Teacher Support Materials 2009

Maths & Statistics GCE

MS/SS1A/W

Question 1

1 A large bookcase contains two types of book: hardback and paperback. The number of books of each type in each of four subject categories is shown in the table.

		Subject category				
		Crime	**Romance**	**Science fiction**	**Thriller**	**Total**
Type	**Hardback**	8	16	18	18	60
	Paperback	16	40	14	30	100
	Total	24	56	32	48	160

(a) A book is selected at random from the bookcase. Calculate the probability that the book is:

(i) a paperback; *(1 mark)*

(ii) not science fiction; *(2 marks)*

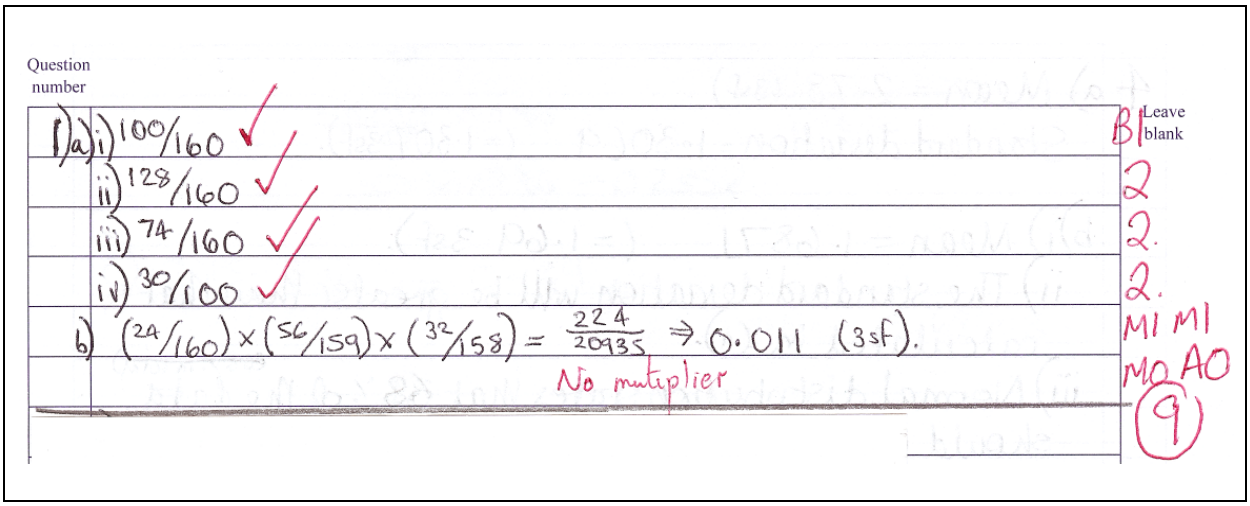
(iii) science fiction or a hardback; *(2 marks)*

(iv) a thriller, given that it is a paperback. *(2 marks)*

(b) Three books are selected at random, without replacement, from the bookcase.

Calculate, to three decimal places, the probability that one is crime, one is romance and one is science fiction. *(4 marks)*

Student Response



Commentary

This illustrates a typical answer. The candidate has answered the 4 parts of part (a) correctly, albeit without simplifying the expressions to fractions in their simplest forms or to decimals (incorrect simplifications were not penalised). The candidate has obtained a correct expression for one permutation (many candidates divided each column total by 160) but has then not realised that there are 3! = 6 possible permutations.

Mark scheme

MS/SS1A/W

Q	Solution	Marks	Total	Comments
1(a)				**In (a), ratios (eg 100:160) are only penalised by 1 mark at first correct answer**
(i)	$P(P) = 100/160 = 50/80 = 25/40 = 10/16$			
	$= 5/8 = 0.625$	B1	1	CAO
(ii)	$P(S') = 1 - \frac{32}{160}$ or $P(S) = \frac{32}{160}$	M1		Or equivalent Ignore labels of S' & S Can be implied by **correct** answer
	$= 128/160 = 64/80 = 32/40 = 16/20 = 8/10$			
	$= 4/5 = 0.8$	A1	2	CAO
(iii)	$P(S \text{ or } H) = P(S \cup H) =$			
	$\frac{60+32-18}{160}$ or $\frac{60+14}{160}$ or $\frac{32+8+16+18}{160}$	M1		Or equivalent Can be implied by **correct** answer
	$= 74/160 = 37/80 = 0.462$ to 0.463	A1	2	CAO/AWFW (0.4625)
(iv)	$P(T \mid P) = \frac{30/160}{\text{(i)}}$	M1		Or equivalent Can be implied by **correct** answer but watch for $18/60$ or $48/160$
	$= 30/100 = 3/10 = 0.3$	A1	2	CAO
(b)	$P(1C \ \& \ 1R \ \& \ 1S) =$			
	$\frac{24}{160} \times \frac{56}{159} \times \frac{32}{158}$	M1		Multiplication of any 3 different given subject totals
		M1		Multiplication of 160, 159 & 158
	$(0.15 \times 0.35220 \times 0.20253)$			Accept 3 dp accuracy
	$\times 6$	M1		Award for $3 \leq$ multiplier ≤ 6
	$= 0.064$ to 0.0644	A1		AWFW (0.0642) Do not accept a fraction as answer A **correct** answer can imply 4 marks
	Special Case: (Any given subject total) ÷ 160 seen anywhere in (b)	(M1)	4	Can award if no marks scored in (b) Accept a decimal equivalent
	Total		**11**	

