

Wednesday 25 May 2016 – Morning

AS GCE MATHEMATICS (MEI)

4766/01 Statistics 1

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:

- Printed Answer Book 4766/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:

- Scientific or graphical calculator

Duration: 1 hour 30 minutes



INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found inside the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book.** If additional space is required, you should use the lined page(s) at the end of this booklet. The question number(s) must be clearly shown.
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets [] at the end of each question or part question on the Question Paper.
- 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.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **8** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER/INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

Section A (36 marks)

- 1 The stem and leaf diagram illustrates the weights in grams of 20 house sparrows.

25		0		
26		0	5	8
27		7	9	
28		1	4	5
29		0	0	2
30		7	7	
31		6		
32		0	4	7
33		3	3	

Key: 27 | 7 represents 27.7 grams

- (i) Find the median and interquartile range of the data. [3]
- (ii) Determine whether there are any outliers. [4]
- 2 In a hockey league, each team plays every other team 3 times. The probabilities that Team A wins, draws and loses to Team B are given below.

- $P(\text{Wins}) = 0.5$
- $P(\text{Draws}) = 0.3$
- $P(\text{Loses}) = 0.2$

The outcomes of the 3 matches are independent.

- (i) Find the probability that Team A does not lose in any of the 3 matches. [1]
- (ii) Find the probability that Team A either wins all 3 matches or draws all 3 matches or loses all 3 matches. [2]
- (iii) Find the probability that, in the 3 matches, exactly two of the outcomes, 'Wins', 'Draws' and 'Loses' occur for Team A. [4]
- 3 (i) There are 5 runners in a race. How many different finishing orders are possible? [You should assume that there are no 'dead heats', where two runners are given the same position.] [1]
- For the remainder of this question you should assume that all finishing orders are equally likely.
- (ii) The runners are denoted by A, B, C, D, E. Find the probability that they either finish in the order ABCDE or in the order EDCBA. [2]
- (iii) Find the probability that the first 3 runners to finish are A, B and C, in that order. [1]
- (iv) Find the probability that the first 3 runners to finish are A, B and C, in any order. [2]

- 4 The probability distribution of the random variable X is given by the formula

$$P(X=r) = \frac{k}{r(r-1)} \text{ for } r = 2, 3, 4, 5, 6.$$

(i) Show that the value of k is 1.2. Using this value of k , show the probability distribution of X in a table. [3]

(ii) Find $E(X)$ and $\text{Var}(X)$. [5]

- 5 Measurements of sunshine and rainfall are made each day at a particular weather station. For a randomly chosen day,

- R is the event that at least 1 mm of rainfall is recorded,
- S is the event that at least 1 hour of sunshine is recorded.

You are given that $P(R) = 0.28$, $P(S) = 0.87$ and $P(R \cup S) = 0.94$.

(i) Find $P(R \cap S)$. [2]

(ii) Draw a Venn diagram showing the events R and S , and fill in the probability corresponding to each of the four regions of your diagram. [3]

(iii) Find $P(R | S)$ and state what this probability represents in this context. [3]

Section B (36 marks)

- 6 An online store has a total of 930 different types of women's running shoe on sale. The prices in pounds of the types of women's running shoe are summarised in the table below.

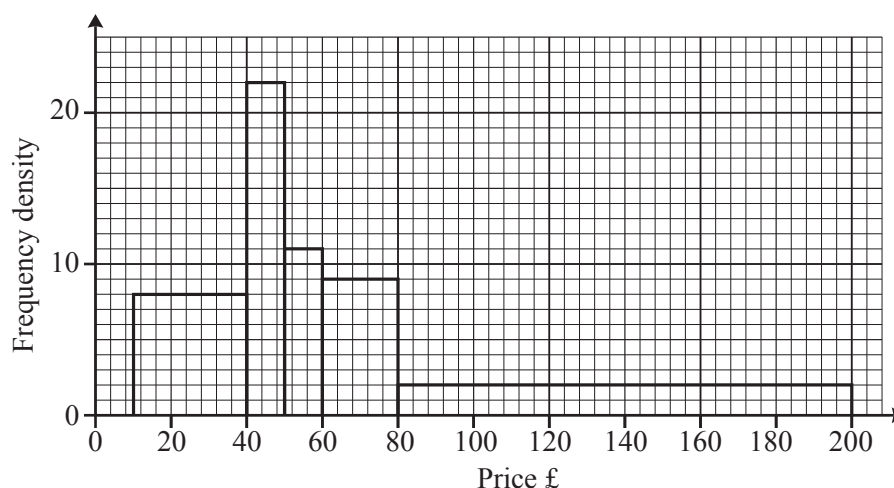
Price (£ x)	$10 \leq x \leq 40$	$40 < x \leq 50$	$50 < x \leq 60$	$60 < x \leq 80$	$80 < x \leq 200$
Frequency	147	109	182	317	175

(i) Calculate estimates of the mean and standard deviation of the shoe prices. [4]

(ii) Calculate an estimate of the percentage of types of shoe that cost at least £100. [3]

(iii) Draw a histogram to illustrate the data. [5]

The corresponding histogram below shows the prices in pounds of the 990 types of men's running shoe on sale at the same online store.



(iv) State the type of skewness shown by the histogram for men's running shoes. [1]

(v) Martin is investigating the percentage of types of shoe on sale at the store that cost more than £100. He believes that this percentage is greater for men's shoes than for women's shoes. Estimate the percentage for men's shoes and comment on whether you can be certain which percentage is higher. [3]

(vi) You are given that the mean and standard deviation of the prices of men's running shoes are £68.83 and £42.93 respectively. Compare the central tendency and variation of the prices of men's and women's running shoes at the store. [2]

7 To withdraw money from a cash machine, the user has to enter a 4-digit PIN (personal identification number). There are several thousand possible 4-digit PINs, but a survey found that 10% of cash machine users use the PIN '1234'.

