Teacher Support Materials 2008

Maths GCE

Paper Reference MS/SS1B

Question 1

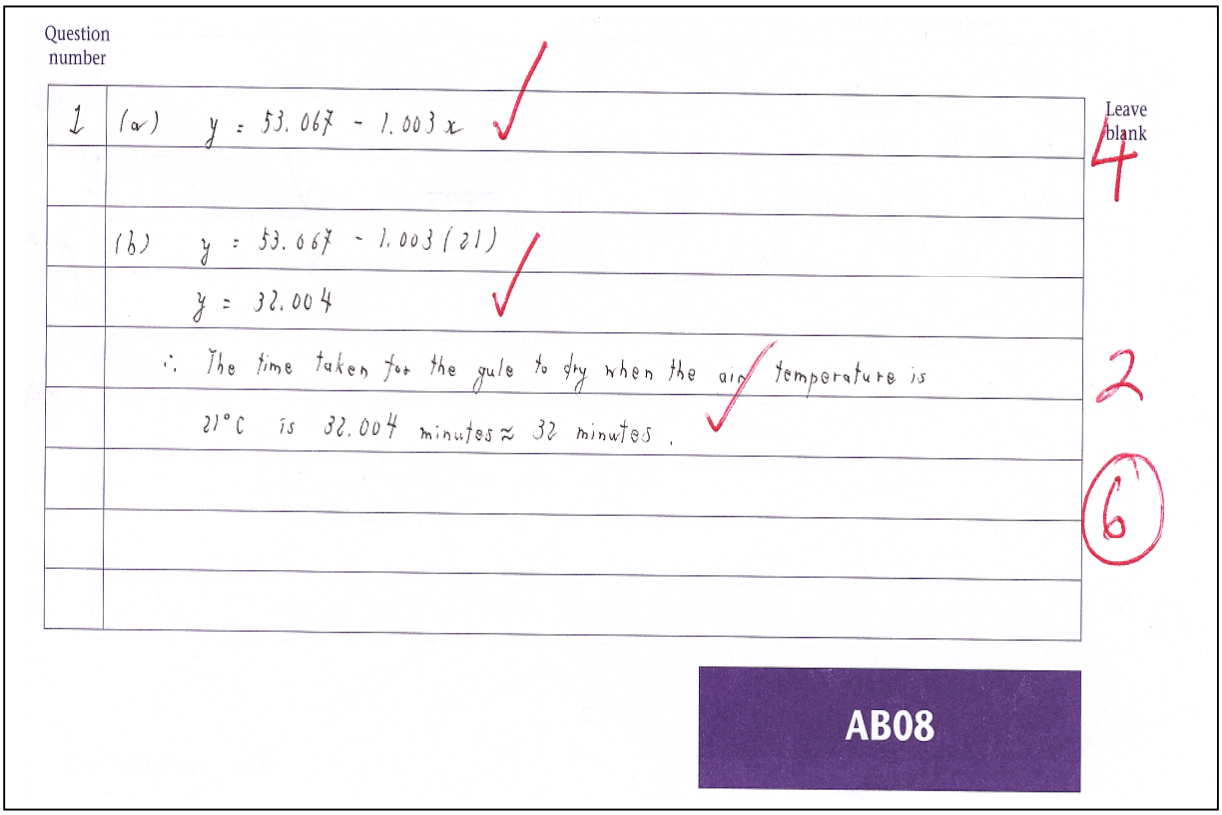
1 The table shows the times taken, y minutes, for a wood glue to dry at different air temperatures, x °C.

x	10	12	15	18	20	22	25	28	30
y	42.9	40.6	38.5	35.4	33.0	30.7	28.0	25.3	22.6

(a) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(b) Estimate the time taken for the glue to dry when the air temperature is 21 °C. *(2 marks)*

Student Response



Commentary

This is a particularly brief but fully-correct answer to part (a) that has clearly been done, as encouraged, using a calculator's inbuilt function. The answer to part (b) shows a clear method and there is a sensible rounding of 32.004 to 32 (minutes).

Mark Scheme

Q	Solution	Marks	Total	Comments
1(a)	b (gradient) = –1.01 to –1(.00) (b (gradient) = –1.05 to –0.95)	B2 (B1)		AWFW (–1.00337)
	a (intercept) = 53(.0) to 53.2 (a (intercept) = 52(.0) to 54(.0))	B2 (B1)	4	AWFW (53.06736)
	OR			
	Attempt at $\sum x$, $\sum x^2$, $\sum y$ and $\sum xy$ **or** Attempt at S_{xx} and S_{xy}	(M1)		180, 3986, 297 and 5552.7 386 and –387.3
	Attempt at **correct** formula for b (gradient)	(m1)		
	b (gradient) = –1.01 to –1(.00) a (intercept) = 53(.0) to 53.2	(A1) (A1)		AWFW AWFW
	Accept a and b interchanged only if then identified correctly in part (b), but B2 in (b) does **not** necessarily imply 4 marks in (a)			
(b)	When x = 21,			
	y = 31.7 to 32.2 (y = 29.9 to 34.1)	B2 (B1)	2	AWFW (32.0) AWFW
	Evidence of use of 21 in c's equation	(M1)		
	Special Cases (if seen):			
	$y = \frac{33.0+30.7}{2}$ = 31.8 to 31.9	(B1)		AWFW; or equivalent
	y = 31.85 without working	(B1)		
	Total		**6**	

Question 2

2 A basket in a stationery store contains a total of 400 marker and highlighter pens. Of the marker pens, some are permanent and the rest are non-permanent. The colours and types of pen are shown in the table.

