



eClassroom

INTERNATIONAL A LEVEL MATHEMATICS • STATISTICS 1

Statistics 1

Exam Questions by Topic: Mark Scheme

Pearson Edexcel International Advanced Level (WST01/01) • Calculator and Statistical Tables allowed

Every question in the booklet, marked line by line the way an examiner marks it

Chapters (tap a chapter to jump to it)	Qs	Marks	Page	My score
1 Mathematical Modelling Not examined on its own: modelling ideas are tested within Chapters 2 to 7	—	—	Not applicable	
2 Measures of Location and Spread Averages, quartiles, interpolation, standard deviation, coding	10	83	3	
3 Representations of Data Histograms, stem and leaf, box plots, outliers, skewness, comparisons	10	80	10	
4 Probability Venn and tree diagrams, conditional probability, independence	10	80	17	
5 Correlation and Regression PMCC, least squares regression, interpretation, coding	10	84	24	
6 Discrete Random Variables Probability and cumulative distributions, $E(X)$, $\text{Var}(X)$, uniform	10	82	32	
7 The Normal Distribution Standardising, tables, finding μ and σ , conditional problems	10	83	39	

[Question booklet](#)

[Free courses](#)

[More worksheets](#)



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

How to use this mark scheme

The same marking rules Pearson examiners use for WST01.

Mark every line, not just the answer. Method marks are given for a correct method even when the final answer is wrong. If you lost marks, find the *first* line where your working differs from the scheme. That is the thing to fix.

Code	Meaning
M1	Method mark: for knowing a correct method and attempting to apply it. Small slips in the working do not lose this mark.
dM1	Dependent method mark: can only be given if the method mark before it has been given.
A1	Accuracy mark: for a correct answer or line of working. It can only be given if the method mark it depends on has been given.
B1	Independent mark: for a correct result or statement, whatever method was used.
*	The answer is given in the question. Every step must be shown, and the answer must follow from correct working (cso).
cao	Correct answer only.
cso	Correct solution only: no errors anywhere in the working.
ft	Follow through: the mark can be given for correct work using your own earlier (wrong) answer.
oe	Or equivalent: any correct equivalent form is accepted.

“Solutions relying on calculator technology are not acceptable.” When a question says this, a correct final answer with no algebraic working scores **no marks**. The notes under each scheme line say when this applies.



2

Chapter 2: Measures of Location and Spread

Mark scheme. Marks are shown on the right of each line.

Question 1

6 marks

Part	Scheme	Marks
(a)	Cumulative frequencies 7, 19, 29, 35, 38, 40; the 20th and 21st values give $\text{median} = 2$ Q_1: 10th and 11th values = 1; Q_3: 30th and 31st values = 3 $\text{IQR} = 2$	B1 M1 A1 (3)
(b)	$\sum fx = 72$, so $\bar{x} = 1.8$ $\sigma^2 = \frac{204}{40} - 1.8^2 = 1.86$ $\sigma = 1.36$ <i>The common error is to forget to subtract $\bar{x}^2$, or to give the variance instead of the standard deviation.</i>	B1 M1 A1 (3)
Total for Question 1: 6 marks		

Question 2

5 marks

Part	Scheme	Marks
(a)	Boundaries 179.5 and 194.5 , midpoint 187	B1 (1)
(b)	Midpoints 154.5, 164.5, 174.5, 187, 202: $\frac{13905}{80}$ 173.8 cm (173.8125)	M1 A1 (2)
(c)	40th value, in the 170–179 class: $169.5 + \frac{40 - 30}{30} \times 10$ 172.8 cm <i>Using 170 instead of the class boundary 169.5 scores M1 A0. Using $n+1$ (40.5th value) is condoned.</i>	M1 A1 (2)
Total for Question 2: 5 marks		



Question 3

5 marks

Part	Scheme	Marks
(a)	$\bar{x} = 46$	B1
	$\sigma^2 = \frac{53850}{25} - 46^2 = 38$	M1
	$\sigma = 6.16$	A1
		(3)
(b)	Corrected $\sum x = 1150 + 20 = 1170$	M1
	46.8 minutes	A1
		(2)
Total for Question 3: 5 marks		

Question 4

6 marks

Part	Scheme	Marks
(a)	$\bar{y} = \frac{64}{16} = 4$ and $\bar{x} = 5\bar{y} + 200$	M1
	220 g	A1
		(2)
(b)	$\sigma_y^2 = \frac{544}{16} - 4^2 = 18$	M1
	$\sigma_x = 5\sigma_y = 5\sqrt{18}$	M1
	21.2 g	A1
	The “−200” does not affect the spread. Using 5^2 on the standard deviation, or dividing by 5, scores M1 M0 A0.	(3)
(c)	$25 \times 18 = 450$	B1
		(1)
Total for Question 4: 6 marks		



Question 5

5 marks

Part	Scheme	Marks
(a)	Totals $12 \times 64 = 768$ and $8 \times 59 = 472$ 62	M1 A1 (2)
(b)	$\sum x^2 = n(\sigma^2 + \bar{x}^2): 12(25 + 4096) + 8(16 + 3481) = 77\,428$ $\sigma^2 = \frac{77\,428}{20} - 62^2 = 27.4$ 5.23	M1 M1 A1 (3)
Total for Question 5: 5 marks		

Question 6

12 marks

Part	Scheme	Marks
(a)	Median: 30th value, $140 + \frac{30 - 21}{18} \times 10 = 145$ $\text{Median} = 145 \text{ g}$ $Q_1 = 120 + \frac{15 - 6}{15} \times 20 = 132$ and $Q_3 = 150 + \frac{45 - 39}{13} \times 20 = 159.23 \dots$ $\text{IQR} = 27.2 \text{ g}$	M1 A1 M1 A1 (4)
(b)	Midpoints 110, 130, 145, 160, 185: $\sum fx = 8780$ $\bar{x} = 146 \text{ g}$ (146.33...) $\sigma^2 = \frac{1\,311\,150}{60} - 146.33 \dots^2 = 439.05 \dots$ $\sigma = 21.0 \text{ g}$	M1 A1 M1 A1 (4)
(c)	$\frac{170 - 160}{20} \times 13 + 8$ 14.5 (accept 14 or 15)	M1 A1 (2)
(d)	Without replacement: $\frac{18}{60} \times \frac{17}{59}$ 0.0864 $\left(\frac{18}{60}\right)^2 = 0.09$ (with replacement) scores M0 A0: the second potato cannot be the first one again.	M1 A1 (2)
Total for Question 6: 12 marks		