Question 2

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

Hermione, who is studying reptiles, measures the length, x cm, and the weight, y grams, of a sample of 11 adult snakes of the same type. Her results are shown in the table.

Snake	A	B	C	D	E	F	G	H	I	J	K
x	46	39	54	79	47	58	73	35	43	51	36
y	55	48	58	88	61	55	82	51	50	66	57

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

(b) Interpret your value in context. *(2 marks)*

(c) On **Figure 1**, complete the scatter diagram for these data. *(2 marks)*

(d) Subsequently it is found that, of the 11 adult snakes, 9 are male and 2 are female.

(i) Given that female adult snakes are generally larger than male adult snakes, identify the 2 snakes which are most likely to be female. *(1 mark)*

(ii) Hence, **without further calculation**, estimate the value of r for the 9 male snakes and revise, as necessary, your interpretation in part (b). *(2 marks)*

Student Response

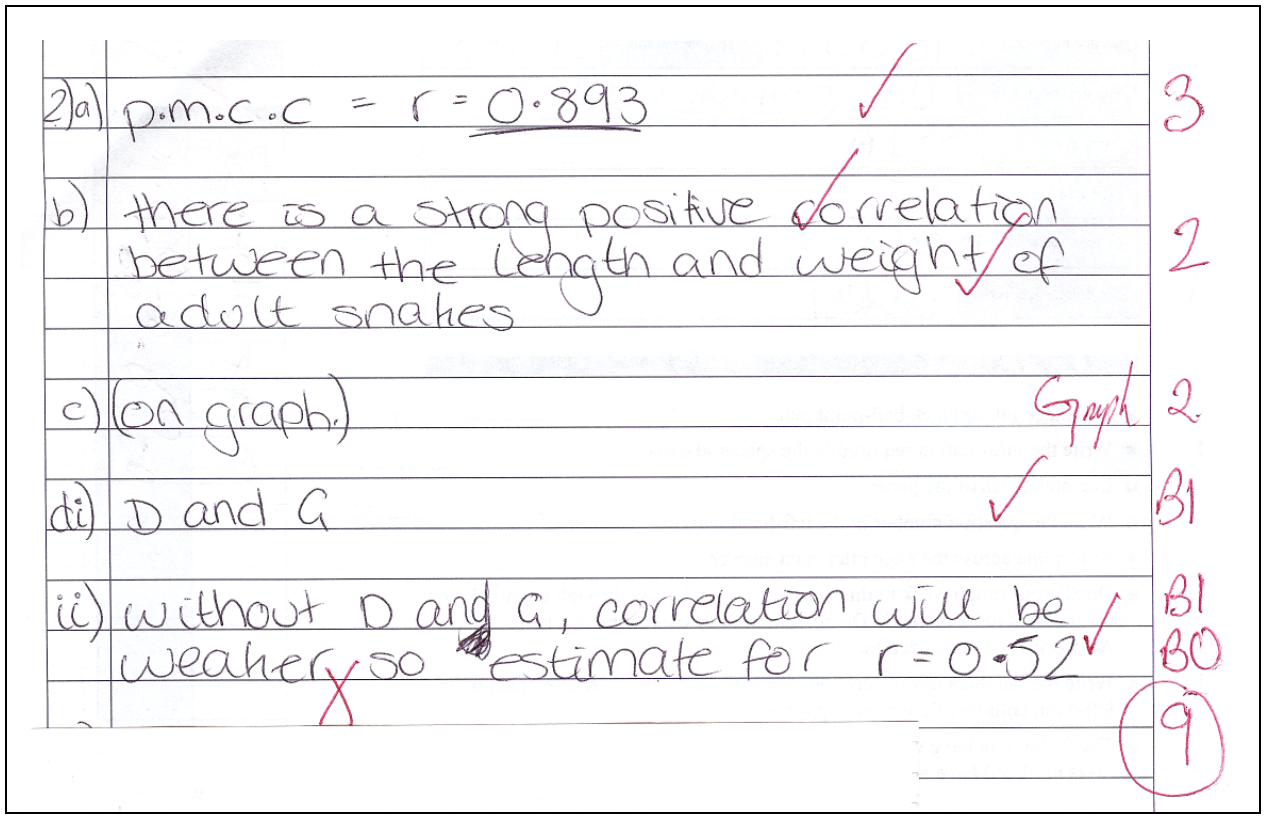
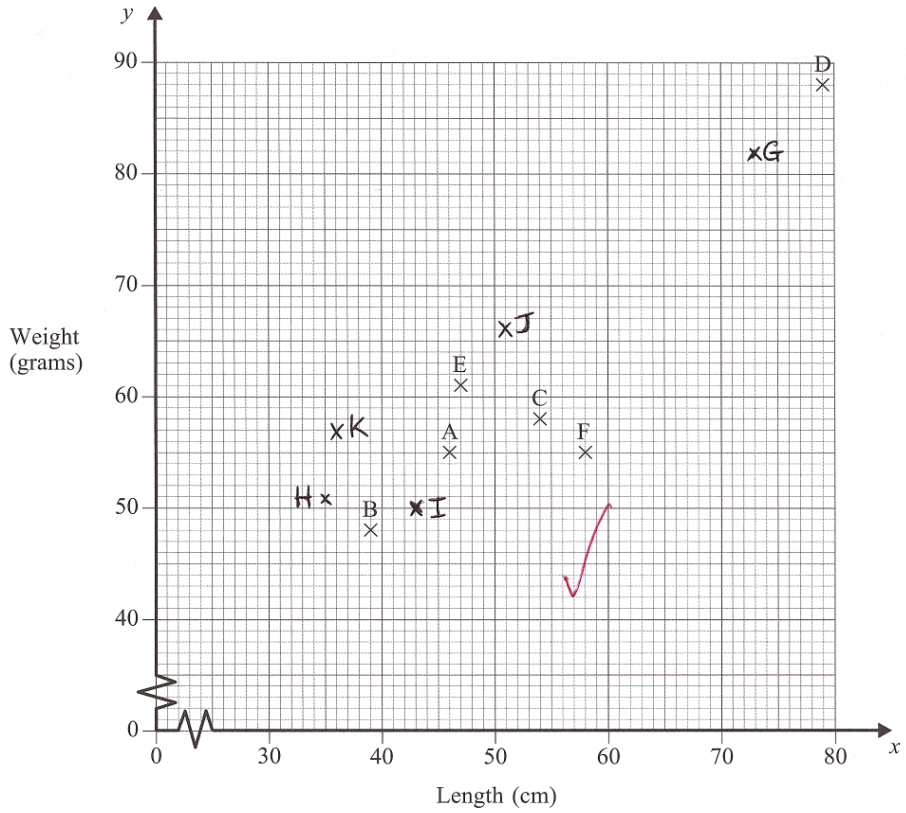


Figure 1 (for use in Question 2)

Lengths and Weights of Snakes



Commentary

This illustrates another typical answer that almost scored full marks. As expected the candidate made good use of a calculator's inbuilt statistical function to obtain the correct value of *r* quoted to an appropriate number of decimal places (extra places were not penalised but quoting to only 2 dp was penalised). Part (b) required a reference to the strength and the sign of the correlation in context; all referenced here. The 5 points were accurately plotted and labelled (candidates were penalised for omitting the latter). The candidate has identified the two most likely female snakes and estimated the value of *r* for the remaining 9 male snakes. However, as was often the case, the candidate's revised interpretation was rather a comparison with that stated in part (b).

Mark Scheme

MS/SS1A/W (cont)