	Colour			
Type	**Black**	**Blue**	**Red**	**Green**
Permanent marker	44	66	32	18
Non-permanent marker	36	53	21	10
Highlighter	0	41	37	42

A pen is selected at random from the basket. Calculate the probability that it is:

(a) a blue pen; *(1 mark)*

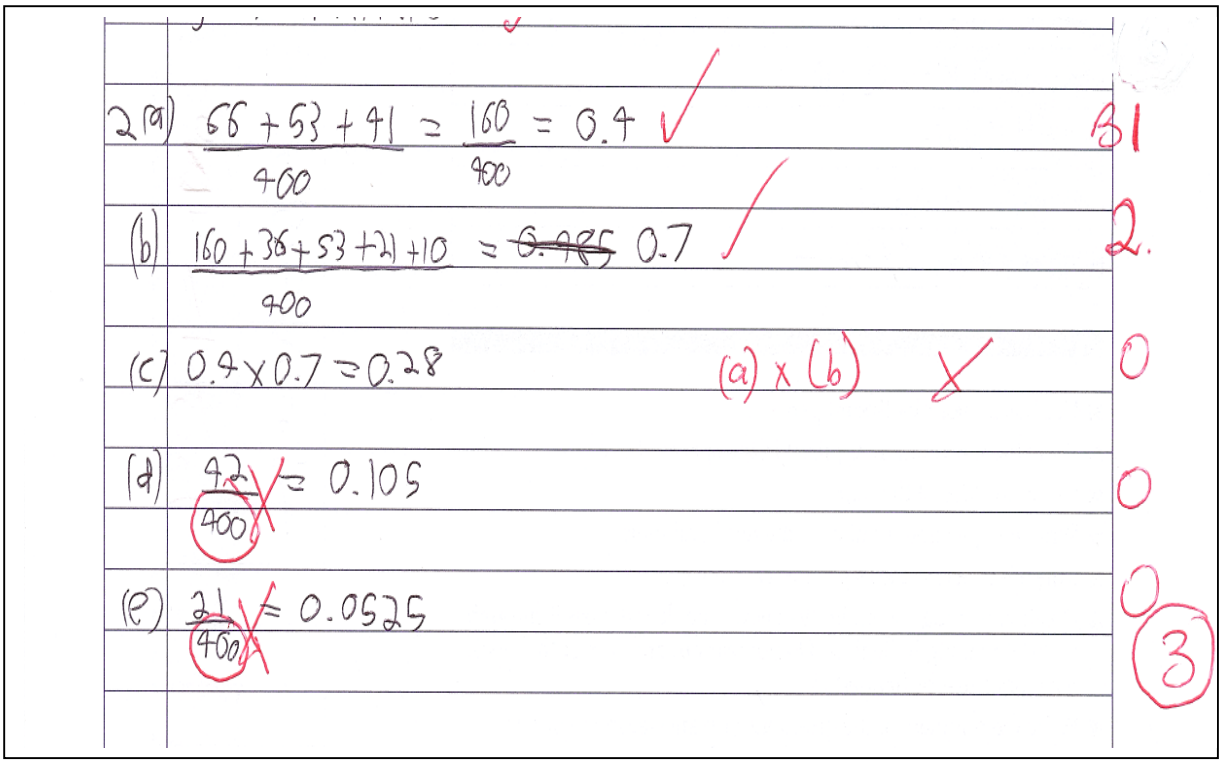
(b) a marker pen; *(2 marks)*

(c) a blue pen or a marker pen; *(2 marks)*

(d) a green pen, given that it is a highlighter pen; *(2 marks)*

(e) a non-permanent marker pen, given that it is a red pen. *(2 marks)*

Student response



Commentary

The candidate has derived (many simply quoted) correct answers to parts (a) & (b). In part (c), the candidate has misinterpreted 'or' as 'and' and also incorrectly assumed independence. In parts (d) & (e), the candidate appears to have no knowledge that the word 'given' infers that conditional probabilities are required. The majority of candidates made fewer, sometimes, no mistakes.

Mark Scheme

Q	Solution	Marks	Total	Comments
2(a)	$\mathrm{P}(Blue) = \frac{160}{400} = 0.4$ or $\frac{2}{5}$ or $\frac{160}{400}$	B1	1	CAO; or equivalent
	In (b) to (e), method marks are for single fractions, or equivalents, only			
(b)	$\mathrm{P}(Marker) = \frac{280}{400}$	M1		$270 \leq$ Numerator ≤ 290 and Numerator $<$ Denominator ≤ 400
	$= 0.7$ or $\frac{7}{10}$ or $\frac{280}{400}$	A1	2	CAO; or equivalent
(c)	$\mathrm{P}(B \text{ or } M) = \mathrm{P}(B \cup M) =$			
	$\frac{160+280-119}{400} = \frac{280+41}{400} = \frac{321}{400}$	M1		$290 \leq$ Numerator ≤ 321 and Numerator $<$ Denominator ≤ 400
	$= 0.802$ to 0.803 or $\frac{321}{400}$	A1	2	AWFW/CAO (0.8025)
(d)	$\mathrm{P}(Green \mid Highlighter) = \mathrm{P}(G \mid H) = \frac{42}{120}$	M1		Numerator $= 42$ and $110 \leq$ Denominator ≤ 120
	$= 0.35$ or $\frac{7}{20}$ or $\frac{42}{120}$	A1	2	CAO; or equivalent
(e)	$\mathrm{P}(Non\text{-}Permanent \mid Red) = \mathrm{P}(P' \mid R) = \frac{21}{90}$	M1		Numerator $= 21$ and $80 \leq$ Denominator ≤ 90
	$= 0.233$ to 0.234 or $\frac{7}{30}$ or $\frac{21}{90}$	A1	2	AWFW/CAO (0.2333)
	Total		**9**	

Question 3

3 [**Figure 1**, printed on the insert, is provided for use in this question.]

