Cambridge International Examinations
Cambridge Pre-U Certificate

Cambridge Pre-U

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

Paper 2 Statistics   **For Examination from 2016**

SPECIMEN PAPER

2 hours

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 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.

DO **NOT** WRITE IN ANY BARCODES.

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 **6** printed pages and **2** blank pages.

                                                       [Turn over

1 The table below shows a sample from the Home Office statistics on the breaching of anti-social behaviour orders (ASBOs), in 13 police areas, between 1 June 2000 and 31 December 2009.

Police area	ASBOs issued, x	ASBOs breached, y
Avon and Somerset	350	187
Bedfordshire	176	73
Cambridgeshire	196	86
Derbyshire	250	131
Essex	287	169
Gloucestershire	132	80
Hampshire and Isle of Wight	551	347
Lancashire	675	428
Merseyside	712	357
Norfolk	229	112
South Wales	360	160
Thames Valley	334	168
Warwickshire	160	89

Source: http://www.homeoffice.gov.uk/

$n = 13$, $\Sigma x = 4412$, $\Sigma x^2 = 1\,939\,552$, $\Sigma y = 2387$, $\Sigma y^2 = 605\,147$, $\Sigma xy = 1\,074\,848$.

It is believed that there is a linear relationship between the number of ASBOs issued and the number breached.

(i) Calculate the values of S_{xx}, S_{yy} and S_{xy}, giving each correct to the nearest integer, and hence show that a linear model is appropriate for the data. [5]

(ii) Calculate the equation of the regression line of y on x, and use it to predict the number of ASBOs breached in Greater London where there were 2203 ASBOs issued between 1 June 2000 and 31 December 2009. [4]

(iii) The actual number of ASBOs breached in Greater London was 1066. Give two distinct reasons that could explain why the prediction was unreliable. [2]

2 The table shows fuel economy figures in miles per gallon (mpg) for some new cars.

Make and model	City (mpg)	Rank	Motorway (mpg)	Rank
Toyota Prius	54	1	48	1
Honda Insight	40	2	43	2
Honda CR-Z	35	3	31	12
Volkswagen Golf	34	4	42	3
Audi A3	33	5	41	4
Volkswagen Jetta	32	6	40	5
Ford Fiesta SFE FWD	31	7	39	6
Mini Cooper	30	8	37	7
Toyota Yaris	27	9	36	8
Nissan Versa	26	10	32	11
Hyundai Sonata	24	11	35	9
Honda Accord	23	12	34	10
Kia Rondo	20	13	27	13
Subaru Impreza	17	14	23	14
Aston Martin DB7	11	15	17	15

Source: http://www.fueleconomy.gov (adapted)

(i) Find the median and quartiles for the mpg of these fifteen cars under city driving conditions and use these values to identify any cars for which the mpg under city driving conditions is an outlier. [5]

(ii) Explain why the calculations in part **(i)** suggest that the mpg values do not come from a normal distribution. [1]

(iii) Calculate the value of Spearman's rank correlation coefficient between the mpgs for these cars under the two driving conditions. [4]

3 A survey into left-handedness found that 13% of the population of the world is left-handed.

   (i) State the assumptions necessary for it to be appropriate to model the number of left-handed children in a class of 20 children using the binomial distribution B(20, 0.13). [2]

   (ii) Assuming that this binomial model is appropriate, calculate the probability that fewer than 13% of the 20 children are left-handed. [4]

A leading academic claims that there are significantly more left-handed people in high-status jobs than would be expected. Seven of the twenty most recent presidents of the United States of America have been left-handed.

   (iii) Test whether this result supports the academic's claim, assuming that a binomial model is valid. Use a 5% level of significance for the test and use values from the statistical table below. [6]

Only one of the first twenty presidents of the United States of America was left-handed.

   (iv) Give a reason that could explain why the number of left-handed presidents in the first twenty is much lower than for the most recent twenty. [1]

CUMULATIVE BINOMIAL PROBABILITIES

$n = 20$	$p = 0.13$
$x = 4$	0.8917
5	0.9630
6	0.9897
7	0.9976
8	0.9995

4 When people exercise, their heart rate increases. Suppose that, for the population as a whole, increases in heart rate follow a normal distribution, $N(\mu, \sigma^2)$. You are given that only 1% of the population have an increase that is more than 120, while 25% of the population have an increase that is less than 84.