Question 7

11 marks

Part	Scheme	Marks
(a)	$\bar{y} = \frac{58}{40} = 1.45$ $\bar{w} = 4 \times 1.45 + 50 = 55.8$ *	M1 A1* (2)
(b)	$\sigma_y^2 = \frac{262}{40} - 1.45^2 = 4.4475$ $\sigma_w = 4\sigma_y$ 8.44 g	M1 M1 A1 (3)
(c)	$\sum w = 40 \times 55.8 = 2232$ $\sum w^2 = 40(71.16 + 55.8^2)$, using $\sigma_w^2 = 16 \times 4.4475 = 71.16$ $\sum w^2 = 127392$	B1 M1 A1 (3)
(d)	(i) The mean is unchanged, as the new value equals the mean (ii) The standard deviation decreases because $\sum (w - \bar{w})^2$ is unchanged but is now divided by 41 instead of 40 <i>The last B1 needs a reason about the spread (the new value adds no deviation but increases n).</i>	B1 B1 B1 (3)
Total for Question 7: 11 marks		



Question 8

10 marks

Part	Scheme	Marks
(a)	Midpoints 2.5, 7.5, 12.5, 17.5, 25: $\sum fx = 602.5$	M1
	$\bar{t} = 12.05 \text{ min}$	A1
	$\sum fx^2 = 9206.25$, so $\sigma^2 = \frac{9206.25}{50} - 12.05^2 = 38.92 \dots$	M1
	$\sigma = 6.24 \text{ min}$	A1
		(4)
(b)	90th percentile is the 45th value	M1
	Cumulative frequency reaches 45 at the end of the $15 \leq t < 20$ class, so $P_{90} = 20$ minutes	A1
	$20 > 18$: more than 10% wait longer than 18 minutes, so the claim is not supported	A1
		(3)
(c)	$\bar{t} + \sigma = 18.29$ (in the $15 \leq t < 20$ class)	M1
	$\frac{20 - 18.29}{5} \times 9 + 5$	M1
	8 customers (8.08...)	A1
		(3)
Total for Question 8: 10 marks		



Question 9

11 marks

Part	Scheme	Marks
(a)	$\sum fx = 4 + 2p + 30 + 28 + 15 = 77 + 2p$ and $\sum f = 24 + p$ $\frac{77 + 2p}{24 + p} = 3$ $77 + 2p = 72 + 3p$, so $p = 5$ *	M1 M1 A1* (3)
(b)	$n = 29$: median is the 15th value, 3 Q_1 is the 8th value = 2 and Q_3 is the 22nd value = 4 IQR = 2	B1 M1 A1 (3)
(c)	$\sum fx^2 = 4 + 20 + 90 + 112 + 75 = 301$ $\sigma^2 = \frac{301}{29} - 3^2$ 1.17	B1 M1 A1 (3)
(d)	$\frac{87 + k}{30} = 3.1$ $k = 6$	M1 A1 (2)
Total for Question 9: 11 marks		



Question 10

12 marks

Part	Scheme	Marks
(a)	$\bar{y} = 1.5$, so $\bar{x} = 501.5 \text{ g}$ $\sigma_y^2 = \frac{245}{20} - 1.5^2 = 10$ $\sigma_x = 3.16 \text{ g}$ (the coding $y = x - 500$ does not change the spread)	B1 M1 A1 (3)
(b)	$\frac{30 \times 502 + 20 \times 501.5}{50}$ 501.8 g	M1 A1 (2)
(c)	$\sum x_A^2 = 30(9 + 502^2) = 7\,560\,390$ $\sum x_B^2 = 20(10 + 501.5^2) = 5\,030\,245$ $\sigma^2 = \frac{12\,590\,635}{50} - 501.8^2 = 9.46$ 3.08 g <i>Rounding the means or the sums of squares early spoils the subtraction: keep all figures.</i>	M1 M1 M1 A1 (4)
(d)	$\frac{30 \times 502 - 10}{30} = 501.7 \text{ g}$ $\text{Less than } 3 \text{ g}$ 511 was 9 g above the mean; 501 is close to the new mean, so the total squared deviation is much smaller	B1 B1 B1 (3)
Total for Question 10: 12 marks		



3

Chapter 3: Representations of Data

Mark scheme. Marks are shown on the right of each line.

Question 1

5 marks

Part	Scheme	Marks
(a)	Width: 2 minutes $\leftrightarrow$ 1 cm, so 8 minutes gives 4 cm Frequency densities: $\frac{20}{2} = 10$ and $\frac{40}{8} = 5$; height $\propto$ frequency density $5 \times \frac{5}{10} = \mathbf{2.5 \text{ cm}}$ <i>Scaling the height by the frequencies alone ($5 \times 2 = 10 \text{ cm}$) ignores the change in class width: M0.</i>	B1 M1 A1 (3)
(b)	$\frac{2}{5} \times 32 + 20 + \frac{3}{8} \times 40$ 48 (47.8)	M1 A1 (2)
Total for Question 1: 5 marks		

Question 2

5 marks

Part	Scheme	Marks
(a)	Median = 30 (10th value) Q_1: $\frac{19}{4} = 4.75$, 5th value 23 Q_3: 15th value 41	B1 B1 B1 (3)
(b)	Limits: $41 + 1.5 \times 18 = 68$ and $23 - 27 = -4$ 72 is above 68 (63 is not), so 72 is the only outlier *	M1 A1* (2)
Total for Question 2: 5 marks		



Question 3

5 marks

Part	Scheme	Marks
(a)	Median 48	B1
	$IQR = 60 - 35 = 25$	B1
		(2)
(b)	Location: school B has a slightly higher median (50 compared with 48)	B1
	Spread: school B's marks are less spread, IQR 13 compared with 25	B1
	<i>Each comparison must be in context (marks) with the values.</i>	(2)
(c)	$Q_3 - Q_2 = 5 < Q_2 - Q_1 = 8$, so negative skew	B1
		(1)
Total for Question 3: 5 marks		

