Cambridge International Examinations
Cambridge Pre-U Certificate

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

Paper 2 Statistics **For Examination from 2016**

SPECIMEN MARK SCHEME

2 hours

MAXIMUM MARK: 80

The syllabus is approved for use in England, Wales and Northern Ireland as a Cambridge International Level 3 Pre-U Certificate.

This document consists of **7** printed pages and **1** blank page.

Mark Scheme Notes

Marks are of the following three types:

M Method mark, awarded for a valid method applied to the problem. Method marks are not lost for numerical errors, algebraic slips or errors in units. However it is not usually sufficient for a candidate just to indicate an intention of using some method or just to quote a formula; the formula or idea must be applied to the specific problem in hand, e.g. by substituting the relevant quantities into the formula. Correct application of a formula without the formula being quoted obviously earns the M mark and in some cases an M mark can be implied from a correct answer.

A Accuracy mark, awarded for a correct answer or intermediate step correctly obtained. Accuracy marks cannot be given unless the associated method mark is earned (or implied).

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

The following abbreviations may be used in a mark scheme:

AG Answer Given on the question paper (so extra checking is needed to ensure that the detailed working leading to the result is valid)
CAO Correct Answer Only (emphasising that no "follow through" from a previous error is allowed)
aef Any equivalent form
art Answers rounding to
cwo Correct working only (emphasising that there must be no incorrect working in the solution)
ft Follow through from previous error is allowed
o.e. Or equivalent

1	(i)	$S_{xx} = 1939552 - \frac{(4412)^2}{13} = 442187$ (to nearest integer)	B1
		$S_{yy} = 605147 - \frac{(2387)^2}{13} = 166857$ (to nearest integer)	B1
		$S_{xy} = 1074848 - \frac{4412 \times 2387}{13} = 264737$ (to nearest integer)	B1
		$r = \frac{264737}{\sqrt{442187 \times 166857}} = 0.975$ (0.9746)	
		Calculating r from *their* S_{xx}, S_{yy} and S_{xy} (numerical working or *their* r value correct to 3 sf or better)	M1
		r is near 1, so a good fit to an upward sloping line Drawing a valid conclusion (confirming that a linear fit is appropriate, as stated in question)	A1
	(ii)	$b = \frac{264737}{442187} = 0.599$ (0.5987) Calculating b from *their* S_{xx} and S_{xy}	M1
		$a \quad = \frac{2387}{13} - 0.5987 \times \frac{4412}{13}$ Calculating a from $\sum x$, $\sum y$ and *their* b $\quad = 183.6 - 0.5987 \times 339.4 = -19.6$	M1
		$y = 0.599x - 19.6$ Line correct with coefficients to 3sf or better	A1
		$x = 2203 \Rightarrow \hat{y} = 1300$ (From their line ($\pm$ 2))	B1
	(iii)	Extrapolation beyond range of data	B1
		Any valid objection Small sample / only based on one sample Sampling method not known / not random sampling London is not typical / London 'is different'	B1
2	(i)	Median = 30 mpg CAO	B1
		Quartiles = 34 mpg and 23 mpg (Accept 33 to 35 and 20 to 24)	B1
		IQR = 11 mpg (*their* IQR calculated)	M1
		Outliers have mpg < 6.5 or > 50.5	A1
		$\Rightarrow$ Toyota Prius	B1
	(ii)	Using median and quartile values appropriately to deduce non-normal	B1
		e.g. The difference between 23 and 30 is much greater than the difference between 30 and 34; this suggests that the distribution is not symmetric.	

	(iii)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 1 2 12 3 4 5 6 7 8 11 9 10 13 14 15 0 0 –9 1 1 1 1 1 1 –1 2 2 0 0 0 Substantially correct calculation of d or $\|d\|$ or d^2 for the ranks $\Sigma d^2 = 96$ $r_s = 1 - \frac{6 \times 96}{15 \times 224} = 1 - 0.17143 = 0.82857$ Substantially correct calculation of d or $\|d\|$ or d^2 for the ranks = 0.829 (3 sf)	 M1 A1 M1 A1
3	**(i)**	Independence between children, (random sample) class are typical of population in respect of left-handedness (Independence Probability 13% (constant probability))	B1 B1
	(ii)	X = number of left-handers $X \sim B(20, 0.13)$ 13% of 20 = 2.6, so want $P(X \leqslant 2)$ $(0.87)^{20} + 20(0.13)(0.87)^{19} + 190(0.13)^2(0.87)^{18}$ Calculating a probability in B(20, 0.13) (At least) three correct probabilities added = 0.061714 + 0.18443 + 0.26181 = 0.50795... = 0.508 to 3sf	 B1 M1 A1 A1
	(iii)	$X \sim B(20, p)$ p = P(left-hander) (may be implied) H_0: $p = 0.13$ H_1: $p > 0.13$ $\alpha = 5\%$ one-tailed test (Omission of p only penalised once. May imply level of test and one-tailed) Assuming H_0, $X \sim B(20, 0.13)$ $P(X \geqslant 7) = 1 - 0.9897 = 0.0103$ or cv = 6 0.0103 < 5% or 7 > 6 Reject H_0 Evidence supports claim, significantly more of the most recent twenty presidents were left-handed than would be expected by chance. (must be in context)	 B1 B1 M1 M1 A1 B1
	(iv)	Any valid reason, either from context or addressing statistical variation e.g. Schools trained pupils to write with their right hand in the past Left-handedness was not recorded accurately in the past Not random samples, could be due to sample variation	B1