(i) 16 cash machine users are selected at random.

(A) Find the probability that exactly 3 of them use 1234 as their PIN. [3]

(B) Find the probability that at least 3 of them use 1234 as their PIN. [2]

(C) Find the expected number of them who use 1234 as their PIN. [1]

An advertising campaign aims to reduce the number of people who use 1234 as their PIN. A hypothesis test is to be carried out to investigate whether the advertising campaign has been successful.

(ii) Write down suitable null and alternative hypotheses for the test. Give a reason for your choice of alternative hypothesis. [4]

(iii) A random sample of 20 cash machine users is selected.

(A) Explain why the test could not be carried out at the 10% significance level. [3]

(B) The test is to be carried out at the $k\%$ significance level. State the lowest integer value of k for which the test could result in the rejection of the null hypothesis. [1]

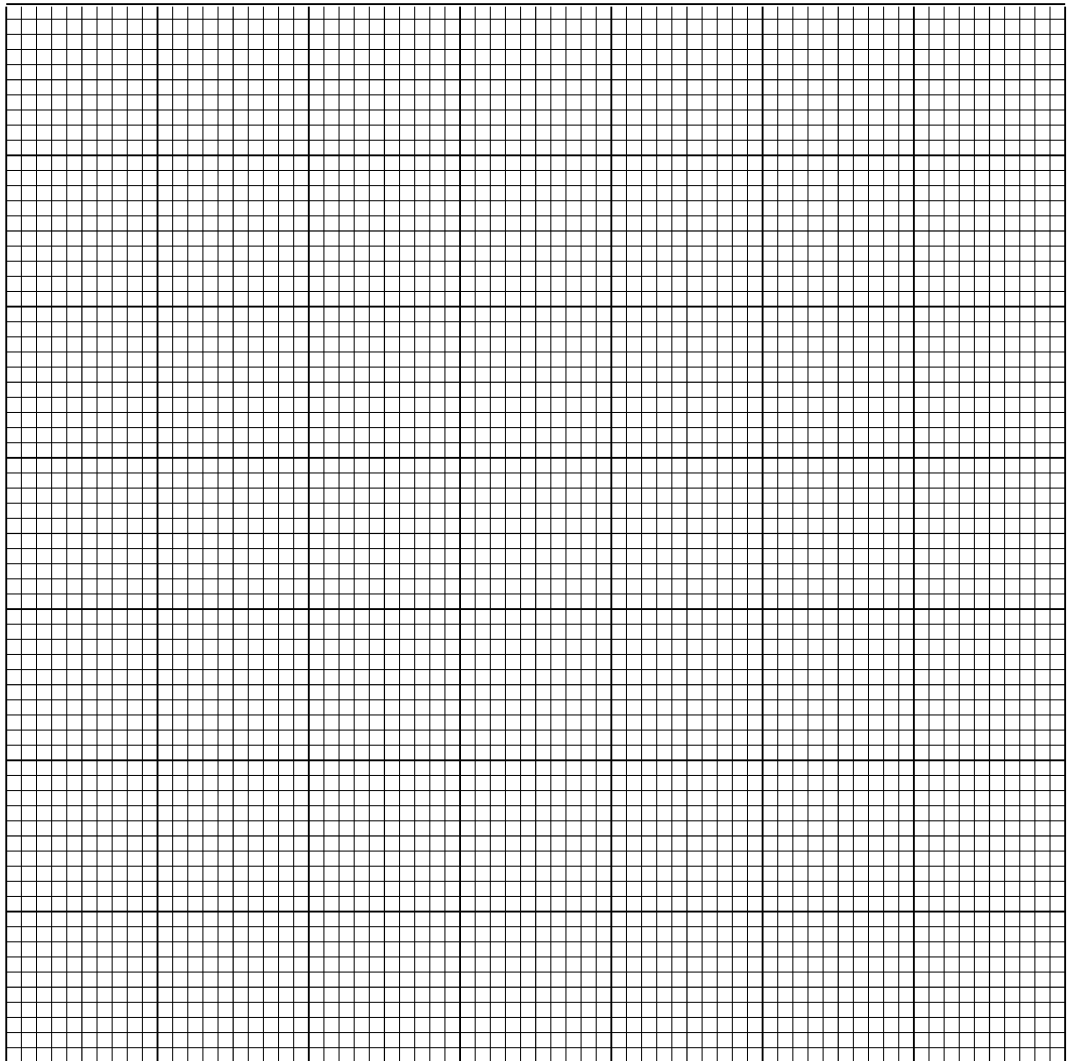
(iv) A new random sample of 60 cash machine users is selected. It is found that 2 of them use 1234 as their PIN. You are given that, if $X \sim B(60, 0.1)$, then (to 4 decimal places)

$$P(X = 2) = 0.0393, \quad P(X < 2) = 0.0138, \quad P(X \leq 2) = 0.0530.$$

Using the same hypotheses as in part (ii), carry out the test at the 5% significance level. [4]

END OF QUESTION PAPER

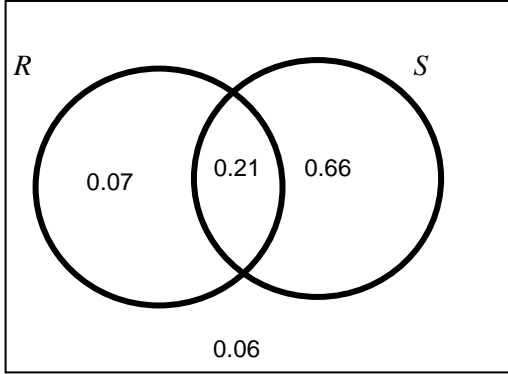
6(iii)



