UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
Cambridge International Level 3 Pre-U Certificate
Short Course

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

Paper 2 Statistics **May/June 2013**

2 hours

7382175073

Additional Materials: Answer Booklet/Paper
Graph Paper
List of Formulae (MF21)

READ THESE INSTRUCTIONS FIRST

If you have been given an Answer Booklet, follow the instructions on the front cover of the Booklet.
Write your Centre number, candidate number and name on all the work you hand in.
Write in dark blue or black pen.
You may use a soft pencil for any diagrams or graphs.
Do not use staples, paper clips, highlighters, glue or correction fluid.

Answer **all** the questions.
Give non-exact numerical answers correct to 3 significant figures, or 1 decimal place in the case of angles in degrees, unless a different level of accuracy is specified in the question.
The use of an electronic calculator is expected, where appropriate.
You are reminded of the need for clear presentation in your answers.

At the end of the examination, fasten all your work securely together.
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 80.

This document consists of **5** printed pages and **3** blank pages.

JC13 06_1347_02/4R

UNIVERSITY of CAMBRIDGE International Examinations

1 The table shows the total population of the UK (in millions) every ten years from 1901 to 2001 and also the increase over each ten-year period. For example, the increase from 1901 to 1911 was 3.81 million.

Year	**1901**	**1911**	**1921**	**1931**	**1941**	**1951**	**1961**	**1971**	**1981**	**1991**	**2001**
Population (millions)	38.33	42.14	44.07	46.07	48.22	50.29	52.81	55.93	56.35	57.81	59.01
Increase (millions)	3.81	1.93	2.00	2.15	2.07	2.52	3.12	0.42	1.46	1.20	

(i) Calculate the mean and median of the ten **increases**, and comment on what these two values tell you about the distribution of increases. [3]

When the population increase from 2001 to 2011 is included, the summary statistics for the increases (in millions) are as follows.

minimum = 0.42, lower quartile = 1.35, median = 2.00, upper quartile = 2.52, maximum = 3.81

(ii) Use these values to show that the increase from 1901 to 1911 was not an outlier. [4]

(iii) By considering the trend of the increases, explain why using the mean of the ten increases in the table does not give a good estimate of the increase between 2001 and 2011. [1]

2 Josh has a set of five cards, numbered 1, 2, 3, 4 and 5. He picks two of the cards at random. The order in which the cards are picked does not matter. He calculates the total, X, of the numbers on the two cards.

(i) List all the possible pairs of cards and hence construct a table to show the probability distribution for X. [5]

(ii) Calculate the expectation, variance and standard deviation of X. [5]

3 The table shows some of the results from a large national birdwatch survey.

	% of gardens where seen	Average number per garden
House Sparrow	65	4.16
Starling	52	3.91
Blackbird	95	3.27
Blue Tit	87	3.16
Chaffinch	57	2.35
Woodpigeon	68	1.93
Great Tit	61	1.56
Goldfinch	34	1.51
Robin	87	1.46
Collared Dove	54	1.34

Source: http://www.rspb.org.uk/birdwatch/

Jack, Emily, Chloe and Oliver took part in the survey by watching the birds in their gardens.

(i) Use an appropriate binomial distribution to calculate the probability that robins were seen in fewer than two of these four gardens. [3]

(ii) State what must be assumed for a binomial model to be valid in this situation. [2]

It has been suggested that the probability of seeing house sparrows is lower in city gardens. 20 city gardens were chosen at random from the responses to the survey. House sparrows were recorded in only 5 of these 20 gardens.

(iii) Use this information to test the hypothesis that the probability of seeing house sparrows in city gardens is less than 65%. Use a 1% level of significance for your test. [7]

4 The males of a certain species of insect 'chirp' by rubbing their wings together. It has been observed that the rate of chirping increases with temperature.

Louise recorded the temperature and the number of chirps in 3 seconds on each of five randomly chosen hot summer evenings.