4	**(i)**	$P(Z > z) = 0.01 \Rightarrow z = 2.326$ $P(Z < z) = 0.25 \Rightarrow z = -0.674$ (both values required for mark) Substantially correct method $2.326 = \frac{120-\mu}{\sigma} \Rightarrow 120 - \mu = 2.326\sigma$ $-0.674 = \frac{84-\mu}{\sigma} \Rightarrow 84 - \mu = -0.674\sigma$ Both correct for *their* z-values, one of which is positive and one negative $\Rightarrow \mu = 92.1$ CAO $\sigma = 12$ CAO	B1 M1 A1 B1 B1
	(ii)	H_0: samples come from same populations H_1: S tend to have larger increases than N (S have smaller rank values than N) Appropriate statement of hypotheses One-tailed test, $\alpha = 5\%$ Rank sum for S = 1 + 3 + 4 + 5 + 6 + 8 $\Rightarrow W = 27$ $m = 6$ $n = 10 \Rightarrow$ critical value for $W = 35$ Reject H_0 (Correct conclusion, in context) At the 5% level the data support the claim that the increases are greater for the smokers than for the non-smokers	 B1 M1 A1 B1 B1
	(iii)	For the smokers, $\sum x = 708 \Rightarrow \bar{x} = 118$ Estimate $\widehat{\mu_s} = 118$ CAO $\sum x^2 = 83864 \Rightarrow S_{xx} = 320 \Rightarrow s^2 = 64$ Sight of one of 83864, 320, 64, 8, 53.3 or 7.30 Estimate $\widehat{\sigma_s^2} = 64$ CAO	 B1 M1 A1
	(iv)	$\overline{X} \sim N(\mu_s, \frac{\sigma_s^2}{n})$ where $\widehat{\sigma_s} = 8$ and $n = 6$ Critical values in t(5) are $\pm$ 2.571 Using t tables to find 2.571 or 2.447 Confidence interval is $118 \pm 2.571 \times \frac{8}{\sqrt{6}}$ Correct method for *their* "t" value and *their* $\bar{x}, \hat{\sigma}$ $= 118 \pm 8.4 = [109.6, 126.4]$ (ft *their* values from part **(iii)**)	 B1 M1 A1

5	**(i)**	$x = 47 \rightarrow z = 0.667$; $x = 51 \rightarrow z = 2.0$ (z values may be implied)	B1
		$P(47 < X \leqslant 51) = 0.9772 - 0.7477$ (may be implied)	M1
		Expected frequency = 0.2295 × 100 = 22.95 AG	A1
	(ii)	Merge classes in tails to make expected frequencies at least 5	B1
		Weight: <43, 43–45, 45–47, >47 Observed frequency: 32, 24, 30, 14 Expected frequency: 25.23, 24.77, 24.77, 25.23 $(O-E)^2/E$: 1.82, 0.02, 1.10, 5.00 Substantially correct calculation of X^2 (with or without merging)	M1
		X^2 calculated = art 7.94 CAO	A1
		H_0: N(45, 9) distribution H_1: some other distribution	B1
		From tables, critical value = 7.815 CAO	B1
		Reject H_0 Data is not consistent with a N(45, 9) distribution	B1
	(iii)	$v = n - 1 = 4 - 1 = 3$	B1
		No need to reduce degrees of freedom for parameters as not estimated from sample data	
	(iv) (a)	Variance cannot be estimated, midpoints cannot be found for first and last classes since boundaries are not known	B1
	(b)	Sign test or binomial test or equivalent (e.g. test proportion that are below 45)	B1
		H_0: median = 45 $\alpha = 5\%$ H_1: median ≠ 45 two-tailed test	
		Y = number of chicks with weight $\leqslant$ 45g Assuming H_0, $Y \sim B(100, 0.5)$	M1
		Approximate by N(50, 25) Critical values are $50 \pm 1.96 \times 5 = 50 \pm 9.8$ = [40.2, 59.8]	A1
		Observed $y = 56$ (or 44 above)	
		Accept H_0 Data are consistent with a distribution with median = 45. No evidence that median is not 45	B1

6	**(i)**	$X \sim N(10, 9)$ approx Correct mean Correct variance *their* mean + 1.645 × *their* sd (with or without continuity correction) Critical value = 10 + 1.645 × 3 + 0.5 = 14.935 + 0.5 = 15.435 Critical value = 16 If the number observed is 15 or fewer, accept H_0 and conclude that p may be 0.10 If number observed is 16 or more, reject H_0 and conclude that p is probably greater than 0.10 P(Type I error) = P(reject H_0 when it is true) (ft their integer cv of 16 or 15) = $P(X \geqslant 16)$ in B(100, 0.10) = $P(X \geqslant 15.5)$ in N(10, 9) approx. = $P(Z \geqslant (15.5 - 10)/3) = P(Z \geqslant 1.833) = 1 - 0.9666$ = 0.0334	 M1 A1 M1 A1 B1 B1 M1 A1
	(ii)	P(Type II error) = P(accept H_0 when it is false) = $P(X \leqslant 15)$ in B(100, 0.20) B(100, 0.20) or N(20, 16) used = $P(X \leqslant 15.5)$ in N(20, 16) approx. ft *their* integer cv of 16 or 15 = $P(Z \leqslant (15.5 - 20)/4) = P(Z \leqslant -1.125) = 1 - 0.8696$ = 0.1304	B1 M1 A1
	(iii)	$P \sim N(0.14, \frac{0.14 \times 0.86}{100})$ = N(0.14, 0.001204) approx. Correct method for *their* distribution 95% CI = $0.14 \pm 1.96\sqrt{0.001204}$ = 0.14 ± 0.068 = [0.072, 0.208] Correct interval, in any form, with or without an attempt at continuity 0.10 and 0.20 are both in this interval	B1 M1 A1 B1

BLANK PAGE