Question 4

5 marks

Part	Scheme	Marks
(a)	$\frac{3(29.4 - 26)}{11.2} = 0.911$	M1
	Positive skew	A1
		(2)
(b)	$\frac{38 - 52 + 21}{17} = \frac{7}{17}$	M1
	0.41	A1
		(2)
(c)	Median and IQR, because the data are skewed (and they are not affected by extreme values)	B1
		(1)
Total for Question 4: 5 marks		



Question 5

5 marks

Part	Scheme	Marks
(a)	Stems 2 to 5 with leaves in order: 2 7 8 9; 3 1 3 4 6 7 9; 4 0 1 1 4 5 8; 5 2 Correct diagram with a key, e.g. 3 1 means 31 minutes	M1 A1 (2)
(b)	$n = 16$: median = $\frac{37 + 39}{2} = 38$ $Q_1 = \frac{31 + 33}{2} = 32$ and $Q_3 = \frac{41 + 44}{2} = 42.5$ IQR = 10.5	B1 M1 A1 (3)
Total for Question 5: 5 marks		

Question 6

11 marks

Part	Scheme	Marks
(a)	Median is halfway between the 15th and 16th values: $\frac{2w + 26}{2} = 25.5$, so $w = 5$ $x = 16$ (8th value) $y = 37$ (23rd value)	B1 B1 B1 (3)
(b)	$37 + 1.0 \times (37 - 16) = 58$ 68 is the only outlier <i>Using 1.5 instead of the rule given (limit 68.5, so no outliers) scores M0 A0.</i>	M1 A1 (2)
(c)	Box from 16 to 37 with the median at 25.5 Whiskers to 7 and 52 Outlier 68 marked with a cross	M1 A1 A1 (3)
(d)	$\frac{(37 - 25.5) - (25.5 - 16)}{37 - 16} = \frac{2}{21}$ 0.095	M1 A1 (2)
(e)	The total is estimated by mean $\times$ number of days; the mean uses all the data (including the high values)	B1 (1)
Total for Question 6: 11 marks		



Question 7

11 marks

Part	Scheme	Marks
(a)	Area of the 5–10 bar represents 40 people, so 1 unit of area ($1 \text{ km} \times 1 \text{ unit of density}$) is 1 person Frequencies 20, 40, 50, 30, 20 160	M1 M1 A1 (3)
(b)	Part of the 5–10 bar: $\frac{2}{5} \times 40 = 16$ Whole 10–20 bar, and half of the 20–30 bar: $50 + \frac{5}{10} \times 30 = 65$ 81	M1 M1 A1 (3)
(c)	80th value, in the 10–20 class $10 + \frac{80 - 60}{50} \times 10$ 14 km	M1 M1 A1 (3)
(d)	Midpoints 2.5, 7.5, 15, 25, 40: $\frac{2650}{160}$ 16.6 km	M1 A1 (2)
Total for Question 7: 11 marks		



Question 8

11 marks

Part	Scheme	Marks
(a)	$Q_2 = 72$ (8th value) $Q_1 = 66$ (4th value) $Q_3 = 79$ (12th value) <i>Read the non-athletes' leaves left to right. Reading the wrong side of the diagram is the common error.</i>	B1 B1 B1 (3)
(b)	Upper limit $79 + 1.5 \times 13 = 98.5$ and lower limit $66 - 19.5 = 46.5$ 101 is above 98.5 and no value is below 46.5, so 101 is the only outlier *	M1 A1* (2)
(c)	The athletes have lower pulse rates: median 56 compared with 72 The athletes' pulse rates are less spread: IQR 9 compared with 13 Both comparisons in context with values	B1 B1 B1 (3)
(d)	Athletes: $Q_3 - Q_2 = 5 > Q_2 - Q_1 = 4$, so positive skew Non-athletes: $Q_3 - Q_2 = 7 > Q_2 - Q_1 = 6$, so positive skew	B1 B1 (2)
(e)	A non-athlete : 69 is above the athletes' Q_3 (61) but within the non-athletes' IQR (66 to 79)	B1 (1)
Total for Question 8: 11 marks		



Question 9

10 marks

Part	Scheme	Marks
(a)	$\bar{x} = \frac{227}{12} = 18.9 \text{ cm}$ $\sigma^2 = \frac{4949}{12} - 18.916\dots^2$ $\sigma = 7.39 \text{ cm}$	B1 M1 A1 (3)
(b)	$\bar{x} + 2\sigma = 18.9\dots + 2 \times 7.38\dots = 33.7$ $42 > 33.7$, so 42 is an outlier *	M1 A1* (2)
(c)	$n = 11, \sum x = 185, \sum x^2 = 4949 - 42^2 = 3185$ Mean 16.8 cm Standard deviation 2.59 cm	M1 A1 A1 (3)
(d)	Tray P seedlings are taller on average (mean 16.8 compared with 15.2 cm) Tray P heights are less spread (sd 2.59 compared with 3.9 cm)	B1 B1 (2)
Total for Question 9: 10 marks		



Question 10

12 marks

Part	Scheme	Marks
(a)	Limits: $22 + 1.5 \times 8 = 34$ and $14 - 12 = 2$ 36 is above 34 and 33 is not; 8 is above 2. So exactly one outlier, 36 *	M1 A1* (2)
(b)	Box from 14 to 22 with the median at 18 Whiskers to 8 and 33 Outlier at 36 marked with a cross	M1 A1 A1 (3)
(c)	Type B grows more: median 24 compared with 18 cm Type B growth is more spread: IQR 11 compared with 8 cm	B1 B1 (2)
(d)	$g > 28$ (above type B's Q_3) and $g \leq 34$ (not an outlier for type A) $28 < g \leq 34$	M1 A1 (2)
(e)	Type B IQR = 11: limits $28 + 16.5 = 44.5$ and $17 - 16.5 = 0.5$ Two ranges $h > 44.5$ or $h < 0.5$ <i>The final mark needs both ranges with the correct inequality signs.</i>	M1 M1 A1 (3)
Total for Question 10: 12 marks		



4

Chapter 4: Probability

Mark scheme. Marks are shown on the right of each line.

