OCR RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE **G241/01**

MEI STATISTICS

Statistics 1 (Z1)

FRIDAY 6 JUNE 2008 Afternoon

Time: 1 hour 30 minutes

Additional materials (enclosed): None

Additional materials (required):

Answer Booklet (8 pages)
Graph paper
MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is **72**.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

This document consists of **6** printed pages and **2** blank pages.

Section A (36 marks)

1 In a survey, a sample of 44 fields is selected. Their areas (x hectares) are summarised in the grouped frequency table.

Area (x)	$0 < x \leqslant 3$	$3 < x \leqslant 5$	$5 < x \leqslant 7$	$7 < x \leqslant 10$	$10 < x \leqslant 20$
Frequency	3	8	13	14	6

(i) Calculate an estimate of the sample mean and the sample standard deviation. [4]

(ii) Determine whether there could be any outliers at the upper end of the distribution. [2]

2 In the 2001 census, people living in Wales were asked whether or not they could speak Welsh. A resident of Wales is selected at random.

- W is the event that this person speaks Welsh.
- C is the event that this person is a child.

You are given that $\mathrm{P}(W) = 0.20$, $\mathrm{P}(C) = 0.17$ and $\mathrm{P}(W \cap C) = 0.06$.

(i) Determine whether the events W and C are independent. [2]

(ii) Draw a Venn diagram, showing the events W and C, and fill in the probability corresponding to each region of your diagram. [3]

(iii) Find $\mathrm{P}(W|C)$. [2]

(iv) Given that $\mathrm{P}(W|C') = 0.169$, use this information and your answer to part **(iii)** to comment very briefly on how the ability to speak Welsh differs between children and adults. [1]

3 In a game of darts, a player throws three darts. Let X represent the number of darts which hit the bull's-eye. The probability distribution of X is shown in the table.

r	0	1	2	3
$\mathrm{P}(X = r)$	0.5	0.35	p	q

(i) (A) Show that $p + q = 0.15$. [1]

(B) Given that the expectation of X is 0.67, show that $2p + 3q = 0.32$. [1]

(C) Find the values of p and q. [2]

(ii) Find the variance of X. [3]

4 A small business has 8 workers. On a given day, the probability that any particular worker is off sick is 0.05, independently of the other workers.

(i) A day is selected at random. Find the probability that

(A) no workers are off sick, [2]

(B) more than one worker is off sick. [3]

(ii) There are 250 working days in a year. Find the expected number of days in the year on which more than one worker is off sick. [2]

5 A psychology student is investigating memory. In an experiment, volunteers are given 30 seconds to try to memorise a number of items. The items are then removed and the volunteers have to try to name all of them. It has been found that the probability that a volunteer names all of the items is 0.35. The student believes that this probability may be increased if the volunteers listen to the same piece of music while memorising the items and while trying to name them.

The student selects 15 volunteers at random to do the experiment while listening to music. Of these volunteers, 8 name all of the items.

