Cambridge Pre-U

Cambridge International Examinations
Cambridge Pre-U Certificate

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

Paper 2 Statistics **May/June 2017**

2 hours

5076125488

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 an HB pencil for any diagrams or graphs.
Do not use staples, paper clips, 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.

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

1 The table shows median weekly earnings (in £) for full-time employees, average house prices (in £000's), and the ratio of average house price to annual earnings, for the 12 regions of the UK.

UK REGION	Earnings, x	House prices, y	Ratio
Greater London	652	402	11.86
South East	531	260	9.42
East Anglia	495	191	7.42
Scotland	488	147	5.79
West Midlands	472	170	6.93
Yorks and Humber	466	150	6.19
East Midlands	464	160	6.63
South West	460	220	9.20
North West	460	150	6.27
Wales	454	146	6.18
North East	452	136	5.79
Northern Ireland	450	136	5.81

$n = 12$ $\Sigma x = 5844$ $\Sigma x^2 = 2\,881\,590$ $\Sigma y = 2268$ $\Sigma y^2 = 493\,502$ $\Sigma xy = 1\,149\,981$

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

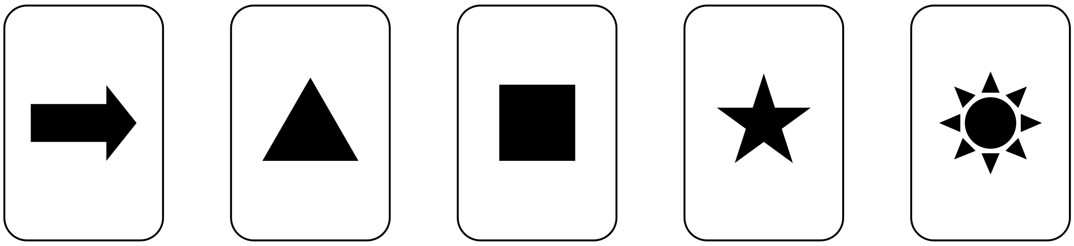
(ii) Calculate the equation of the regression line of y on x. [3]

(iii) Use your regression line to comment on house prices in relation to earnings in

(a) Greater London, [2]

(b) the South West. [2]

2 Joanne is testing whether she can make people believe in psychic powers. She carries out an experiment using pairs of volunteers and a set of five cards.



One pair of volunteers are Alex and Billie. Joanne asks Alex to pick a card at random, and concentrate on it, while Billie tries to guess which card was chosen. Alex then has to say whether the guess was right or not. This is repeated to give ten results.

(i) Use an appropriate binomial distribution to calculate the probability that Billie will make four or more correct guesses just by chance. [3]

Joanne wants to test whether Billie is more successful than could be accounted for by chance.

(ii) State the null and alternative hypotheses for the test. [2]

(iii) Find the critical value for this test at the 1% significance level, showing the values of any relevant probabilities. [4]

Unknown to Billie, Alex decides to say that the guess is correct not only when it is correct, but also on alternate occasions when it is wrong.

(iv) Find the probability that Alex says that a guess is correct when Billie is guessing entirely by chance. Hence calculate the probability of a Type II error in the hypothesis test in part **(iii)**. State what this probability represents in the context of the experiment. [4]

3 The table shows the lengths of the four longest rivers in each of five regions, together with the total of the four lengths in each case.

AFRICA	**km**
Nile	6650
Congo	4700
Niger	4200
Zambezi	2693
TOTAL	**18243**

AMERICAS	**km**
Amazon	6400
Mississippi	6275
Rio de la Plata	4880
Mackenzie	4240
TOTAL	**21795**

ASIA	**km**
Yangtze	6300
Yellow	5464
Mekong	4910
Lena	4400
TOTAL	**21074**

AUSTRALASIA	**km**
Murray	2374
Murrumbidgee	1485
Darling	1472
Lachlan	1340
TOTAL	**6671**

EUROPE	**km**
Volga	3692
Danube	2860
Ural	2428
Dneiper	2290
TOTAL	**11270**

(i) For these twenty lengths, calculate

(a) the mean, [2]

(b) the median, [2]

(c) the interquartile range. [3]

When the lengths of the 100 longest rivers in the world are used, the upper quartile is 3058 km and the lower quartile is 1625 km.