The table shows, for each of a sample of 12 handmade decorative ceramic plaques, the length, x millimetres, and the width, y millimetres.

Plaque	x	y
A	232	109
B	235	112
C	236	114
D	234	118
E	230	117
F	230	113
G	246	121
H	240	125
I	244	128
J	241	122
K	246	126
L	245	123

(a) Calculate the value of the product moment correlation coefficient between x and y. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

(c) On **Figure 1**, complete the scatter diagram for these data. *(3 marks)*

(d) In fact, the 6 plaques A, B, ..., F are from a different source to the 6 plaques G, H, ..., L.

With reference to your scatter diagram, **but without further calculations**, estimate the value of the product moment correlation coefficient between x and y for **each** source of plaque. *(2 marks)*

Student Response

Question number

Leave blank

3a) $S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 681575 - \frac{(2859)^2}{12}$ ✓ ✓

$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 170342 - \frac{(1428)^2}{12}$ ✓ ✓

$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 340555 - \frac{(2859 \times 1428)}{12}$ ✓

$r = 0.807$ (3dp) ✓ 3

3b) There is a strong ✓ positive correlation between the length ~~of~~ and ✓ width of the handmade ceramic plaques. As the length (x) increases, the width (y) also increases. B2

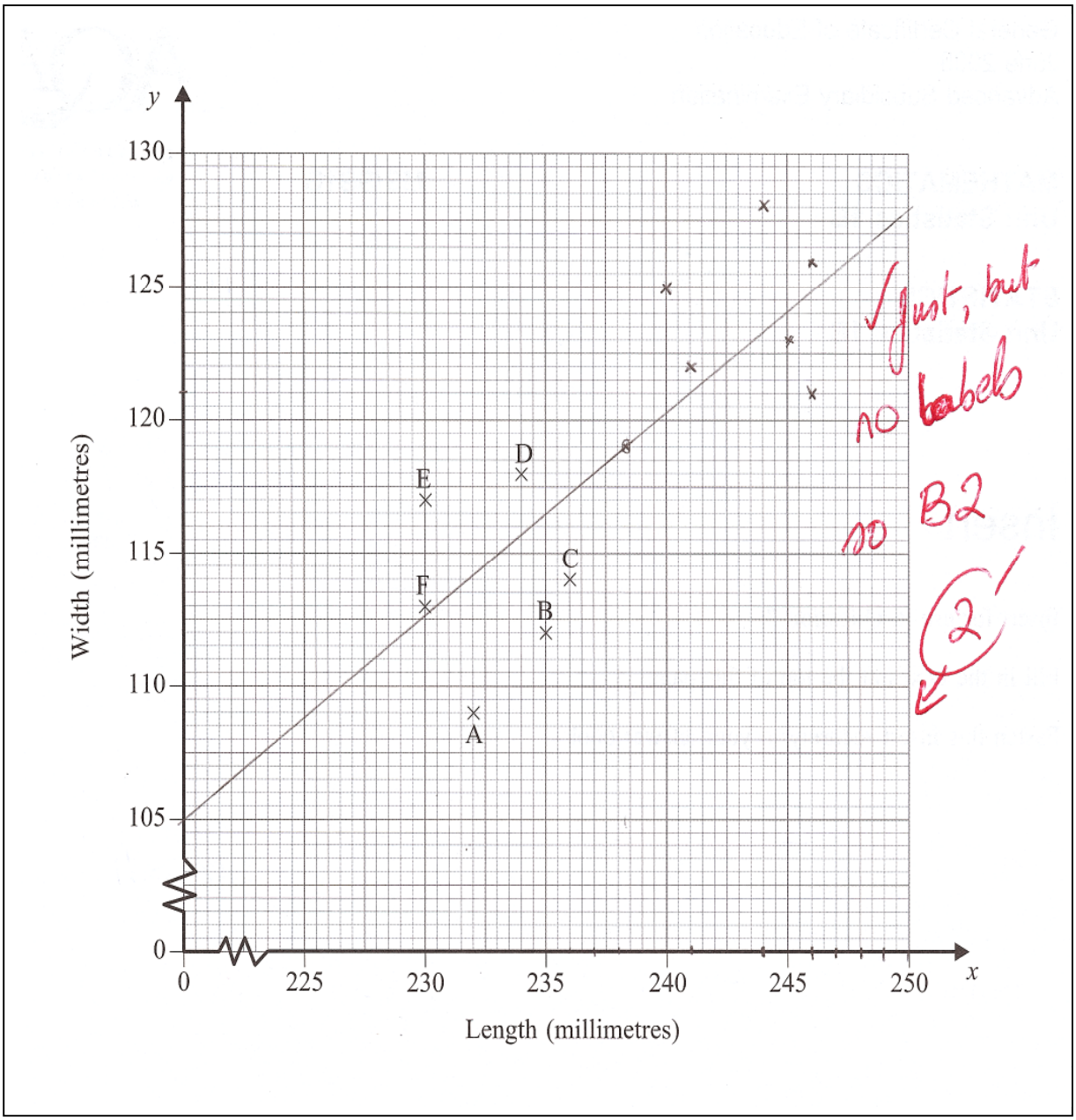
c) $\bar{x} = 238.25$

$\bar{y} = 119$ 2 ✓

d) Plaques A-F = + (positive) 0.7 ✗ ~~(0 < r < 1)~~ B0

Plaques G-L = + (positive) 0.75 ✗ (0 < r < 1) B0

7



Commentary

Most candidates scored the 3 marks in (a) simply using their calculators' inbuilt function. In cases as illustrated here, working **may** score marks even if the answer is incorrect. The points are plotted correctly on the graph but a mark is lost for no labels. The line thereon is unnecessary and so is ignored. As was sadly often the norm, the candidate appears to have no idea that, for **each** source, the points are so scattered as to indicate virtually no correlation so inferring that $r \approx 0$ for each.

Mark Scheme

Q	Solution	Marks	Total	Comments
3(a)	r = 0.806 to 0.807	B3	3	AWFW (0.80656)
	(r = 0.8(0) to 0.81)	(B2)		AWFW
	(r = 0.7 to 0.9)	(B1)		AWFW
	OR			
	Attempt at $\sum x$, $\sum x^2$, $\sum y$, $\sum y^2$ and $\sum xy$			2859, 681575, 1428, 170342 and 340555
	or	(M1)		
	Attempt at S_{xx}, S_{yy} and S_{xy}			418.25, 410 and 334
	Attempt at **correct** formula for r	(m1)		
	r = 0.806 to 0.807	(A1)		AWFW
(b)	**Moderate/fairly strong/strong positive correlation** (relationship/association)	B1		Or equivalent; must qualify strength and indicate positive B0 for some/average/medium/very strong/etc