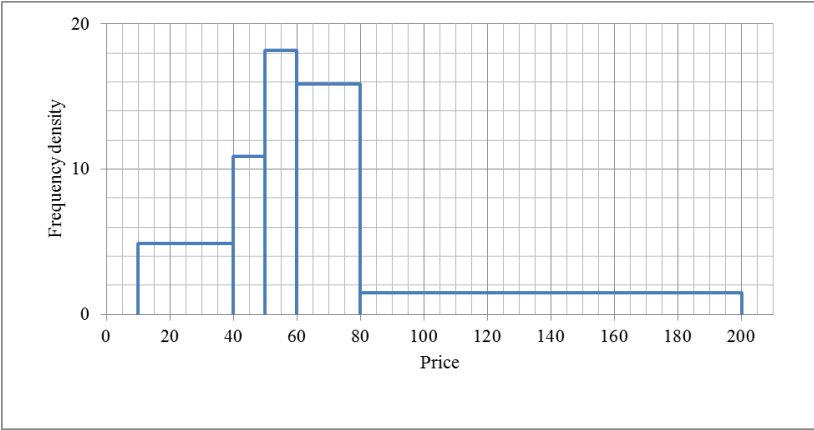
Question			Answer	Marks	Guidance	
1	(i)		Median = 29.0 IQR = 31.8 – 27.8 = 4.0	B1 M1 A1 [3]	For either quartile – allow alternative definitions of quartiles	Condone wrong method Allow 27.75 and 31.9 leading to 4.15 Do not allow 27.7, 27.9, 31.6, 32.0
1	(ii)		Lower limit = 27.8 - 1.5 × 4.0 = 21.8 27.75, 31.9 lead to 21.525 and 38.125 27.7, 31.6 lead to 21.85 and 37.45 Upper limit = 31.8 + 1.5 × 4.0 = 37.8 So there are no outliers (at either end of the distribution)	M1 A1 A1 B1 [4]	Method for either FT sensible quartiles and IQR FT sensible quartiles and IQR Dep on at least one A1 Use of median scores 0/4	For use of mean (29.44) and SD (2.516765...) 29.44 ± 2×2.516765 M1 Lower Limit = 24.4 A1 Upper limit = 34.5 A1 So no outliers B1
2	(i)		P(Does not lose any match) = $0.8^3 = 0.512 = \frac{64}{125}$	B1 [1]		
2	(ii)		P(Wins all 3 or draws all 3 or loses all 3) = $0.5^3 + 0.3^3 + 0.2^3$ = 0.16 = $\frac{4}{25}$	M1 A1 [2]	Including addition	
2	(iii)		P(all three outcomes occur) = $3! \times 0.5 \times 0.3 \times 0.2$ = 0.18 Required probability = 1 – ‘0.18’ – 0.16 = 0.66 = $\frac{33}{50}$	M1* A1 *M1 dep A1 [4]	CAO	Allow M1 for $k \times 0.5 \times 0.3 \times 0.2$ even if k = 1 Even if cubed Not if cubed
			OR:			
			P(WWW') + P(DDD') + P(LL'L')	M1	For any one product (no need for '3 ×')	Even if cubed
			$3 \times 0.5^2 \times 0.5 + 3 \times 0.3^2 \times 0.7 + 3 \times 0.2^2 \times 0.8$	M1	For '3 ×'	Dep on at least 1 correct term
			0.375 + 0.189 + 0.096	M1	For sum of three correct terms (no need for '3 ×') And no incorrect terms	NB common wrong answer of 0.22 from omitting '3 ×' or 0.44 from '2×' scores M1M0M1A0 Not if cubed
			0.66	A1	CAO	

Question			Answer	Marks	Guidance	
			OR:			
			$P(WWD) + P(WWL) + P(DDW) + P(DDL) + P(LLW) + P(LLD)$	M1	For any one product (no need for '3 ×')	Even if cubed
			$3 \times 0.5^2 \times 0.3 + 3 \times 0.5^2 \times 0.2 + 3 \times 0.3^2 \times 0.5 + 3 \times 0.3^2 \times 0.2 + 3 \times 0.2^2 \times 0.5 + 3 \times 0.2^2 \times 0.3$	M1	For '3 ×'	Dep on at least 1 correct term
			$0.225 + 0.15 + 0.135 + 0.054 + 0.06 + 0.036$	M1	For sum of six correct terms (no need for '3 ×') And no incorrect terms	Not if cubed
			0.66	A1	CAO	
3	(i)		Number of ways = $5! = 120$	B1 [1]		
3	(ii)		Probability = $2/120$ $= 1/60$ or 0.0167 or 0.016	M1 A1 [2]	For division by their 120 CAO	M1 for $\frac{k}{120}$ Condone final answer of $2/120$ Do not allow 0.016
3	(iii)		$\frac{1}{5} \times \frac{1}{4} \times \frac{1}{3} = \frac{1}{60}$ or 0.0167 or 0.016	B1 [1]	Condone $2/120$ for B1	Do not allow 0.016
3	(iv)		$\frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} = \frac{1}{10}$ or $1/5 C_3 = 1/10$	M1 A1 M1 A1 [2]	For $3/5 \times$ CAO For division by 5C_3	Listing options gives $12/120$ Or $(3! / 120) \times 2$ Or $({}^3P_3 \times {}^2P_2) / 120$ SC2 for $3! \times$ their part (iii) or $6 \times$ their part (iii)

