OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 4767

Statistics 2

Thursday **9 JUNE 2005** Morning 1 hour 30 minutes

Additional materials:
Answer booklet
Graph paper
MEI Examination Formulae and Tables (MF2)

TIME 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- Final answers should be given to a degree of accuracy appropriate to the context.
- The total number of marks for this paper is 72.

This question paper consists of 5 printed pages and 3 blank pages.

Registered Charity 1066969
[Turn over

1 A student is collecting data on traffic arriving at a motorway service station during weekday lunchtimes. The random variable X denotes the number of cars arriving in a randomly chosen period of ten seconds.

(i) State two assumptions necessary if a Poisson distribution is to provide a suitable model for the distribution of X. Comment briefly on whether these assumptions are likely to be valid. [4]

The student counts the number of arrivals, x, in each of 100 ten-second periods. The data are shown in the table below.

x	0	1	2	3	4	5	>5
Frequency, f	18	39	20	12	8	3	0

(ii) Show that the sample mean is 1.62 and calculate the sample variance. [3]

(iii) Do your calculations in part **(ii)** support the suggestion that a Poisson distribution is a suitable model for the distribution of X? Explain your answer. [1]

For the remainder of this question you should assume that X may be modelled by a Poisson distribution with mean 1.62.

(iv) Find $P(X = 2)$. Comment on your answer in relation to the data in the table. [4]

(v) Find the probability that at least ten cars arrive in a period of 50 seconds during weekday lunchtimes. [3]

(vi) Use a suitable approximating distribution to find the probability that no more than 550 cars arrive in a randomly chosen period of one hour during weekday lunchtimes. [4]

2 The fuel economy of a car varies from day to day according to weather and driving conditions. Fuel economy is measured in miles per gallon (mpg).

The fuel economy of a particular petrol-fuelled type of car is known to be Normally distributed with mean 38.5 mpg and standard deviation 4.0 mpg.

(i) Find the probability that on a randomly selected day the fuel economy of a car of this type will be above 45.0 mpg. [4]

(ii) The manufacturer wishes to quote a fuel economy figure which will be exceeded on 90% of days. What figure should be quoted? [3]

The daily fuel economy of a similar type of car which is diesel-fuelled is known to be Normally distributed with mean 51.2 mpg and unknown standard deviation σ mpg.

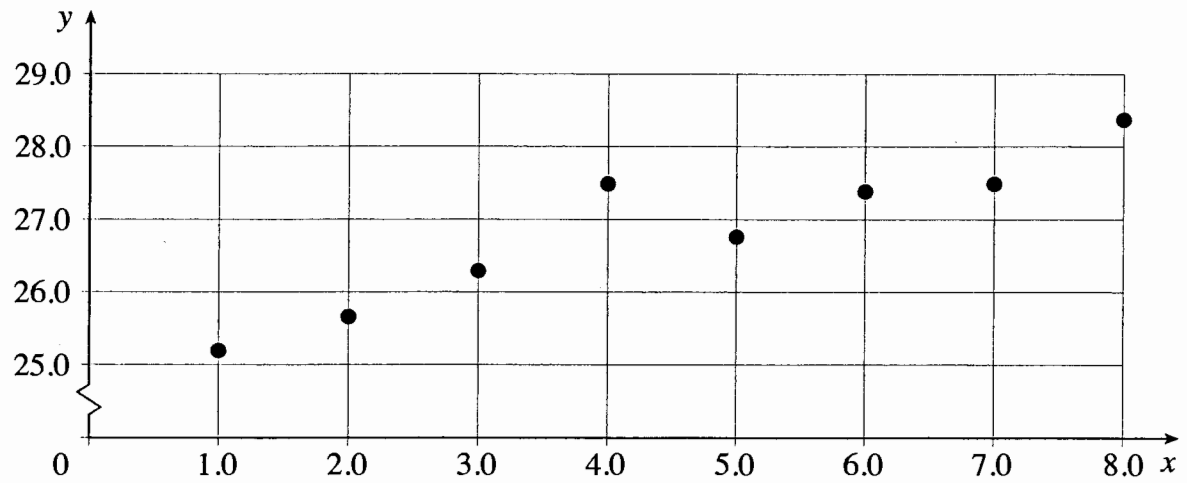
(iii) Given that on 75% of days the fuel economy of this type of car is below 55.0 mpg, show that $\sigma = 5.63$. [3]

(iv) Draw a sketch to illustrate both distributions on a single diagram. [4]

(v) Find the probability that the fuel economy of either the petrol or the diesel model (or both) will be above 45.0 mpg on a randomly selected day. You may assume that the fuel economies of the two models are independent. [4]

[Turn over

3 In a triathlon, competitors have to swim 600 metres, cycle 40 kilometres and run 10 kilometres. To improve her strength, a triathlete undertakes a training programme in which she carries weights in a rucksack whilst running. She runs a specific course and notes the total time taken for each run. Her coach is investigating the relationship between time taken and weight carried. The times taken with eight different weights are illustrated on the scatter diagram below, together with the summary statistics for these data. The variables x and y represent weight carried in kilograms and time taken in minutes respectively.