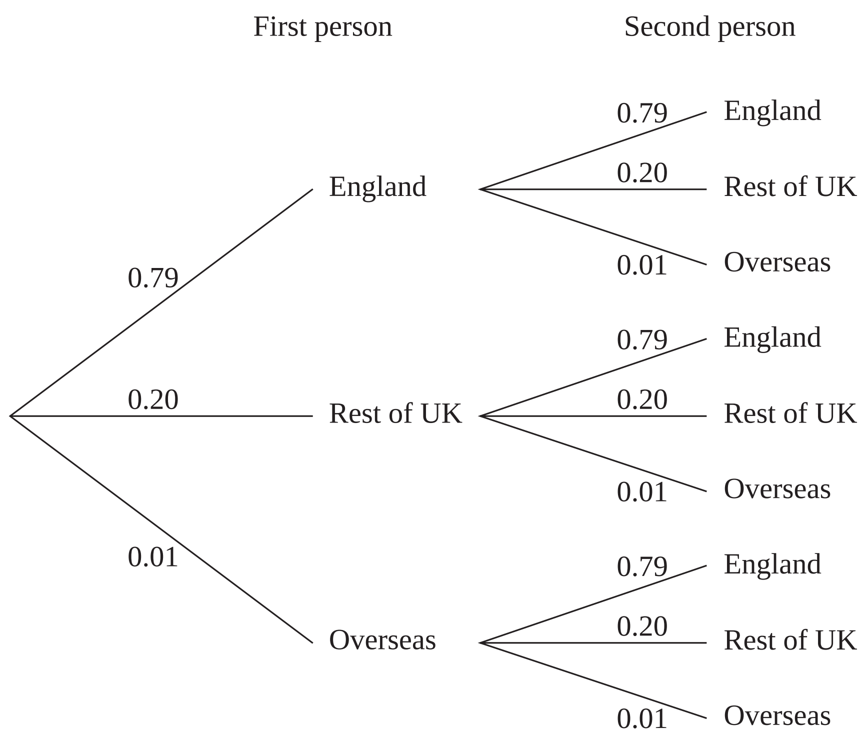
(i) Write down suitable hypotheses for a test to determine whether there is any evidence to support the student's belief, giving a reason for your choice of alternative hypothesis. [4]

(ii) Carry out the test at the 5% significance level. [4]

Section B (36 marks)

6 In a large town, 79% of the population were born in England, 20% in the rest of the UK and the remaining 1% overseas. Two people are selected at random.

You may use the tree diagram below in answering this question.



(i) Find the probability that

(*A*) both of these people were born in the rest of the UK, [2]

(*B*) at least one of these people was born in England, [3]

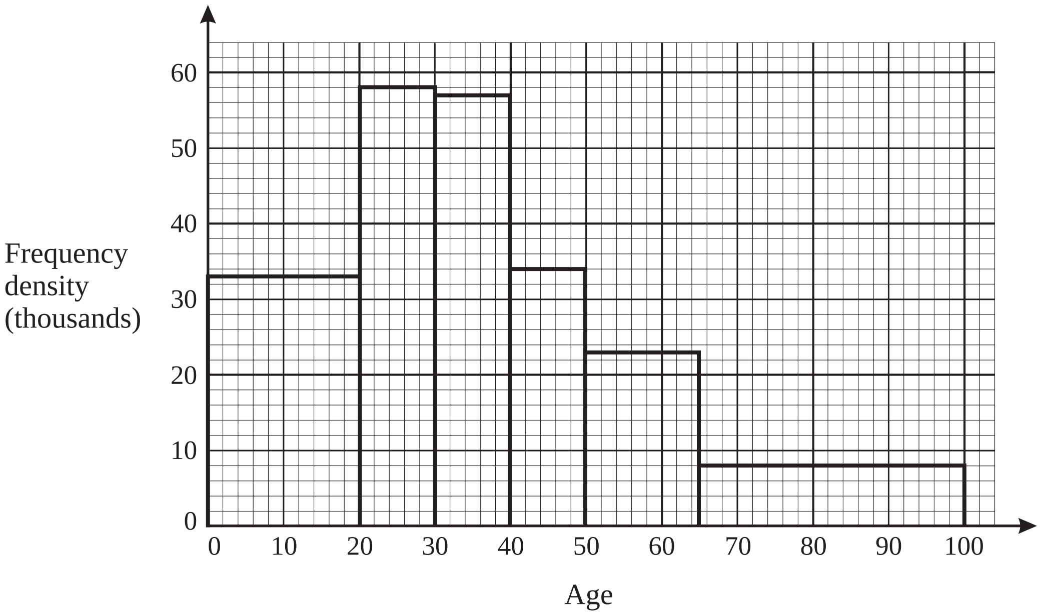
(*C*) neither of these people was born overseas. [2]

(ii) Find the probability that both of these people were born in the rest of the UK given that neither was born overseas. [3]

(iii) (*A*) Five people are selected at random. Find the probability that at least one of them was not born in England. [3]

(*B*) An interviewer selects n people at random. The interviewer wishes to ensure that the probability that at least one of them was not born in England is more than 90%. Find the least possible value of n. You must show working to justify your answer. [3]

7 The histogram shows the age distribution of people living in Inner London in 2001.



Data sourced from the 2001 Census, www.statistics.gov.uk

(i) State the type of skewness shown by the distribution. [1]

(ii) Use the histogram to estimate the number of people aged under 25. [3]

(iii) The table below shows the cumulative frequency distribution.