Question		Answer	Marks	Guidance													
4	(i)	$k/2 + k/6 + k/12 + k/20 + k/30 = 1$ $(30 + 10 + 5 + 3 + 2)k/60 = 1$ $50k = 60$ $k = 1.2$ <table border="1"> <tr> <td>r</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr> <tr> <td>$P(X = r)$</td><td>0.6 $= \frac{3}{5}$</td><td>0.2 $= \frac{1}{5}$</td><td>0.1 $= \frac{1}{10}$</td><td>0.06 $= \frac{3}{50}$</td><td>0.04 $= \frac{1}{25}$</td></tr> </table>	r	2	3	4	5	6	$P(X = r)$	0.6 $= \frac{3}{5}$	0.2 $= \frac{1}{5}$	0.1 $= \frac{1}{10}$	0.06 $= \frac{3}{50}$	0.04 $= \frac{1}{25}$	M1 A1 B1 [3]	For correct equation including = 1 Need one further intermediate step after equation NB Answer Given Complete correct table in fraction or decimal form NOT in terms of k	Allow substitution of $k = 1.2$ to show probabilities add to 1 with convincing working which must be more than just $1.2/2 + 1.2/6 + 1.2/12 + 1.2/20 + 1.2/30 = 1$ This latter gets M1A0 Must tabulate probabilities, though may be seen in part(ii) If fractions any denominator is ok provided numerators are integers
r	2	3	4	5	6												
$P(X = r)$	0.6 $= \frac{3}{5}$	0.2 $= \frac{1}{5}$	0.1 $= \frac{1}{10}$	0.06 $= \frac{3}{50}$	0.04 $= \frac{1}{25}$												
4	(ii)	$E(X) = (2 \times 0.6) + (3 \times 0.2) + (4 \times 0.1) + (5 \times 0.06) + (6 \times 0.04)$ $E(X) = 2.74$ or $\frac{137}{50}$	M1 A1	For $\sum rp$ (at least 3 terms correct) Provided 5 reasonable probabilities seen. CAO If probs wrong but sum = 1 allow max M1A0M1M1A1. If sum \neq 1 allow max M1A0M1M1A0 (provided all probabilities ≥ 0 and < 1)	Use of $E(X - \mu)^2$ gets M1 for attempt at $(x - \mu)^2$ should see $(-0.74)^2, 0.26^2, 1.26^2, 2.26^2, 3.26^2$, (if $E(X)$ wrong FT their $E(X)$) (all 5 correct for M1), then M1 for $\sum p(x - \mu)^2$ (at least 3 terms correct with their probabilities)												
		$E(X^2) = (4 \times 0.6) + (9 \times 0.2) + (16 \times 0.1) + (25 \times 0.06) + (36 \times 0.04)$ $= 8.74$ or $\frac{437}{50}$ $\text{Var}(X) = 8.74 - 2.74^2 = 1.23$ or 1.232 or $\frac{3081}{2500}$	M1* M1* dep A1 [5]	For $\sum r^2 p$ (at least 3 terms correct) for – their $E(X)^2$ FT their $E(X)$ provided $\text{Var}(X) > 0$	Division by 5 or other spurious value at end and/or rooting final answer gives max M1A1M1M1A0, or M1A0M1M1A0 if $E(X)$ also divided by 5. Unsupported correct answers get 5 marks (Probably from calculator) Condone 1.2324 despite the fact that this is over-specified since it is the exact answer												

Question			Answer	Marks	Guidance	
5		(i)	$P(R \cap S) = P(R) + P(S) - P(R \cup S)$ $= 0.28 + 0.87 - 0.94$ $= 0.21$	M1 A1 [2]	For correct use of formula	Or $0.28 - x + 0.87 - x + x = 0.94$
5		(ii)		G1 G1 G1 [3]	For two labelled intersecting circles For at least 2 correct probabilities. FT their $P(R \cap S)$ For remaining probabilities. FT their $P(R \cap S)$	Allow labels such as $P(R)$ and $P(S)$ Allow other sensible shapes in place of circles Allow their $P(R \cap S)$ rounded to 2dp For both G1 marks FT their 0.21 provided < 0.28 For FT if $P(R \cap S) = x$ then others are $0.28 - x$, $0.87 - x$, $x - 0.15$ 0.2436 leads to 0.0364 , 0.6264 , 0.0936
5		(iii)	$P(R S) = \frac{P(R \cap S)}{P(S)} = \frac{0.21}{0.87} = \frac{21}{87} = 0.241$ <p>Exact answer 0.241379...</p> <p>This is the probability that (on a randomly selected day) there is at least 1 mm of rain, given that there is at least 1 hour of sun.</p>	M1 A1 E1 [3]	for fraction CAO FT their part (i) (for M1 only) but M0 if their answer to part (i) is $P(R) \times P(S)$ Need more than just probability of rain given sun Must include 'probability' or 'chance' oe Do not allow just $P(\text{at least 1 mm of rain, given that there is at least 1 hour of sun})$	Allow $\frac{7}{29}$ or $\frac{21}{87}$ as final answer Allow 0.24 with working Condone 'if' or 'when' for 'given that' but not the words 'and' or 'because' or 'due to' for E1. E1 (independent of M1): the order/structure must be correct i.e. no reverse statement Allow 'The probability that on a randomly selected day when there is at least 1 hour of sun there is at least 1 mm of rain.' oe