Q	Solution	Marks	Total	Comments
2(a)	r = 0.893 to 0.8933	B3		AWFW (0.89319)
	r = 0.89 to 0.896	(B2)		AWFW
	r = 0.8 to 0.95	(B1)		AWFW
	or			
	Attempt at $\sum x$ $\sum x^2$ $\sum y$ $\sum y^2$ & $\sum xy$			561 30667 671 42613 & 35882 (all 5 attempted)
	or	(M1)		
	Attempt at S_{xx} S_{yy} & S_{xy}			2056 1682 & 1661 (all 3 attempted)
	Attempt at **correct** corresponding formula for r	(m1)		
	r = 0.893 to 0.8933	(A1)	3	AWFW
(b)	Fairly strong / strong / very strong positive (linear) correlation / relationship / association / link (but not 'trend')	B1dep		Or equivalent; must qualify strength and indicate positive Dependent on $0.8 \leq r \leq 0.95$ B0 for some/average/medium/etc
	between **length and weight** of adult snakes	B1	2	Context; providing $0 < r < 1$
(c)	Figure 1: 5 correct labelled points	B2		Deduct 1 mark if > 1 point not labelled
	4 or 3 correct labelled points	(B1)	2	
(d)(i)	D and G	B1	1	Both CAO
(ii)	r = 0.25 to 0.75	B1		AWFW (0.48790) No penalty for calculation Accept a range only if whole of it falls within 0.25 to 0.75
	Fairly weak / weak / some / moderate positive (linear) correlation / relationship / association / link	B1dep		Or equivalent; must qualify strength and indicate positive Dependent on $0.25 \leq r \leq 0.75$ B0 for very weak/little/slight/ hardly any/fair/average/medium/ anything involving strong/etc
	Do not accept comparison with value in (a) or statement in (b)		2	
	Total		**10**	

Question 3

3 The weight of gravel, in kilograms, collected by a power shovel may be modelled by a normal distribution with unknown mean, μ, and standard deviation 50.

The weights of a random sample of 20 collections have a mean of 1030 kg.

(a) Construct a 98% confidence interval for μ, giving the limits to four significant figures. *(4 marks)*

(b) Comment on a claim that the power shovel is, on average, collecting more than 1000 kg of gravel. *(2 marks)*

Student Response

Question number

Leave blank

3) Mean $= \mu$ $\sigma = 50$

Sample $n = 20$ $\bar{x} = 1030$ kg

a) 98% confidence interval (interval 90%) 1% each side

$\mu = \bar{x} \pm Z\left(\frac{\sigma}{\sqrt{n}}\right)$ $\mu = 1030 \pm 2.3263\left(\frac{50}{\sqrt{20}}\right)$

$1030 + 2.3263 \times \left(\frac{50}{\sqrt{20}}\right) = 1056.0088\ldots$

$1030 - 2.3263 \times \left(\frac{50}{\sqrt{20}}\right) = 1003.9911\ldots$

μ (1004, 1056) kg

4

b) Collecting more than 1000 kg

There is a 98% certainty that on average the power shovel collects >1000 as the lower limit is larger than 1000 so it must collect >1000 on average.

2

6

Commentary

This illustrates one of the better explained answers to this question. Whilst many candidates scored the 4 marks available in part (a), few produced such a well-documented solution. In attempting part (b), most candidates failed to match their CI to an average stating rather that the shovel always collects more than 1000 kg. Here the candidate has clearly pointed out that LCL > 1000 so claim of more than 1000 kg on average appears valid.

Mark Scheme

Q	Solution	Marks	Total	Comments
3(a)	98% (0.98) $\Rightarrow$ z = 2.32 to 2.33	B1		AWFW (2.3263)
	CI for μ is $\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$	M1		Used Must have $\sqrt{n}$ with $n > 1$
	Thus $1030 \pm 2.3263 \times \frac{50}{\sqrt{20}}$	A1F		F on z only
	Hence 1030 ± 26 **or** (1004, 1056)	A1	4	CAO & AWRT AWRT
(b)	Whole of confidence interval is **above 1000** so	B1F		F on (a) Or equivalent
	Agree with claim	B1F dep	2	F on (a) Or equivalent Dependent on previous B1F
		Total	**6**	

Question 4

4 A survey of all the households on an estate is undertaken to provide information on the number of children per household.

The results, for the 99 households with children, are shown in the table.

Number of children (x)	1	2	3	4	5	6	7
Number of households (f)	14	35	25	13	9	2	1

(a) For these 99 households, calculate values for the mean and the standard deviation. *(3 marks)*

(b) In fact, 163 households were surveyed, of which 64 contained no children.