Age	20	30	40	50	65	100
Cumulative frequency (thousands)	660	1240	1810	*a*	2490	2770

(*A*) Use the histogram to find the value of *a*. [2]

(*B*) Use the table to calculate an estimate of the median age of these people. [3]

The ages of people living in Outer London in 2001 are summarised below.

Age (x years)	$0 \leqslant x < 20$	$20 \leqslant x < 30$	$30 \leqslant x < 40$	$40 \leqslant x < 50$	$50 \leqslant x < 65$	$65 \leqslant x < 100$
Frequency (thousands)	1120	650	770	590	680	610

(iv) Illustrate these data by means of a histogram. [5]

(v) Make two brief comments on the differences between the age distributions of the populations of Inner London and Outer London. [2]

(vi) The data given in the table for Outer London are used to calculate the following estimates.

Mean 38.5, median 35.7, midrange 50, standard deviation 23.7, interquartile range 34.4.

The final group in the table assumes that the maximum age of any resident is 100 years. These estimates are to be recalculated, based on a maximum age of 105, rather than 100. For each of the five estimates, state whether it would increase, decrease or be unchanged. [4]

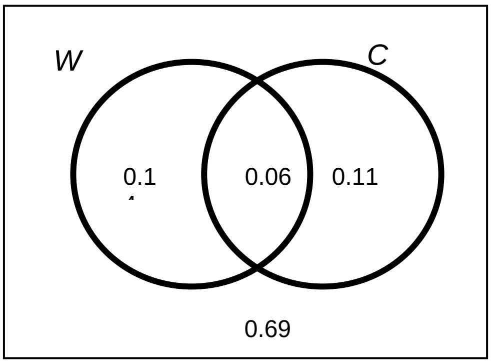
BLANK PAGE

BLANK PAGE

Acknowledgements:

Q.2 & Q.7 Data sourced from the 2001 Census, www.statistics.gov.uk

G241 Statistics 1

1		(i)	Mean = 7.35 (or better) Standard deviation: 3.69 – 3.70 (awfw) Allow s^2 = 13.62 to 13.68 Allow rmsd = 3.64 – 3.66 (awfw) **After B0, B0 scored** then if at least 4 correct mid-points seen or used. {**1.5, 4, 6, 8.5, 15**} Attempt of their mean $=\frac{\sum fx}{\mathbf{44}}$, with $301 \le fx \le 346$ and fx strictly from mid-points not class widths or top/lower boundaries.	B2cao $\sum fx = 323.5$ B2cao $\sum fx^2 = 2964.25$ (B1) for variance s.o.i.o (B1) for rmsd (B1) mid-points (B1) $6.84 \le \text{mean} \le 7.86$	4
		(ii)	Upper limit = 7.35 + 2 × 3.69 = 14.73 or 'their sensible mean' + 2 × 'their sensible s.d.' So there could be one or more outliers	M1 (with s.d. < mean) E1**dep** on B2, B2 earned and comment	2
				TOTAL	**6**
2		(i)	P(*W*) × P(*C*) = 0.20 × 0.17 = 0.034 P(*W*∩*C*) = 0.06 (given in the question) Not equal so not independent (Allow 0.20 × 0.17 ≠ 0.06 or ≠ p (W ∩ C) so not independent).	M1 for multiplying or 0.034 seen A1 (numerical justification needed)	2
		(ii)	 The last two G marks are independent of the labels	G1 for two **overlapping** circles labelled G1 for 0.06 and either 0.14 or 0.11 in the **correct places** G1 for all 4 **correct** probs in the **correct** places **(including the 0.69)** NB No credit for Karnaugh maps here	3
		(iii)	$P(W\|C) = \frac{P(W \cap C)}{P(C)} = \frac{0.06}{0.17} = \frac{6}{17} = 0.353$ (awrt 0.35)	M1 for 0.06 / 0.17 A1 cao	2