Question		Answer	Marks	Guidance	
6	(i)	<p>Mean =</p> $\frac{(25 \times 147) + (45 \times 109) + (55 \times 182) + (70 \times 317) + (140 \times 175)}{930}$ $\frac{750 \times 7 + 1250 \times 22 + 1750 \times 26 + 2500 \times 18 + 4000 \times 7}{80} = \frac{151250}{80} = (£)70.19 \text{ or } (£)70.2$ <p>$\Sigma x^2 f =$</p> $(25^2 \times 147) + (45^2 \times 109) + (55^2 \times 182) + (70^2 \times 317) + (140^2 \times 175)$ $= 91875 + 220725 + 550550 + 1553300 + 3430000$ $= 5846450$ $S_{xx} = 5846450 - \frac{65280^2}{930} = 1264215.161 \text{ or } 5846450 - 930 \times 70.19^2$ $s = \sqrt{\frac{1264215}{929}} = \sqrt{1360.83} = 36.89 \text{ or } (£)36.9$ <p>Allow any answer between 36.87 and 36.90 without checking working</p>	<p>M1</p> <p>A1</p> <p>M1</p> <p>A1</p> <p>[4]</p>	<p>For midpoints (at least 3 correct) (allow 25.005, 45.005 etc leading to answer 70.20)</p> <p>CAO (exact answer 70.19355...) Correct answers obtained from use of calculator statistical functions gain full marks Condone answer of (£)70.20</p> <p>For attempt at S_{xx} Should include sum of at least 3 correct multiples $fx^2 - \Sigma x^2/n$ Do not FT their incorrect mean for A1 (exact answer 36.88949...) Condone answer of (£)36.90 If both mean and sd over-specified, just deduct one mark</p>	<p>M0A0M0A0 unless using midpoints Answer must NOT be left as improper fraction as this is an estimate Accept correct answers for mean and sd from calculator even if eg wrong S_{xx} given</p> <p>For use of midpoints 25.5, 45.5, 55.5, 70.5, 140.5 allow SC1 for £70.69 and SC1 for 36.89</p> <p>If using $(x - \bar{x})^2$ method, B2 if 36.9 or better, otherwise B0 Allow use of 70.2 in calculation of $S_{xx} = 1263372.8$ leading to 36.87719... Condone RMSD of 36.87 (36.86985...) since n is so large</p>
6	(ii)	<p>$100/120 \times 175 = 145.83$ $145.83/930 = 0.1568$ So 15.7%</p>	<p>M1*</p> <p>*M1dep</p> <p>A1</p> <p>[3]</p>	<p>For 175/120</p>	<p>Or $20/120 \times 175 = 29.166$ oe $(175 - 29.166)/930$ Accept 16% with working</p>

Question			Answer	Marks	Guidance																									
6	(iii)		<table><thead><tr><th>Price</th><th>Frequency</th><th>Group width</th><th>Frequency density</th></tr></thead><tbody><tr><td>$10 \leq x \leq 40$</td><td>147</td><td>30</td><td>4.90</td></tr><tr><td>$40 < x \leq 50$</td><td>109</td><td>10</td><td>10.90</td></tr><tr><td>$50 < x \leq 60$</td><td>182</td><td>10</td><td>18.20</td></tr><tr><td>$60 < x \leq 80$</td><td>317</td><td>20</td><td>15.85</td></tr><tr><td>$80 < x \leq 200$</td><td>175</td><td>120</td><td>1.46</td></tr></tbody></table>	Price	Frequency	Group width	Frequency density	$10 \leq x \leq 40$	147	30	4.90	$40 < x \leq 50$	109	10	10.90	$50 < x \leq 60$	182	10	18.20	$60 < x \leq 80$	317	20	15.85	$80 < x \leq 200$	175	120	1.46	M1 A1	<p>For fds - at least 3 correct Accept any suitable unit for fd such as eg freq per cm.</p> <p>Allow 15.9 and 1.5 and condone 1.45</p>	<p>M1 can be also be gained from freq per 10 – 4.9, 10.9, 18.2, 15.35, 0.146 (at least 3 correct) or similar.</p> <p>If fd not explicitly given, M1 A1 can be gained from all heights correct (within \leq one square) on histogram (and M1A0 if at least 3 correct)</p>
Price	Frequency	Group width	Frequency density																											
$10 \leq x \leq 40$	147	30	4.90																											
$40 < x \leq 50$	109	10	10.90																											
$50 < x \leq 60$	182	10	18.20																											
$60 < x \leq 80$	317	20	15.85																											
$80 < x \leq 200$	175	120	1.46																											
			<div></div> <p>NB If not using fd's only mark available is B1 for width of bars</p> <p>Heights must be within ≤ 1 square of overlay (only for scales 2cm = 4 units (blue) or 5 units (red)) – otherwise check heights. Note that you must make sure that the overlay is aligned correctly with the vertical axis.</p>	B1 B1 B1 [5]	<p>linear scales on both axes and label on both axes (Allow horizontal axis labelled x) Vertical scale starting from zero (not broken - but can get final mark for heights if broken)</p> <p>width of bars (within half a square) (NO GAPS ALLOWED)</p> <p>height of bars</p>	<p>Linear scale and label on vertical axis IN RELATION to first M1 mark ie fd or frequency density or if relevant freq/10, etc (NOT eg fd/10). However allow scale given as fd\times10, or similar Accept f/w or f/cw (freq/width or freq/class width) Can also be gained from an accurate key G0 if correct label but not fd's.</p> <p>Must have linear scale. Condone starting at 10 rather than 0. For inequality labels see additional notes below.</p> <p>Height of bars – must be linear vertical scale. FT of heights dep on at least 3 heights correct and all must agree with their fds If fds not given and 3 or 4 heights correct then max M1A0G1G1G0</p>																								