Question 1

5 marks

Part	Scheme	Marks
(a)	$\frac{29}{80}$	B1 (1)
(b)	$\frac{12}{12+16}$ $\frac{3}{7}$	M1 A1 (2)
(c)	$P(A)P(M) = \frac{35}{80} \times \frac{28}{80} = 0.153\dots$ and $P(A \cap M) = \frac{12}{80} = 0.15$ $0.15 \neq 0.153$, so not independent <i>Alternatively $P(A M) = \frac{3}{7} = 0.429 \neq P(A) = 0.4375$.</i>	M1 A1 (2)
Total for Question 1: 5 marks		

Question 2

5 marks

Part	Scheme	Marks
(a)	$0.61 = 0.4 + 0.35 - P(A \cap B)$ $P(A \cap B) = 0.14$	M1 A1 (2)
(b)	$P(A) \times P(B) = 0.4 \times 0.35 = 0.14$ $= P(A \cap B)$, so A and B are independent *	M1 A1* (2)
(c)	$0.35 - 0.14 = 0.21$	B1 (1)
Total for Question 2: 5 marks		



Question 3

5 marks

Part	Scheme	Marks
(a)	First counter: $\frac{3}{8}$ for blue	B1
	Second counter: $\frac{4}{7}$, $\frac{3}{7}$ after red; $\frac{2}{7}$ for blue after blue	B1
		(2)
(b)	$\frac{5}{8} \times \frac{4}{7} + \frac{3}{8} \times \frac{2}{7}$	M1
	$\frac{13}{28}$	A1
		(2)
(c)	$1 - \frac{5}{8} \times \frac{4}{7} = \frac{9}{14}$	B1
		(1)
Total for Question 3: 5 marks		

Question 4

5 marks

Part	Scheme	Marks
(a)	$P(A \cup B) = 3p + p - 3p \times p$ (addition rule with independence)	M1
	$4p - 3p^2 = 0.68$, so $3p^2 - 4p + 0.68 = 0$ *	A1*
		(2)
(b)	$p = \frac{4 \pm \sqrt{16 - 8.16}}{6} = 0.2 \text{ or } 1.13 \dots$	M1
	$P(B) = 0.2$: $p = 1.13$ is rejected because a probability cannot exceed 1	A1
		(2)
(c)	$1 - 0.68 = 0.32$	B1
		(1)
Total for Question 4: 5 marks		



Question 5

5 marks

Part	Scheme	Marks
(a)	$P(A) = 0.3 + x$ and $P(B) = 0.2 + x$ Independence: $x = (0.3 + x)(0.2 + x)$ $x = 0.06 + 0.5x + x^2$, so $x^2 - 0.5x + 0.06 = 0$ *	B1 M1 A1* (3)
(b)	$(x - 0.2)(x - 0.3) = 0$, and $P(A \cup B) = 0.5 + x$ $0.5 + x < 0.75$, so $x = 0.2$ Both roots are valid probabilities; the condition on $P(A \cup B)$ rejects $x = 0.3$ (which gives 0.8).	M1 A1 (2)
Total for Question 5: 5 marks		

Question 6

11 marks

Part	Scheme	Marks
(a)	$P(F \cup G) = P(F) + P(G) - P(F \cap G)$ Independence: $P(F \cap G) = \frac{2}{5}P(G)$ $\frac{7}{10} = \frac{2}{5} + P(G) - \frac{2}{5}P(G)$, so $\frac{3}{5}P(G) = \frac{3}{10}$ $P(G) = \frac{1}{2}$ * Using $\frac{1}{2}$ to show that $\frac{1}{2}$ works (a circular argument) scores at most M1 M1 M0 A0.	M1 M1 M1 A1* (4)
(b)	K lies inside F : $P(K' \cap F) = \frac{2}{5} - \frac{1}{10} = \frac{3}{10}$, then $\div \frac{2}{5}$ $\frac{3}{4}$	M1 A1 (2)
(c)	K drawn inside F and not overlapping G ; F and G overlapping $F \cap G = \frac{1}{5}$ $K = \frac{1}{10}$ and the rest of F : $\frac{2}{5} - \frac{1}{5} - \frac{1}{10} = \frac{1}{10}$ G only: $\frac{1}{2} - \frac{1}{5} = \frac{3}{10}$ Outside all three: $1 - \frac{7}{10} = \frac{3}{10}$	B1 B1 B1 B1 (5)
Total for Question 6: 11 marks		



Question 7

11 marks

Part	Scheme	Marks
(a)	Good: 0.97 0.05 (faulty, not identified); 0.04 and 0.96 after good	B1 B1 (2)
(b)	$0.03 \times 0.95 + 0.97 \times 0.04$ 0.0673	M1 A1 (2)
(c)	$\frac{P(\text{faulty} \cap \text{identified})}{P(\text{identified})}$ 0.0285 0.0673 0.423	M1 M1 A1 (3)
(d)	Faulty and identified twice: $0.03 \times 0.95^2 = 0.027075$ Good and identified twice: $0.97 \times 0.04^2 = 0.001552$ $\frac{0.027075}{0.027075 + 0.001552}$ 0.946 <i>Each item has three stages on the tree: the numerator is a product of three probabilities.</i>	M1 M1 M1 A1 (4)
Total for Question 7: 11 marks		



Question 8

10 marks

Part	Scheme	Marks
(a)	Total: $0.08 + 0.12 + 0.04 + p + q + 0.3 = 1$, so $p + q = 0.46$ $P(A) = 0.24 + p$, $P(B) = 0.04 + p + q = 0.5$, $P(A \cap B) = 0.04 + p$ Independence: $0.04 + p = 0.5(0.24 + p)$ $p = 0.16$, $q = 0.3$	M1 M1 M1 A1 (4)
(b)	$P(A \cap B') + P(A' \cap B) = (0.08 + 0.12) + 0.3$ 0.5	M1 A1 (2)
(c)	$P(C) = 0.16$, $P(B) = 0.5$: $P(B)P(C) = 0.08$, but $P(B \cap C) = 0.04$ $0.04 \neq 0.08$, so not independent	M1 A1 (2)
(d)	$P(C) = 0.16$, so the survey has $\frac{96}{0.16} = 600$ people $P(A \cap B) = 0.2$: $0.2 \times 600 = 120$ $A \cap B$ includes the part of C inside B: $0.04 + 0.16 = 0.2$.	M1 A1 (2)
Total for Question 8: 10 marks		