(i) Calculate the values of μ and σ that are consistent with this information. [5]

An investigation was carried out into the effects of smoking on health. The heart rates of 16 volunteers were measured before and after exercise, and the increases in heart rate were recorded. After the data had been collected the volunteers revealed whether they were smokers or not.

The results are given below.

Volunteer number	Age	Male or female	Increase in heart rate	Smoker or not
1	30	F	132	Smoker
2	18	F	122	Not
3	20	M	120	Smoker
4	18	M	118	Smoker
5	17	F	116	Smoker
6	17	F	114	Smoker
7	27	M	110	Not
8	29	M	108	Smoker
9	20	M	106	Not
10	26	F	98	Not
11	25	M	88	Not
12	32	M	84	Not
13	29	F	80	Not
14	17	F	72	Not
15	31	F	64	Not
16	21	M	60	Not

(ii) The volunteers have been ranked by increase in heart rate, with the largest increase having rank 1. Carry out a Wilcoxon rank sum test, at the 5% level, to test whether the increases in heart rate are greater for smokers than for non-smokers. [5]

Now consider just the population of all smokers.

(iii) Calculate unbiased estimates of the mean and variance increase in heart rate for the population of all smokers. [3]

(iv) Assuming that increases in heart rate for the population of all smokers follow a normal distribution, $N(\mu_s, \sigma_s^2)$, calculate an approximate 95% confidence interval for μ_s. [3]

5 The table shows the weights, x grams, of 100 cockatiel chicks two weeks after they were hatched.

x	$x \leqslant 39$	$39 < x \leqslant 43$	$43 < x \leqslant 45$	$45 < x \leqslant 47$	$47 < x \leqslant 51$	$51 < x$
Number of chicks	5	27	24	30	12	2

It has been claimed that the weights follow a normal distribution with mean 45 and variance 9.

The expected frequencies from this distribution, correct to 2 decimal places, are as follows.

x	$x \leqslant 39$	$39 < x \leqslant 43$	$43 < x \leqslant 45$	$45 < x \leqslant 47$	$47 < x \leqslant 51$	$51 < x$
Expected frequency	2.28	22.95	24.77	24.77	22.95	2.28

   (i) Show how the expected frequency for $47 < x \leqslant 51$ was calculated. [3]

   (ii) Use a χ^2 goodness of fit test at the 5% level with 3 degrees of freedom to test whether the sample is consistent with a N(45, 9) distribution. [6]

   (iii) Explain why 3 degrees of freedom were used in the goodness of fit test. [1]

Suppose that it was known that the weights should follow a normal distribution. The sample is to be tested to find out if it is consistent with the distribution having a mean of 45.

   (iv) (a) Explain what difficulty would arise in attempting to carry out a χ^2 goodness of fit test. [1]

       (b) Name an alternative test that could be used, and carry out this test at the 5% level of significance. [4]

6 The authors of a report claim that the likelihood of a certain river flooding its banks in any year is 10%. The report gives the number of years over the last century in which the river has flooded.

Let X represent the number of years, out of one hundred, in which the river floods. The authors of the report assume that X can be modelled using the distribution B(100, p), where p is the probability that the river floods in a year.

A group of residents who live near the river say that the likelihood quoted in the report is too small. They want to test this at the 5% level using H$_0$: $p = 0.10$ against H$_1$: $p > 0.10$.

   (i) Use a normal approximation to the binomial distribution to find the critical value for the test. State how the residents should use this critical value to interpret the data. Calculate the probability that a Type I error occurs. [8]

It is later found that the likelihood of the river flooding should have been given as 20%.

   (ii) Calculate the probability that, when the residents carried out their test, a Type II error occurred. [3]

The data that the authors of the report made available showed that the river had flooded in 14 of the last hundred years.

   (iii) Use this result to calculate a 95% confidence interval for p and show that both 10% and 20% fall within the confidence interval. [4]

BLANK PAGE

BLANK PAGE