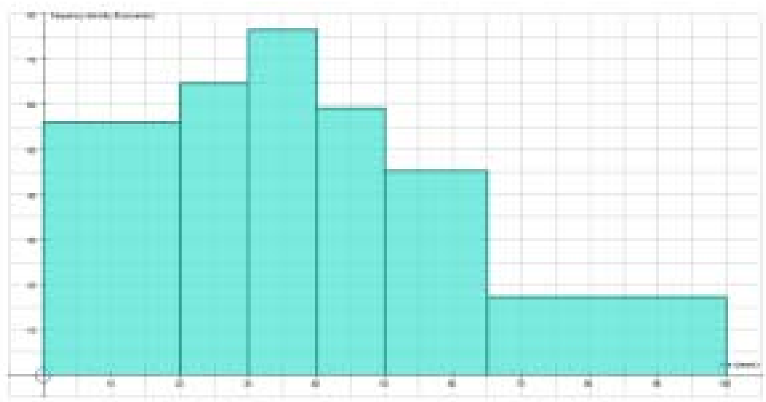
		(iv)	Children are more likely than adults to be able to speak Welsh or 'proportionally more children speak Welsh than adults' Do not accept: **'**more Welsh children speak Welsh than adults'	E1FT Once the correct idea is seen, apply ISW	**1**
				TOTAL	**8**

3		(i)	(A) $0.5 + 0.35 + p + q = 1$ so $p + q = 0.15$ (B) $0 \times 0.5 + 1 \times 0.35 + 2p + 3q = 0.67$ so $2p + 3q = 0.32$ (C) from above $2p + 2q = 0.30$ so $q = 0.02, p = 0.13$	B1 p + q in a correct equation before they reach p + q =0.15 B1 2p + 3q in a correct equation before they reach 2p + 3q = 0.32 (B1) for any 1 correct answer B2 for both correct answers	1 1 2
		(ii)	E(X^2) = $0 \times 0.5 + 1 \times 0.35 + 4 \times 0.13 + 9 \times 0.02 = 1.05$ Var(X) = 'their 1.05' – 0.67^2 = 0.6011 (awrt 0.6) (M1, M1 can be earned with their p^+ and q^+ but not A mark)	M1 $\Sigma x^2 p$ (at least 2 non zero terms correct) M1dep for (– 0.67^2), provided Var(X) > 0 A1 cao (No n or n-1 divisors)	3
				TOTAL	7
4		(i)	$X \sim$ B(8, 0.05) (A) P(X = 0) = $0.95^8 = 0.6634$ 0.663 or better *Or* using tables P(X = 0) = 0.6634 (B) P(X = 1) = $\binom{8}{1} \times 0.05 \times 0.95^7 = 0.2793$ $P(X > 1) = 1 - (0.6634 + 0.2793) = 0.0573$ *Or* using tables P(X > 1) = 1 – 0.9428 = 0.0572	M1 0.95^8 A1 CAO Or B2 (tables) M1 for P(X = 1)(allow 0.28 or better) M1 for 1 – P($X \le 1$) must have both probabilities A1cao (0.0572 – 0.0573) M1 for P($X \le 1$) 0.9428 M1 for 1 – P($X \le 1$) A1 cao (must end in…2)	2 3
		(ii)	Expected number of days = 250 × 0.0572 = 14.3 awrt	M1 for 250 x prob(B) A1 FT but no rounding at end	2
				TOTAL	7
5		(i)	Let p = probability of remembering or naming all items (for population) (whilst listening to music.) H_0: $p = 0.35$ H_1: $p > 0.35$ H_1 has this form since the student believes that the	B1 for definition of p B1 for H_0 B1 for H_1 E1dep on p>0.35 in	