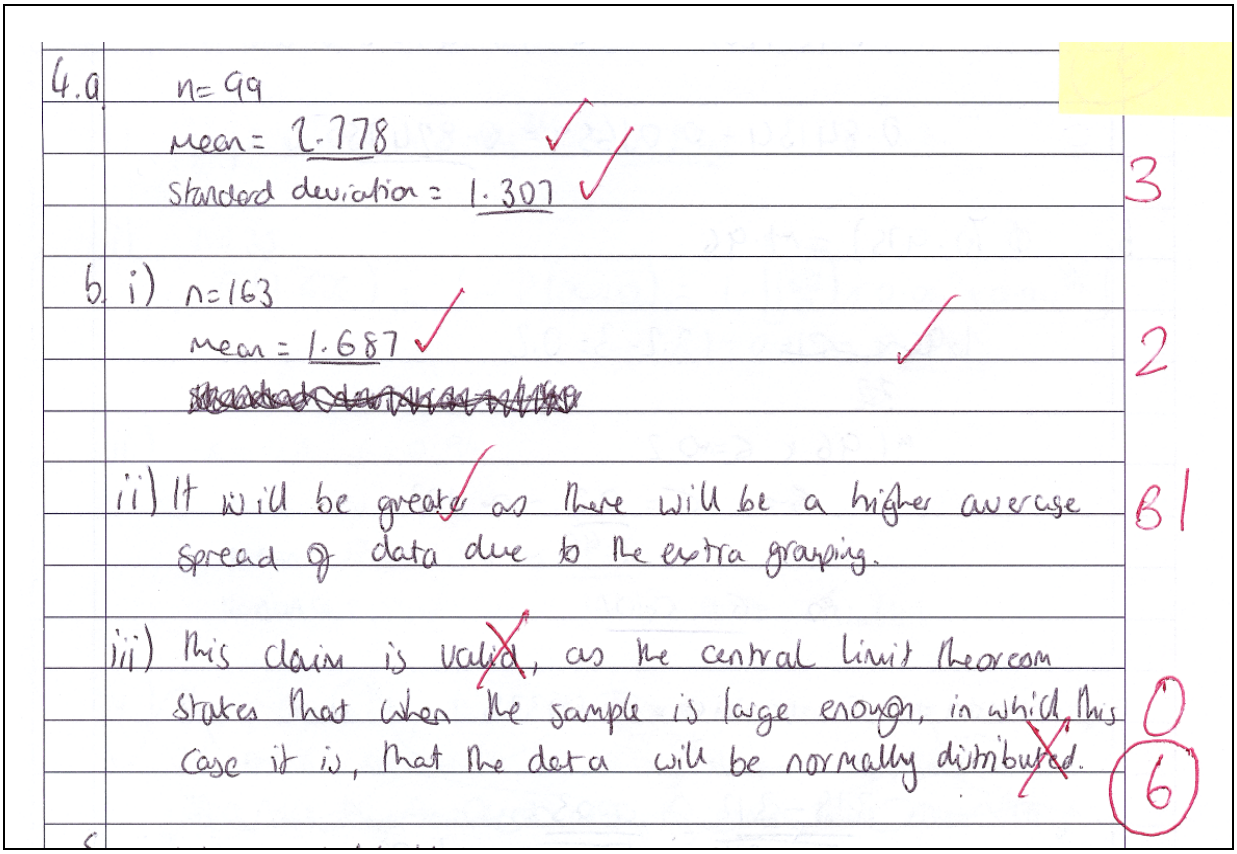
(i) For all 163 households, calculate the value for the mean number of children per household. *(2 marks)*

(ii) State whether the value for the standard deviation, when calculated for all 163 households, will be smaller than, the same as, or greater than that calculated in part (a). *(1 mark)*

(iii) It is claimed that, for all 163 households on the estate, the number of children per household may be modelled approximately by a normal distribution.

Comment, with justification, on this claim. Your comment should refer to a fact other than the discrete nature of the data. *(2 marks)*

Student Response



Commentary

Many candidates scored the first 6 marks; again making use of their calculator's inbuilt statistical functions in part (a) and often also in part (b). This is to be encouraged. As here, most candidates scored the mark for 'greater' in part (b)(ii) with some obviously calculating its value to make sure! The candidate, in common with the vast majority, simply appears to not understand the CLT. As a result, all too often 0 marks were scored. **The CLT is irrelevant here as it deals with the *distribution of the sample mean*; not the distribution of the sample or population!** What was needed here was a reference to the non symmetry of the population of children per household or to the fact that (mean) – 2 × (standard deviation) < 0.

