight line of best fit with positive gradient, passing between (6.5, 45), (6.5, 59) and (12, 70), (12, 80) or a vertical line drawn from 10 A1 for answer in range 62-70 or ft from single straight "line of best fit" with positive gradient

Q2.

Question	Working	Answer	Mark	Notes
(a)	See scatter graph	(12,14) and(17,20) plotted	1	B1 for correct plotting of both points
(b)			1	B1 Positive (accept a correct relationship)
(c)		16 to 18	2	B2 16 – 18 (M1 for a single line segment from m=11 to m = 16 within overlay or a vertical line drawn from m = 15 A1 for an answer in the range 16-18 or ft their line of best fit ± 2mm)

Q3.

Question	Working	Answer	Notes
(a)		improvement	C1 appropriate improvement eg do not have axes starting at (0, 0)
(b)		explanation	C1 explanation eg pine cone has a very short width for its length

Q4.

5MB1H/01 June 2015				
Question	**Working**	**Answer**	**Mark**	**Notes**
(a)		Description	1	B1 description eg Taller trees are older. Accept positive correlation.
(b)		20	1	B1 19 – 21
(c)		2	2	M1 for evidence of taking readings at two points from Sandy line, or increase excluding start eg 24÷10, 14÷5 A1 for answer 1.8 to 2.2
(d)		Comparison	2	B2 for a complete explanation e.g.Trees grow at approximately twice the rate on sandy soil (B1 for a partial explanation e.g. Trees grow faster on sandy soil)

Q5.

Question	Working	Answer	Mark	Notes
(a)(i)		480 – 500	B1 B1	for line of best that can be used to estimate time of flight for 480 – 500 or ft lobf
(a)(ii)		reason	C1	for reason, e.g. lobf can vary, data is only a sample, scale cannot be read exactly
(b)(i)		9.4 – 9.8	M1 A1	for method to find gradient, e.g. triangle drawn with "change in distance ÷ change in time" for 9.4 – 9.8 or ft lobf
(b)(ii)		speed	C1	for speed (in miles per minute) oe

Q6.

Question Number	Scheme	Marks
(a)	The explanatory variable is the age of each coin. This is because the age is set and the weight varies.	B1 B1 (2)
(b) (i)	awrt 11.5	B1
(ii)	Decrease(in weight of coin of 0.1052 g) = 0.1 or -0.1 or increase of -0.1 awrt(-0.1)	B1 (2)
(c)	Decrease; removing the fake will result in a better linear fit	B1;B1 (2) Total 6

Silver Questions

Calculator

The total mark for this section is 28

Q1

A random sample of 15 days is taken from the large data set for Perth in June and July 1987.

The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

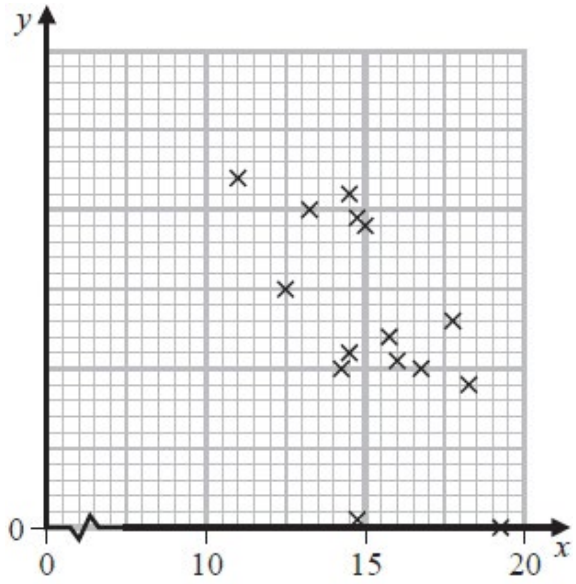


Figure 1

(a) Describe the correlation.

(1)

The variable on the x-axis is Daily Mean Temperature measured in °C.

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

(i) suggest which variable is on the y-axis,
(ii) state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a negative correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

(c) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

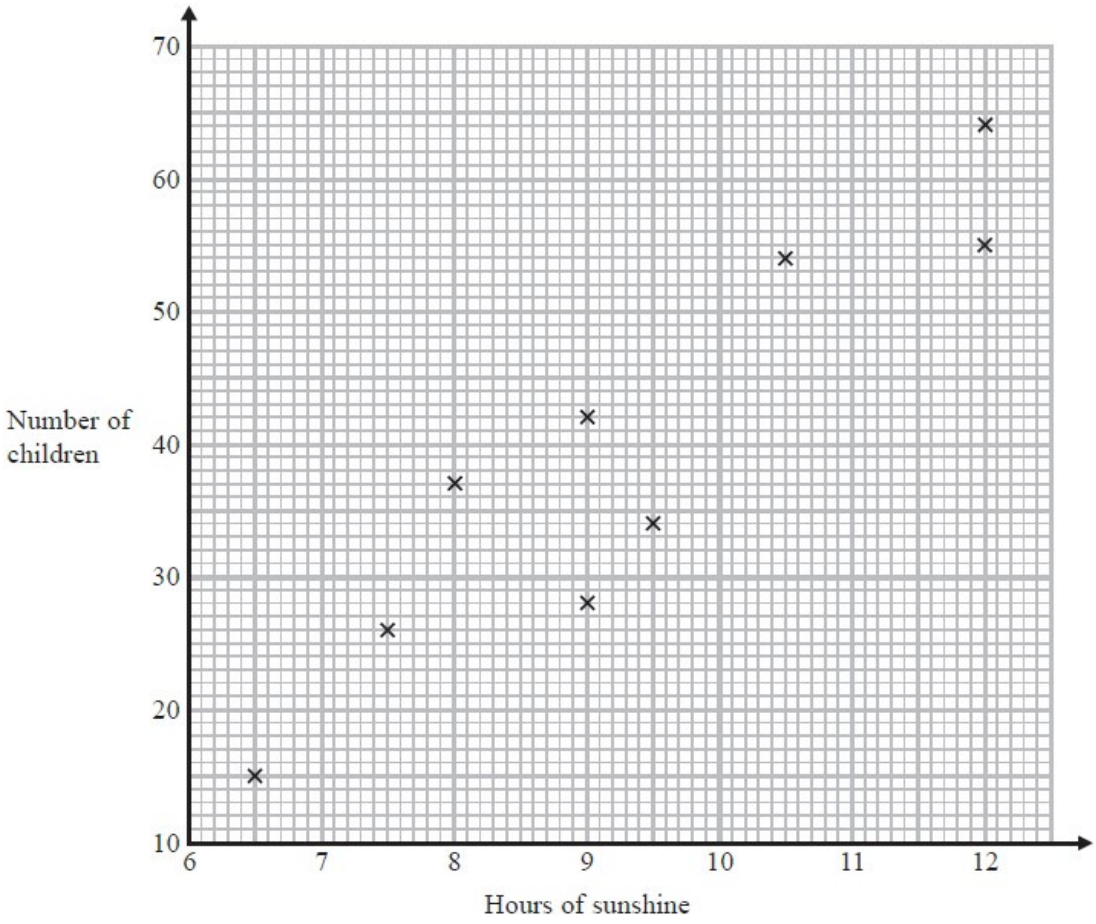
(Total for Question 1 is 4 marks)

Q2

Sally looks after a children's paddling pool in a park.

Each day, Sally records the number of hours of sunshine and the number of children who use the paddling pool.

The scatter graph shows this information.



(a) Describe the correlation between the number of children who use the paddling pool and the number of hours of sunshine.

(1)

One day there were 10 hours of sunshine.

(b) Estimate how many children used the paddling pool.

(2)

On another day, there were 6.5 hours of sunshine and 45 children used the pool.

(c) (i) Show this information on the scatter graph.

This point is isolated on the scatter graph.

(ii) Explain what may have happened on this day.

(2)

(Total for Question 2 is 5 marks)

Q3

A scientist is researching whether or not birds of prey exposed to pollutants lay eggs with thinner shells. He collects a random sample of egg shells from each of 6 different nests and tests for pollutant level, p, and measures the thinning of the shell, t. The results are shown in the table below.

p	3	8	30	25	15	12
t	1	3	9	10	5	6

(a) Draw a scatter diagram to represent these data.

(2)

(b) Explain why a linear regression model may be appropriate to describe the relationship between p and t.

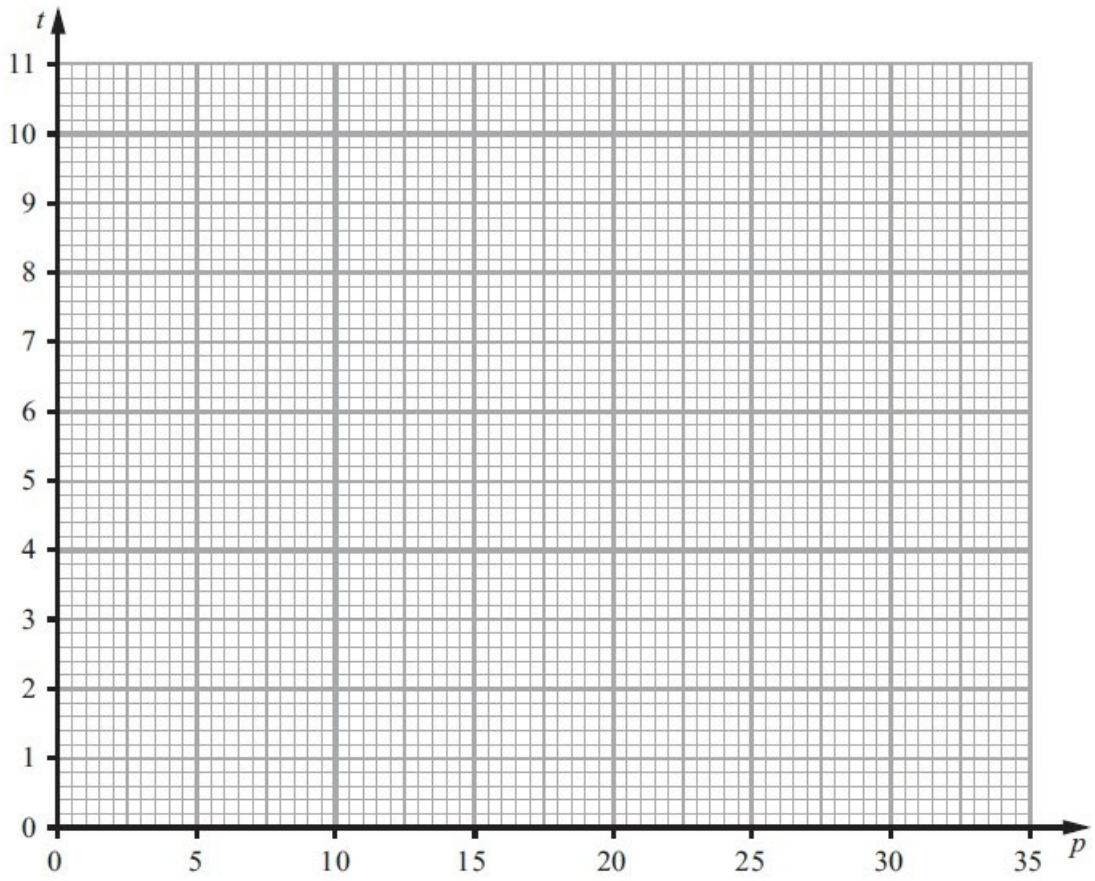
(1)

The scientist reviews similar studies and finds that pollutant levels above 16 are likely to result in the death of a chick soon after hatching.

Given that $t = 0.741+0.318p$

(c) Estimate the minimum thinning of the shell that is likely to result in the death of a chick.

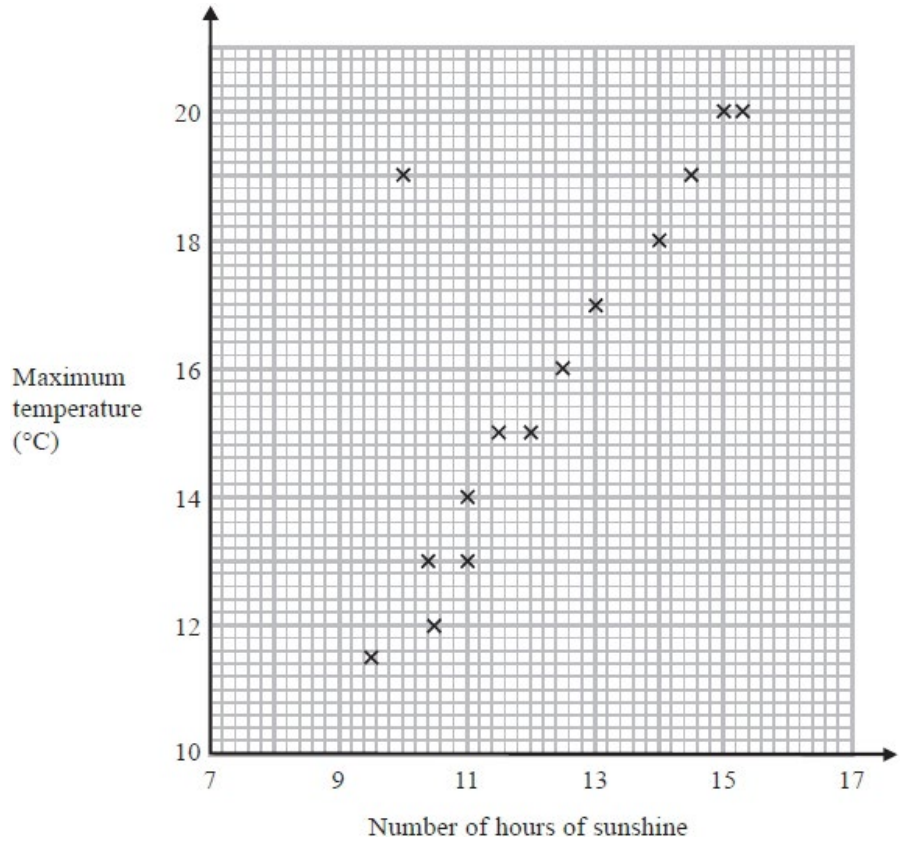
(2)



(Total for Question 3 is 5 marks)

Q4

The scatter graph shows the maximum temperature and the number of hours of sunshine in fourteen British towns on one day.