			probability will be increased/ improved/ got better /gone up.	H_0 In words not just because p > 0.35	4
		(ii)	Let $X \sim B(15, 0.35)$ ***Either***: $P(X \geq 8) = 1 - 0.8868 = 0.1132 > 5\%$ Or $0.8868 < 95\%$ So not enough evidence to reject H_0 (Accept H_o) Conclude that there is not enough evidence to indicate that the probability of remembering all of the items is improved / improved/ got better /gone up. (when listening to music.) --- ***Or***: Critical region for the test is {9,10,11,12,13,14,15} 8 does not lie in the critical region. So not enough evidence to reject H_0 Conclude that there is not enough evidence to indicate that the probability of remembering all of the items is improved / improved/ got better /gone up. (when listening to music.) --- ***Or***: The smallest critical region that 8 could fall into is {8, 9, 10, 11, 12, 13, 14, and 15}. The size of this region is 0.1132 $0.1132 > 5\%$ So not enough evidence to reject H_0 Conclude that there is not enough evidence to indicate that the probability of remembering all of the items is improved (when listening to music)	***Either:*** M1 for probability (0.1132) M1**dep** for comparison A1**dep** E1**dep on all previous marks** for conclusion in **context** --- ***Or***: M1 for correct CR(no omissions or additions) M1**dep** for 8 does not lie in CR A1**dep** E1**dep on all previous marks** for conclusion in **context** --- **Or**: M1 for CR{8,9,…15}**and** size = 0.1132 M1 **dep** for comparison A1**dep** E1**dep on all previous marks** for conclusion in **context**	4
				TOTAL	**8**

			Section B		
6		(i)	(*A*) P(both rest of UK) = 0.20×0.20 = 0.04	M1 for multiplying A1cao	2
			(*B*) *Either:* **All 5 case** P(at least one England) = (0.79 x 0.20) + (0.79 x 0.01) + (0.20 x 0.79) + (0.01 x 0.79) + (0.79 x 0.79) = 0.158 + 0.0079 + 0.158 + 0.0079 + 0.6241 = 0.9559 *Or* P(at least one England) = 1 – P(neither England) = 1 – (0.21×0.21) = 1 – 0.0441 = 0.9559 or listing all = 1 –{ (0.2 x 0.2) + (0.2 x 0.01) + (0.01 x 0.20) + (0.01x 0.01)} = 1 – (**) = 1 – { 0.04 + 0.002 + 0.002 + 0.0001) = 1 – 0.0441 = 0.9559 Or: **All 3 case** P(at least one England) = = $0.79 \times 0.21 + 0.21 \times 0.79 + 0.79^2$ = 0.1659 + 0.1659 + 0.6241 = 0.9559	M1 for any correct term (3case or 5case) M1 for correct sum of all 3 (or of all 5) with **no extras** A1cao (condone 0.96 www) *Or* M1 for 0.21×0.21 or for (**) fully enumerated or 0.0441 seen M1**dep** for 1 – (1st part) A1cao See above for 3 case	3
			(*C*)*Either* 0.79 x 0.79 + 0.79 x 0.2 + 0.2 x 0.79 + 0.2 x 0.2 = 0.9801 *Or* 0.99×0.99 = 0.9801 Or 1 – { 0.79 x 0.01 + 0.2 x 0.01 + 0.01 x 0.79 + 0.01 x 0.02 + 0.01^2 } = 1 – 0.0199 = 0.9801	M1 for sight of all 4 correct terms summed A1 cao (condone 0.98 www) *or* M1 for 0.99 x 0.99 A1cao ***Or*** M1 for everything 1 – {…..} A1cao	2