	between **length and width** of plaques	B1	2	Context; providing $0 < r < 1$
(c)	**Figure 1:**			
	6 correct labelled points	B3	3	
	(5 correct labelled points)	(B2)		Deduct 1 mark if not labelled
	(4 correct labelled points)	(B1)		
(d)	A to F: r = –0.2 to +0.2	B1		AWFW (–0.0275) No penalties for calculations
	Accept 'Zero' but not 'No' correlation			Statements must include a **single value** within range
	G to L: r = –0.2 to +0.2	B1	2	AWFW (–0.0196)
	Special Cases:			
	r = –0.2 to +0.2 with **no** sources	(B1)		AWFW
	r = –0.2 to +0.2 for **each/both** source(s)	(B2)		AWFW; or equivalent identification
	If B0 B0 but both values of r = –0.4 to +0.4	(B1)		AWFW
	Total		**10**	

Question 4

4 The runs scored by a cricketer in 11 innings during the 2006 season were as follows.

47 63 0 28 40 51 a 77 0 13 35

The exact value of a was unknown but it was greater than 100.

(a) Calculate the median and the interquartile range of these 11 values. *(4 marks)*

(b) Give a reason why, for these 11 values:

(i) the mode is **not** an appropriate measure of average;

(ii) the range is **not** an appropriate measure of spread. *(2 marks)*

Student Response

(4) 0, 0, 13, 28, 35, 40, 47, 51, 63, 77, a

a) Median = $\frac{11+1}{2}$ = 6th term = 40 ✓

b) LQ = $\frac{11+1}{4}$ = 3rd term = 13 ✓

UQ = 3 x 3 = 9th term = 63 ✓

IQR = 63 - 13 = 50

∴ The inter-quartile range is 50 ✓ 4

b)

i) Because none of the values repeate, so mode is not valid. ✗ 0

ii) Range is also not valid as we do not know the value of the top end of the range ✓ as a is not given. 1

5

Commentary

The candidate has ranked the 11 values and then identified correct values for the median, (quartiles) and the interquartile range. As was often the case when answering part (b)(i), the candidate has stated 'none of the values repeat', this despite listing two values of zero in part (a)! Part (b)(ii) was answered correctly by indicating that the maximum value, a, is unknown.

Mark Scheme

Q	Solution	Marks	Total	Comments
4(a)	Ordering: 0 0 13 28 35 40 47 51 63 77 a	M1		May be implied by 40 and/or 63 and 13
	Median (6^{th}) = 40	B1		CAO
	IQR = $Q_3(9^{th})$ – $Q_1(3^{rd})$			
	= 63 – 13	(B1)		Identification of 63 and 13
	= 50	B2	4	CAO
(b)(i)	*Mode*: Zero is not representative / sensible reason Wide range of (known) values Small number of values mostly different	B1		Or equivalent
(ii)	*Range*: Largest value, a, is unknown Cannot be calculated	B1	2	Or equivalent
	Total		**6**	

Question 5

5 When a particular make of tennis ball is dropped from a vertical distance of 250 cm on to concrete, the height, X centimetres, to which it first bounces may be assumed to be normally distributed with a mean of 140 and a standard deviation of 2.5.