Question 9

11 marks

Part	Scheme	Marks
(a)	$P(A): x + y + 0.05 + 0.05 = 0.4$, so $x + y = 0.3$ $P(A \cup B): x + y + z + 0.05 + 0.05 + 0.1 = 0.7$, so $z = 0.7 - 0.2 - 0.3 = 0.2$ *	M1 A1* (2)
(b)	$P(B) = y + z + 0.05 + 0.1 = y + 0.35$ and $P(A \cap B) = y + 0.05$ Independence: $y + 0.05 = 0.4(y + 0.35)$ $0.6y = 0.09$, so $y = 0.15$ $x = 0.15$	M1 M1 A1 A1ft (4)
(c)	$1 - (0.15 + 0.15 + 0.2 + 0.05 + 0.05 + 0.1 + 0.1) = 0.2$	B1 (1)
(d)	$P(B) = 0.5$, $P(C) = 0.3$ and $P(B \cap C) = 0.05 + 0.1 = 0.15$ $0.5 \times 0.3 = 0.15 = P(B \cap C)$, so B and C are independent	M1 A1 (2)
(e)	$P(B \cup C) = 1 - (x + r) = 0.65$ and $P(A \cap (B \cup C)) = y + 0.05 + 0.05 = 0.25$ $\frac{5}{13}$ (0.385)	M1 A1 (2)
Total for Question 9: 11 marks		



Question 10

12 marks

Part	Scheme	Marks
(a)	$P(\text{late}) = 0.4x + 0.1(1 - x)$ $0.4x + 0.1 - 0.1x = 0.19$ $0.3x = 0.09$, so $x = 0.3$ *	M1 M1 A1* (3)
(b)	$\frac{0.3 \times 0.4}{0.19}$ 0.632 $\left(\frac{12}{19}\right)$	M1 A1 (2)
(c)	$2 \times 0.19 \times 0.81$ 0.3078	M1 A1 (2)
(d)	$0.19 \times 0.5 + 0.81 \times 0.05$ 0.1355	M1 A1 (2)
(e)	Rain and Tom late: $0.3 \times 0.4 \times 0.5 + 0.3 \times 0.6 \times 0.05 = 0.069$ $\frac{0.069}{0.1355}$ 0.509 <i>Both routes through "rain" are needed: the bus late (then Tom walks) and the bus on time.</i>	M1 M1 A1 (3)
Total for Question 10: 12 marks		



5

Chapter 5: Correlation and Regression

Mark scheme. Marks are shown on the right of each line.

Question 1

5 marks

Part	Scheme	Marks
(a)	$S_{xx} = 385 - \frac{55^2}{10} = 82.5$	B1
	$S_{yy} = 10650 - \frac{320^2}{10} = 410$	B1
	$S_{xy} = 1920 - \frac{55 \times 320}{10} = 160$	B1
		(3)
(b)	$\frac{160}{\sqrt{82.5 \times 410}}$	M1
	0.870	A1
		(2)
Total for Question 1: 5 marks		

Question 2

5 marks

Part	Scheme	Marks
(a)	$b = \frac{96}{40} = 2.4$	B1
	$a = 20 - 2.4 \times 6 = 5.6$	M1
	$y = 5.6 + 2.4x$	A1
		(3)
(b)	22.4	B1
		(1)
(c)	25 is outside the range of the data (2 to 10): extrapolation	B1
		(1)
Total for Question 2: 5 marks		



Question 3

5 marks

Part	Scheme	Marks
(a)	Strong positive correlation: students who revise for more hours tend to get higher test scores	B1 (1)
(b)	Hours of revision, since the test score depends on it	B1 (1)
(c)	Correlation does not show causation: other factors (e.g. ability) could explain the scores	B1 (1)
(d)	Very weak negative (almost no) linear correlation between revision time and score	B1 (1)
(e)	It only measures <i>linear</i> correlation (or: it is affected by outliers; it does not show cause)	B1 (1)
Total for Question 3: 5 marks		

Question 4

5 marks

Part	Scheme	Marks
(a)	0.72	B1 (1)
(b)	$\frac{y}{10} = 2.1 + 0.35(x - 100)$ $y = 21 + 3.5x - 350$ $y = 3.5x - 329$	M1 M1 A1 (3)
(c)	For each increase of 1 in x , y increases by 3.5 (on average)	B1 (1)
Total for Question 4: 5 marks		



Question 5

5 marks

Part	Scheme	Marks
(a)	For each extra employee, spending increases by 0.156 thousand dollars i.e. by \$156 per month <i>\$0.156 or \$1.56 scores M1 A0: c is in thousands of dollars.</i>	M1 A1 (2)
(b)	$12.5 + 0.156 \times 80 = 24.98$ \$24 980	M1 A1 (2)
(c)	$p = 0$ is outside the range of the data (20 to 120), so this is extrapolation	B1 (1)
Total for Question 5: 5 marks		



Question 6

13 marks

Part	Scheme	Marks
(a)	$S_{tg} = 5160 - \frac{168 \times 235}{8} = 225$ $S_{tt} = 168$ $S_{gg} = 341.875$	B1* B1 B1 (3)
(b)	$\frac{225}{\sqrt{168 \times 341.875}}$ 0.939	M1 A1 (2)
(c)	Strong positive correlation: on hotter days the café tends to sell more ice creams	B1 (1)
(d)	$b = \frac{225}{168} = 1.339 \dots$ $\bar{t} = 21, \bar{g} = 29.375$ $a = 29.375 - 1.339 \dots \times 21 = 1.25$ $g = 1.25 + 1.34t$	M1 B1 M1 A1 (4)
(e)	For each 1 °C rise, about 13 (1.34 tens) more ice creams are sold <i>The unit "tens" must be used: "1.34 more ice creams" scores B0.</i>	B1 (1)
(f)	(i) $1.25 + 1.339 \dots \times 25 = 34.7$ tens, i.e. about 347 ice creams (ii) Unreliable: 40 °C is outside the range of temperatures (14 to 28 °C), extrapolation	B1 B1 (2)
Total for Question 6: 13 marks		