Mark Scheme

4(a)	Mean $= \frac{\sum fx}{\sum f} = \frac{275}{99} = 2.77$ to 2.78	B1		AWFW (2.778)
	If not identified, assume order is $\bar{x}$ then s			Treat rounding to integers as ISW
	SD $\left(\sum fx^2 = 933\right) = 1.3(0)$ to 1.32	B2		AWFW (1.307 & 1.314)
	Special Case:			
	Evidence of $\frac{\sum fx}{99}$	(M1)	3	Can award if no marks scored in (ii)
(b\)(i)	$\text{Mean}_{163} = \frac{99 \times \text{Mean}_{99}}{163}$ or $\frac{\sum fx \text{ from(a)(ii)}}{163}$	M1		Or equivalent; may be implied by an answer within range
	$= 1.68$ to 1.69	A1	2	AWFW (1.687)
(ii)	Increase	B1	1	CAO; or equivalent (1.696) Ignore any working (1.702)
(iii)	Data is (positively/negatively) skewed / not symmetric / bimodal / not bell-shaped from frequency distribution / given table			
	or	B1		Or equivalent
	[C's mean in (b)(i)] – 2 × [C's SD in (a)(ii)] <			
	or			(–1.75 to –0.90)
	[C's mean in (b)(i)] – 2 × [1.69 to 1.71] < 0			
	Thus claim appears **not valid**	B1 dep	2	Or equivalent Dependent upon previous B1
	Total		**8**	

Question 5

5 A machine fills boxes with wine. The volume, W litres, of wine delivered by the machine into a box may be modelled by a normal distribution with mean 3.12 and standard deviation σ.

(a) Given that $\sigma = 0.08$, determine $P(2.95 < W < 3.20)$. *(4 marks)*

(b) Assuming that the value of the mean remains unchanged, determine the value of σ necessary to ensure that at most 2.5% of boxes filled by the machine contain less than 3 litres of wine. *(4 marks)*

(c) After an adjustment to the machine, W can be modelled by a normal distribution with mean 3.12 and variance 0.00375.

Determine the probability that the **mean** volume of wine in 5 boxes, selected at random from those filled by the machine, is less than 3.15 litres. *(3 marks)*

Student Response

Leave blank

5a. $\mu = 3.12 \quad \sigma = 0.08$

$z = \frac{3.2 - 3.12}{0.08} = 1$ ✓ $\quad p(z<1) = 0.84134$ ✓

$z = \frac{2.95 - 3.12}{0.08} = -2.125$ ✓ $\quad p(z > -2.125) = 0.98341$ ✓

$1 - 0.98341 = 0.01659$ ✓

$p(2.95 < W < 3.20) = 0.84134 - 0.01659$ ✓

$= 0.82475$

$= 0.825$ 4

b. 97.5% = 1.96 ✓ $P(W > 3.)$ B1

~~98% = 2.3263~~

1.96 ✓

$2.3263 = \frac{3 - 3.12}{\sigma}$ M1 A1

1.96

$2.3263\sigma = -0.12$

$\sigma = \frac{-0.12}{1.96} = -0.061$ Minus sign! A0

$\sigma = 0.061$

c. $\mu = 3.12 \quad V = 0.00375 \quad \sigma = 0.61237$ B0

No $\sqrt{n}$

$z = \frac{3.15 - 3.12}{0.61237} = 0.0489$ M0

$p(z < 0.49) = 0.68793$ ✗ A0

7

Commentary

As expected, the candidate has answered part (a) correctly with pleasing amount of detail. Such detail can benefit a candidate whose answer is incorrect. As was often the case, the candidate appears to have tried to trick the examiner by losing a minus sign; some hope! The initial use of +1.96 often gave it away. In part (c), the critical common error of standardising using the given standard deviation of $\sqrt{0.00375}$ rather than $\sqrt{\frac{0.00375}{5}}$ lost all 3 marks.

Mark Scheme

Q	Solution	Marks	Total	Comments
5(a)	$W \sim N(3.12,\ 0.08^2)$			
	$P(2.95 < X < 3.20) = P\left(\frac{2.95-3.12}{0.08} < Z < \frac{3.20-3.12}{0.08}\right)$	M1		Standardising (2.945, 2.95 or 2.955) or (3.195, 3.20 or 3.205) with 3.12 and ($\sqrt{0.08}$, 0.08 or 0.08^2) and/or $(3.12 - x)$
	$= P(-2.125 < Z < 1)$	A1		Either; CAO 1 AWFW –2.13 to –2.12
	$= P(Z < 1) - [1 - P(Z < 2.125)]$	m1		Area change; may be implied
	$= 0.84134 - [1 - (0.98300 \text{ to } 0.98341)]$			
	$= 0.824$ to 0.825	A1	4	AWFW (0.82455) (1 – answer) ⇒ M1 A1 max
(b)	2.5% (0.975) ⇒ $z = -1.96$	B1		AWRT; ignore sign (–1.9600)
	$z = \frac{3-3.12}{\sigma}$	M1		Standardising 3 with 3.12 and σ; allow $(3.12 - 3)$
	$= -1.96$	A1		Only allow: ±1.96 ±1.64 to ±1.65
	$\sigma = 0.06$ to 0.0613	A1		AWFW (0.06122)
	Note: $\frac{3-3.12}{\sigma} = 1.96 \Rightarrow \sigma = 0.06122$ ⇒ B1 M1 A1 A0		4	Or equivalent inconsistent signs
(c)	$W \sim N(3.12,\ 0.00375)$			
	Variance of $\bar{W}_5 = 0.00375/5 = 0.00075$			CAO
	SD of $\bar{W}_5 = \sqrt{0.00375}/\sqrt{5}$ $= 0.0273$ to 0.0275	B1		Stated or used AWFW
	$P(\bar{W}_5 < 3.15) = P\left(Z < \frac{3.15-3.12}{\sqrt{0.00375/5}}\right)$	M1		Standardising 3.15 with 3.12 and $\sqrt{0.00075}$ or equivalent; allow $(3.12 - 3.15)$
	$= P(Z < 1.09 \text{ to } 1.1) = 0.862$ to 0.865	A1	3	AWFW (0.86334) (1 – answer) ⇒ B1 M1 max
		Total	**11**	