One of the points is an outlier.

(a) Write down the coordinates of this point.

(1)

(b) For all the other points write down the type of correlation.

(1)

On the same day, in another British town, the maximum temperature was 16.4°C.

(c) Estimate the number of hours of sunshine in this town on this day.

(2)

A weatherman says,

"Temperatures are higher on days when there is more sunshine."

(d) Does the scatter graph support what the weatherman says?

Give a reason for your answer.

(1)

(Total for Question 4 is 5 marks)

Q5

Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £ w, and the average weekly temperature, t °C, for 8 weeks during the summer.
There is a negative correlation for these data.

(a) Suggest a possible reason for this correlation.

(1)

Tessa suggests that a linear regression model could be used to model these data.

(b) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion.

(1)

(c) State, giving a reason, which variable would be the explanatory variable.

(1)

Tessa calculated the linear regression equation as $w = 10\ 755 - 171t$

(d) Give an interpretation of the gradient of this regression equation.

(1)

(Total for Question 5 is 4 marks)

Q6

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.

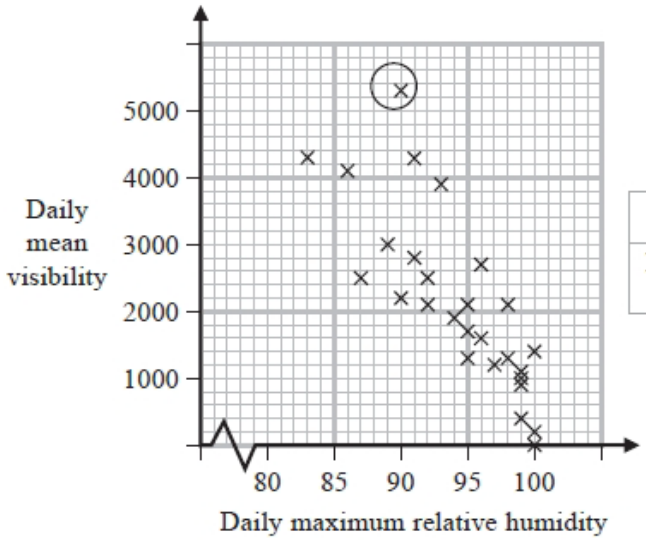


Figure 2

	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.

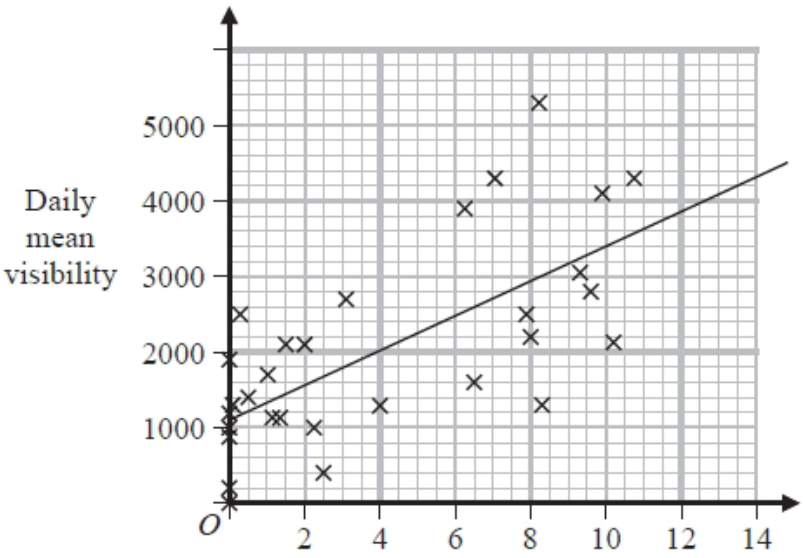


Figure 3

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

(1)

(Total for Question 6 is 6 marks)

Silver Mark Scheme

Q1.

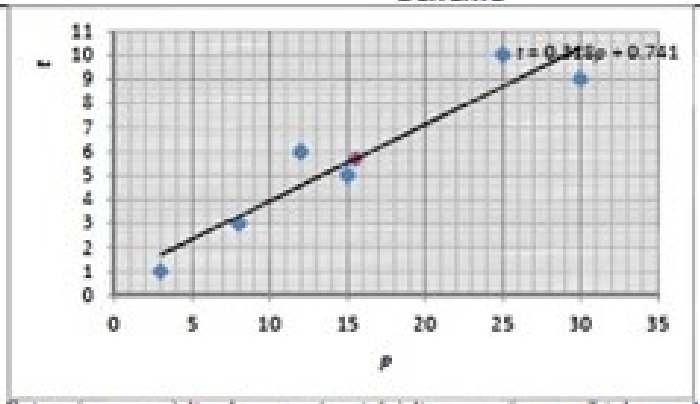
	Scheme	Marks	AO
(a)	Negative	B1 (1)	1.2
(b)(i) (ii)	Rainfall mm or Pressure hPa or Pascals or hectopascals or mb or millibars	B1 B1ft (2)	2.2b 1.1b
(c)	Humidity is high and there is evidence of negative correlation So expect amount of sunshine to be lower than the average for Heathrow(oe)	B1 (1)	2.2b
		(4 marks)	

Q2.

Question	Working	Answer	Mark	Notes
(a)		Positive	1	B1 for positive
(b)		40 - 50	2	M1 for suitable line of best fit drawn or evidence of vertical line at 10 or a point indicated at (10, y) where $40 \leq y \leq 50$ A1 for 40 – 50
(c)(i) (ii)		Point plotted	2	B1 for point plotted at (6.5, 45) B1 for explanation of outlier eg holiday, party, special event etc

Q3.

Q4.

Question	Scheme	Marks
(a)		Use overlay B1 B1 (2)
(b)	Points (appear to) lie close to a (straight) line or "strong /high correlation"	B1 (1)
(c)	$t =$ "0.741" + "0.318" × 16 = 5.825... awrt 5.8	M1 A1 (2) [5]

	Notes
(a)	B2 for all 6 data points plotted correctly. B1 for any 5 correct. Points not wholly outside the circles.
(c)	M1 for clear use of their line (equation or on graph) and $p = 16$ to estimate t. This may be an expression or lines marked on the diagram A1 for awrt 5.8, even if their line is not fully correct. Accept "$t > 5.8$"(oe). Answer only 2/2

Q5.

Question	Working	Answer	Mark	Notes
(a)		10,19	B1	cao
(b)		positive	C1	positive (correlation)
(c)		12 to 13	M1	for an appropriate line of best fit drawn, or a point marked at (x, 16.4) or a horizontal line drawn from 16.4 across to (x, 16.4) where x is in the range 12 to 13
			A1	hours given in the range 12 to 13
(d)		explanation	C1	(yes) e.g. as the majority of points for high temperature appear when there are more hours of sunshine (positive correlation)

Qu	Scheme	Marks	AO
(a)	e.g. As temperature increases people spend more time on the beach and less time shopping (o.e.)	B1 (1)	2.4
(b)	Since r is close to -1, it is consistent with the suggestion	B1 (1)	2.4
(c)	t will be the explanatory variable since sales are likely to depend on the temperature	B1 (1)	2.4
(d)	Every degree rise in temperature leads to a drop in weekly earnings of £171	B1 (1)	3.4
		(4 marks)	

	Notes
(a)	B1 for a suitable reason to explain negative correlation using the context given. e.g. "As temperature drops people are more likely to go shopping (than to the beach)" e.g. "As temperature increases people will be outside rather than in shops" A mere description in context of negative correlation is B0 SO e.g. "As temperature increases people don't want to go shopping/buy clothes" is B0 e.g. "Less clothes needed as temp increases" is B0
(b)	B1 for a suitable reason e.g. "strong"/"significant"/"near perfect" "correlation", $\|r\|$ close to 1 and saying it is consistent with the suggestion. Allow "yes" followed by the reason.
(c)	B1 For identifying t and giving a suitable reason. Need idea that "w depends on t" or "w responds to t" or "t affects w" (o.e.) Allow t (temperature) affects the other variable etc Just saying "t is the independent variable" or "t explains change in w" is B0 N. B. Suggesting causation is B0 e.g. "t causes w to decrease"
(d)	B1 for a description that conveys the idea of rate per degree Celsius. Must have 171, condone missing "£" sign.

Q6.

Question	Scheme	Marks	AOs
(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)

Gold Questions

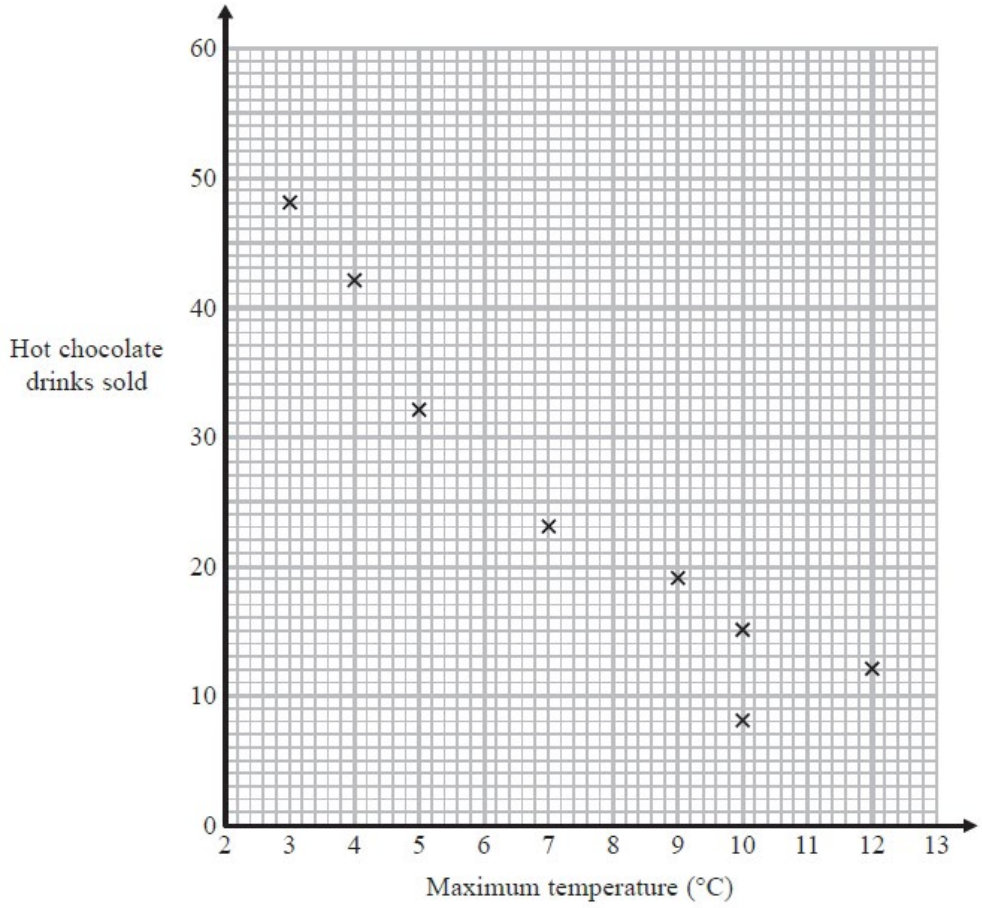
Calculator

The total mark for this section is 26

Q1

Carlos has a cafe in Clacton.
Each day, he records the maximum temperature in degrees Celsius (°C) in Clacton and the number of hot chocolate drinks sold.

The scatter graph shows this information.



On another day the maximum temperature was 6 °C and 35 hot chocolate drinks were sold.

(a) Show this information on the scatter graph.

(1)

(b) Describe the relationship between the maximum temperature and the number of hot chocolate drinks sold.

(1)

(c) Draw a line of best fit on the scatter diagram.

(1)

One day the maximum temperature was 8 °C.

(d) Use your line of best fit to estimate how many hot chocolate drinks were sold.