		(ii)	P(both the rest of the UK | neither overseas) $= \dfrac{\text{P(the rest of the UK } \textit{and} \text{ neither overseas)}}{\text{P(neither overseas)}}$ $= \dfrac{0.04}{0.9801} = 0.0408$ {Watch for: $\dfrac{answer(A)}{answer(C)}$ as evidence of method (p <1)}	M1 for numerator of 0.04 or 'their answer to (i)(A)' M1 for denominator of 0.9801 or 'their answer to (i) (C)' A1 FT (0 < p < 1) 0.041 at least	**3**
		(iii)	(A) Probability $= 1 - 0.79^5$ $= 1 - 0.3077$ $= 0.6923$ (accept awrt 0.69) see additional notes for alternative solution (B) $1 - 0.79^n > 0.9$ **EITHER:** $1 - 0.79^n > 0.9$ or $0.79^n < 0.1$ (condone = and ≥ throughout) but not reverse inequality $n > \dfrac{\log \mathbf{0.1}}{\log \mathbf{0.79}}$, so n > 9.768… Minimum n = 10 Accept n ≥ 10 --- **OR** (using trial and improvement): Trial with 0.79^9 or 0.79^{10} $1 - 0.79^9 = 0.8801$ (< 0.9) or $0.79^9 = 0.1198$ (> 0.1) $1 - 0.79^{10} = 0.9053$ (> 0.9) or $0.79^{10} = 0.09468$ (< 0.1) Minimum n = 10 Accept n ≥ 10 --- NOTE: n = 10 unsupported scores SC1 only	M1 for 0.79^5 or 0.3077... M1 for $1 - 0.79^5$ dep A1 CAO M1 for equation/inequality in n (accept either statement opposite) M1(indep) for process of using logs i.e. $\dfrac{\log a}{\log b}$ A1 CAO ---------------------------- M1(indep) for sight of 0.8801 **or** 0.1198 M1(indep) for sight of 0.9053 **or** 0.09468 A1 dep **on both M's** cao ------------------------------	 **3** **3**
				TOTAL	**16**

7		(i)	Positive	B1	1
		(ii)	Number of people = 20 × 33 (000) + 5 × 58 (000) = 660 (000) + 290 (000) = 950 000	M1 first term M1(indep) second term A1 cao NB answer of 950 scores M2A0	3
		(iii)	(A) a = 1810 + 340 = 2150	M1 for sum A1 cao 2150 or 2150 thousand but not 215000	2
			(B) Median = age of 1 385 (000[th]) person or 1385.5 (000) Age 30, cf = 1 240 (000); age 40, cf = 1 810 (000) Estimate median = (30) + $\frac{\mathbf{145}}{\mathbf{570}} \times \mathbf{10}$ Median = 32.5 years (32.54...) If no working shown then 32.54 or better is needed to gain the M1A1. If 32.5 seen with **no previous working** allow SC1	B1 for 1 385 (000) or 1385.5 M1 for attempt to interpolate $\frac{\mathbf{145}k}{\mathbf{570}k} \times 10$ (2.54 or better suggests this) A1 cao min 1dp	3
		(iv)	Frequency densities: 56, 65, 77, 59, 45, 17 *(accept 45.33 and 17.43 for 45 and 17)* 	B1 for any one correct B1 for all correct (soi by listing or from histogram) **Note: all G marks below *dep* on attempt at frequency density, NOT frequency** G1 Linear scales on both axes (no inequalities) G1 Heights FT their listed fds or all must be correct. **Also widths. All blocks joined** G1 Appropriate label for vertical	5

				scale eg 'Frequency density (thousands)', 'frequency (thousands) per 10 years', 'thousands of people per 10 years'. (allow key). **OR f.d.**	
		(v)	Any two suitable comments such as: Outer London has a greater proportion (or %) of people under 20 (or almost equal proportion) The modal group in Inner London is 20-30 but in Outer London it is 30-40 Outer London has a greater proportion (14%) of aged 65+ **<u>All</u>** populations in **<u>each</u>** age group are higher in Outer London Outer London has a more evenly spread distribution or balanced distribution (ages) o.e.	E1 E1	**2**
		(vi)	Mean increase ↑ median unchanged (-) midrange increase ↑ standard deviation increase ↑ interquartile range unchanged. (-)	Any one correct B1 Any two correct B2 Any three correct B3 All **five** correct B4	**4**
				TOTAL	**20**