Question			Answer	Marks	Guidance	
6	(iv)		Positive skewness	B1 [1]	Allow +ve	
6	(v)		Area for men from 100 to 200 = $100 \times 2 = 200$ $200/990 = 0.202$ So 20.2% Cannot be certain as both figures are estimates	M1 A1 E1 [3]	Independent	Or $^{100}/_{120} \times 240$ 20% with working Allow comments such as ‘grouped data so cannot be certain’ or ‘Values are not exact so cannot be certain’ oe or ‘midpoints have been used so cannot be certain’ oe
6	(vi)		Men’s running shoes have a lower average price than women’s (as their mean is only £68.83 compared to £70.19). Or equivalent for women Men’s running shoes have a more variation in price than women’s (as their sd is £42.93 compared to £36.89). Or equivalent for women	E1 E1 [2]	FT their mean Do NOT condone lower central tendency or lower mean FT their SD	Allow ‘on the whole’ or similar in place of ‘average’. Allow ‘more spread’ or similar but not ‘higher range’ or ‘higher variance’ or ‘less distributed’ Condone less consistent
7	(i)	(A)	$X \sim B(16, 0.1)$ $P(X = 3) = 0.25^4 \times 0.75^{16} \binom{20}{4} \times 0.25^4 \times 0.75^{16} = 0.1423$ Or: From tables $P(X \leq 3) - P(X \leq 2) = 0.9316 - 0.7892 = 0.1424$	M1 M1 A1 M2 A1 [3]	For $0.1^3 \times 0.9^{13}$ For $\binom{16}{3} \times p^3 \times q^{13}$ CAO For $0.9316 - 0.7892$ CAO	With $p + q = 1$ Also for $560 \times 0.000254..$ Allow 0.14 or better
7	(i)	(B)	$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.7892 = 0.2108$	M1 A1 [2]	For 0.7892 CAO	If calculating $P(X = 0) + P(X = 1) + P(X = 2)$ allow M1 for 0.79 or better and A1 for 0.21 or better.

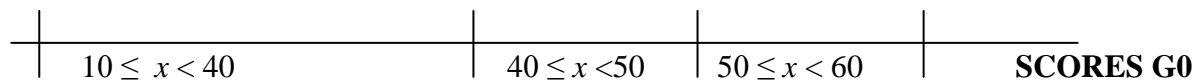
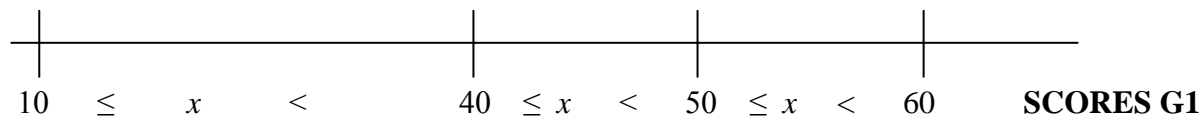
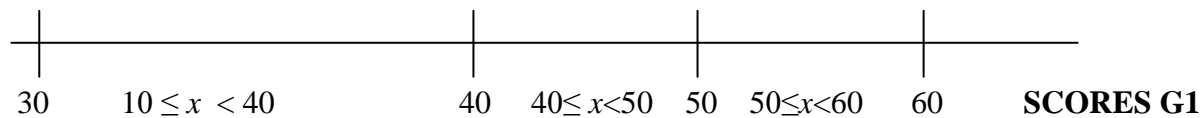
Question			Answer	Marks	Guidance	
7	(i)	(C)	Expected number = $16 \times 0.1 = 1.6$	B1 [1]		Do not allow final answer of 1 or 2 even if correct 1.6 given earlier
7	(ii)		Let p = probability of a randomly chosen person using 1234 as their PIN (in the population) $H_0: p = 0.1$ $H_1: p < 0.1$ The alternative hypothesis has this form as the advertising campaign aims to reduce the proportion of the population who use 1234 as their PIN.	B1 B1 B1 B1 [4]	For definition of p (in context) For H_0 For H_1 Dep on < 0.1 used in H_1 Do Not allow just 'proportion will be lower' or similar.	Do NOT allow <u>number</u> in place of probability. See below for additional notes For use of 0.9 as P(do not use 1234), contact team leader. E0 for simply stating H_1 in words Condone number instead of proportion.
7	(iii)	(A)	For $n = 20$, $P(X \leq 0) = 0.1216$ $0.1216 > 0.10$ So no point in carrying out the test as H_0 could not be rejected (even if nobody in the sample uses 1234 as their PIN). oe	M1* *M1dep A1 [3]	For sight of 0.1216 For > 0.10 or $> 10\%$ Do NOT FT wrong H_1 or state 'There is no critical region' oe For A1 need $P(X \leq 0)$ or $P(X = 0)$ somewhere oe	Condone $P(X = 0)$ in place of $P(X \leq 0)$ Need to see a comparison with 0.1 or 10% explicitly, not just mentioning significance level. Allow SC2 for clearly indicating use of $B(20, 0.1)$ but with no mention of 0.1216 with convincing reasoning and final answer correct Allow CR is empty but NOT CR is zero
		(B)	Lowest value of k is 13	B1 [1]	Or 13%	
7	(iv)		$P(X \leq 2) = 0.0530$ $0.0530 > 0.05$ So not significant. Do not reject H_0 Conclude that there is not enough evidence to support the suggestion that the advertising campaign has been successful. Reminder: When you mark this question part, if you 'fit to height' you can check the last page for working or mark it BP if there none	B1 M1 A1* *E1 dep [4]	For <u>use</u> of $P(X \leq 2)$ only For comparison of 0.0530 with 5% Also allow $P(X \leq 2) > 0.05$, $(P(X \leq 1) < 0.05)$ so CR is $\{0, 1\}$ for first two marks then A1E1 as usual Condone 'number of people' in conclusion	No marks unless H_1 correct If B0 then no further marks Allow 'accept H_0 ' or 'reject H_1 ' Must include 'insufficient evidence to suggest that' or something similar i.e. an element of doubt either in the A or E mark.