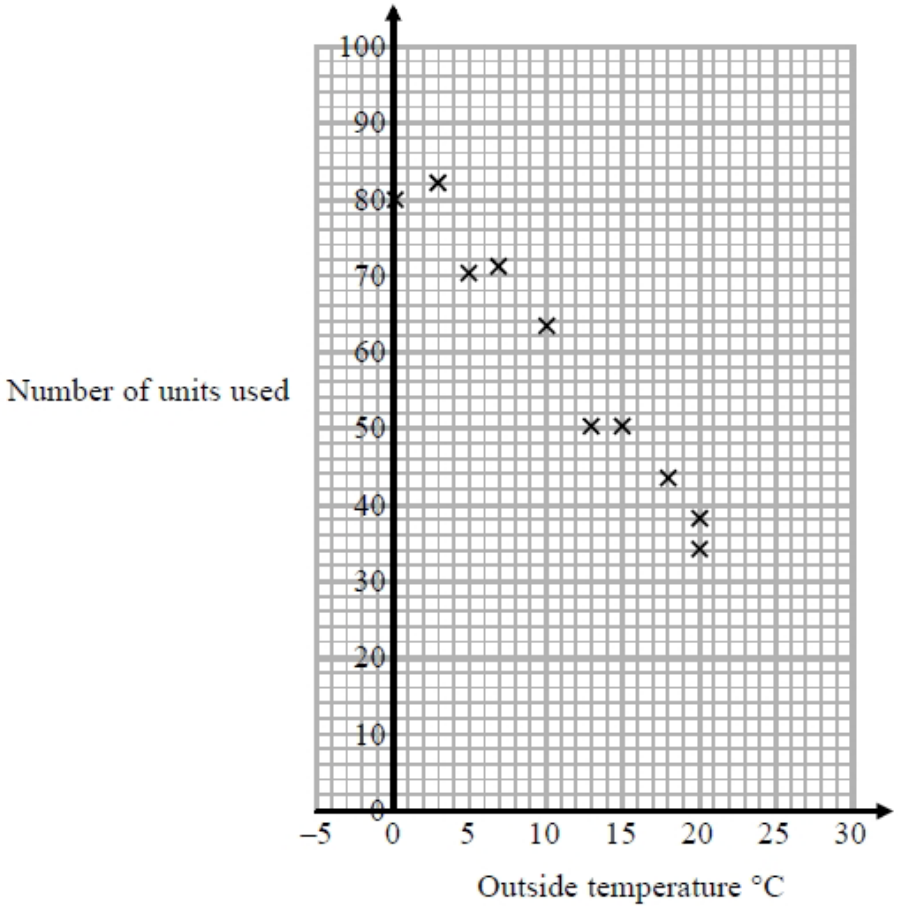
(1)

(Total for Question 1 is 4 marks)

Q2

In a survey, the outside temperature and the number of units of electricity used for heating were recorded for ten homes.

The scatter diagram shows this information.



Molly says,

> "On average the number of units of electricity used for heating decreases by 4 units for each °C increase in outside temperature."

(a) Is Molly right?
Show how you get your answer.

(3)

(b) You should **not** use a line of best fit to predict the number of units of electricity used for heating when the outside temperature is 30 °C.

Give one reason why.

(1)

(Total for Question 2 is 4 marks)

Q3

A meteorologist believes that there is a positive correlation between the daily mean windspeed, w kn, and the daily mean temperature, t °C. A random sample of 9 consecutive days is taken from past records from a town in the UK in July and the relevant data is given in the table below.

t	13.3	16.2	15.7	16.6	16.3	16.4	19.3	17.1	13.2
w	7	11	8	11	13	8	15	10	11

(a) Explain why a linear regression model based on these data is unreliable on a day when the mean temperature is 24 °C

(1)

Using the same 9 days a location from the large data set gave $\overline{t} = 27.2$ and $\overline{w} = 3.5$

(b) Using your knowledge of the large data set, suggest, giving your reason, the location that gave rise to these statistics.

(1)

(Total for Question 3 is 2 marks)

Q4

A large company is analysing how much money it spends on paper in its offices every year. The number of employees, x, and the amount of money spent on paper, p (£ hundreds), in 8 randomly selected offices are given in the table below.

x	8	9	12	14	7	3	16	19
p (£ hundreds)	40.5	36.1	30.4	39.4	32.6	31.1	43.4	45.7

(a) Given that $p = 28.3 + 0.824x$

Estimate the amount of money spent on paper in an office with 10 employees.

(2)

(b) Explain the effect each additional employee has on the amount of money spent on paper.

(1)

Later the company realised it had made a mistake in adding up its costs, p. The true costs were actually half of the values recorded. The product moment correlation coefficient and the equation of the linear regression line are recalculated using this information.

(c) Write down the new value of the gradient of the regression line

(1)

(Total for Question 4 is 4 marks)

Q5

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)

(Total for Question 5 is 3 marks)

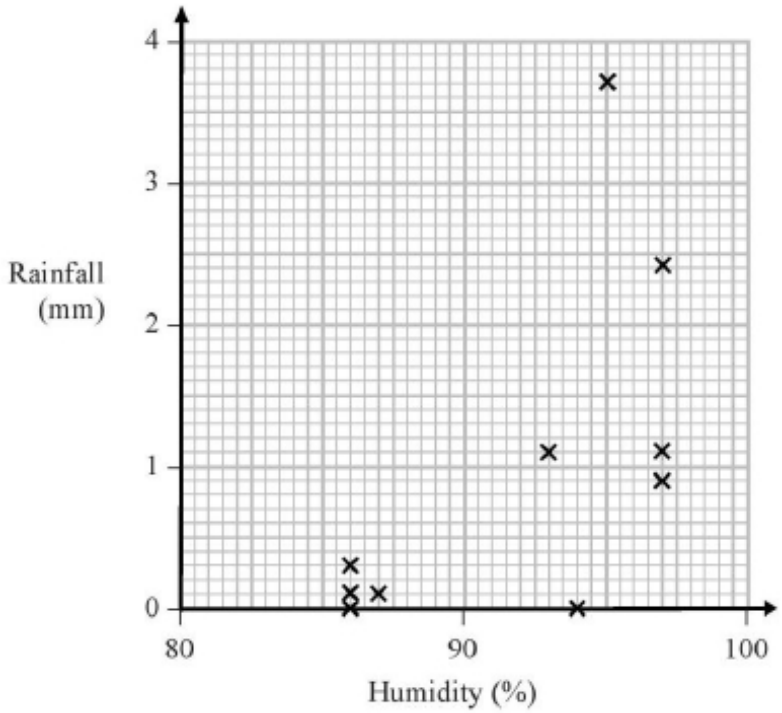
Q6

Sara was studying the relationship between rainfall, r mm, and humidity, h %, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of r and h.



(a) Give an interpretation of the correlation between rainfall and humidity.

(1)

The equation of the regression line of r on h for these 10 days is $r = -12.8 + 0.15h$

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

(1)

(c) (i) Comment on the suitability of Sara's sampling method for this study.

(ii) Suggest how Sara could make better use of the large data set for her study.

(2)

(Total for Question 6 is 4 marks)

Q7

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)

(Total for Question 7 is 5 marks)

Gold Mark Scheme

Q1.

Paper_5MB1H_01					
Question		**Working**	**Answer**	**Mark**	**Notes**
	(a)		Point plotted	1	B1 for point plotted at (6,35)
	(b)			1	B1 for description of dynamic relationship or negative correlation
	(c)			1	B1 for single straight line of best fit which could be used to take readings
	(d)		21 - 26	1	B1 for answer in the range 21 - 26 or ft from single straight line segment (if previous B0)

Q2.

Paper 1MA1: 3H				
Question	**Working**	**Answer**	**Notes**	
(a)	Draws LOBF Finds ht ÷ base = $\frac{85-20}{0-25} = -2.6$	No + reason	M1 M1 C1	Interpret question eg. draw line of best fit Start to test eg. gradient e.g. $\frac{85-20}{0-25} = -2.6$ Gradient within range $\pm(2-3)$ and 'no'
(b)		The LOBF would have to be used outside the data	C1	Convincing explanation

Q3.

Question	Scheme	Marks	AOs
(a)	e.g. It requires extrapolation so will be unreliable (o.e.)	B1	1.2
		(1)	
(b)	Higher $\bar{t}$ suggests overseas and not Perth…lower wind speed so perhaps not close to the sea so suggest **Beijing**	B1	2.4
		(1)	
			(2 marks)

Notes:

(a)
B1: for a correct statement (unreliable) with a suitable reason

(b)
B1: for suggesting Beijing with some supporting reason based on t or w
Allow Jacksonville with a reason based just on higher $\bar{t}$

Q4.

Question Number	Scheme	Marks
(a)	$p = 28.3.... + 0.824.... \times 10 = 36.57552....$ awrt £3700	M1 A1 (2)
(b)	Goes up £82.40	B1 (1)
(c)	$b = 0.412$	B1 (1)
		4

Q5.

Qu	Scheme	Marks	AO
(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)	

	Notes
(a)	B1 for "positive". Allow an interpretation e.g. "as points increase pay increases" is B1 Read whole answer: contradictory comments such as "positive correlation, as points increase pay decreases" scores B0
(b)	B1 for any correct comment conveying idea of £s per point and including a correct value; must have idea of rate. Can condone missing £ sign. Accept 4.5 e.g. "every 10 points earns an extra (or increase) of £45" is B1 BUT "every point earns £4.5(0)" is B0 *doesn't have idea of rate*
(c)	B1 for a suitable comment mentioning "points" or "pay" (o.e. e.g. "amount") or commenting on "small sample" or "range of points" used to find line The following examples would score B1 Can say that n points (for $n < 10.\dot{4}$) would give negative pay so not suitable Any comment suggesting that some jobs would end up with negative pay Don't know the range of points used to find the regression line A small sample of size 8 may not be representative to cover all jobs B0 for a focus on "qualifications" or "hours" worked only The following examples would score B0 Some jobs require no (or low) skills or qualifications (*need negative pay*)

Q6.

Question	Scheme	Marks	AOs
(a)	e.g. "as humidity increases rainfall increases"	B1	2.2b
		(1)	
(b)	e.g. a 10% increase in humidity gives rise to a 1.5 mm increase in rainfall or represents 0.15mm of rainfall per percentage of humidity	B1	3.4
		(1)	
(c)(i)	Not a good method since only uses 11 days from one location in one month.	B1	2.4
(ii)	e.g. She should use data from more of the UK locations and more of the months or using a spreadsheet or computer package she could use all of the available UK data	B1	2.4
		(2)	
		(4 marks)	

Q7.

Question	Scheme	Marks	AOs
(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)

Pearson

Topic 5

Probability

Bronze, Silver, Gold

Worksheets for

AS Level Mathematics

Teacher Notes

These Bronze, Silver and Gold worksheets are designed to be used either straight after the content has been taught or as part of a skills gap analysis, especially as students move into year 13.

They are drawn from the latest specification questions and legacy questions. The papers are between 25 and 35 marks.

The topic number on this worksheet relates to the corresponding chapter number in the 'Pearson Edexcel AS and A Level Mathematics: Statistics and Mechanics Year 1/AS' textbook.

Quick Links

(Press Ctrl, as you click with your mouse to follow these links)

- Bronze Questions
- Bronze Mark Scheme
- Silver Questions
- Silver Mark Scheme
- Gold Questions
- Gold Mark Scheme

Extension and Enrichment

If you have students that have enjoyed the challenge of the Gold questions, then they should have a go at the more challenging question from our Advanced Extension Award (AEA) papers. The Mathematics AEA is a single, 3 hour non-calculator paper, taken at the end of year 13. It helps students to develop high level problem solving and proof skills. It is entirely based on the content of the A Level Mathematics Course. No extra material needs to be covered to take the AEA in Mathematics. A second important difference is that marks are awarded for the clarity and quality of their solution. Developing this key skill, alongside the extra problem-solving experience, can pay dividends in the way they approach A Level Mathematics and Further Mathematics problems.

More information about the Advanced Extension Award can be found here on the Pearson Edexcel Website, or here on the Maths Emporium

Bronze Questions

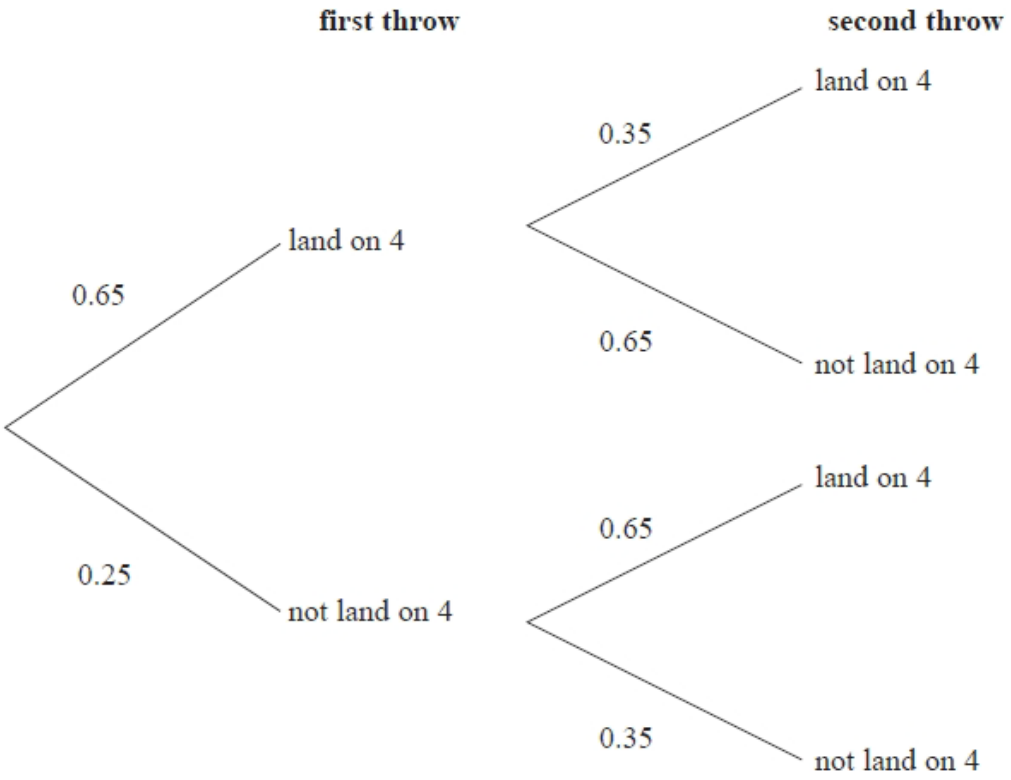
Calculator

The total mark for this section is 29

Q1

When a biased 6-sided dice is thrown once, the probability that it will land on 4 is 0.65
The biased dice is thrown twice.