Question 6

6 Mr Alott and Miss Fewer work in a postal sorting office.

(a) The number of letters per batch, R, sorted incorrectly by Mr Alott when sorting batches of 50 letters may be modelled by the distribution $B(50, 0.15)$.

Determine:

(i) $P(R < 10)$;

(ii) $P(5 \leqslant R \leqslant 10)$. *(4 marks)*

(b) It is assumed that the probability that Miss Fewer sorts a letter incorrectly is 0.06, and that her sorting of a letter incorrectly is independent from letter to letter.

(i) Calculate the probability that, when sorting a batch of **22** letters, Miss Fewer sorts exactly 2 letters incorrectly. *(3 marks)*

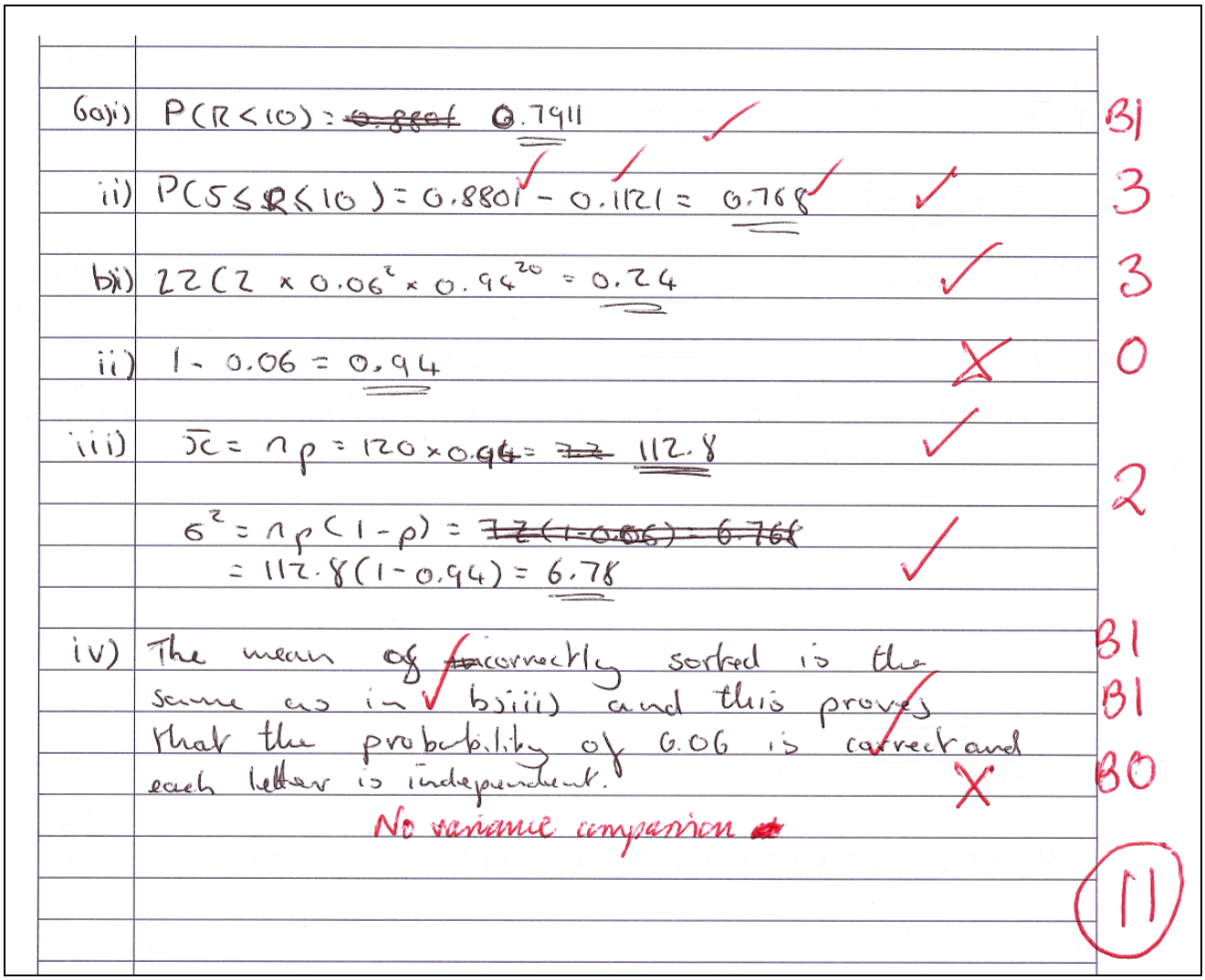
(ii) Calculate the probability that, when sorting a batch of **35** letters, Miss Fewer sorts at least 1 letter incorrectly. *(2 marks)*