Question 7

11 marks

Part	Scheme	Marks
(a)	$\frac{11\,670}{\sqrt{586 \times 234\,889}}$ 0.995	M1 A1 (2)
(b)	$b = \frac{11\,670}{586} = 19.914\dots$ $\bar{L} = 35 \text{ and } \bar{m} = 431.11\dots$ $a = 431.11\dots - 19.914\dots \times 35 = -265.90\dots$ $m = -265.9 + 19.91L, \text{ so } m = -266 + 19.9L^*$ <i>Values to one more figure than those given must be seen for the final mark.</i>	M1 B1 M1 A1* (4)
(c)	0.36 m = 36 cm: $-266 + 19.9 \times 36$ 450 g (451 using the unrounded line) <i>Substituting $L = 0.36$ gives a negative mass: M0.</i>	M1 A1 (2)
(d)	The line gives $m = -67$ (a negative mass), which is impossible 10 cm is outside the range of lengths (22 to 48 cm), so the linear model may not hold	B1 B1 (2)
(e)	Each extra cm of length adds about 19.9 g to the mass	B1 (1)
Total for Question 7: 11 marks		



Question 8

13 marks

Part	Scheme	Marks
(a)	$S_{xx} = 55 - \frac{15^2}{6} = 17.5$	B1
	$S_{yy} = 371.9 - \frac{41^2}{6} = 91.73$	B1
	$S_{xy} = 141.8 - \frac{15 \times 41}{6} = 39.3$	B1
		(3)
(b)	$\frac{39.3}{\sqrt{17.5 \times 91.73 \dots}}$	M1
	0.981	A1 (2)
(c)	0.981: linear coding does not change the correlation coefficient	B1 (1)
(d)	$b = \frac{39.3}{17.5} = 2.245 \dots$	M1
	$a = \frac{41}{6} - 2.245 \dots \times \frac{15}{6} = 1.219 \dots$	M1
	$y = 1.22 + 2.25x$	A1
		(3)
(e)	$\frac{p - 200}{10} = 1.219 \dots + 2.245 \dots \times \frac{t - 2000}{4}$	M1
	$p = 212 + 5.61(t - 2000)$	A1 (2)
(f)	$212.19 + 5.614 \times 30 = \text{£}381$	B1
	Unreliable: 2030 is outside the years in the data (2000 to 2020), extrapolation	B1
		(2)
Total for Question 8: 13 marks		



Question 9

10 marks

Part	Scheme	Marks
(a)	$\frac{1350}{\sqrt{2250 \times 1200}}$ 0.822	M1 A1 (2)
(b)	$b = \frac{1350}{2250} = 0.6$ $a = 130 - 0.6 \times 45$ $s = 103 + 0.6a$	B1 M1 A1 (3)
(c)	Blood pressure rises by about 0.6 mmHg for each extra year of age	B1 (1)
(d)	$0.6x = 9$ 15 years <i>Substituting $s = 9$ into the whole equation is the common error: it is the change in s that is 9.</i>	M1 A1 (2)
(e)	The line of s on a predicts s from a ; estimating a needs the regression line of a on s (age is the explanatory variable)	B1 (1)
(f)	Correlation does not imply causation (other factors, e.g. lifestyle, may be involved)	B1 (1)
Total for Question 9: 10 marks		



Question 10

12 marks

Part	Scheme	Marks
(a)	$S_{xx} = 440 - 360 = 80$, $S_{yy} = 13\,050 - 12\,250 = 800$, $S_{xy} = 2280 - 2100 = 180$ $\frac{180}{\sqrt{80 \times 800}}$ 0.712	M1 M1 A1 (3)
(b)	$\sum y = 365$ $\sum y^2 = 13\,050 - 900 + 2025 = 14\,175$ $\sum xy = 2280 - 240 + 360 = 2400$	B1 B1 B1 (3)
(c)	$S_{yy} = 14\,175 - \frac{365^2}{10} = 852.5$ and $S_{xy} = 2400 - \frac{60 \times 365}{10} = 210$ $\frac{210}{\sqrt{80 \times 852.5}} = \text{0.804}$ The correlation is stronger (0.804 compared with 0.712)	M1 A1 B1 (3)
(d)	$b = \frac{210}{80} = 2.625$ $a = 36.5 - 2.625 \times 6 = 20.75$ $y = 20.8 + 2.63x$	M1 M1 A1 (3)
Total for Question 10: 12 marks		



6

Chapter 6: Discrete Random Variables

Mark scheme. Marks are shown on the right of each line.

Question 1

5 marks

Part	Scheme	Marks
(a)	$k(2 + 3 + 4 + 5) = 1$ $14k = 1$, so $k = \frac{1}{14}$ *	M1 A1* (2)
(b)	$\frac{1 \times 2 + 2 \times 3 + 3 \times 4 + 4 \times 5}{14}$ $\frac{20}{7}$	M1 A1 (2)
(c)	$\frac{4+5}{14} = \frac{9}{14}$	B1 (1)
Total for Question 1: 5 marks		

Question 2

5 marks

Part	Scheme	Marks
(a)	$a + b = 0.6$ (probabilities add to 1) $0.1 + 2a + 0.9 + 4b = 2.9$, i.e. $2a + 4b = 1.9$	B1 B1 (2)
(b)	Solves simultaneously $a = 0.25, b = 0.35$	M1 A1 (2)
(c)	$a + 0.3 = \frac{9}{14}$	B1ft (1)
Total for Question 2: 5 marks		



Question 3

5 marks

Part	Scheme	Marks
(a)	Differences of consecutive values of F $0.15, 0.2, 0.25, 0.25, 0.15$ for $x = 1, \dots, 5$	M1 A1 (2)
(b)	0.35	B1 (1)
(c)	$P(2 \leq X \leq 4) = F(4) - F(1)$ 0.7	M1 A1 (2)
Total for Question 3: 5 marks		

Question 4

5 marks

Part	Scheme	Marks
(a)	$5 - 2 \times 2.4 = 0.2$	B1 (1)
(b)	$(-2)^2 \times 1.3$ 5.2 <i>The 5 does not affect the variance, and -2 is squared: -5.2 or 2.6 scores M0 A0.</i>	M1 A1 (2)
(c)	$E(X^2) = \text{Var}(X) + [E(X)]^2 = 1.3 + 2.4^2$ 7.06	M1 A1 (2)
Total for Question 4: 5 marks		