Amir draws this probability tree diagram.
The diagram is **not** correct.



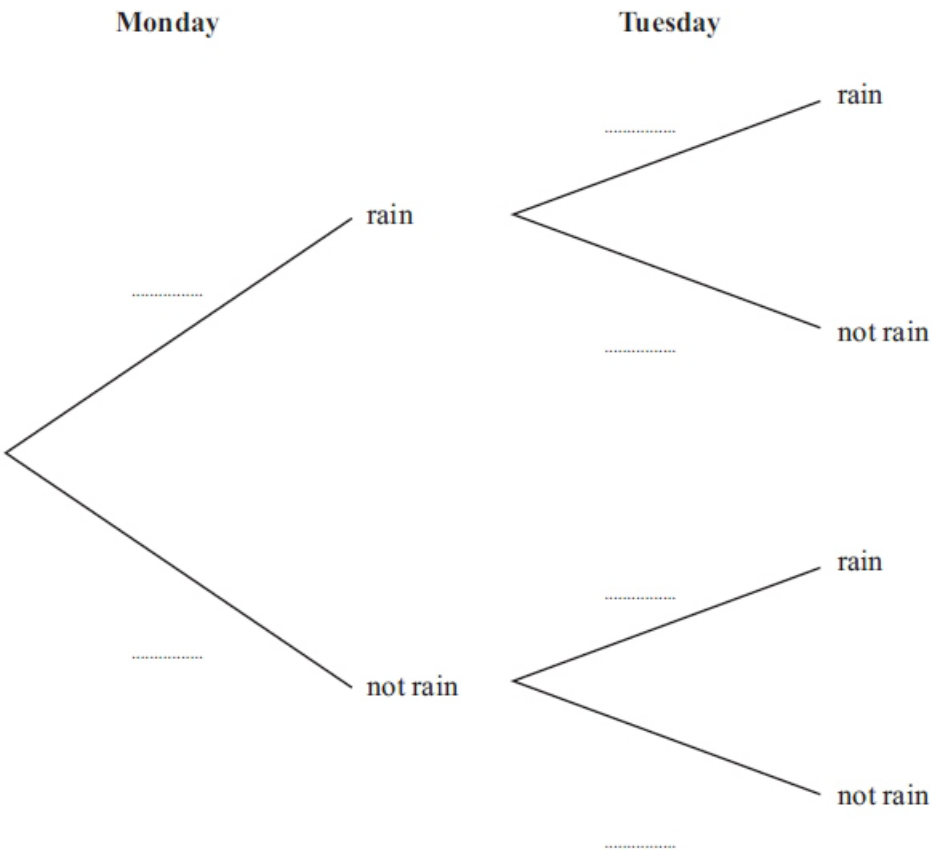
Write down **two** things that are wrong with the probability tree diagram.

(Total for Question 1 is 2 marks)

Q2

The probability that it will rain on Monday is 0.6

When it rains on Monday, the probability that it will rain on Tuesday is 0.8
When it does **not** rain on Monday, the probability that it will rain on Tuesday is 0.5



(a) Complete the probability tree diagram.

(2)

(b) Work out the probability that it will rain on both Monday and Tuesday.

.

(2)

(c) Work out the probability that it will rain on at least one of the two days.

(3)

(Total for Question 2 is 7 marks)

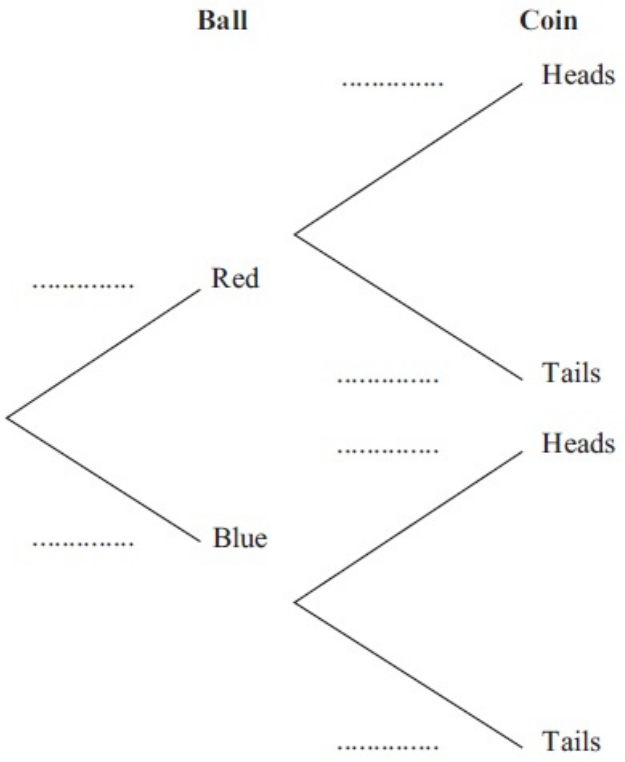
Q3

An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability $\frac{2}{3}$ of landing heads is spun.

When a blue ball is selected a fair coin is spun.

(a) Complete the tree diagram below to show the possible outcomes and associated probabilities.



(2)

Shivani selects a ball and spins the appropriate coin.

(b) Find the probability that she obtains a head.

(2)

Shivani and Tom each repeat this experiment.

(c) Find the probability that the colour of the ball Shivani selects is the same as the colour of the ball Tom selects.

(3)

(Total for Question 3 is 7 marks)

Q4

Sami asked 50 people which drinks they liked from tea, coffee and milk.

All 50 people like at least one of the drinks

19 people like all three drinks.

16 people like tea and coffee but do **not** like milk.

21 people like coffee and milk.

24 people like tea and milk.

40 people like coffee.

1 person likes only milk.

Sami selects at random one of the 50 people.

(a) Work out the probability that this person likes tea.

(4)

(b) Given that the person selected at random from the 50 people likes tea, find the probability that this person also likes exactly one other drink.

(2)

(Total for Question 4 is 6 marks)

Q5

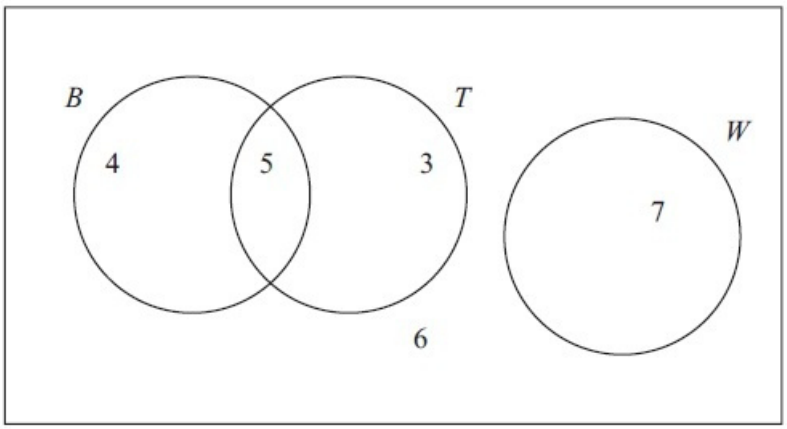


Figure 1

Figure 1 shows how 25 people travelled to work.

Their travel to work is represented by the events

B bicycle
T train
W walk

(a) Write down 2 of these events that are mutually exclusive. Give a reason for your answer.

(2)

(b) Determine whether or not *B* and *T* are independent events.

(3)

One person is chosen at random.

Find the probability that this person

(c) Walks to work,

(1)

(d) Travels to work by bicycle and train.

(1)

(Total for Question 5 is 7 marks)

Bronze Mark Scheme

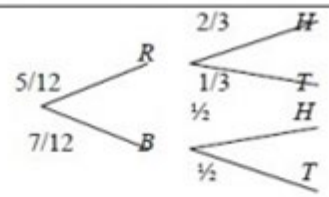
Q1.

Question	Answer	Mark	Mark scheme	Additional guidance
	Probabilities should sum to 1	C1	for stating that the probabilities should total 1 eg 0.25 should be 0.35	
	0.35 and 0.65 reversed	C1	for recognising that the 0.35 and 0.65 in the first branches for the 2nd throw should be reversed eg, "for the second throw, the probability it lands on 4 should be 0.65"	Can be shown on the diagram

Q2.

Question	Working	Answer	Mark	Notes
(a)		0.8 0.6 0.2 0.4 0.5 0.5	2	B2 for all 6 correct probabilities (B1 for two correct probabilities)
(b)	0.6 × 0.8	0.48	2	M1 for '0.6' × '0.8' ft probability tree diagram A1 cao
(c)	1 – (0.4 × 0.5) 1 – 0.2 OR 0.6 × 0.8 + 0.6 × 0.2 + 0.4 × 0.5	0.8	3	M2 for 1 – '0.4' × '0.5' ft probability tree diagram A1 cao OR M2 for '0.6' × '0.8' + '0.6' × '0.2' + '0.4' × '0.5' (M1 for any two of '0.6' × '0.8', '0.6' × '0.2', '0.4' × '0.5'added) A1 cao

Q3.

Question Number	Scheme	Marks
(a)	 P(R) and P(B)	B1
	2^{nd} set of probabilities	B1
		(2)
(b)	$P(H) = \frac{5}{12}\times\frac{2}{3}+\frac{7}{12}\times\frac{1}{2}.=\frac{41}{72}$ or awrt 0.569	M1 A1
		(2)
(c)	$\left(\frac{5}{12}\right)^2+\left(\frac{7}{12}\right)^2$	M1 A1ft
	$=\frac{25}{144}+\frac{49}{144}=\frac{74}{144}$ or $\frac{37}{72}$ or awrt 0.514	A1
		(3)
		7

(a) 1^{st} B1 for the probabilities on the first 2 branches. Accept $0.41\dot{6}$ and $0.58\dot{3}$

2^{nd} B1 for probabilities on the second set of branches. Accept $0.\dot{6}$, $0.\dot{3}$, 0.5 and $\frac{1.5}{3}$

Allow exact decimal equivalents using clear recurring notation if required.

(b) M1 for an expression for P(H) that follows through their sum of two products of **probabilities** from their tree diagram

(c) M1 for $\left(\frac{5}{12}\right)^2$ or $\left(\frac{7}{12}\right)^2$ can follow through their equivalent values from tree diagram

1^{st} A1 for both values correct or follow through from their original tree and +

2^{nd} A1 for a correct answer

Special Case $\frac{5}{12}\times\frac{4}{11}$ or $\frac{7}{12}\times\frac{6}{11}$ seen award M1A0A0

Q4.

Paper 1MA1: 3H			
Question	**Working**	**Answer**	**Notes**
(a)	Draws correct Venn diagram	$\frac{44}{50}$	M1 Begin to interpret given information eg. 3 overlapping labelled ovals with central region correct M1 Extend interpretation of given information eg. 3 overlapping labelled ovals with at least 5 regions correct M1 Method to communicate given information eg. 3 overlapping labelled ovals with all regions correct including outside A1 oe
(b)		$\frac{21}{44}$	P1 For correct process to identify correct regions in Venn diagram and divide by '44' A1

Q5.

