UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
Pre-U Certificate

MARK SCHEME for the May/June 2012 question paper

for the guidance of teachers

1347 MATHEMATICS (STATISTICS WITH PURE MATHEMATICS)

1347/02 Paper 2 (Statistics), maximum raw mark 80

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 must be read in conjunction with the question papers and the report on the examination.

- Cambridge will not enter into discussions or correspondence in connection with these mark schemes.

Cambridge is publishing the mark schemes for the May/June 2012 question papers for most IGCSE, Pre-U, GCE Advanced Level and Advanced Subsidiary Level syllabuses and some Ordinary Level syllabuses.

UNIVERSITY of CAMBRIDGE
International Examinations

Note: since there were no candidates this session, this mark scheme is a draft, and has not been modified in light of candidates' responses.

1 (i)	$S_{xx} = 1939552 - \frac{(4412)^2}{13} = 442187$	B1	442187 to nearest integer
	$S_{yy} = 605147 - \frac{(2387)^2}{13} = 166857$	B1	166857 to nearest integer
	$S_{xy} = 1074848 - \frac{4412 \times 2387}{13} = 264737$	B1	264737 to nearest integer
	$r = \frac{264737}{\sqrt{442187 \times 166857}} = 0.975\ (0.9746)$	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 near 1, so a good fit to an upward sloping line	A1 **[5]**	Drawing a valid conclusion (confirming that a linear fit is appropriate, as stated in question)
(ii)	$b = \frac{264737}{442187} = 0.599\ (0.5987)$	M1	Calculating b from their S_{xx}, S_{xy}
	$a = \frac{2387}{13} - 0.5987 \times \frac{4412}{13}$		
	$= 183.6 - 0.5987 \times 339.4 = -19.6$	M1	Calculating a from $\sum x$, $\sum y$ and their b
	$y = 0.599x - 19.6$	A1	Line correct with coefficients to 3sf or better
	$x = 2203 \Rightarrow \hat{y} = 1300$	B1 **[4]**	From their line ($\pm$ 2)
(iii)	Extrapolation beyond range of data	B1	Extrapolation
	Small sample / only based on one sample Sampling method not known / not random sampling London is not typical / London 'is different'	B1 **[2]**	Any valid objection

2 (i)	Median = 30 mpg Quartiles = 34 mpg and 23 mpg IQR = 11 mpg Outliers have mpg < 6.5 or > 50.5 ⇒ Toyota Prius	B1 B1 M1 A1 B1 **[5]**	30 cao Accept 33 to 35 and 20 to 24 Their IQR calculated Fences calculated for their quartiles Identified by name in any way (follow through their fences to at most three outliers)
(ii)	The difference between 23 and 30 is much greater than the difference between 30 and 34, this suggests that the distribution is not symmetric	B1 **[1]**	Using median and quartile values appropriately to deduce non-normal
(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 $\Sigma d^2 = 96$ $r_s = 1 - \frac{6 \times 96}{15 \times 224} = 1 - 0.17143 = 0.82857$ = 0.829 (3 sf)	M1 A1 M1 A1 **[4]**	Substantially correct calculation of d or $\|d\|$ or d^2 for the ranks Correct calculation of r_s for their Σd^2 Correct value, to 3 sf or better

3 (i)	Independence between children, class are typical of population in respect of left-handedness	B1 B1 **[2]**	Independence (random sample) Probability 13% (constant probability)
(ii)	X = number of left-handers $X \sim \text{B}(20, 0.13)$ 13% of 20 = 2.6, so want $\text{P}(X \le 2)$ $(0.87)^{20} + 20(0.13)(0.87)^{19} +$ $190(0.13)^2(0.87)^{18}$ = 0.061714 + 0.18443 + 0.26181 = 0.50795... = 0.508 to 3sf	 B1 M1 A1 A1 **[4]**	 $\text{P}(X \le 2)$ is needed Calculating a probability in B(20, 0.13) (At least) three correct probabilities added 0.508 or better
(iii)	$X \sim \text{B}(20, p)$ p = P(left-hander) H_0: $p = 0.13$ H_1: $p > 0.13$ $\alpha = 5\%$ one-tailed test Assuming H_0, $X \sim \text{B}(20, 0.13)$ $\text{P}(X \ge 7) = 1 - 0.9897 = 0.0103$ 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	 B1 B1 M1 M1 A1 B1 **[6]**	May imply definition of p Null hypothesis $p = 0.13$ Alternative hypothesis $p > 0.13$ Omission of p only penalised once May imply level of test and one-tailed Calculating $\text{P}(X \le 7)$ or critical value (cv = 7 for a two-sided H_1) Compare with 5% (or $2\frac{1}{2}$% for a two-sided H_1) or compare cv with observed value 7 Reject H_0, from correct calculations Correct conclusion in context
(iv)	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 **[1]**	Any valid reason, either from context (reasons why sample from past data may be different from sample from current data) or addressing statistical variation (random fluctuation)