Question 5

5 marks

Part	Scheme	Marks
(a)	Discrete uniform: $E(X) = \frac{13}{2} = 6.5$ and $\text{Var}(X) = \frac{12^2 - 1}{12} = \frac{143}{12}$ 6.5 and 11.9 (11.916...)	B1 B1 (2)
(b)	$E(Y) = 4 \times 6.5 - 3 = \text{23}$ $\text{Var}(Y) = 16 \times \frac{143}{12} = \text{190.7}$ ($\frac{572}{3}$)	B1 B1 (2)
(c)	$4X - 3 > 30 \Rightarrow X > 8.25$, so $X = 9, \dots, 12$: $\frac{1}{3}$	B1 (1)
Total for Question 5: 5 marks		

Question 6

12 marks

Part	Scheme	Marks
(a)	$a = 0.2$ $c = 0.55$ $b = 0.3$	B1 B1 B1 (3)
(b)	$0.1 + 0.4 + 0.75 + 1.2 + 0.75$ 3.2	M1 A1 (2)
(c)	$E(X^2) = 0.1 + 0.8 + 2.25 + 4.8 + 3.75 = 11.7$ 11.7 correct $\text{Var}(X) = 11.7 - 3.2^2 = 1.46$ *	M1 A1 A1* (3)
(d)	$E(Y^2) = \text{Var}(Y) + [E(Y)]^2 = 9 \times 1.46 + 8.6^2$ (or $E(9X^2 - 6X + 1)$) 87.1	M1 A1 (2)
(e)	$P(\text{odd}) = 0.1 + 0.25 + 0.15 = 0.5$ $0.5^2 = \text{0.25}$	M1 A1 (2)
Total for Question 6: 12 marks		



Question 7

12 marks

Part	Scheme	Marks
(a)	Pairs (R, B) : $(1, 2), (2, 2), (3, 2), (1, 4)$ $\frac{4}{16} = \frac{1}{4}$	M1 A1 (2)
(b)	$R = 1: B = 2, 4, 6, 8; R = 2: B = 6, 8; R = 3, 4$: none 6 outcomes out of 16 $\frac{3}{8}$	M1 A1 A1 (3)
(c)	$F(1) = \frac{5}{16}$ $F(4) = \frac{3}{4}$	B1 B1 (2)
(d)	Probabilities $\frac{2}{16}, \frac{3}{16}, \frac{3}{16}, \frac{2}{16}, \frac{2}{16}, \frac{2}{16}, \frac{1}{16}, \frac{1}{16}$ for $d = 0, \dots, 7$ $\frac{0+3+6+6+8+10+6+7}{16}$ $\frac{23}{8}$	M1 M1 A1 (3)
(e)	The four outcomes in (a) have $D = 1, 0, 1, 3$ $\frac{3}{4}$	M1 A1 (2)
Total for Question 7: 12 marks		



Question 8

11 marks

Part	Scheme	Marks
(a)	$2p + 0.5 = 1$: $p = 0.25$	B1 (1)
(b)	$E(X) = 0.25 + 1 + 0.75 = 2$ * $E(X^2) = 0.25 + 2 + 2.25 = 4.5$ 4.5 correct $\text{Var}(X) = 4.5 - 2^2 = 0.5$ *	B1* M1 A1 A1* (4)
(c)	$E(Y) = 3$ $\frac{5^2 - 1}{12} = 2$	B1 B1 (2)
(d)	$\text{Var}(W) = a^2 \times 0.5 = 2$ $a = 2$ (as $a > 0$) $E(W) = 2 \times 2 + b = 3$ $b = -1$	M1 A1 M1 A1 (4)
Total for Question 8: 11 marks		



Question 9

10 marks

Part	Scheme	Marks
(a)	$E(X) = 0.5 + 0.6 = 1.1$ $E(X^2) = 0.5 + 9 \times 0.2 = 2.3$ $\text{Var}(X) = 2.3 - 1.1^2$ 1.09	B1 M1 M1 A1 (4)
(b)	He needs at least 3 points from 2 rounds: at least one round scoring 3 $1 - 0.8^2$ 0.36 <i>Listing: (3, any) or (0 or 1, 3): $0.2 + 0.8 \times 0.2 = 0.36$. Scores of 1 and 1 give only 2 points.</i>	M1 M1 A1 (3)
(c)	$P(3 \text{ and } 3) = 0.2^2 = 0.04$ $\frac{0.04}{0.36}$ $\frac{1}{9}$	B1 M1 A1 (3)
Total for Question 9: 10 marks		



Question 10

12 marks

Part	Scheme	Marks
(a)	$F(4) = 1: (4+k)^2 = 36$ $4+k = 6$ (as $k > 0$), so $k = 2$ *	M1 A1* (2)
(b)	$F(1) = \frac{9}{36}, F(2) = \frac{16}{36}, F(3) = \frac{25}{36}$ and uses differences At least two correct $\frac{9}{36}, \frac{7}{36}, \frac{9}{36}, \frac{11}{36}$ for $x = 1, 2, 3, 4$	M1 A1 A1 (3)
(c)	$\frac{9+14+27+44}{36}$ $\frac{47}{18}$	M1 A1 (2)
(d)	$2X > 7$, so $X > 3.5$, i.e. $X = 4$ $\frac{11}{36}$	M1 A1 (2)
(e)	$E(X^2) = \frac{9+28+81+176}{36} = \frac{49}{6}$ $\text{Var}(X) = \frac{49}{6} - \left(\frac{47}{18}\right)^2 = 1.348\dots$ $9 \times 1.348\dots = 12.1$	M1 M1 A1 (3)
Total for Question 10: 12 marks		



7

Chapter 7: The Normal Distribution

Mark scheme. Marks are shown on the right of each line.

