Cambridge International Examinations
Cambridge Pre-U Certificate

MATHEMATICS (STATISTICS WITH PURE MATHEMATICS) **1347/02**

Paper 2 Statistics **May/June 2017**

MARK SCHEME

Maximum Mark: 80

Published

This mark scheme is published as an aid to teachers and candidates, to indicate the requirements of the examination. It shows the basis on which Examiners were instructed to award marks. It does not indicate the details of the discussions that took place at an Examiners' meeting before marking began, which would have considered the acceptability of alternative answers.

Mark schemes should be read in conjunction with the question paper and the Principal Examiner Report for Teachers.

Cambridge will not enter into discussions about these mark schemes.

Cambridge is publishing the mark schemes for the May/June 2017 series for most Cambridge IGCSE®, Cambridge International A and AS Level and Cambridge Pre-U components, and some Cambridge O Level components.

This document consists of **8** printed pages.

MARK SCHEME NOTES

The following notes are intended to aid interpretation of mark schemes in general, but individual mark schemes may include marks awarded for specific reasons outside the scope of these notes.

Types of mark

M Method marks, awarded for a valid method applied to the problem.

A Accuracy mark, awarded for a correct answer or intermediate step correctly obtained. For accuracy marks to be given, the associated Method mark must be earned or implied.

B Mark for a correct result or statement independent of Method marks.

When a part of a question has two or more 'method' steps, the M marks are in principle independent unless the scheme specifically says otherwise; and similarly where there are several B marks allocated. The notation '**dep**' is used to indicate that a particular M or B mark is dependent on an earlier mark in the scheme.

Abbreviations

AEF	any equivalent form
art	answers rounding to
cao	correct answer only
dep	dependent
FT	follow through after error
oe	or equivalent
rot	rounded or truncated
SC	Special Case
soi	seen or implied
www	without wrong working

Question	Answer	Marks	Guidance
1(i)	$S_{xx} = 2881590 - \frac{(5844)^2}{12} = 35562$	B1	
	$S_{yy} = 493502 - \frac{(2268)^2}{12} = 64850$	B1	
	$S_{xy} = 1149981 - \frac{5844 \times 2268}{12} = 45465$	B1	
	$r = \frac{45465}{\sqrt{35562 \times 64850}} = 0.9467$ (= 0.947)	M1	Calculating r from *their* S_{xx}, S_{yy} and S_{xy} (numerical working or *their* r value correct to 3 sf or better)
	r is very near 1, so a good fit to (an upward sloping) line	A1	Drawing a valid conclusion (confirming that a linear fit is appropriate, as stated in question)
1(ii)	$b = \frac{45465}{35562} = 1.278$ (= 1.28)	M1	Calculating b from their S_{xx}, S_{xy} (calculation seen or 1.27 to 1.28)
	$a = \frac{2268}{12} - 1.278 \times \frac{5844}{12}$ $= 189 - 1.278 \times 487 = -433.4$	M1	Calculating a from Σx, Σy and their b (soi) or –433.0 to –434.5 for a (www)
	$y = 1.278x - 433.4$	A1	Line correct with coefficients 1.27 to 1.28 and –433.0 to –434.5 used correctly
1(iii)(a)	$x = 652 \Rightarrow \hat{y} = 400$	M1	Anything rounding to 400 or 401 or **FT** *their* line
	Actual price (402) is close to prediction	A1	An appropriate comment
1(iii)(b)	$x = 460 \Rightarrow \hat{y} = 154.5$	M1	Anything rounding to 154 or 155 or **FT** *their* line
	Actual price (220) is substantially higher than prediction	A1	An appropriate comment
2(i)	X = number of correct guesses $X \sim B(10, 0.2)$	B1	Parameters 10 and 0.2 seen or implied
	$P(X \geqslant 4) = 1 - P(X \leqslant 3) = 1 - 0.8791$	M1	Finding $P(X \leqslant 3)$ or $P(X \leqslant 4)$
	$= 0.1209$	A1	[0.9672 or 0.0328 $\Rightarrow$ M1, A0]