NOTE RE OVER-SPECIFICATION OF ANSWERS

If answers are grossly over-specified, deduct the final answer mark. Note in Q6i only deduct 1 mark even if both mean and sd over-specified. Probabilities should also be rounded to a sensible degree of accuracy. In general final non probability answers should not be given to more than 4 significant figures. Allow probabilities given to 5 sig fig.

ANNOTATION RULES

See note 12 above and particularly 12a. Remember to put full annotation on all practice and standardisation scripts unless the candidate has scored full marks or zero. In addition for all marking in Q6(iii), if the candidate has not scored full marks then show which B marks have been awarded in the right hand margin, in the same order as they are given in the mark scheme. You should indicate any errors made.

Additional notes re Q6 part iii G mark**BUT**

Additional notes re Q7 part ii

Minimum needed for B1 is p = probability of using 1234.

Allow $p = P(\text{using 1234})$

Definition of p must include word probability (or chance or proportion or percentage or likelihood but NOT possibility, number or amount).

Preferably given as a separate comment. However can be at end of H_0 as long as it is a clear definition ' p = the probability of using 1234.'

Do NOT allow ' p = the probability of using 1234 is different'

Allow $p=10\%$, allow only p or θ or π or ρ . However allow any single symbol if defined (including x)

Allow $H_0 = p=0.1$, Allow $H_0 : p=1/10$

Allow NH and AH in place of H_0 and H_1

Do not allow $H_0 : P(X=x) = 0.1$

Do not allow $H_0 : =0.1, =10\%, P(0.1), p(x)=0.1, x=0.1$ (unless x correctly defined as a probability)

Do not allow H_0 and H_1 reversed

For hypotheses given in words allow Maximum B0B1B1

Hypotheses in words must include probability (or chance or proportion or percentage) and the figure 0.1 or

Thus eg $H_0 : P(\text{using 1234}) = 0.1, H_1 : P(\text{using 1234}) < 0.1$ gets B0B1B1

4766 Statistics 1

General Comments:

As last year, the majority of candidates coped very well with this paper and a large number scored at least 60 marks out of 72. There was no evidence of candidates being unable to complete the paper in the allocated time. Most candidates had adequate space in the answer booklet without having to use additional sheets. Candidates who did need additional space often used the last page of the answer book, but a number did not, presumably not realising that it was available. It is also pleasing to report that losing marks due to over-specification was less of a problem than in previous years. Some candidates lost a mark in question 6 part (i) due to giving their answer to the mean and/or the standard deviation to 5 significant figures. A very small number gave probabilities to more than 5 significant figures thus again losing a mark.

Candidates usually scored very well on question 4 on frequency distributions, question 6 parts (i) and (iii) on data measures and histograms and on question 7 parts (i) on the binomial distribution. Even part (ii) of question 7, where candidates had to state hypotheses and define p , was very well answered. This is very pleasing, as up until recently, this topic has caused many candidates problems.

Questions on which candidates did not score so highly included question 3 parts (iii) and (iv) on probability, question 5 part (iii) on conditional probability and the latter parts of question 7 on hypothesis testing. In question 1, a surprisingly large number of candidates did not know how to find the quartiles correctly from a stem and leaf diagram, although most then knew the definition of outliers, so were able to gain marks in part (ii). Questions 2 and 3 caused difficulties to some candidates who scored highly in other areas of the paper. These two questions appeared to be a little less routine and required engagement with the scenario.

Candidates sometimes did not read the question carefully enough, so explanations did not always answer what was asked or explanations were missed out entirely. Candidates have, in most cases, been well prepared for calculations required in the paper but less so for analysing their findings. Although many candidates gave well written explanations, the poor handwriting, grammar and/or use of English made it difficult to work out what some others were trying to say. There were also problems with fours that looked like sixes or nines and ones and sevens that were difficult to tell apart.

Comments on Individual Questions:

Q1(i) The vast majority of candidates found the median correctly. A small minority misread/ignored the key to the stem and leaf diagram and gave an incorrect answer of 290. However under half of candidates found the quartiles correctly, with many using 5th and 15th values, which was penalised.

Q1(ii) Most candidates gained full marks, often on follow through from quartiles which were slightly out. The most common error was to use the median in calculations. A few candidates started from scratch and calculated mean and standard deviation. Some managed this successfully, but others made errors in their calculations, or incorrectly used a combination of both methods such as mean $\pm 1.5 \times$ interquartile range.

Q2(i) Many candidates thought that not losing meant winning, and hence gave the common wrong answer of $0.5^3 = 0.125$. Others tried to consider combinations of wins and draws, often without success. The fact that the question part was only worth 1 mark should have been a clue to the fact that there was an easier approach.

Q2(ii) This was generally well answered although a few candidates interpreted this as ‘find three separate probabilities which they did, and listed them but with no addition thus scoring zero.

Q2(iii) Only a small proportion of candidates used the most elegant approach (the first method in the mark scheme) and of those who did, many forgot to multiply by 6. Most candidates gave lists or tree diagrams to show $P(WWL)$ etc., but many then did not multiply by 3, so the most common answer was 0.22, rather than the correct 0.66.