Question	Scheme	Marks
(a)	B, W or T, W [accept $B \cup T, W$ or $B \cap T, W$] [Condone P(B), P(W) etc]	B1
	Since there is no overlap between the events or cannot happen together (o.e.)	B1 (2)
	(Accept comment in context e.g. "no one walks and takes the train")	
(b)	e.g. $P(B)=\frac{9}{25}$, $P(T)=\frac{8}{25}$, $P(B \cap T)=\frac{5}{25}$	M1
	$P(B \cap T) \neq P(B) \times P(T)$ [$0.2 \neq 0.36 \times 0.32 = 0.1152$ o.e.]	M1
	So B and T are not independent	A1cso (3)
(c)	[P(W) =] $\frac{7}{25}$ or 0.28	B1 (1)
(d)	[$P(B \cap T)$ =] $\frac{5}{25}$ or $\frac{1}{5}$ or 0.2	B1 (1)
		7

	Notes
(a)	1st B1 for a suitable pair. Do not accept universally exclusive pairs such as B and B' etc
	2nd B1 for any correct statement. Accept use of symbols e.g.: $B \cap W = \varnothing$ or $P(T \cap W) = 0$ etc
	But $T \cap W = 0$ is B0 (since it is not a correct statement)
(b)	1st M1 for an attempt at all required probabilities with labels for a suitable test (allow one error). Accept use of A and B as long as they can be identified as B and T by correct probabilities
	Must be probabilities not integers such as 5, 9, 8 etc for both these M marks
	2nd M1 for $P(B) \times P(T)$ evaluated (correct for their probabilities)
	or $P(B \cap T) \neq P(B) \times P(T)$ stated or implied in symbols or using their probabilities.
	or $P(B \mid T) \neq P(B)$ or $P(T \mid B) \neq P(T)$ stated or implied in symbols or using their probabilities.
	A1 for a conclusion of not independent. Requires all probabilities used to be correct and seen. This A mark is dependent on both Ms
	NB $P(B \mid T)=\frac{5}{8}$ & $P(B)=\frac{9}{25}$ or $P(T \mid B)=\frac{5}{9}$ & $P(T)=\frac{8}{25}$ seen, followed by a correct conclusion scores 3/3

Silver Questions

Calculator

The total mark for this section is 31

Q1

A jar contains 2 red, 1 blue and 1 green bead. Two beads are drawn at random from the jar without replacement.

(a) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities. State your probabilities clearly.

(3)

(b) Find the probability that a blue bead and a green bead are drawn from the jar.

(2)

(Total for Question 1 is 5 marks)

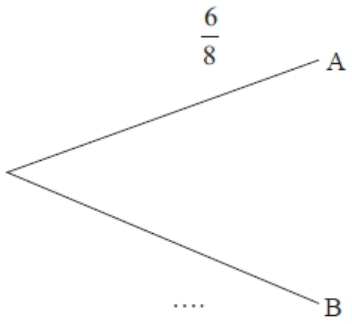
Q2

There are 8 counters in a box.
The letter A is on 6 of the counters.
The letter B is on the other 2 counters.

Sally takes at random a counter from the box.
She keeps the counter.
Then Tina takes at random a counter from the box.

(a) Complete the probability tree diagram.

Sally Tina



(3)

(b) Work out the probability that both Sally and Tina take a counter with the letter A on it.

(2)

(c) Work out the probability that at least one counter with the letter A on it is taken.

(3)

(Total for Question 2 is 8 marks)

Q3

The following shows the results of a wine tasting survey of 100 people.

96 like wine A,
93 like wine B,
96 like wine C,
92 like A and B,
91 like B and C,
93 like A and C,
90 like all three wines.

(a) Draw a Venn Diagram to represent these data.

(6)

Find the probability that a randomly selected person from the survey likes

(b) None of the three wines,

(1)

(c) Wine A but not wine B,

(2)

(d) Any wine in the survey except wine C,

(2)

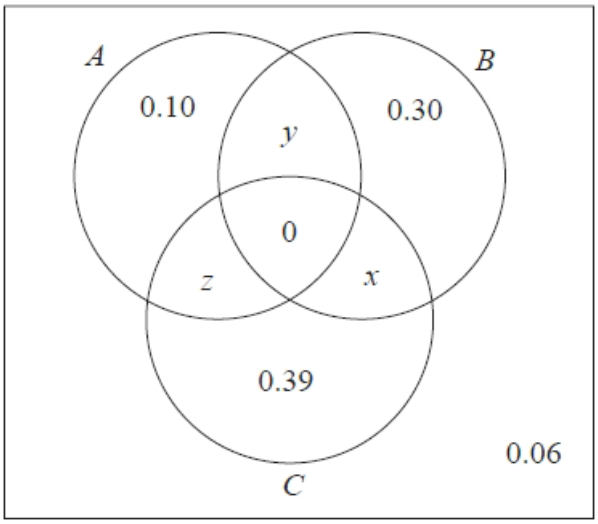
(e) Exactly two of the three kinds of wine.

(2)

(Total for Question 3 is 3 marks)

Q4

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



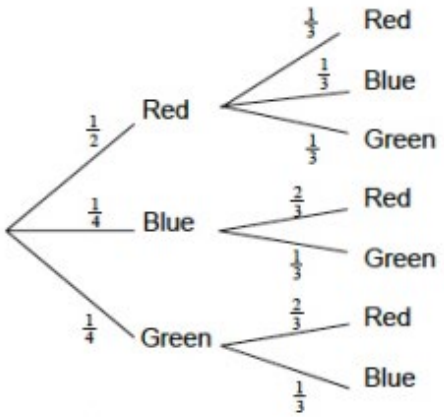
Events B and C are mutually exclusive.
Events A and C are independent.

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

(Total for Question 4 is 5 marks)

Silver Mark Scheme

Q1.

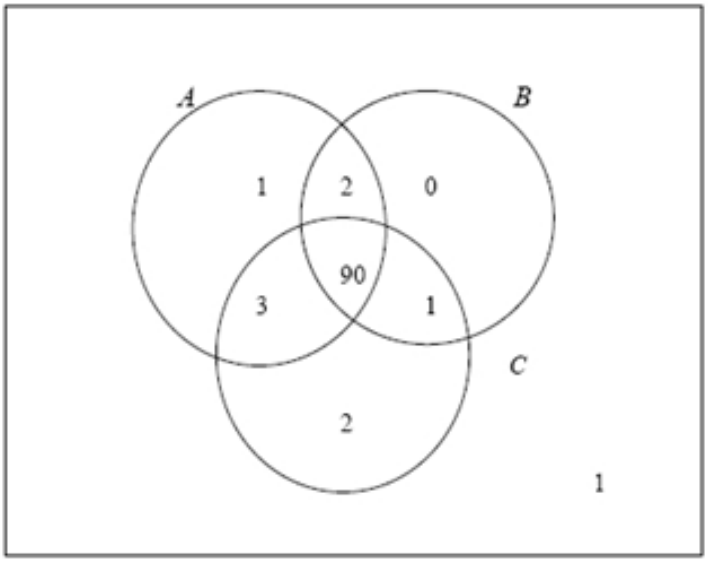
Question Number	Scheme	Marks
(a)		M1 A1 A1 (3)
(b)	P(Blue bead and a green bead) $=\left(\frac{1}{4}\times\frac{1}{3}\right)+\left(\frac{1}{4}\times\frac{1}{3}\right)=\frac{1}{6}$ (or any exact equivalent)	M1 A1 (2) **Total [5]**

(a)	M1 for shape and labels: 3 branches followed by 3,2,2 with some *R*, *B* and *G* seen Allow 3 branches followed by 3, 3, 3 if 0 probabilities are seen implying that 3, 2, 2 intended Allow blank branches if the other probabilities imply probability on blanks is zero Ignore further sets of branches 1st A1 for correct probabilities and correct labels on 1st set of branches. 2nd A1 for correct probabilities and correct labels on 2nd set of branches. (accept 0.33, 0.67 etc or better here)
(b)	M1 for identifying the 2 cases *BG* and *GB* and adding 2 products of probabilities. These cases may be identified by their probabilities e.g. $\left(\frac{1}{4}\times\frac{1}{3}\right)+\left(\frac{1}{4}\times\frac{1}{3}\right)$ NB $\frac{1}{6}$ (or exact equivalent) with no working scores 2/2
Special Case	With Replacement (This oversimplifies so do not apply Mis-Read: max mark 2/5) (a) B1 for 3 branches followed by 3, 3, 3 with correct labels and probabilities of $\frac{1}{2},\frac{1}{4},\frac{1}{4}$ on each. (b) M1 for identifying 2, possibly correct cases and adding 2 products of probabilities but A0 for wrong answer $\left[\left(\frac{1}{4}\times\frac{1}{4}\right)+\left(\frac{1}{4}\times\frac{1}{4}\right)\right]$ will be sufficient for M1A0 here but $\frac{1}{4}\times\frac{1}{2}+\ldots$ would score M0

Q2.

Question	Working	Answer	Mark	Notes
(a)		Probability tree	3	B1 for $\frac{2}{8}$ in the correct place B1 for $\frac{5}{7},\frac{2}{7};\frac{6}{7},\frac{1}{7}$ in the correct place on a probability tree B1 complete probability tree with labelling eg A, B etc.
(b)		$\frac{30}{56}$	2	M1 $\frac{6}{8}\times\frac{5}{7}$ A1 oe eg 0.5357.... or $\frac{15}{28}$
(c)		$\frac{54}{56}$	3	M1 for $\frac{6}{8}\times\frac{5}{7}$ or $\frac{6}{8}\times\frac{2}{7}$ or $\frac{2}{8}\times\frac{6}{7}$ oe eg 0.5357... or 0.214... M1 for $\frac{6}{8}\times\frac{5}{7}+\frac{6}{8}\times\frac{2}{7}+\frac{2}{8}\times\frac{6}{7}$ or $1-\left(\frac{2}{8}\times\frac{1}{7}\right)$ A1 oe eg 0.964.... or $\frac{27}{28}$

Q3.

Question Number	Scheme	Marks
(a)	**Diagram may be drawn with $B \subset (A \cup C)$ or with the 0 for $B \cap (A \cup C)'$ simply left blank** Accept decimals or probs. in Venn diagram  3cc **M1** 90,3,2,1 **A1** 1,(0),2 **M1A1** 1 outside **A1** Box **B1**	**(6)**
(b)	P(none)=0.01	**B1ft** **(1)**
(c)	P(A but not B)=0.04	**M1 A1ft** **(2)**
(d)	P(any wine but C)=0.03	**M1A1ft** **(2)**
(e)	P(exactly two)=0.06	**M1A1ft** **(2)**
		13

(a)	1st M1	for 3 closed, labelled curves that overlap. A1 for the 90, 3, 2 and 1	For M marks in (c) to (e) they must have a fraction
	2nd M1	for one of 1, 0 or 2 correct or a correct sum of 4 values for A, B or C	
	2nd A1	for all 7 values correct. Accept a blank instead of 0.	
	NB final mark is a B1 for the box not an A mark as on EPEN		
	In parts (b) to (f) full marks can be scored for correct answers or correct ft		
(b)	B1ft	Follow through their '1' from outside divided by 100	
(c)	M1	for correct expression eg $P(A \cup B) - P(B)$ or calculation e.g. 3 + 1 or 4 on top	
	A1	for a correct probability, follow through with their '3+1' from diagram	
(d)	M1	for correct expression or calculation e.g. 1+2+0 or 99-96 or 3 on top	
	A1	for a correct probability, follow through their '2+1+0' from diagram	
(e)	M1	for a correct expression or calculation e.g. 3+2+1 or 6 on top	

Q4.

Question	Scheme	Marks	AOs
	$x = 0$	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
	$[\sum p = 1]$ $0.06 + 0.3 + 0.39 + 0.1 + z + y[+x] = 1 \rightarrow [z + y[+x] = 0.15]$	M1	1.1b
	Solving (simultaneously) leading to $\underline{z = 0.13}$ $\underline{y = 0.02}$	A1	1.1b
			(5 marks)

Notes	
	B1: for $x = 0$, may be seen on Venn diagram
	M1: Identifying the probabilities required for independence and at least 2 correct These must be labelled If there are no labels, then this may be implied by $z = (0.1 + z + y)(0.39 + z\ [+x])$, allow one numerical slip Allow e.g. $P(A') = 0.39 + 0.30 + 0.06[+x] \quad P(C) = 0.39 + z[+x] \quad P(A' \text{ and } C) = 0.39$ [Not on spec. but you may see use of conditional probabilities]
	M1: Use of independence equation with their labelled probabilities in terms y, z [and x] All their probabilities must be substituted into a correct formula Sight of a correct equation e.g. $z = (0.1 + z + y)(0.39 + z\ [+x])$ scores M1M1
	M1: Using $\Sigma p = 1$ Implied by $[x +]\ y + z = 0.15$ or their $x + y + z = 0.15$ where x, y, and z are all probabilities or e.g. $P(A) = 0.25$
	A1: both $y = 0.02$ and $z = 0.13$

Gold Questions

Calculator

The total mark for this section is 31

Q1

On a randomly chosen day the probability that Bill travels to school by car, by bicycle or on foot is $\frac{1}{2}$, $\frac{1}{6}$ and $\frac{1}{3}$ respectively. The probability of being late when using these methods of travel is $\frac{1}{5}$, $\frac{2}{5}$ and $\frac{1}{10}$ respectively.

(a) Draw a tree diagram to represent this information.

(3)

(b) Find the probability that on a randomly chosen day

(i) Bill travels by foot and is late,

(ii) Bill is not late.

(4)

(Total for Question 1 is 7 marks)

Q2

There are 180 students at a college following a general course in computing. Students on this course can choose to take up to three extra options.

112 take systems support,

70 take developing software,

81 take networking,

35 take developing software and systems support,

28 take networking and developing software,

40 take systems support and networking,

4 take all three extra options.

(a) Draw a Venn diagram to represent this information.