(ii) By identifying any outliers, find which of the rivers listed above are exceptionally long amongst the 100 longest rivers in the world. [4]

Amongst the 100 longest rivers in the world, nine of the rivers that are longer than 3058 km (the upper quartile) are in the Americas.

(iii) By considering the number of rivers that are longer than 3058 km in Africa, Australasia and Europe, find how many of the rivers that are longer than 3058 km are in Asia. [2]

4 Cuckoos put their eggs in the nests of other birds and expect the host bird to incubate the eggs. Sometimes the host birds recognise the cuckoo eggs and reject them.

Some students investigated the rejection of cuckoo eggs by host birds. The local bird sanctuary let the students place 100 cuckoo eggs in some nests and the students recorded whether or not each host bird rejected the cuckoo egg.

The table records the results.

Species of host bird	Number of cuckoo eggs placed by students	Number of cuckoo eggs rejected by host bird
Warbler	30	19
Thrush	30	16
Blackbird	20	10
Wagtail	20	15

Let p denote the probability that a cuckoo egg is rejected by the host bird. It has been claimed that this probability is independent of the species of the host bird.

(i) Use the data in the table to estimate a value for p. Hence calculate the number of cuckoo eggs that you would expect to be rejected by each species of host bird, assuming the claim to be true. [3]

(ii) Carry out a chi-squared goodness of fit test to test the claim at the 10% level of significance. [7]

5 Cambridge Intermittent Exercisers entered six runners for each of two races. One race took place in January and the other in February. The finishing times, in minutes, were as follows.

January	44	45	59	62	63	67
February	38	39	43	48	51	52

You may assume that the times are independent and form a random sample of their population, and that the population distributions are symmetric and have the same shape.

(i) Rank the 12 times, in increasing order, and assign the label J or F to each, according to whether the time was achieved in the January race or the February race. Use a Wilcoxon rank sum test, at the 5% level, to test whether the population medians are the same. [7]

(ii) Suppose that 120 times had been available instead of 12 (60 from each race instead of 6). Calculate the critical value for a Wilcoxon rank sum test, at the 5% level, to test whether the population medians are the same. Describe how this value would be used to decide whether to accept or reject H_0 once a W value had been calculated. [6]

The individual times achieved in each race by the six runners from Cambridge Intermittent Exercisers are given in the following table.

Runner	Time in January (mins)	Time in February (mins)
Amir	67	52
Bakari	45	48
Chantal	63	43
Dante	44	51
Elsa	62	39
Fatima	59	38

(iii) Calculate the six differences (January time minus February time) and carry out a Wilcoxon signed rank test, at the 5% level, to test whether the times have improved from January to February. [6]

6 The size, X, of a primary school class in Wales was recorded for a random sample of 100 classes. The results can be summarised as follows.

$$n = 100 \qquad \Sigma x = 3000 \qquad \Sigma(x - \bar{x})^2 = 1584$$

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

(ii) Explain why choosing one class from each of 100 randomly chosen primary schools in Wales would not generate a random sample of primary school classes in Wales. [1]

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

(iii) Calculate an approximate 90% confidence interval for μ. Explain why it is **not** necessary to use a t-distribution. [5]

The size, Y, of a primary school class in Anglesey can be modelled using a normal distribution. For a random sample of size 5, it is found that an unbiased estimate of the population variance is 27.5.

(iv) Use a suitable t-distribution to show that the probability that the sample mean is more than 6.5 greater than the population mean, for primary school classes in Anglesey, is approximately 0.025. [4]

BLANK PAGE