G241 Statistics 1

General Comments

The standard this summer was variable. There were some excellent scripts seen by the examiners reflecting the hard work and dedication of teachers, lecturers and candidates. On the other hand there were a substantial number of candidates who seemed totally out of their depth who struggled to make any real progress.

Candidates should be reminded to work with total accuracy and not to round their answers severely as they progress through a calculation.

It was pleasing to see that a number of centres had acted on comments made in previous reports particularly with regard to the definition of p in the construction of hypotheses.

Comments on Individual Questions

1) The calculation of an estimate of the mean and standard deviation of grouped data presented unexpected problems for a sizeable number of candidates. Often 2 or more mid-points of the classes were incorrect thus throwing out any possibility of achieving the accuracy required. A common error even by the better candidates was to use mid-points of 1.5, 4.5, 6.5, 8.5 and 15. Some candidates had little idea how to obtain the mid-points and thought that the mean could be somehow calculated from multiplying the frequencies by the class widths or the frequencies by one of the boundary values. It was disturbing to see many candidates attempting to work out the standard deviation without using any frequencies. This is clearly a topic which deserves more attention to precision and process for the future.

   The concept of finding the upper boundary for any outliers was well known in terms of mean + 2 standard deviations but several tried to argue the case with Q_3 + 1.5 IQR (not that these data were available) or insisted using mean + 1.5 standard deviations. Candidates should be careful not to make rash statements such as 'there **are** outliers in the data' but instead be more circumspect and claim that 'there could be or may be some outliers in the final class'.

2) The work on testing for independent events was pleasing with a variety of methods used by candidates. Most went down the route of showing numerically that $P(W) \times P(C) \neq P(W \cap C)$ and hence the events were not independent. Some tried their luck with non numerical or qualitative attempts but to little avail.

   The Venn diagram was, unfortunately, often lacking in credibility. There are still too many candidates filling in the various regions with the incorrect probabilities. The region $W \cap C'$ was often given as 0.2 instead of the correct 0.14 and likewise the other region $C \cap W'$ was written as 0.17 instead of the correct 0.11. The region $W \cap C$ was invariably correct as 0.06. A curious number of candidates often labelled the region $W' \cap C'$ as 0.63 instead of the correct 0.69. Again, this is an area that deserves the attention of candidates for future examinations.

   The calculation of P (W/C) was well attempted and most scored 2 marks. The conclusion was usually sound but many did not choose their words carefully and quoted '**more** children speak Welsh' when really they meant 'the proportion of children speaking Welsh is higher.'

3) (i)(*A*) Many candidates had difficulty composing an equation which included p + q and a summation to 1.

(*B*) A little better, with some realising that the equation for E(X) must now include 2p + 3q.

(*C*) The solution of the resulting simultaneous equations seemed to be off the mathematical radar for many candidates with many struggling to find solutions for p and q.

(ii) The variance was usually calculated correctly bearing in mind that a generous follow through was applied for those candidates who did not find the exact values of p and q earlier. The only common error was the omission of 0.67^2 leaving an answer of 1.07.

4) This was a popular question which was well answered by many candidates. In (i) part (A) most gained the correct answer of 0.6634 but then did themselves no favours by unnecessarily rounding the answer to 0.66. Part (B) was well answered but there was some confusion about the meaning of P(X>1). Some believed it to be 1 – P(X=0) rather than the correct form of 1 – {P(X=0) + P(X=1)}. In the last part, most knew the E(X) = np formula and gained the marks, even on follow through.