4 (i)	$P(Z > z) = 0.01 \Rightarrow z = 2.326$ $P(Z < z) = 0.25 \Rightarrow z = -0.674$ $2.326 = \frac{120-\mu}{\sigma} \Rightarrow 120 - \mu = 2.326\sigma$ $-0.674 = \frac{84-\mu}{\sigma} \Rightarrow 84 - \mu = -0.674\sigma$ $\Rightarrow \mu = 92.1, \sigma = 12$	B1 M1 A1 B1 B1 **[5]**	2.326 and 0.674 from tables Substantially correct method (either) Both correct for their z-values, one of which is positive and one negative 92.1 or 92 (cao) 12 (cao)
(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) 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 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 **[5]**	Appropriate statement of hypotheses Attempt to sum ranks (27 or 109) $W = 27$ from correct working Critical value 35 Correct conclusion, in context, from use of Wilcoxon rank-sum
(iii)	For the smokers, $\sum x = 708 \Rightarrow \bar{x} = 118$ Estimate $\widehat{\mu_s} = 118$ $\sum x^2 = 83864 \Rightarrow S_{xx} = 320 \Rightarrow s^2 = 64$ Estimate $\widehat{\sigma_s^2} = 64$	B1 M1 A1 **[3]**	Mean 118 cao Sight of one of 83864, 320, 64, 8, 53.3 or 7.30 Variance 64 cao
(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 Confidence interval is $118 \pm 2.571 \times \frac{8}{\sqrt{6}}$ $= 118 \pm 8.4 = [109.6, 126.4]$	B1 M1 A1 **[3]**	Using t tables to find 2.571 or 2.447 Correct method for their "t" value and their $\bar{x}, \hat{\sigma}$ Interval correct, in any appropriate form (follow through their values from part (iii))

5 (i)	$x = 47 \rightarrow z = 0.667$; $x = 51 \rightarrow z = 2.0$ $P(47 < X \le 51) = 0.9772 - 0.7477$ Expected frequency = 0.2295 × 100 = 22.95	B1 M1 A1 **[3]**	$z = 2$ and 0.667 seen or implied 0.9772 and 0.7477 seen or implied Subtract and multiply by 100 (22.95 given in question)
(ii)	Merge classes in tails to make expected frequencies at least 5 Weight <43 43-45 45-47 >47 Obs freq 32 24 30 14 Exp freq 25.23 24.77 24.77 25.23 $(O-E)^2/E$ 1.82 0.02 1.10 5.00 X^2 calc = 7.94 H_0: N(45, 9) distribution H_1: some other distribution From tables, critical value = 7.815 Reject H_0 Data is not consistent with a N(45, 9) distribution	B1 M1 A1 B1 B1 B1 **[6]**	Merging tails correctly Substantially correct calculation of X^2 (with or without merging) 7.94, or art 7.94, from correct method , cao Any correct statement of H_0 Accept 'normal distribution' 7.815, from tables, cao Correct conclusion, and interpretation in words, for their X^2 value
(iii)	$v = n - 1 = 4 - 1 = 3$ No need to reduce df for parameters as not estimated from sample data	B1 **[1]**	4 classes – 1 restriction (total), or equivalent
(iv)	**(a)** Variance cannot be estimated, midpoints cannot be found for first and last classes since boundaries are not known	B1 **[1]**	Variance unknown since grouped data. Estimate of variance unlikely to be very accurate.
	(b) Sign test H_0: median = 45 $\alpha = 5\%$ H_1: median ≠ 45 two-tailed test Y = number of chicks with weight ≤ 45g Assuming H_0, $Y \sim B(100, 0.5)$ Approximate by N(50, 25) Critical values are $50 \pm 1.96 \times 5 = 50 \pm 9.8$ = [40.2, 59.8] 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 M1 A1 B1 **[4]**	Sign test or binomial test or equivalent (e.g. test proportion that are below 45) Or using proportions An appropriate distribution or approximating distribution Critical values correct (or with proportions) Or test observed proportions (or values) to give a tail probability of 0.115 (or 0.136) Accept H_0, follow through their critical values from substantially correct method , or interpretation in words

6 (i)	$X \sim N(10, 9)$ approx 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) = $P(X \geq 16)$ in B(100, 0.10) = $P(X \geq 15.5)$ in N(10, 9) approx = $P(Z \geq (15.5 - 10)/3) = P(Z \geq 1.833) = 1 - 0.9666$ = 0.0334	M1 A1 M1 A1 B1 B1 M1 A1 **[8]**	Correct mean Correct variance Their mean + 1.645 × their sd (with or without continuity correction) 14.5 to 15.5, from correct working or giving cv as 16 or 15 Correct description of accept or reject from their critical value of 16 or 15 May be worded in terms of x < their critical value in binomial or normal Understanding what a Type I error is Allow for P(Type I error) = 5% Follow through their integer cv of 16 or 15 $P(X \geq 15.5)$ or $P(X \geq 16)$ in their N(10, 9) 0.03 to 0.035 or 3% to 3.5% Note: cv = 15 gives 0.0668 (0.065 to 0.07)
(ii)	P(Type II error) = P(accept H_0 when it is false) = $P(X \leq 15)$ in B(100, 0.20) = $P(X \leq 15.5)$ in N(20, 16) approx = $P(Z \leq (15.5 - 20)/4) = P(Z \leq -1.125) = 1 - 0.8696$ = 0.1304	B1 M1 A1 **[3]**	Understanding what a Type II error is B(100, 0.20) or N(20, 16) used Follow through their integer cv of 16 or 15 0.13 to 0.135 or 13% to 13.5% Note: cv = 15 gives 0.0845 (0.08 to 0.085)
(iii)	$P \sim N(0.14, \frac{0.14 \times 0.86}{100})$ = N(0.14, 0.001204) approx 95% CI = $0.14 \pm 1.96\sqrt{0.001204}$ = 0.14 ± 0.068 = [0.072, 0.208] 0.10 and 0.20 are both in this interval	B1 M1 A1 B1 **[4]**	Mean 0.14 and variance $\frac{0.14 \times 0.86}{100}$ Correct method for their distribution Correct interval, in any form, with or without an attempt at continuity This statement, or equivalent, provided true