(a) Determine:

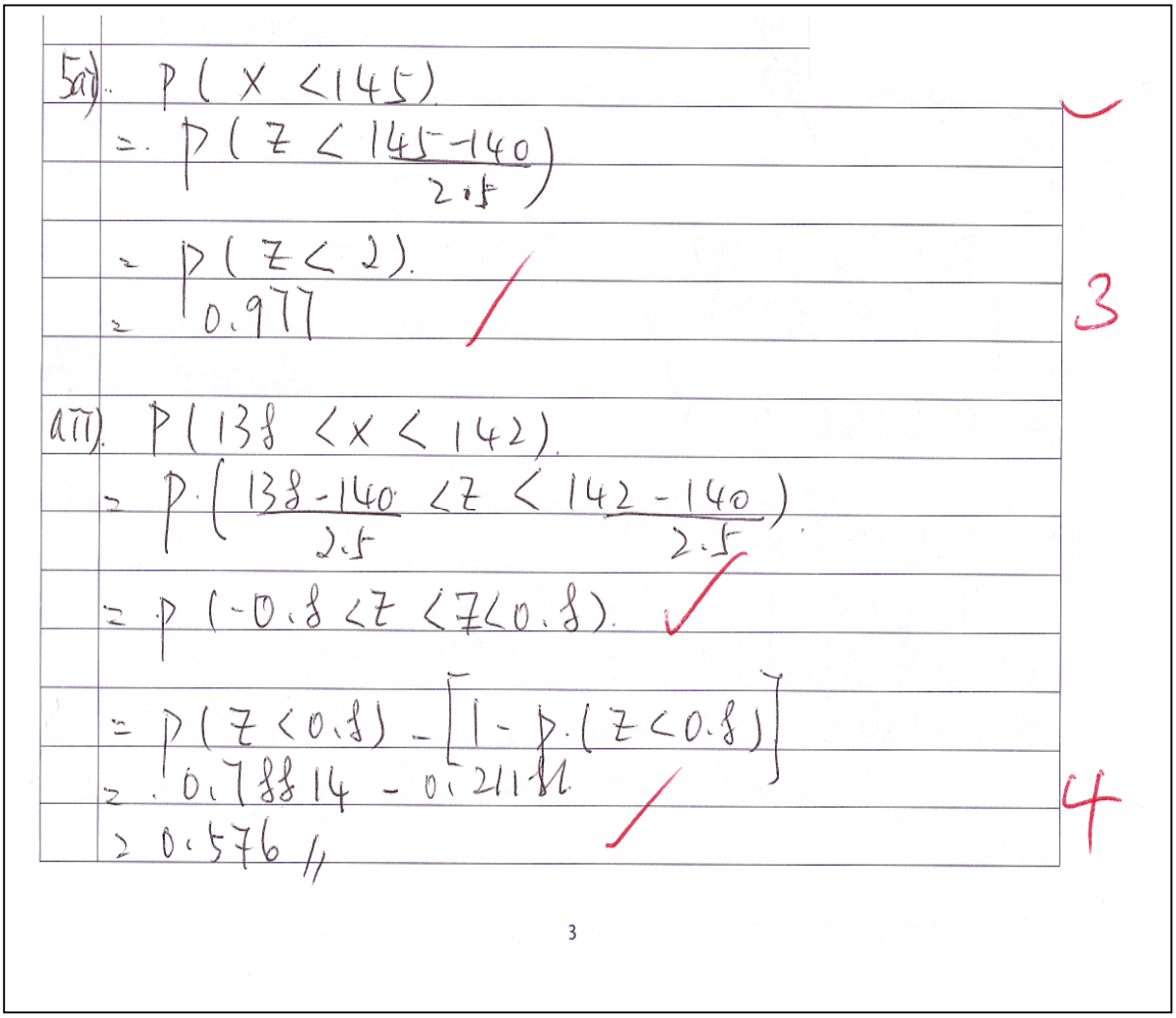
(i) $P(X < 145)$; *(3 marks)*

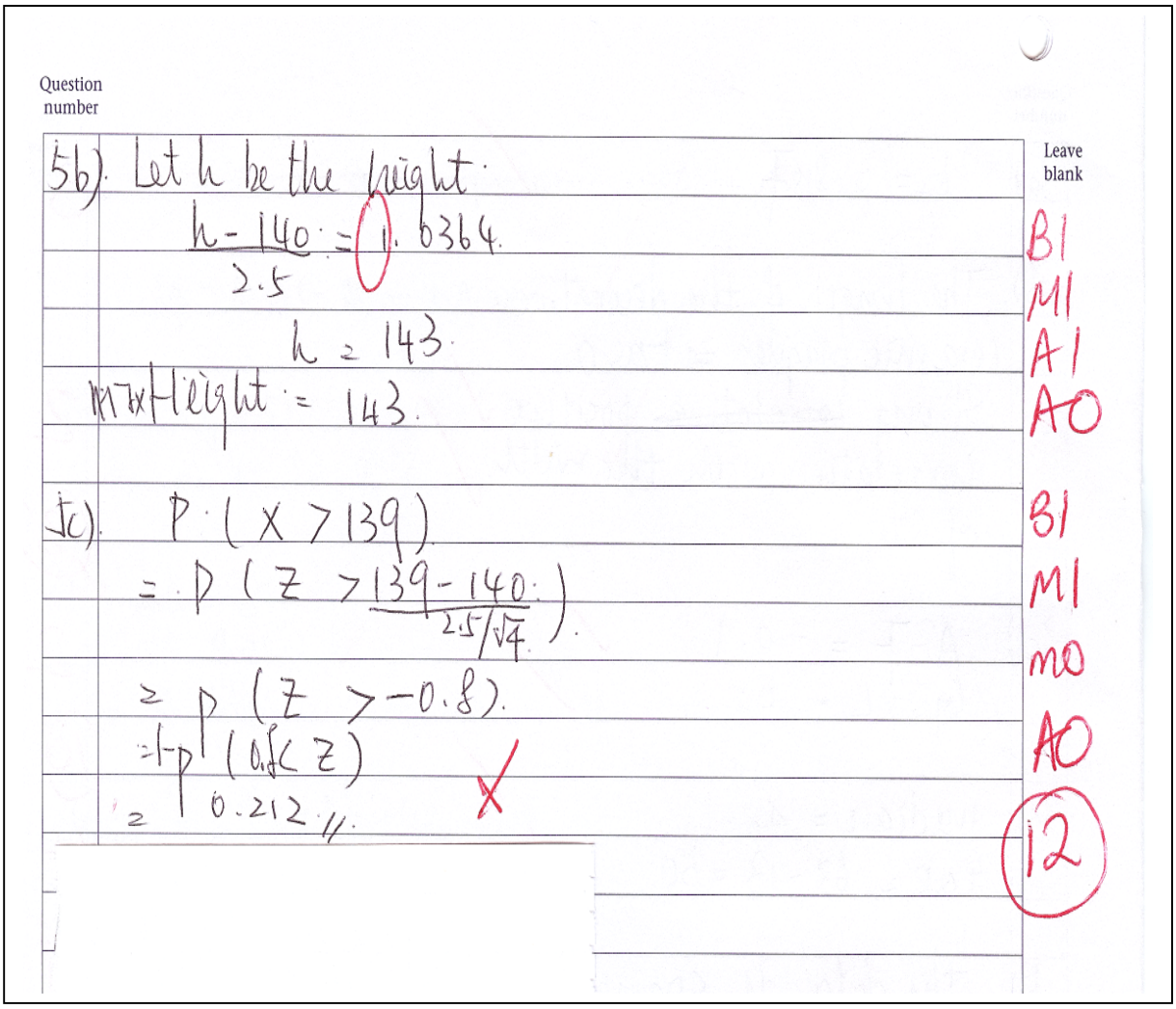
(ii) $P(138 < X < 142)$. *(4 marks)*

(b) Determine, to one decimal place, the maximum height exceeded by 85% of first bounces. *(4 marks)*

(c) Determine the probability that, for a random sample of 4 first bounces, the mean height is greater than 139 cm. *(4 marks)*

Student Response





Commentary

This is a typical less than fully-correct answer. The very standard parts (a)(i) & (ii) are answered correctly for $3 + 4 = 7$ marks. In part (b), as here, the majority of candidates opted for $85\% \Rightarrow z = (+)1.03$ to $(+)1.04$ and so obtained an answer greater than the mean of 140. Either a little thought or a sketch should have suggested that the answer must be less than 140? I part (c), the candidate has made the correct start of finding the standard error, then standardising correctly to $P(Z > -0.8)$ but has then made the common error of finding the equivalent of $P(Z < -0.8)$. Again a little thought or a sketch should have suggested that the answer must be greater than 0.5.

Mark Scheme

Q	Solution	Marks	Total	Comments
5	Height $X \sim N(140, 2.5^2)$			
(a)(i)	$P(X < 145) = P\left(Z < \frac{145-140}{2.5}\right) =$	M1		Standardising (144.5, 145 or 145.5) with 140 and ($\sqrt{2.5}$, 2.5 or 2.5^2) and/or (140 – x)
	$P(Z < 2) =$	A1		2 CAO; ignore sign
	0.977 to 0.98(0)	A1	3	AWFW (0.97725)
(ii)	$P(138 < X < 142) =$ $P(X < 142) - P(X < 138) =$	M1		Difference (142 – 138)