Question 1

5 marks

Part	Scheme	Marks
(a)	$P\left(Z < \frac{56 - 50}{4}\right) = \Phi(1.5)$ 0.9332	M1 A1 (2)
(b)	Standardises both: $z = -1.5$ and $z = 2$ $\Phi(2) - \Phi(-1.5) = 0.9772 - 0.0668$ 0.9104	M1 M1 A1 (3)
Total for Question 1: 5 marks		

Question 2

5 marks

Part	Scheme	Marks
(a)	$z = \frac{99 - 120}{15} = -1.4, \text{ so } P(Z > -1.4) = \Phi(1.4)$ 0.9192	M1 A1 (2)
(b)	$\frac{a - 120}{15} = 1.6449$ 1.6449 from the percentage points table $a = 144.7$ (144.67...) <i>Using 1.64 or 1.65 scores M1 B0 A1 if the answer rounds correctly.</i>	M1 B1 A1 (3)
Total for Question 2: 5 marks		



Question 3

5 marks

Part	Scheme	Marks
(a)	$\Phi(z) = 0.8413$ gives $z = 1$ $\frac{36 - 30}{\sigma} = 1$ $\sigma = 6$	B1 M1 A1 (3)
(b)	0.1587 By symmetry: 24 is the same distance below the mean as 36 is above it	B1 B1 (2)
Total for Question 3: 5 marks		

Question 4

5 marks

Part	Scheme	Marks
(a)	$P(Z > z) = 0.2$ gives $z = 0.8416$ $\frac{40 - \mu}{5} = 0.8416$ $\mu = 35.8$ (35.792)	B1 M1 A1 (3)
(b)	$P\left(Z < \frac{30 - 35.792}{5}\right) = P(Z < -1.16)$ 0.123 (0.1230 from tables; 0.1234 by calculator)	M1 A1 (2)
Total for Question 4: 5 marks		

Question 5

5 marks

Part	Scheme	Marks
(a)	$Q_1: P(Z < z) = 0.25$ gives $z = -0.6745$ (accept -0.67) $Q_1 = 71.9$ $IQR = 16.2$ ($2 \times 0.6745 \times 12$)	M1 A1 A1 (3)
(b)	Limit = $Q_3 + 1.5 \times IQR = 88.09 + 24.28 = 112.4$, i.e. $z = 2.70$ $1 - \Phi(2.70) = 0.0035$	M1 A1 (2)
Total for Question 5: 5 marks		



Question 6

12 marks

Part	Scheme	Marks
(a)	$P(Z < -1) = 1 - 0.8413$ 0.1587	M1 A1 (2)
(b)	$P(Z > 1.5) = 1 - 0.9332$ 0.0668	M1 A1 (2)
(c)	Three positions for the large apple: $3 \times 0.0668 \times 0.9332^2$ Correct expression 0.175	M1 A1 A1 (3)
(d)	$P(170 < X < 180) = \Phi(1.5) - \Phi(1) = 0.9332 - 0.8413 = 0.0919$ $\frac{0.0919}{0.9332}$ 0.0985 <i>The numerator is the range 170 to 180 (more than 170 and not large).</i>	M1 M1 A1 (3)
(e)	$\frac{w - 150}{20} = -1.2816$ $w = 124.4$ (124.368)	M1 A1 (2)
Total for Question 6: 12 marks		



Question 7

12 marks

Part	Scheme	Marks
(a)	$P(T < 46) = 0.1$ gives $z = -1.2816$ $\frac{46 - \mu}{\sigma} = -1.2816$, so $46 = \mu - 1.2816\sigma$ * $70 = \mu + 0.8416\sigma$	B1 B1* B1 (3)
(b)	Subtracts: $24 = 2.1232\sigma$ $\sigma = 11.3$ (11.30...) $\mu = 60.5$ (60.49)	M1 A1 A1 (3)
(c)	$z = \frac{50 - 60.49}{11.30} = -0.93$ and $z = \frac{65 - 60.49}{11.30} = 0.40$ $\Phi(0.40) - \Phi(-0.93) = 0.6554 - 0.1762$ 0.479 (0.478 using unrounded values)	M1 M1 A1 (3)
(d)	$P(T > 65) = 0.3446$ and $P(T > 50) = 0.8238$ $\frac{0.3446}{0.8238}$ 0.418 (awrt 0.42)	M1 M1 A1 (3)
Total for Question 7: 12 marks		



Question 8

11 marks

Part	Scheme	Marks
(a)	By symmetry $P(X < 30) = 0.9$ $\frac{30 - 25}{\sigma} = 1.2816$ $\sigma = 3.901 \dots = 3.90 \text{ *}$	M1 M1 A1* (3)
(b)	$P\left(Z > \frac{7}{3.901}\right) = P(Z > 1.79)$ 0.0367 (0.0364 by calculator)	M1 A1 (2)
(c)	$\frac{P(X > 30)}{P(X > 25)} = \frac{0.1}{0.5}$ 0.2	M1 A1 (2)
(d)	$1 - 0.9^4$ 0.3439	M1 A1 (2)
(e)	$\frac{k}{\sigma} = 0.6745 \text{ (accept 0.67)}$ $k = 2.63$	M1 A1 (2)
Total for Question 8: 11 marks		



Question 9

11 marks

Part	Scheme	Marks
(a)	$z = \frac{49.5 - 50}{0.4} = -1.25$ and $z = 1.25$ $2\Phi(1.25) - 1 = 2 \times 0.8944 - 1$ 0.7888	M1 M1 A1 (3)
(b)	$z = \frac{49.5 - 50.2}{0.5} = -1.4$ and $z = \frac{50.5 - 50.2}{0.5} = 0.6$ $\Phi(0.6) - \Phi(-1.4) = 0.7257 - 0.0808$ 0.6449	M1 M1 A1 (3)
(c)	$0.6 \times 0.7888 + 0.4 \times 0.6449$ 0.731	M1 A1 (2)
(d)	$P(A \cap \text{acceptable}) = 0.6 \times 0.7888 = 0.4733$ $\frac{0.4733}{0.7312 \dots}$ 0.647	M1 M1 A1 (3)
Total for Question 9: 11 marks		



Question 10

12 marks

Part	Scheme	Marks
(a)	$P(Z < z) = 0.0228$ gives $z = -2$ $\frac{30 - \mu}{\mu/4} = -2$ $30 - \mu = -\frac{\mu}{2}$, so $\mu = 60$ * $\sigma = 15$	B1 M1 A1* B1 (4)
(b)	$P(X > 75) = 1 - \Phi(1) = 0.1587$ and $P(X > 60) = 0.5$ $\frac{0.1587}{0.5}$ 0.3174	M1 M1 A1 (3)
(c)	$60 + 1.6449 \times 15$ 84.7	M1 A1 (2)
(d)	$P(45 < X < 75) = 2\Phi(1) - 1 = 0.6826$ $200 \times 0.6826 = 136.5$ About 68% of a normal distribution lies within one standard deviation of the mean	M1 A1 B1 (3)
Total for Question 10: 12 marks		