Question	Answer	Marks	Guidance
2(ii)	H_0: P(correct guess) = 0.2	**B1**	'Probability of a correct guess' – p may be defined words in answer to part (i) 'H_0: Billie is making random guesses', or equivalent, is insufficient ⇒ B0
	H_1: P(correct guess) > 0.2	**B1**	or > 20% or > $\frac{1}{5}$
2(iii)	Assuming H_0: $X \sim B(10, 0.2)$ Using tables,	**M1**	Using B(10, 0.2), or implied from subsequent working
	$P(X \leqslant 4) = 0.9672$ so $P(X \geqslant 5) = 0.0328 > 1\%$	**A1**	0. 0.9672 or 0.0328 seen, may round to 3 sf
	$P(X \leqslant 5) = 0.9936$ so $P(X \geqslant 6) = 0.0064 < 1\%$	**A1**	0.9936 or 0.0064 seen, may round 0.9936 to 3 sf
	Critical value is 6	**B1**	cao
2(iv)	$0.2 + (0.5 \times 0.8) = 0.6$	**B1**	
	In B(10, 0.6),	**M1**	$P(X <$ their cv) in B(10, their 0.6) (but not B(10, 0.2)
	$P(X < 6) = 0.3669$	**A1**	0.367 (or better) or 36.7% (or better), **FT** *their* cv and *their* p if possible.
	This is the probability of falsely concluding that the probability of Billie being told a guess is correct is 0.2	**B1**	May say things like 'concluding that Billie is guessing' or 'concluding that Billie is not psychic'. Follow through wording of their H_0 if reasonable.
3(i)(a)	Mean = 79053 ÷ 20 = 3952.65	**M1**	Method substantially correct
	= 3950 to 3sf	**A1**	3950 or better
3(i)(b)	Median = (4240 + 4200) ÷ 2	**M1**	Method substantially correct
	= 4220	**A1**	
3(i)(c)	Upper quartile = (5464 + 4910)/2 = 5187	**M1**	Accept any value from 4910 to 5464 (incl)
	Lower quartile = (2428 + 2374)/2 = 2401	**M1**	Accept any value from 2374 to 2428 (incl)
	Interquartile range = 5187 – 2401 = 2786	**A1**	IQR correct for their quartiles (within range)

<table>
<tr><th>Question</th><th>Answer</th><th>Marks</th><th>Guidance</th></tr>
<tr><td rowspan="4">3(ii)</td><td>Interquartile range = 1433</td><td>M1</td><td>1433 soi in UQ + 1.5×IQR</td></tr>
<tr><td>$3058 + 1.5 \times 1433 = 5207.5$
(may also see 908.5)</td><td>A1</td><td>5207.5 or (a given length – 3058) ÷ 1433, oe</td></tr>
<tr><td>A valid and correct calculation together with Nile and Amazon</td><td>M1</td><td></td></tr>
<tr><td>Nile, Amazon, Mississippi, Yangtze, Yellow</td><td>A1</td><td>These five and no others</td></tr>
<tr><td rowspan="2">3(iii)</td><td>$25 - (9 + 3 + 1)$</td><td>M1</td><td>Valid first step, e.g. 25 minus …</td></tr>
<tr><td>= 12</td><td>A1</td><td></td></tr>
<tr><td rowspan="3">4(i)</td><td>$p = (19 + 16 + 10 + 15) \div (30 + 30 + 20 + 20)$
$= 60/100 = 0.60$</td><td>B1</td><td>0.60 or 60%</td></tr>
<tr><td>At least two correct values for their p</td><td>M1</td><td></td></tr>
<tr><td><table><tr><th>Species</th><th>Expected number of rejected eggs</th></tr><tr><td>Warbler</td><td>18</td></tr><tr><td>Thrush</td><td>18</td></tr><tr><td>Blackbird</td><td>12</td></tr><tr><td>Wagtail</td><td>12</td></tr></table></td><td>A1</td><td>All correct for $p = 0.60$</td></tr>
</table>

Question	Answer	Marks	Guidance
4(ii)	H_0: $p = 0.6$	**B1**	Stating an appropriate H_0 Allow: p is independent of species, p is constant, etc. Follow through *their* expected frequencies
	$\frac{(19-18)^2}{18}+\frac{(16-18)^2}{18}+\frac{(10-12)^2}{12}$ $+\frac{(15-12)^2}{12}$	**M1**	Any one of these terms correct, for *their* expected frequencies
	$=\frac{1}{18}+\frac{2}{9}+\frac{1}{3}+\frac{3}{4}$ or 0.056 + 0.222 + 0.333 + 0.75 $=\frac{49}{36}$ or 1.361	**A1**	$\frac{49}{36}$ or $1\frac{13}{36}$ or 1.36 (or better)
	$\nu = 2$	**B1**	
	$\Rightarrow$ cv = 4.605	**B1**	**FT** *their* ν at 10% level ($\nu = 3 \Rightarrow$ cv = 6.251)
	$1.361 < 4.605 \Rightarrow$ accept H_0	**M1**	Accept H_0, or equivalent, correct for *their* cv and *their* 1.36
	Insufficient evidence to conclude that probability varies for different species	**A1ft**	A correct conclusion in context, for their values
5(i)	Rank 1 2 3 4 5 6 Time 38 39 43 44 45 48 Race F F F J J F Rank 7 8 9 10 11 12 Time 51 52 59 62 63 67 Race F F J J J J	**B1**	Times ranked correctly
	Rank sum J = 51	**M1**	R_m = 51 or 27
	Rank sum F = 27	**A1**	$m(m+n+1) - R_m$ = the other of 27, 51
	$W = 27$	**B1**	cao
	H_0: population medians same, $\alpha = 5\%$ H_1: different medians two-sided $m = 6$, $n = 6 \Rightarrow$ cv = 26	**B1**	cao
	$27 > 26 \Rightarrow$ accept H_0	**M1**	Accept H_0 (**FT** *their* W and cv)
	Insufficient evidence to claim that the population medians are not the same	**A1**	**FT** Correct conclusion in words, from correct working for *their* W and cv.