	$P(Z < 0.8) - P(Z < -0.8) =$	B1		0.8 CAO
	$P(Z < 0.8) - \{1 - P(Z < 0.8)\} =$ $(0.78814) - (1 - 0.78814) =$	m1		Correct area change
	0.576 to 0.58(0)	A1	4	AWFW (0.57628)
(b)	0.85 (85%) $\Rightarrow z = -1.03$ to -1.04	B1		AWFW; ignore sign (–1.0364)
	$z = \frac{x-140}{2.5}$	M1		Standardising x with 140 and 2.5; allow (140 – x)
	$= \pm 1.03$ to $\pm$ 1.04	A1		Equating z-term to the z-value
	Hence x = 137.3 to 137.5	A1	4	AWFW; CSO (137.41)
(c)	Variance of $\bar{X}_4 = \frac{2.5^2}{4} = 1.56(25)$ SD of $\bar{X}_4 = \frac{2.5}{2} = 1.25$	B1		CAO; stated or used
	$P(\bar{X}_4 > 139) = P\left(Z > \frac{139-140}{\sqrt{2.5^2/4}}\right) =$	M1		Standardising 139 with 140 and 1.25; allow (140 – 139)
	$P(Z > -0.8) = P(Z < 0.8) =$	m1		Correct area change
	0.788 to 0.79(0)	A1	4	AWFW (0.78814)
	Total		**15**	

Question 6

6 For the adult population of the UK, 35 per cent of men and 29 per cent of women do not wear glasses or contact lenses.

(a) Determine the probability that, in a random sample of 40 men:

(i) at most 15 do not wear glasses or contact lenses; *(3 marks)*

(ii) more than 10 but fewer than 20 do not wear glasses or contact lenses. *(3 marks)*

(b) Calculate the probability that, in a random sample of 10 women, exactly 3 do not wear glasses or contact lenses. *(3 marks)*

(c) (i) Calculate the mean and the variance for the number who **do** wear glasses or contact lenses in a random sample of 20 women. *(3 marks)*

(ii) The numbers wearing glasses or contact lenses in 10 groups, each of 20 women, had a mean of 16.5 and a variance of 2.50.