(5)

A student from the course is chosen at random.
Find the probability that this student takes

(b) None of the three extra options,

(1)

(c) Networking only.

(1)

(Total for Question 2 is 7 marks)

Q3

A manufacturer carried out a survey of the defects in their soft toys. It is found that the probability of a toy having poor stitching is 0.03 and that a toy with poor stitching has a probability of 0.7 of splitting open. A toy without poor stitching has a probability of 0.02 of splitting open.

(a) Draw a tree diagram to represent this information.

(3)

(b) Find the probability that a randomly chosen soft toy has exactly one of the two defects, poor stitching or splitting open.

(3)

The manufacturer also finds that soft toys can become faded with probability 0.05 and that this defect is independent of poor stitching or splitting open. A soft toy is chosen at random.

(c) Find the probability that the soft toy has none of these 3 defects.

(2)

(d) Find the probability that the soft toy has exactly one of these 3 defects.

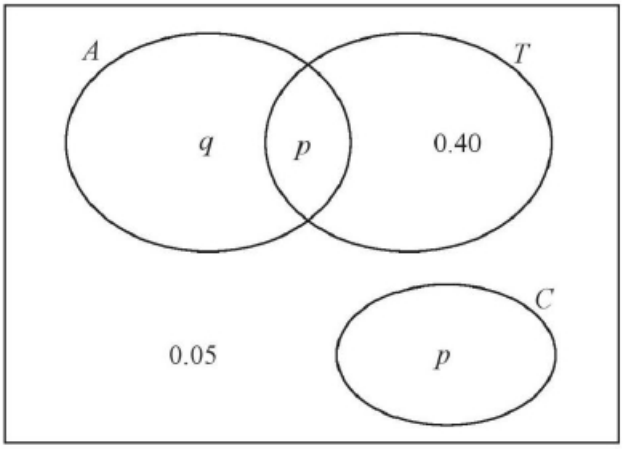
(4)

(Total for Question 3 is 12 marks)

Q4

The Venn diagram shows the probabilities for students at a college taking part in various sports.

A represents the event that a student takes part in Athletics.
T represents the event that a student takes part in Tennis.
C represents the event that a student takes part in Cricket.
p and q are probabilities.



The probability that a student selected at random takes part in Athletics or Tennis is 0.75

(a) Find the value of p.

(1)

(b) State, giving a reason, whether or not the events A and T are statistically independent. Show your working clearly.

(3)

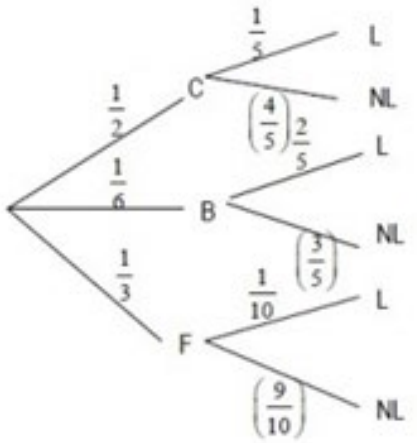
(c) Find the probability that a student selected at random does not take part in Athletics or Cricket.

(1)

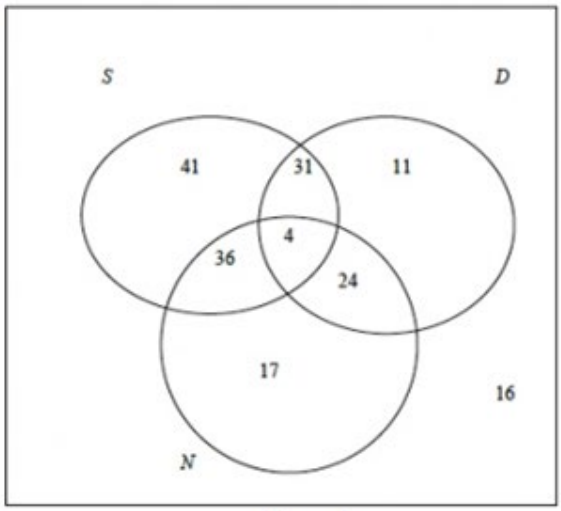
(Total for Question 4 is 5 marks)

Gold Mark Scheme

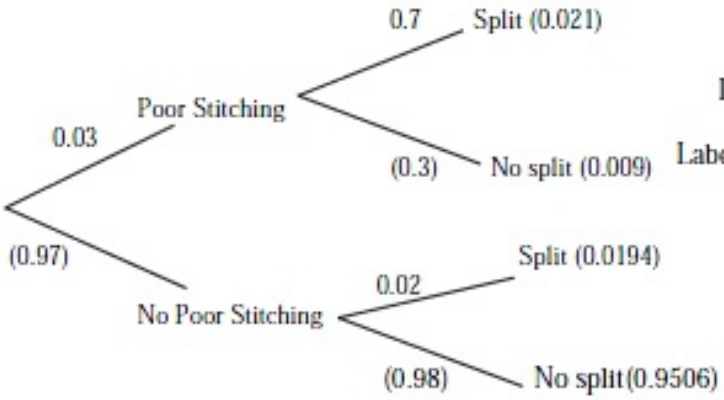
Q1.

Question Number	Scheme	Marks
(a)	 Correct tree All labels Probabilities on correct branches	B1 B1 B1 (3)
(b)(i)	$\frac{1}{3}\times\frac{1}{10}=\frac{1}{30}$ or equivalent	M1 A1 (2)
(ii)	$CNL + BNL + FNL = \frac{1}{2}\times\frac{4}{5}+\frac{1}{6}\times\frac{3}{5}+\frac{1}{3}\times\frac{9}{10}$ $= \frac{4}{5}$ or equivalent	M1 A1 (2) 7
Notes	Exact decimal equivalents required throughout if fractions not used e.g. 2(b)(i) 0.03 Correct path through their tree given in their probabilities award Ms 2(a) All branches required for first B1. Labels can be words rather than symbols for second B1. Probabilities from question enough for third B1 i.e. bracketed probabilities not required. Probabilities and labels swapped i.e. labels on branches and probabilities at end can be awarded the marks if correct. 2(b)(i) Correct answer only award both marks. 2(b)(ii) At least one correct path identified and attempt at adding all three multiplied pairs award M1	

Q2.

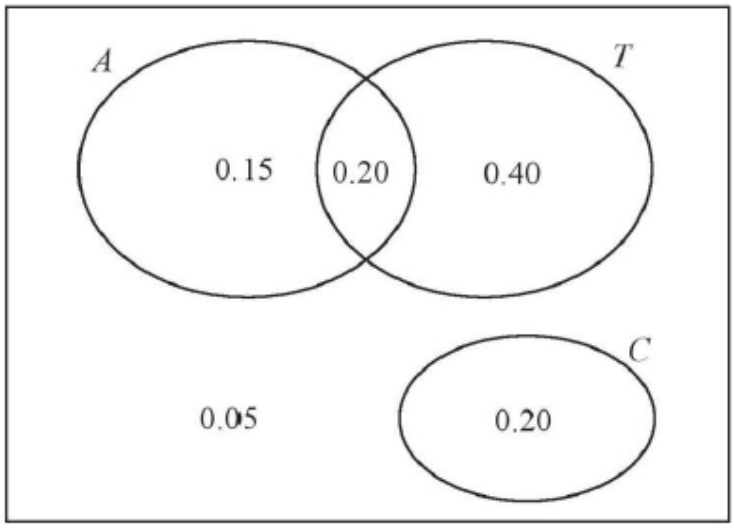
Question Number	Scheme	Marks
(a)	 3 closed curves and 4 in centre Evidence of subtraction 31,36,24 41,17,11 Labels on loops, 16 and box	M1 M1 A1 A1 B1 (5)
(b)	P(None of the 3 options)$=\frac{16}{180}=\frac{4}{45}$	B1ft (1)
(c)	P(Networking only)$=\frac{17}{180}$	B1ft (1) **7**
(a)	2nd M1 There may be evidence of subtraction in "outer" portions, so with 4 in the centre then 35, 40 28 (instead of 31,36,24) along with 33, 9, 3 can score this mark but A0A0 N.B. This is a common error and their "16" becomes 28 but still scores B0 in part (a)	
(b)	B1ft for $\frac{16}{180}$ or any exact equivalent. Can ft their "16" from their box. If there is no value for their "16" in the box only allow this mark if they have shown some working.	
(c)	B1ft ft their "17". Accept any exact equivalent	

Q3.

Question	Scheme	Marks
(a)		B1 B1 B1 (3)
(b)	P(Exactly one defect) = $0.03\times0.3+0.97\times0.02$ or $P(PS\cup Split)-2P(PS\cap Split)$ = [0.009 + 0.0194 =] **0.0284**	M1A1ft A1 cao (3)
(c)	P(No defects) = $(1-0.03)\times(1-0.02)\times(1-0.05)$ (or better) = 0.90307 **awrt 0.903**	M1 A1 cao (2)
(d)	P(Exactly one defect) = (b)$\times(1-0.05)$ + $(1-0.03)\times(1-0.02)\times0.05$ = "0.0284"$\times0.95+0.97\times0.98\times0.05$ = [0.02698 + 0.04753] = 0.07451 **awrt 0.0745**	M1 M1 A1ft A1 cao (4) **[12]**

	Notes
	Allow MR of 0.2 for 0.02 or 0.3 for 0.03 on tree diagram to score all M and A1ft marks only
(a)	1st B1 for 2 branch then 4 branch shape 2nd dB1 dep. on 1st B1 for labels showing stitching (accept letters) and 0.03 value correctly placed 3rd dB1 dep. on 1st B1 for labels showing splitting and 0.7 and 0.02 correctly placed [probabilities shown in brackets are not required and any such values given can be ignored in (a)]
(b)	M1 for $0.03\times p+0.02\times q$ where p and q follow from their tree diagram. Extra terms is M0 1st A1ft for a fully correct expression. Accept $1-0.7$ for 0.3 and $1-0.03$ for 0.97 Follow through 0.2 and 0.3 MR only
MR	0.2 for 0.02 $\rightarrow$0.203 or 0.3 for 0.03 $\rightarrow$ 0.104 or both $\rightarrow$ 0.23 should score M1A1A0 2nd A1 cao for 0.0284 only (or exact equivalent such as $\frac{71}{2500}$)
	Do not allow 0.5 as MR of 0.05 so no M or A marks in (c) or (d)
(c)	M1 for (their 0.97)$\times$(their 0.98)$\times(1-0.05)$ (or better) f.t. values from their tree diagram A1 cao for awrt 0.903
(d)	1st M1 for one correct triple (or correct ft from their tree) of: $[0.03\times0.3\times(1-0.05)]+[0.97\times0.02\times(1-0.05)]+[0.97\times0.98\times0.05]$ 2nd M1 for two correct triples or correct ft from their tree and adding or their (b) $\times(1-0.05)$ 1st A1ft for a fully correct expression or f.t. their (b) and 0.2 or 0.3 MR only
MR	0.2 for 0.02 $\rightarrow$0.23165 or 0.3 for 0.03 $\rightarrow$0.1331 or both $\rightarrow$0.2465 (or awrt 3sf) scores M1M1A1A0 2nd A1 cao for awrt 0.0745

Q4.

Question	Scheme	Marks	AOs
(a)	$p = [1 - 0.75 - 0.05 =]$ **0.20**	B1	1.1b
		(1)	
(b)	$q =$ **0.15**	B1ft	1.1b
	P(A) = 0.35 P(T) = 0.6 P(A and T) = 0.20 P(A) × P(T) = 0.21	M1	2.1
	Since 0.20 ≠ 0.21 therefore A and T are **not** independent	A1	2.4
		(3)	
			
(c)	P(not [A or C]) = **0.45**	B1	1.1b
		(1)	
			(5 marks)

Part	Notes
(a)	B1cao for $p = 0.20$
(b)	B1ft for use of their p and P(A or T) to find q i.e. 0.75 – "p" – 0.40 or $q = 0.15$
	M1 for the statement of all probabilities required for a suitable test and sight of any appropriate calculations required.
	A1 All probabilities correct, correct comparison and suitable comment.
(c)	B1cao for 0.45

Pearson

Topics 6 and 7

Statistical Distributions and Hypothesis Testing

Bronze, Silver, Gold

Worksheets for

AS Level Mathematics

ALWAYS LEARNING

PEARSON

Teacher Notes

These Bronze, Silver and Gold worksheets are designed to be used either straight after the content has been taught or as part of a skills gap analysis, especially as students move into year 13.

They are drawn from the latest specification questions and legacy questions. The papers are between 25 and 35 marks.

The topic number on this worksheet relates to the corresponding chapter number in the 'Pearson Edexcel AS and A Level Mathematics: Statistics and Mechanics Year 1/AS' textbook.

Quick Links

(Press Ctrl, as you click with your mouse to follow these links)

- Bronze Questions
- Bronze Mark Scheme
- Silver Questions
- Silver Mark Scheme
- Gold Questions
- Gold Mark Scheme

Extension and Enrichment

If you have students that have enjoyed the challenge of the Gold questions, then they should have a go at the more challenging question from our Advanced Extension Award (AEA) papers. The Mathematics AEA is a single, 3 hour non-calculator paper, taken at the end of year 13. It helps students to develop high level problem solving and proof skills. It is entirely based on the content of the A Level Mathematics Course. No extra material needs to be covered to take the AEA in Mathematics. A second important difference is that marks are awarded for the clarity and quality of their solution. Developing this key skill, alongside the extra problem-solving experience, can pay dividends in the way they approach A Level Mathematics and Further Mathematics problems.