Question	Answer	Marks	Guidance
5(ii)	$W \sim \mathrm{N}\left(\frac{1}{2}m(m+n+1), \frac{1}{12}mn(m+n+1)\right)$	**M1**	Substantially correct method (but, e.g. using 120 instead of 60)
	$m = n = 60 \Rightarrow W \sim \mathrm{N}(3630, 36300)$ approx	**A1**	N(3630, 36300), cao
	$\alpha = 5\%$, two-sided $\Rightarrow z = \pm 1.96$ in N(0, 1) cv tests smaller value, so –1.96	**B1**	1.96 or –1.96 or ±1.96
	$3630 - 1.96\sqrt{36300} = 3256.57$	**M1**	Their $3630 - 1.96 \times \sqrt{\textit{their}\ 36300}$
	$\Rightarrow$ cv = 3256	**A1**	**FT** *their* 3630, 36300 and round down to an integer
	If the calculated value of W is greater than 3256 then accept H_0, otherwise reject H_0	**B1**	A valid description
5(iii)	Athlete: A, B, C, D, E, F J: 67, 45, 63, 44, 62, 59 F: 52, 48, 43, 51, 39, 38 |J-F|: 15, 3, 20, 7, 23, 21 Sign: +, –, +, –, +, + Rank: 3, 1, 4, 2, 6, 5	**B2**	**B1** Differences or absolute differences correct **B1** *Their* absolute differences ranked correctly
	$P = 18, Q = 3 \Rightarrow T = 3$	**B1**	cao
	H_0: symmetric about 0 $\alpha = 5\%$ H_1: times have improved one-sided $m = 6 \Rightarrow$ cv = 2	**B1**	Differences symmetric decreased, mostly positive differences cao
	$3 > 2 \Rightarrow$ accept H_0	**M1**	Accept H_0 (**FT** *their* T and cv)
	Insufficient evidence to claim that the times have improved	**A1ft**	Correct conclusion in words, from correct working for *their* T and cv.
6(i)	Estimate μ using $\bar{x} = \frac{3000}{100} = 30$	**B1**	
	1584 ÷ either 99 or 100	**M1**	
	Estimate σ^2 using $s^2 = \frac{1584}{99} = 16$	**A1**	
6(ii)	Some schools have more classes than others With this method you cannot get two classes from the same school	**B1**	Sampling units are classes not schools

Question	Answer	Marks	Guidance
6(iii)	$n = 100$ is large enough that we may use N(0, 1) $z = \pm 1.645$	**B1**	1.645 or ± 1.645
	$30 \pm 1.645\times\sqrt{(16/100)}$	**B1**	16/100 soi (**FT** *their* 30 and *their* 16)
	$= 30 \pm 0.658$	**M1**	Correct method for *their* z (or t) value
	$= (29.3, 30.7)$	**A1**	Confidence limits correct (29.342 and 30.658), 3sf or better
	t-distribution not necessary because sample size is not small	**B1**	$n = 100$
6(iv)	Variance has been estimated (and sample is small) so a t-distribution is appropriate Use t-distribution with 4 degrees of freedom	**B1**	$t(4)$ used
	$\mathrm{P}(\overline{Y} > \mu + 6.5) = \mathrm{P}\left(T > \dfrac{6.5}{\sqrt{(27.5/5)}}\right)$	**B1**	$\sqrt{(27.5/5)}$ used
	$= \mathrm{P}(T > 2.77)$ $= 1 - 0.975$ approx,	**M1**	Substantially correct method
	$= 0.025$ as given	**A1**	2.77 and then showing how this leads to the given answer of 0.025