Comment on the claim that these 10 groups were **not** random samples. *(3 marks)*

Student Response

6a (i) men → 35% women → 29%

~~$X \sim B(40, 0.64)$~~ $X \sim B(40, 0.35)$

$P(x \leq 15) = 0.6946$

$= 0.695$,,

3

(ii) $P(10 < x < 20) = P(11 \leq x \leq 20)$

$= P(x \leq 20) - P(x \leq 10)$

$= 0.9827 - 0.1215$

$= 0.8612$

$= 0.861$,,

M2

A0

6b $X \sim B(10, 0.29)$

$P(X = 3) = {}^{10}C_3 (0.29)^3 (0.71)^7$

$= 0.266$,,

3

~~$X \sim B(20, 0.29)$~~

~~mean $= 20(0.29) = 5.8$~~

~~variance $= 20(0.29)(0.71) = 4.12$,,~~

WR

Question number

6c (i) $X \sim B(20, 0.71)$

mean $= 20(0.71) = 14.2$,,

variance $= 20(0.71)(0.29) = 4.12$,,

3

(ii) As the mean and variance are different, BOD therefore the claim is ~~[illegible]~~ valid, i.e. These 10 groups were not random samples.

3

14

Leave blank

Commentary

After a correct answer to part (a)(i), many candidates dropped at least 1 mark, as illustrated here, by not using the tables correctly for $P(10 < M < 20)$. The formula for $B(10, 0.29)$ was used correctly to find $P(F = 3)$ in part (b). In part (c)(i), the candidate has noted the emboldened word '**do**' and so moved to $B(10, 0.71)$ to find correct values for the mean and variance. As a result, correct comparisons are made for the results stated in part (c)(ii).

Mark Scheme

Q	Solution	Marks	Total	Comments
6	Binomial distribution	M1		Used somewhere in question
(a)(i)	$M \sim B(40, 0.35)$	A1		Used; may be implied
	$P(M \leq 15) = 0.69(0)$ to 0.696	A1	3	AWFW (0.6946)
(ii)	$P(10 < M < 20) =$			
	0.9637 or 0.9827	M1		Accept 3 dp accuracy
	minus 0.1215 or 0.0644	M1		Accept 3 dp accuracy
	$= 0.84(0)$ to 0.843	A1	3	AWFW (0.8422)
	OR			
	B(40, 0.35) expressions stated for **at least 3** terms within $10 \leq M \leq 20$	(M1)		Or implied by a correct answer
	Answer $= 0.84(0)$ to 0.843	(A2)		AWFW
(b)	$W \sim B(10, 0.29)$	B1		Used; may be implied
	$P(W = 3) = \binom{10}{3}(0.29)^3(0.71)^7$	M1		Stated; may be implied
	$= 0.266$ to 0.2665	A1	3	AWFW (0.2662) **Note**: $B(10, 0.3) \Rightarrow 0.2668$
(c)(i)	$n = 20 \quad p = 0.71$	B1		Stated or used; may be implied by 14.2
	Mean, $\mu = np = 14.2$	B1		CAO
	Variance, $\sigma^2 = np(1-p) = 4.11$ to 4.12	B1	3	AWFW (4.118)
(ii)	Mean of 16.5 is greater/different **or** $16.5/20 = 0.825$ is greater/different to 0.71	B1dep		Dependent on $\mu = 14.2$
	Means and variances are different	*(B2,1 dep)*		
	Variance of 2.50 is smaller/different	B1dep		Dependent on $\sigma^2 = 4.11$ to 4.12
	Suggests **claim** that groups are not random samples **is justified**	B1dep	3	Dependent on previous 2 marks Or equivalent
	Total		**15**	

Question 7

7 Vernon, a service engineer, is expected to carry out a boiler service in one hour.

One hour is subtracted from each of his actual times, and the resulting differences, x minutes, for a random sample of 100 boiler services are summarised in the table.

Difference	Frequency
$-6 \leqslant x < -4$	4
$-4 \leqslant x < -2$	9
$-2 \leqslant x < 0$	13
$0 \leqslant x < 2$	27
$2 \leqslant x < 4$	21
$4 \leqslant x < 6$	15
$6 \leqslant x < 8$	7
$8 \leqslant x \leqslant 10$	4
Total	**100**

(a) (i) Calculate estimates of the mean and the standard deviation of these differences. *(4 marks)*

(ii) Hence deduce, in minutes, estimates of the mean and the standard deviation of Vernon's actual service times for this sample. *(3 marks)*

(b) (i) Construct an approximate 98% confidence interval for the mean time taken by Vernon to carry out a boiler service. *(4 marks)*

(ii) Give a reason why this confidence interval is approximate rather than exact. *(1 mark)*

(c) Vernon claims that, more often than not, a boiler service takes more than an hour and that, on average, a boiler service takes much longer than an hour.