More information about the Advanced Extension Award can be found here on the Pearson Edexcel Website, or here on the Maths Emporium

Bronze Questions

Calculator

The total mark for this section is 31

Q1

The probability of a bolt being faulty is 0.3. Find the probability that in a random sample of 20 bolts there are

(a) Exactly 2 faulty bolts,

(2)

(b) More than 3 faulty bolts.

(2)

These bolts are sold in bags of 20. John buys 10 bags.

(c) Find the probability that exactly 6 of these bags contain more than 3 faulty bolts.

(3)

(Total for Question 1 is 7 marks)

Q2

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 $P(4 \leq X < 8)$

(5)

(Total for Question 2 is 6 marks)

Q3

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.

(Total for Question 3 is 6 marks)

Q4

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)

(Total for Question 4 is 7 marks)

Q5

A test statistic has a distribution B(25, p).

Given that

$$H_0 : p = 0.5 \quad H_1 : p \neq 0.5$$

(a) Find the critical region for the test statistic such that the probability in each tail is as close as possible to 2.5%.

(3)

(b) State the probability of incorrectly rejecting H_0 using this critical region.

(2)

(Total for Question 5 is 5 marks)

Bronze Mark Scheme

Q1.

Question Number	Scheme Marks	
(a)	Let X be the random variable the number of faulty bolts	
	$P(X \leq 2) - P(X \leq 1) = 0.0355 - 0.0076 = 0.0279$ or $(0.3)^2(0.7)^{18}\frac{20!}{18!2!} = 0.0278$	M1 A1 (2)
(b)	$1 - P(X \leq 3) = 1 - 0.1071 = 0.8929$ **or** $1-(0.3)^3(0.7)^{17}\frac{20!}{17!3!}-(0.3)^2(0.7)^{18}\frac{20!}{18!2!}-(0.3)(0.7.)^{19}\frac{20!}{19!1!}-(0.7)^{20}$	M1 A1 (2)
(c)	$\frac{10!}{4!6!}(0.8929)^6(0.1071)^4 = 0.0140.$	M1A1√A1 (3)

Notes	
(a)	**M1** Either attempting to use $P(X \leq 2) - P(X \leq 1)$ or attempt to use binomial and find $p(X = 2)$. Must have $(p)^2(1-p)^{18}\frac{20!}{18!2!}$, with a value of p **A1** awrt 0.0278 or 0.0279.
(b)	**M1** Attempting to find $1 - P(X \leq 3)$ **A1** awrt 0.893
(c)	**M1** for $k\ (p)^6(1-p)^4$. They may use any value for p and k can be any number or ${}^nC_6p^6(1-p)^{n-6}$ **A1**√ $\frac{10!}{4!6!}(\textit{their part b})^6(1-\textit{their part b})^4$ may write ${}^{10}C_6$ or ${}^{10}C_4$ **A1** awrt 0.014

Q2.

Question	Scheme	Marks	AOs
(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)

Notes	
(a)	Continuous uniform is B0
(b)	**B1:** for identifying correct model, B(28, 0.2) allow B, bin or binomial may be implied by one correct answer or sight one correct probability i.e. awrt 0.678, awrt 0.818 or awrt 0.160 B(0.2, 28) is B0 unless it is used correctly
(i)	**M1:** Writing or using $1 - P(X \leq 6)$ or $1 - P(X < 7)$ **A1:** awrt 0.322 (correct answer only scores M1A1)
(ii)	**M1:** Writing or using $P(X \leq 7) - P(X \leq 3)$ or $P(X < 8) - P(X < 4)$ or $P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)$ Condone P(4) as $P(X = 4)$, etc. **A1:** awrt 0.658 (correct answer only scores M1A1)

Q3.

Question	Scheme	Marks	AOs
	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)

Notes		
	M1:	A fully correct method with all the required steps. For gaining 2 correct equations with at least one correct(allow if unsimplified). Attempting to solve to find a value of c followed by **correct method** to find the probability
	B1:	Forming a correct equation from the information given in the question
	B1:	A correct equation using the sum of the probabilities equals 1
	M1:	Correct method for solving 2 equations to find c Implied by $c=\frac{1}{12}$
	M1:	Recognising the ways to get a total of 6. Condone missing arrangments or repeats. Do not ignore extras written unless ignored in the calculation. May be implied by $m\times"\frac{1}{12}"\times0.15+n\times0.1^2$ where m and n are positive integers
	A1cso:	Cao 0.035, $\frac{7}{200}$ oe

Q4.

Qu	Scheme	Marks	AO
(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)	

Q5.

Question Number	Scheme	Marks
(a)	$X \sim B(25,0.5)$ may be implied by calculations in part a or b	M1
	$P(X \leq 7) = 0.0216$	
	$P(X \geq 18) = 0.0216$	
	CR $X \leq 7; \cup X \geq 18$	A1,A1 (3)
(b)	P(rejecting H_0) = 0.0216 + 0.0216	M1
	= 0.0432 awrt 0.0432/0.0433	A1 (2)
		Total 5

Notes

2(a) M1 - Using B(25,0.5) – may be implied by a correct critical region or by calculations in part a or b
Note Just seeing either $P(X \leq 7)$ or $P(X \geq 18)$ scores M1 A0 A0.
You may need to check their probabilities in the tables for values other than 7 or 18.
1st A1 – also allow $X < 8$ or $[0,7]$ or $0 \leq X \leq 7$ or $0 \leq X < 8$ oe e.g. [0, 8) or a full list
DO NOT allow CRs given as $P(X \leq 7)$ or 7 – 0 for the A mark.
2nd A1 – also allow $X > 17$ or $[18,25]$ or $18 \leq X \leq 25$ or $17 < X \leq 25$ oe e.g. (17, 25] or a full list
DO NOT allow CRs given as $P(X \geq 18)$ or 18 - 25 for the A mark.
SC $7 \geq X \geq 18$ gains M1 A1 A0.

(b) M1 – adding their two critical regions' probabilities together or may be awarded for awrt 0.0432
If they add their critical regions' probabilities and then go on and get a different probability as their answer then it is M0A0
e.g. 0.0216 + 0.0216 = 0.0432 then 0.05 – 0.0432 = 0.0068 gets M0 A0
e.g. 0.0216 + 0.0216 = 0.0432 < 0.05 reject H_0 gets M1 A1
e.g. 0.0216 + 0.0216 = 0.0432 so probability of rejecting H_0 is 1 – 0.0432 = 0.9568 gets M0 A0

Silver Questions

Calculator

The total mark for this section is 34

Q1

Linda regularly takes a taxi to work five times a week. Over a long period of time she finds the taxi is late once a week. The taxi firm changes her driver and Linda thinks the taxi is late more often. In the first week, with the new driver, the taxi is late 3 times.

You may assume that the number of times a taxi is late in a week has a Binomial distribution.

Test, at the 5% level of significance, whether or not there is evidence of an increase in the proportion of times the taxi is late. State your hypotheses clearly.

(Total for Question 1 is 7 marks)

Q2

A potter believes that 20% of pots break whilst being fired in a kiln. Pots are fired in batches of 25.

(a) Let X denote the number of broken pots in a batch. A batch is selected at random. Using a 10% significance level, find the critical region for a two tailed test of the potter's belief. You should state the probability in each tail of your critical region.

(4)

The potter aims to reduce the proportion of pots which break in the kiln by increasing the size of the batch fired. He now fires pots in batches of 50. He then chooses a batch at random and discovers there are 6 pots which broke whilst being fired in the kiln.

(b) Test, at the 5% level of significance, whether or not there is evidence that increasing the number of pots in a batch has reduced the percentage of pots that break whilst being fired in the kiln. State your hypotheses clearly.

(5)

(Total for Question 2 is 9 marks)

Q3

In a manufacturing process 25% of articles are thought to be defective. Articles are produced in batches of 20

(a) A batch is selected at random. Using a 5% significance level, find the critical region for a two tailed test that the probability of an article chosen at random being defective is 0.25
You should state the probability in each tail which should be as close as possible to 0.025

(5)

The manufacturer changes the production process to try to reduce the number of defective articles. She then chooses a batch at random and discovers there are 3 defective articles.

(b) Test at the 5% level of significance whether or not there is evidence that the changes to the process have reduced the percentage of defective articles. State your hypotheses clearly.

(5)

(Total for Question 3 is 10 marks)

Q4

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)

(Total for Question 4 is 8 marks)

Silver Mark Scheme

Q1.

Question Number	Scheme	Marks
	One tail test Method 1 $H_0 : p = 0.2$ $H_1 : p > 0.2$	B1 B1
	$X \sim B(5, 0.2)$ may be implied	M1
	$P(X \geq 3) = 1 - P(X \leq 2)$ $[P(X \geq 3) = 1 - 0.9421 = 0.0579]$ att $P(X \geq 3)$ $P(X \geq 4)$ $= 1 - 0.9421$ $P(X \geq 4) = 1 - 0.9933 = 0.0067$	M1
	$= 0.0579$ CR $X \geq 4$ awrt 0.0579	A1
	$0.0579 > 0.05$ $3 \leq 4$ or 3 is not in critical region or 3 is not significant	M1
	(Do not reject H_0.) There is insufficient evidence at the 5% significance level that there is an increase in the number of times the taxi/driver is late. **Or** Linda's claim is not justified	B1 (7) (Total 7)
	Method 2 $H_0 : p = 0.2$ $H_1 : p > 0.2$	B1 B1
	$X \sim B(5, 0.2)$ may be implied	M1
	$P(X < 3) =$ $[P(X < 3) = 0.9421]$ att $P(X < 3)$ $P(X < 4)$ $P(X < 4) = 0.9933$	
	0.9421 CR $X \geq 4$ awrt 0.942	M1A1
	$0.9421 < 0.95$ $3 \leq 4$ or 3 is not in critical region or 3 is not significant	M1
	(Do not reject H_0.) There is insufficient evidence at the 5% significance level that there is an increase in the number of times the taxi/driver is late. **Or** Linda's claim is not justified	B1 (7)

Two tail test

Method 1

$H_0 : p = 0.2$

$H_1 : p \neq 0.2$

B1

B0

M1

$X \sim X \sim B(5, 0.2)$ may be implied

M1

$P(X \geq 3) = 1 - P(X \leq 2)$ $[P(X \geq 3) = 1 - 0.9421 = 0.0579]$ att $P(X \geq 3)$ $P(X \geq 4)$

$= 1 - 0.9421$ $P(X \geq 4) = 1 - 0.9933 = 0.0067$

A1

$= 0.0579$ CR $X \geq 4$ awrt 0.0579

M1

$0.0579 > 0.025$ $3 \leq 4$ or 3 is not in critical region or 3 is not significant

B1

(Do not reject H_0.) There is insufficient evidence at the 5% significance level that there is an increase in the number of times the taxi/driver is late.

Or Linda's claim is not justified

(7)

Method 2

$H_0 : p = 0.2$

$H_1 : p \neq 0.2$

B1

B0

M1

$X \sim X \sim B(5, 0.2)$ may be implied

$P(X < 3) =$ $[P(X < 3) = 0.9421]$ att $P(X < 3)$ $P(X < 4)$

$P(X < 4) = 0.9933$

0.9421 CR $X \geq 4$ awrt 0.942

M1A1

$0.9421 < 0.975$ $3 \leq 4$ or 3 is not in critical region or 3 is not significant

M1

Do not reject H_0. There is insufficient evidence at the 5% significance level that there is an increase in the number of times the taxi/driver is late.

Or Linda's claim is not justified

B1

(7)

Special Case

If they use a probability of $\frac{1}{7}$ throughout the question they may gain B1 B1 M0 M1 A0 M1 B1.

NB they must attempt to work out the probabilities using $\frac{1}{7}$

Q2.