(iii) Calculate the mean and the variance for the number of letters sorted **correctly** by Miss Fewer when she sorts a batch of **120** letters. *(2 marks)*

(iv) Miss Fewer sorts a random sample of 20 batches, each containing 120 letters. The number of letters sorted **correctly** per batch has a mean of 112.8 and a variance of 56.86.

Comment on the assumptions that the probability that Miss Fewer sorts a letter incorrectly is 0.06, and that her sorting of a letter incorrectly is independent from letter to letter. *(3 marks)*

Student Response



Commentary

This is a typical above average answer to this final question on the binomial distribution. Unlike many candidates who lost marks for 0.8801 in part (a)(i) and/or the use of 0.2194 in part (a)(ii), the candidate scored all 4 marks. The candidates correct answer to part (b)(i) was mirrored by most candidates. Unfortunately the same applies to part (b)(ii) where very few candidates equated P(at least 1) to 1 – P(0). However, things usually improved through correct answers in part (b)(iii); perhaps '**correctly**' helped? The answer above to part (b)(iv) is better than most seen. All too often candidates failed to compare means (equal) and variances (different) and so scored no marks.

Mark Scheme

Q	Solution	Marks	Total	Comments
6(a)	$R \sim B(50,\ 0.15)$			
(i)	$P(R < 10) = 0.791$	B1		AWRT (0.7911)
(ii)	$P(5 \le R \le 10) = 0.8801$ or 0.7911 (p_1)	M1		Accept 3 dp accuracy $(1 - p_2) - p_1 \Rightarrow$ M0 M0 A0 $p_1 - (1 - p_2) \Rightarrow$ M1 M0 A0 only providing result > 0
	minus 0.1121 or 0.2194 (p_2)	M1		Accept 3 dp accuracy
	$= 0.768$	A1		AWRT (0.7680)
	or B(50, 0.15) expressions stated for **at least 3** terms within $4 \le R \le 10$ gives probability	M1		Can be implied by correct answer
	$= 0.768$	A2	4	AWRT
(b)	**Confusion** of 22, 35, 120 and/or 0.15, 0.06			Do **not** treat as misreads
(i)	$S \sim B(22,\ 0.06)$	M1		Used in (b)(i) as evidenced by any correct binomial term for $S > 0$
	$P(S = 2) = \binom{22}{2}(0.06)^2(0.94)^{20}$	A1		Can be implied by correct answer Ignore any additional terms
	$= 0.24$ to 0.242	A1	3	AWFW (0.24125)
(ii)	$P(S \ge 1) = 1 - q^{35}$ where $0.84 \le q \le 0.96$	M1 (B1)		Can be implied by correct answer Award for $(0.94)^{35}$ seen in an expression but not if accompanied by a multiplier $\ne 1$
	$= 0.885$ to 0.89	A1	2	AWFW (0.88532)
(iii)	Mean $= np = 120 \times 0.94 = 112.8$ or 113 If not identified, assume order is μ then σ^2	B1		Either
	Variance $= np(1-p)$ $= 120 \times 0.94 \times 0.06 = 6.76$ to 6.78	B1	2	Must clearly state variance value AWFW (6.768)
(iv)	**Means** are (approximately) the **same** stated or **Variances** are (very) **different** stated	B1		Must have scored 1st B1 in (iii) Must have scored 2nd B1 in (iii)
	Agree with P(sorts letter incorrectly) = 0.06	B1 dep		Dependent on 'means same' stated
	Disagree with independent from letter to letter	B1 dep	3	Dependent on 'variances different' stated
	Total		**14**	