5) Candidates need to be reminded that a hypothesis test on the binomial distribution requires an initial set up of the following conditions.

- The definition of the parameter p, in context
- The use of the correct notation for H_0 and H_1, namely in the case of this question that H_0: p = 0.35 and H_1: p > 0.35
- A clear explanation, in context, of why H_1 takes the form that it does.

Unfortunately, many omit the requirements of the first and last bullet points, thus losing 2 valuable marks. It is worth reminding centres again that sloppy or poor notation such as H_0: P(x = 0.35) and H_1: P(x > 0.35) is penalised by the examiners. Too many candidates are prone to this form of notation.
Many otherwise worthy initial set ups were spoilt by candidates using point probabilities or selecting the wrong tail. It was not uncommon to see P(X $\geq$8) = 0.0422 when, in fact this was P(X $\geq$9). The correct solution required P(X $\geq$8) = 1 – P(X$\leq$7) = 1 - 0.8868 = 0.1132. Some candidates wrote ridiculous statements along the lines of 0.9578 > 5%. It must be emphasised, once again, that the tail probability must be compared with the significance level of the test. All further marks in the question are dependent on this important fact. The next stage is to accept or reject H_0 and then reach a valid conclusion in context.

6) There were many successful attempts to the first half of this question. Candidates were able to demonstrate a good understanding of probability calculations using their tree diagrams.

Part (A) was invariably correct as 0.04. Most were able to achieve 0.9559 in part (B) by adding the 5 separate probabilities but very few candidates realised the quick way to achieve the answer by $1 – 0.21^2$ =0.9559. A common error in part (B) was the omission of the 0.79^2 term giving 0.3318 as an answer.

In part (C) most candidates preferred to list and add the 4 probability terms to gain 0.9801. Relatively few spotted the quick way of 0.99^2 would reach the same answer. Some candidates made the error of believing that neither of the people was born overseas could be calculated from $1 – 0.01^2 = 0.9999$. The conditional probability in part (ii) elicited some very good responses with most realising the correct method although some did write (0.04 x 0.9801)/0.9801with depressing regularity.

Only the better candidates made any progress in part (iii) with many finding $1 – 0.79^5$. Some candidates had become muddled by this stage and it was not uncommon to see $1 – 0.21^5$ or even $1 – 0.9559^5$. The latter two methods did, however, attract a partial award. Part (iii) (B) was often well attempted by the better candidates with equally as many opting for using logarithms as for using a trial and improvement method.

7) Part (i) was almost invariably correct with the response of positive skewness.

Part (ii) was well tackled with many achieving the answer of 950 000 but some candidates left their answer as 950 and lost a mark.

Many reached the required cumulative frequency of 2150 (thousands) via 1810 + 340 but there were instances of 1810 + 345 seen by the examiners. Almost all candidates were able to locate the position of the median as the 1385 or 1385½ value. Only the very talented candidates were then able to carry out the linear interpolation of
$30 + \frac{145}{570} \times 10 = 32.54$, to achieve the median age.

It was pleasing to see many successful attempts at finding the frequency densities in part (iv). Without doubt, the frequency divided by class width was the most popular method but other strange but nevertheless correct methods were seen. The resulting histogram was well drawn but some candidates did make life difficult for themselves by choosing a bizarre scaling (e.g. 3cm = 10 units on the vertical axis).

The comments in part (iv) were often not what the examiners were looking for. Many opted to compare numbers across the two histograms but it should have been evident that **all** the populations for **each** age group were higher in Outer London than Inner London. Some candidates did pick up on the salient points of the two histograms by comparing the different modal classes (20 – 30 for Inner London; 30 -40 for Outer London). In making comparisons it is advisable that candidates mention proportions rather than refer to ‘more than’ or ‘less than’ statements.

Part (vi) elicited some positive responses with many realising that the mean, midrange and standard deviation would all increase in the light of the new information. Some thought the standard deviation would decrease rather than increase but most knew the interquartile range would be unchanged.