Comment, with a justification, on **each** of these claims. *(2 marks)*

Student Response (next page)

Question number

Leave blank

7.

a i. $\frac{-6-4}{2} = -5$ $\frac{-4-2}{2} = -3$

$\frac{-2-0}{2} = -1$ $\frac{0+2}{2} = 1$ $\frac{2+4}{2} = 3$

$\frac{4+6}{2} = 5$ $\frac{6+8}{2} = 7$ $\frac{8+10}{2} = 9$

$\bar{x} = 1.9$ minutes ✓

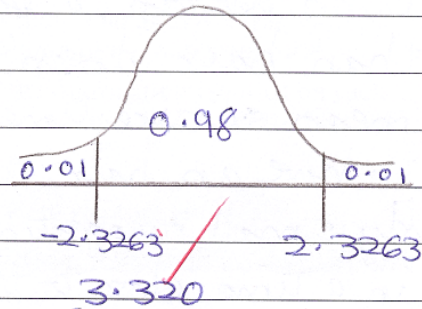
$\sigma_{n-1} = 3.320$ minutes ✓ 4

ii. $1.9 + 60 = 61.9$ minutes = mean

~~$= 3.320$ minutes~~ ✓

standard deviation ~~$3.320 + 60 = 63.32$ minutes~~ 3

b i. $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$



3.320

$61.9 \pm 2.3263 \times \frac{3.32}{\sqrt{100}}$ ✓

~~$[47.17, 76.63]$~~ $= [61.13, 62.67]$ ✓ 4

ii. the mean and standard deviation ~~is~~ are estimates because they have been calculated from a grouped frequency table ~~rather~~ using the

Question number

Leave blank

midpoints of the groups rather than the actual values. ~~This~~ These estimates of the mean and standard deviation were then used to calculate the confidence interval so this is ~~an~~ also an approximation. B1

c. His claim that more often than not a boiler ~~takes more than~~ service takes more than hour seems to be valid as from the sample, ~~37~~ 74 out of 100 boiler services took longer than an hour. The confidence interval shows that ~~you~~ 98% of the time the boiler service will take between 61.13 and 62.67 minutes so virtually every service will take longer than an hour as both these limits are above an hour. ✗ Common misunderstanding B1

His claim that, on average, a boiler service takes much longer than hour doesn't seem to be valid as his mean service time is 61.9 minutes which is just above an hour and his standard deviation is fairly small suggesting that he consistently services boilers in a time close to that mean. Only 1% of the boilers serviced should take longer than 62.67 minutes according to the confidence interval. ~~and~~ This ~~claim~~ claim is not valid ✓ OR Here B1

14

Commentary

This is an unusually fully correct answer to this final question; in fact from a 'perfect' script! Whilst correct answers to part (a)(i) were not unusual, far too many candidates could not answer part (a) (ii) correctly; usually through the addition of 1 or 100 to both answers in part (a)(i). Surely candidates at this level should know that one hour is 60 minutes? The two answers in part (b) are again correct and show a clear understanding of the technique needed. The verbose answer in part (c) does include the common misunderstanding that a confidence interval is for values rather than a mean but, in this instance, this error is just 'excused' in view of the other two salient points.

Mark Scheme

Q	Solution	Marks	Total	Comments
7(a)(i)	x: −5 −3 −1 1 3 5 7 9 f: 4 9 13 27 21 15 7 4			
	Mean $(\bar{x})$ = 1.9 (0.9 to 2.9)	B2 (B1)		CAO (190) AWFW
	Standard deviation $(s_{n-1}$ or $\sigma_n)$ = 3.3(0) to 3.32 (3(.00) to 3.5(0))	B2 (B1)	4	(1452) AWFW (3.31967) AWFW (3.30303)
	If no marks scored but $\sum fx$ attempted and result divided by 100	(M1)		
(ii)	Mean = 60 + $\bar{x}$ = 61.9	M1 A1√		√ on (a)(i)
	Standard deviation = 3.3(0) to 3.32	B1√	3	√ on (a)(i); accept 'same as' only providing answer in (a)(i)
(b)(i)	98% ⇒ z = 2.32 to 2.33 (⇒ t = 2.36 to 2.37)	B1		AWFW (2.3263) AWFW (2.364)
	CI for μ is $\bar{x} \pm z/t \times \frac{s_{n-1} \text{ or } \sigma_n}{\sqrt{n \text{ or } n-1}}$	M1		Used; must have $\sqrt{n}$ with $n > 1$
	Thus $61.9 \pm 2.3263 \times \frac{3.3 \text{ to } 3.32}{\sqrt{100 \text{ or } 99}}$	A1√		√ on (a)(ii) and z/t only
	Hence 61.9 ± (0.7 to 0.8) **or** (61.1 to 61.2, 62.6 to 62.7)	A1	4	Accept 1.03 ± (0.012 to 0.013) AWFW Accept (1.01 to 1.02, 1.04 to 1.05)
(ii)	Mean and SD based upon grouped data SD (**not mean**) calculated from a sample CLT used / Times (may) not (be) normal	B1	1	Actual times/values unknown Or equivalent
(c)	*S > 1 hour or 60 minutes*: **Valid** as 74/100 or 0.74 or 74% > 50%	B1		Must use 74 etc Or equivalent
	$\bar{S}$ >> *1 hour or 60 minutes*: **Not valid** as UCL ≈ 1 hour (Accept Both limits ≈ 1 hour)	B1dep	2	Dependent on UCL = 62.6 to 62.7 or UCL = 1.04 to 1.05
	Total		**14**	
	TOTAL		**75**	