Temperature (°C), x	31	22	34	29	27
Number of chirps in 3 seconds, y	10	7	11	9	9

$$n = 5 \quad \Sigma x = 143 \quad \Sigma x^2 = 4171 \quad \Sigma y = 46 \quad \Sigma y^2 = 432 \quad \Sigma xy = 1342$$

(i) Calculate the values of S_{xx}, S_{yy} and S_{xy}, and hence show that a linear model fits Louise's data well. [5]

(ii) Calculate the equation of the regression line of y on x, and use it to predict the number of chirps in 3 seconds on a night when the temperature is 20°C. Give two statistical reasons why this estimate may be unreliable. [7]

5 Reaction times, in seconds, for an experiment are known to follow a normal distribution with variance 0.001. It has been stated that the mean reaction time is 0.2. Sam claims that the mean is greater than this and wants to test his claim. He carries out the experiment once and records the reaction time.

(i) State null and alternative hypotheses for Sam's test. [2]

(ii) Calculate the critical reaction time for Sam's test at the 5% level of significance, and explain how Sam should use this and his recorded reaction time to reach a conclusion. [3]

It is later found that the mean reaction time has been rounded, and that the true value is 0.24.

(iii) Calculate the probability that Sam's test results in a Type II error. [3]

6 As part of her AS Psychology course, Sarah carried out an experiment in which pictures of five men and five women were shown to a group of ten volunteers. The volunteers were asked to rate the 'attractiveness' of the men and women in the pictures, using a score of 1, 2 or 3, where the higher the score the more attractive the person was rated.

The men and women in the pictures were actually five married couples. Sarah used a spreadsheet to calculate the chi-squared statistic to test whether there is any association between the scores given to the men and their wives. The spreadsheet printout is given below.

				Woman			
		OBS	1	2	3		
		1	13	3	0	16	
	Man	2	4	10	4	18	
		3	0	3	13	16	
			17	16	17	50	
		EXP	1	2	3		
		1	5.44	5.12	5.44	16	
		2	6.12	5.76	6.12	18	
		3	5.44	5.12	5.44	16	
			17	16	17	50	
		CHI SQ	1	2	3		
		1	10.51	0.88	5.44	16.83	
		2	0.73	3.12	0.73	4.58	
		3	5.44	0.88	10.51	16.83	
			16.68	4.88	16.68	38.24	

(i) State the null and alternative hypotheses for the test. [1]

(ii) Show how the value 5.76 for the expected frequency of both the man and the woman having a score of 2 is calculated. [2]

(iii) Show how the value 3.12 for the contribution to the chi-squared total from the cell where the man and woman both have a score of 2 is calculated. [2]

(iv) Carry out the test at the 1% level of significance and state the conclusion. [4]

7 Ten people took part in a test of a new diet. Seven of the people (N1 to N7) were randomly chosen to follow the new diet and the other three (E8 to E10) followed an established diet. The people did not know which diet they were using. The table shows the weight, in kg, of the ten people before and after their diets, and their weight losses.

Person	N1	N2	N3	N4	N5	N6	N7	E8	E9	E10
Before	76	85	89	98	102	114	127	68	90	146
After	68	70	72	88	81	90	95	62	71	100
Weight loss	8	15	17	10	21	24	32	6	19	46

By giving the smaller weight losses the lower ranks, carry out a Wilcoxon rank-sum test, at the 5% level, to test the null hypothesis that the weight losses on the two diets follow the same distribution against the alternative that they are different. [7]

8 At a certain place on the east coast of England, the coastline is being eroded. The annual erosion, X centimetres, was measured in 16 randomly chosen years. The results can be summarised as follows.

$$n = 16 \qquad \Sigma x = 2144 \qquad \Sigma(x - \bar{x})^2 = 8640$$

(i) Use these values to calculate unbiased estimates of the population mean and variance of X. [3]

It may be assumed that X has the distribution $\mathrm{N}(\mu, \sigma^2)$.

(ii) Explain why it is appropriate to use a t-distribution to calculate a confidence interval for the population mean μ. [2]

(iii) Calculate an approximate 95% confidence interval for μ. [5]

A published report claims that the actual distribution of X is $\mathrm{N}(145, 600)$.

(iv) Assuming that this claim is correct, and continues to be correct, state the distribution of the sample mean annual erosion, $\overline{X}$ centimetres, over the next 6 years. Hence calculate the probability that the mean annual erosion over the next 6 years is less than 125.0 centimetres. [4]

BLANK PAGE

BLANK PAGE

BLANK PAGE