Q3(i) This was generally well answered.

Q3(ii) Again this was generally well answered, although some candidates truncated their decimal rather than correctly rounding. The use of fractions was preferable here.

Q3(iii) Only around two thirds of candidates scored the mark here. The most common error was to use combinations.

Q3(iv) Again about two thirds of candidates scored both marks, with many scoring the marks for their answer to part (iii), correct or not, multiplied by 6, rather than for the fairly simple $\frac{3}{5} \times \frac{2}{4} \times \frac{1}{3}$.

Q4(i) Generally the solution of the equation to find k was carried out well, although not always entirely efficiently. Most candidates seemed to know that the sum of the probabilities should be 1, although not all thought it necessary to say so. However some did not include addition signs and so lost the marks, and others seemed to have no idea about addition of fractions, thinking that the sum of the probabilities was $\frac{5k}{70}$. Those who used $k = 1.2$ and verified the sum got some very easy marks, although not all were convincing enough to get full credit. The vast majority used the correct probabilities in part (ii), although not all tabulated them in a table to get the third mark in part (i).

Q4(ii) This part was very well answered with about 80% of candidates scoring full marks. A few candidates only found $E(X^2)$ and a few used spurious division. Very few candidates attempted to find $E(X - \mu)^2$ and those who did were rarely successful.

Q5(i) This part was very well answered but a considerable number of candidates assumed that the probabilities were independent and calculated $P(R) \times P(S)$. Some were more confused about the correct formula to use and calculated $P(R) \times P(R \cup S)$.

Q5(ii) The idea of the Venn diagram was well understood, and most candidates produced a fully correct solution (often following through from an error in part (i)). Very few noticed the contradiction produced by their wrong answer, which gave the outer zone as 0.0936 instead of 0.06 from the question.

Q5(iii) Among those who had not made the independence error in part (i), the correct answer was quite common. The explanation of what the probability means was usually correct but sometimes lacked sufficient detail. There were a few candidates who ‘reversed’ the statement and gave an explanation of $P(S|R)$.

Q6(i) This part was fairly well answered with over half of candidates gaining full credit. A few had no idea how to proceed, but most used correct midpoints, although some made slips with them or occasionally used figures such as 25.5, 45.5, etc. The standard deviation proved more difficult for a number of candidates with a variety of wrong methods seen. Very few used the statistical functions on their calculator to do this question, despite this being the recommended method. A few candidates over-specified either or both of their final answers and so lost a mark.

Q6(ii) Candidates found this part rather more challenging, although almost half scored full marks. Trying to establish the proportion they were after was the biggest stumbling block. However, some were then unsure what to do with the figure of 145.83 once they had found it. Some rounded down to 145 (probably the most common mistake of those who understood what they needed to do) and others failed to finish by finding the percentage, just giving the final answer as a decimal 0.157.

Q6(iii) This part was again well answered with around 80% of candidates gaining at least 4 marks out of 5. Various errors were seen, but none very commonly. The most frequently seen were: using frequency rather than frequency density, using a non-linear scale on one of the axes (usually the horizontal axis), stopping the horizontal axis at 120, and labelling the horizontal axis 'Class width'.

Q6(iv) Over 90% of candidates scored the one mark available here.

Q6(v) A good number of candidates achieved full marks, and the question was answered better than question 6 part (ii) which is a similar calculation. Of those who got the calculation incorrect most started with $240/990$ or $20/990$, rather than $200/990$. The explanation over certainty was well answered with most candidates achieving this mark, whether or not they got the first 2 marks.

Q6(vi) Although this is essentially a simple question, almost a third of candidates scored zero. Candidates struggled to provide acceptable comparisons, with many relying on terms such as "central tendency" when comparing the means, and relatively few discussing averages. A more encouraging proportion of candidates were able to provide a good interpretation for the differences in the standard deviations. Some thought that central tendency was something to do with variation. A number of candidates were unable to construct a proper, legible, grammatically correct sentence.

Q7(i)(A) This was very well answered.

Q7(i)(B) Again this was well answered, usually by use of tables, although some candidates did calculate the three probabilities, add and subtract from 1. A few candidates forgot to subtract from 1, and a few just subtracted $P(X = 2)$ from 1.

Q7(i)(C) The majority of the candidates found this part straightforward, but a small minority lost the mark when they rounded their final answer to 1 or 2.

Q7(ii) As in recent years, candidates did well on this part, with over 80% gaining at least 3 marks out of 4. Most candidates scored the first two marks for the hypotheses, with many knowing that they needed to define p , thus scoring the third mark. A valid explanation of the reason for the form of the alternative hypothesis was usually given, even if not always very well worded.

Q7(iii)(A) Only about half of the candidates scored any marks here at all. Many candidates did not use any numbers so could not gain all the marks, but were awarded special case 2 if they gave a very convincing explanation. Some of those that did state that $P(X \leq 0) = 0.1216$ then failed to show a comparison with 10% or 0.1 and so only scored 1 mark.

Q7(iii)(B) Those candidates who had 0.1216 in the previous part usually gave the correct answer of 13%. Some who did not get marks in the previous part did give the correct answer, so they probably simply did not know how to verbalise the previous answer. However under half scored the mark, some by simply failing to round to an integer.

Q7(iv) There were many good, clear answers to this part of the question but there were still a good proportion of candidates that were tempted to use point probabilities. A significant number who did use the correct probability (or probabilities if using a critical region method) failed to give the conclusion of the test in context. Some lost the final mark for commenting that the proportion had not changed instead of had not reduced and some gave a conclusion which was too assertive.