Question Number	Scheme		Marks
	Allow any letter instead of X or c for this question		
(a)	$X \sim B(25, 0.2)$	**M1** Writing or using B(25,0.2)or B(25,1/5) [allow Po(5)] May be written in full or implied by a correct CR (allow written as a probability statement)	M1
	$[P(X \geq 9) =] 0.0468$ $[P(X \leq 1) =] 0.0274$	**1st A1** both awrt 0.0468 and awrt 0.0274 seen.	A1
	$X = [0 \leq] \ X \leq 1$	**2nd A1** $X \leq 1$ or $X < 2$ or $0 \leq X \leq 1$ or [0,1] or 0,1 or equivalent statements. $X \leq c$ and $c = 1$	A1
	$9 \leq X \ [\leq 25]$	**3rd A1d** dependent on seeing a probability from the B(25, 0.2) and $X \geq 9$ or $X > 8$ or $9 \leq X \leq 25$ or 9,10,11,12,13,14,15,16,17,18,19,20,21,22, 23,24,25 or [9,25] or equivalent statements. $X \geq c$ and $c = 9$	A1d
	NB These two final 2 A marks must be for statements with "X" only(or list) – not in probability statements SC If a probability from the B(25, 0.2) is seen **and** they either have both CR correct but written as probability statements **or** the CR is written as $1 \geq X \geq 9$ they get A1 A0 for final 2 marks		**(4)**
(b)	$H_0: p = 0.2$ $H_1: p < 0.2$	**B1** both hypotheses with p or π and clear which is H_0 and which is H_1	**B1**
	$P(X \leq 6) = 0.1034$ or CR $X \leq 5$	**1st M1** writing or using B(50, 0.2) and writing or using $P(X \leq 6)$ or $P(X \geq 7)$ on its own. May be implied by a correct CR	**M1**
		1st A1 awrt 0.103. Allow CR $X \leq 5$ or $X < 6$. or if not using CR allow awrt 0.897.	**A1**
	Insufficient evidence to reject H_0, Accept H_0, Not significant. 6 does not lie in the Critical region.	**2nd M1** dependent on previous M being awarded. A correct statement (do not allow if there are contradicting non-contextual statements). ft their Prob/CR compared with 0.05/6/(0.95 if using 0.8979). Do not follow through their hypotheses	**M1d**
	No evidence that increasing the batch size has **reduced** the **percentage** of broken **pots (oe)** **or** evidence that there is **no change** in the **percentage** of broken **pots (oe)**	**2nd A1cso** Conclusion must contain the words **reduced/ no change/not affect oe number/percentage/proportion/ probability oe**, and **pots**. All previous marks must be awarded for this mark to be awarded. Do **not** allow the potters claim /belief is wrong/true **NB** Correct contextual statement on its own scores M1A1	**A1cso** **(5)**
			(Total 9)

Q3.

Question Number	Scheme	Marks
(a)	$X \sim B(20,0.25)$ $P(X \geq 10) = 1 - 0.9861 = 0.0139$ $P(X \leq 1) = 0.0243$ $(0 \leq)X \leq 1 \cup 10 \leq X (\leq 20)$	M1 A1 A1 A1A1 (5)
(b)	$H_0: p = 0.25$ $H_1: p < 0.25$ $X \sim B(20,0.25)$ $P(X \leq 3) = 0.2252$ or CR $X \leq 1$ Insufficient evidence to reject H_0, Accept H_0, Not significant. 3 does not lie in the Critical region. No evidence that the **changes** to the process have **reduced** the **percentage** of **defective articles (oe)**	 B1 M1A1 M1d A1cso (5) Total 10 marks
	Notes	
(a)	M1 using B(20,0.25) may be implied by a correct CR (allow written as a probability statement) 1st A1 awrt 0.0139 2nd A1 awrt 0.0243 3rd A1 $X \leq 1$ or $0 \leq X \leq 1$ or [0,1] or 0,1 or equivalent statements 4th A1 $X \geq 10$ or $10 \leq X \leq 20$ or 10,11,12,13,14,15,16,17,18,19,20 or [10,20] or equivalent statements **NB** These two A marks must be for statements with X (any letter) only – not in probability statements and **SC** for CR written as $1 \geq X \geq 10$ gets A1 A0	
(b)	B1 both hypotheses with p 1st M1 using B(20, 0.25) and finding $P(X \leq 3)$ or $P(X \geq 4)$ may be implied by a correct CR 1st A1 0.2252 (allow 0.7748) if not using CR or CR $X \leq 1$ or $X < 2$ 2nd M1dependent on previous M being awarded. A correct statement (do not allow if there are contradicting non contextual statements) A1cso Conclusion must contain the words **changes/new process oe**, **reduced oe number/percentage oe** , and **defective articles/defectives**. There must be no incorrect working seen.	

Q4.

Question	Scheme	Marks	AOs
(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)

Notes		
(a)	**M1:**	For selecting the right model
	A1:	awrt 0.178
(b)	**M1:**	For $5 \times ("\text{their}(a)") \times ("1 - \text{their}(a)")^4$
	A1:	awrt 0.406
(c)	**B1:**	for correctly stating both hypotheses in terms of p or π Accept $p = 0.16\dot{}$
	M1:	For selecting a suitable model. May be implied by a correct probability or CR
	A1:	Correct probability statement and answer of 0.02 or better (0.02318…) (CR $R \geq 11$ and either $P(R \leq 9) = 0.9450$ or $P(R \leq 10) = 0.9768$ or $1 - P(R \leq 10) = 0.0232$)
	A1:	Dependent on M1A1 but can ignore hypotheses. For conclusion in context supporting **Rowan's** belief / **Rowan** is a better sales person
		Do not accept Rowan can reject H_0

Gold Questions

Calculator

The total mark for this section is 31

Q1

(a) The discrete random variable $X \sim B(40, 0.27)$

Find $P(X \geq 16)$

(2)

Past records suggest that 30% of customers who buy baked beans from a large supermarket buy them in single tins. A new manager suspects that there has been a change in the proportion of customers who buy baked beans in single tins. A random sample of 20 customers who had bought baked beans was taken.

(b) Write down the hypotheses that should be used to test the manager's suspicion.

(1)

(c) Using a 10% level of significance, find the critical region for a two-tailed test to answer the manager's suspicion. You should state the probability of rejection in each tail, which should be less than 0.05

(3)

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

(1)

One afternoon the manager observes that 12 of the 20 customers who bought baked beans, bought their beans in single tins.

(e) Comment on the manager's suspicion in the light of this observation.

(1)

Later it was discovered that the local scout group visited the supermarket that afternoon to buy food for their camping trip.

(f) Comment on the validity of the model used to obtain the answer to part (e), giving a reason for your answer.

(1)

(Total for Question 1 is 9 marks)

Q2

The proportion of houses in Radville which are unable to receive digital radio is 25%. In a survey of a random sample of 30 houses taken from Radville, the number, X, of houses which are unable to receive digital radio is recorded.

(a) Find $P(5 \leqslant X < 11)$

(3)

A radio company claims that a new transmitter set up in Radville will reduce the proportion of houses which are unable to receive digital radio. After the new transmitter has been set up, a random sample of 15 houses is taken, of which 1 house is unable to receive digital radio.

(b) Test, at the 10% level of significance, the radio company's claim. State your hypotheses clearly.

(5)

(Total for Question 2 is 8 marks)

Q3

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.

$H_0 : p = 0.15$
$H_1 : p \geq 0.15$
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)

(Total for Question 3 is 6 marks)

Q4

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 $P(Y - X \leq 4)$

(3)

(Total for Question 4 is 8 marks)

Gold Mark Scheme

Q1.

Question	Scheme	Marks	AOs
(a)	$P(X \geqslant 16) = 1 - P(X \leqslant 15)$	M1	1.1b
	$= 1 - 0.949077\ldots$ = awrt **0.0509**	A1	1.1b
		(2)	
(b)	$H_0 : p = 0.3$ $H_1 : p \neq 0.3$ (Both correct in terms of p or π)	B1	2.5
		(1)	
(c)	$[Y \sim B(20, 0.3)]$ sight of $P(Y \leqslant 2) = 0.0355$ or $P(Y \leqslant 9) = 0.9520$	M1	2.1
	Critical region is $\{Y \leqslant 2\}$ or (o.e.)	A1	1.1b
	$\{Y \geqslant 10\}$ (o.e.)	A1	1.1b
		(3)	
(d)	$[0.0355 + (1 - 0.9520)] = 0.0835$ or **8.35%**	B1ft	1.1b
		(1)	
(e)	(Assuming that the 20 customers represent a random sample then) 12 is in the CR so the manager's suspicion is supported	B1ft	3.2a
		(1)	
(f)	e.g. (e) requires the 20 customers to be a random sample or independent and the members of the scout group may invalidate this so binomial distribution would not be valid (and conclusion in (e) is probably not valid)	B1	3.5a
		(1)	
			(9 marks)

Part	Notes
(a)	M1 for dealing with $P(X \geqslant 16)$ – they need to use cumulative prob. function on calc.
	A1 awrt 0.0509 (from calculator)
(b)	B1 for both hypotheses in terms of p or π and H_1 must be 2-tail
(c)	M1 for correct use of tables to find probability associated with critical value.
	1st A1 for the correct lower limit of the CR. Do not award for $P(Y \leqslant 2)$
	2nd A1 for the correct upper limit.
(d)	B1ft ft on their 0.0355 and (1 – their 0.9520) provided each probability is less than 0.05
(e)	B1ft for a comment that relates 12 to their CR and makes a consistent comment relating this to the manager's suspicion
(f)	B1 for a comment that: gives a suitable reason based on lack of independence or the sample not being random so the binomial model is not valid

Q2.

Question Number	Scheme		Marks
(a)		**notes**	
	$X \sim B(30, 0.25)$	B1: using B(30, 0.25)	B1
	$P(X \leq 10) - P(X \leq 4) = 0.8943 - 0.0979$	M1: using $P(X \leq 10) - P(X \leq 4)$ or $P(X \geq 5) - P(X \geq 11)$ oe	M1 A1
	$= 0.7964$	A1: awrt 0.796	
	NB a correct answer gains full marks		

(b)	$H_0 : p = 0.25$ $H_1 : p < 0.25$	B1: Both hypotheses correct, labelled H_0 or NH or H_n and H_1or AH or H_a , must use p or $p(x)$ or π	B1
	B(15, 0.25)	M1: for using B(15, 0.25)	M1 A1
	$P(X \leq 1) = 0.0802$	A1: awrt 0.0802 or CR $X \leq 1$ (allow $P(X \geq 2) = 0.9198$)	
	NB: Allow M1 A1 for a correct CR with no incorrect working		
	Reject H_0 or Significant or 1 1ies in the critical region	M1: A correct statement – do not allow contradictory non contextual statements. Follow through their Probability/CR (for 1 or 2 tail test). If no H_1 given then M0. Ignore their comparison. For a probabillity < 0.5, statement must be correct compared to 0.1 for 1 tail test and 0.05 for 2 tailed test or if the probability > 0.5, statement must be correct compared to 0.9 for 1 tail test and 0.95 for 2 tailed test.	dM1 A1cso
	There is evidence that the radio **company's** claim is true. **Or** The new transmitter will reduce the proportion of houses unable to receive **radio**	A1: cso (all previous marks awarded) and a correct statement containing the word **company** if writing about the claim or **radio** if full context.	

Q3.

Question	Scheme	Marks	AOs
(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)

Notes	
(a)	**B1:** Identifying that $\geq$ should be $>$ in the alternative hypothesis **B1:** Identifying that $P(X = 8)$ should be $P(X \geq 8)$ Stating $P(X = 8)$ is incorrect on its own is insufficient Check for errors identified and corrected next to the question
(b)	**B1:** Will affect conclusion **and** correct supporting reason
(c)	**M1:** For use of tables to find probability associated with critical value [$P(X \leq 8)$ or $P(X \geq 9)$ with B(30, 0.15) (may be implied by either correct probability awrt 0.97 or awrt 0.03) or by the correct CR] **A1:** $[30 \geq]X \geq 9$ o.e. e.g. $X > 8$ Allow '9 or more' or 'CR $\geq$ 9'
(d)	**B1ft:** awrt 0.0278 (allow awrt 2.78%) or correct ft their one-tailed upper CR from B(30, 0.15) to 3s.f.

Q4.

Qu	Scheme	Marks	AO
(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 $A \sim B(60, "0.35")$	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)	$Y-X\leqslant 4 \Rightarrow \frac{12}{X}-X\leqslant 4$ <u>or</u> $12-X^2\leqslant 4X$ (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
	$P(Y-X\leqslant 4)=P(X\geqslant 2)=0.35+0.15+0.35$ = **<u>0.85</u>**	A1 (3)	3.2a
		(8 marks)	

Notes

(a) M1 for using the given information to obtain $P(X=4)$
Award for statement $P(X=4)=P(X=2)$ <u>or</u> writing $P(X=4)=0.35$
A1 for getting fully correct distribution (any form that clearly identifies probs)
e.g. can be list $P(X=1)=0.15$, $P(X=3)=\ldots$etc or as a probability function
$P(X=x)=\begin{cases}0.15 & x=1,3\\ 0.35 & x=2,4\end{cases}$
[Condone missing $P(X=2)$ as this is given in QP]

(b) B1 for selecting a suitable model, sight of B(60, their 0.35) o.e. in words
f.t. their $P(X=4)$ from part (a).
Can be implied by $P(A\leqslant 30)$ = awrt 0.9941 or final answer = awrt 0.00589
M1 for using their model and interpreting "more than half"
Need to see $1-P(A\leqslant 30)$. Can be implied by awrt 0.00589
Can ignore incorrect LHS such as $P(A\geqslant 30)$
A1 for awrt 0.00589

(c) 1st M1 for translating the prob. problem into a <u>correct</u> mathematical inequality
Just an inequality in 1 variable. May be inside a probability statement.
ALT **Table of values:**

X	1	2	3	4
Y	12	6	4	3

<u>or</u> values of $Y-X=11, 4, 1, -1$

2nd M1 for solving the inequality leading to a range of values, allow 1 or 2 slips
May be a quadratic or cubic but must lead to a set of values of X or $Y-X$
ALT **Table or values:** They must state clearly which values are required
Both Ms can be implied by a correct answer (or correct ft of their distb'n)
A1 for interpreting the inequality and solving the problem i.e. 0.85 cao