Summary statistics: $n = 8, \Sigma x = 36, \Sigma y = 214.8, \Sigma x^2 = 204, \Sigma y^2 = 5775.28, \Sigma xy = 983.6$.

(i) Calculate the equation of the regression line of y on x. [5]

On one of the eight runs, the triathlete was carrying 4 kilograms and took 27.5 minutes. On this run she was delayed when she tripped and fell over.

(ii) Calculate the value of the residual for this weight. [3]

(iii) The coach decides to recalculate the equation of the regression line without the data for this run. Would it be preferable to use this recalculated equation or the equation found in part **(i)** to estimate the delay when the triathlete tripped and fell over? Explain your answer. [2]

The triathlete's coach claims that there is positive correlation between cycling and swimming times in triathlons. The product moment correlation coefficient of the times of twenty randomly selected competitors in these two sections is 0.209.

(iv) Carry out a hypothesis test at the 5% level to examine the coach's claim, explaining your conclusions clearly. [5]

(v) What distributional assumption is necessary for this test to be valid? How can you use a scatter diagram to decide whether this assumption is likely to be true? [2]

4 **(a)** The selling prices of semi-detached houses in the suburbs of a particular city are known to be Normally distributed with mean £166 500 and standard deviation £14 200. A householder on one large estate claims that houses on her estate have a higher mean selling price. The selling prices of six randomly selected houses on her estate are

£180 000, £152 000, £156 500, £172 000, £189 000, £169 000.

(i) State suitable null and alternative hypotheses to test her claim. [2]

(ii) Carry out the test at the 5% level of significance, stating your conclusions clearly. You may assume that the standard deviation of the selling prices of houses on this estate is £14 200. [6]

(b) The manager of a restaurant undertakes a survey of the numbers and types of drinks ordered by a random sample of 400 customers. Customers are categorized as business, tourist or local. The drinks are categorized as alcoholic or soft drinks. A table of results of the survey is as follows.

		Type of drink		**Row totals**
		Alcoholic	Soft drinks	
Type of customer	Business	54	63	117
	Tourist	95	41	136
	Local	71	76	147
Column totals		220	180	400

Carry out a test at the 5% level of significance to examine whether there is any association between type of customer and type of drink. State carefully your null and alternative hypotheses.

[10]

OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 4767

Statistics 2

Wednesday **25 JANUARY 2006** Morning 1 hour 30 minutes

Additional materials:
8 page answer booklet
Graph paper
MEI Examination Formulae and Tables (MF2)

TIME 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is 72.

This question paper consists of 5 printed pages and 3 blank pages.

Registered Charity 1066969
[Turn over

1 A roller-coaster ride has a safety system to detect faults on the track.

(i) State conditions for a Poisson distribution to be a suitable model for the number of faults occurring on a randomly selected day. [2]

Faults are detected at an average rate of 0.15 per day. You may assume that a Poisson distribution is a suitable model.

(ii) Find the probability that on a randomly chosen day there are

(*A*) no faults,

(*B*) at least 2 faults. [4]

(iii) Find the probability that, in a randomly chosen period of 30 days, there are at most 3 faults. [3]

There is also a separate safety system to detect faults on the roller-coaster train itself. Faults are detected by this system at an average rate of 0.05 per day, independently of the faults detected on the track. You may assume that a Poisson distribution is also suitable for modelling the number of faults detected on the train.

(iv) State the distribution of the total number of faults detected by the two systems in a period of 10 days. Find the probability that a total of 5 faults is detected in a period of 10 days. [4]

(v) The roller-coaster is operational for 200 days each year. Use a suitable approximating distribution to find the probability that a total of at least 50 faults is detected in 200 days. [5]

2 The drug EPO (erythropoetin) is taken by some athletes to improve their performance. This drug is in fact banned and blood samples taken from athletes are tested to measure their 'hematocrit level'. If the level is over 50 it is considered that the athlete is likely to have taken EPO and the result is described as 'positive'. The measured hematocrit level of each athlete varies over time, even if EPO has not been taken.

(i) For each athlete in a large population of innocent athletes, the variation in measured hematocrit level is described by the Normal distribution with mean 42.0 and standard deviation 3.0.

(*A*) Show that the probability that such an athlete tests positive for EPO in a randomly chosen test is 0.0038. [3]

(*B*) Find the probability that such an athlete tests positive on at least 1 of the 7 occasions during the year when hematocrit level is measured. (These occasions are spread at random through the year and all test results are assumed to be independent.) [3]

(*C*) It is standard policy to apply a penalty after testing positive. Comment briefly on this policy in the light of your answer to part **(i)**(*B*). [2]

(ii) Suppose that 1000 tests are carried out on innocent athletes whose variation in measured hematocrit level is as described in part **(i)**. It may be assumed that the probability of a positive result in each test is 0.0038, independently of all other test results.

(*A*) State the exact distribution of the number of positive tests. [2]

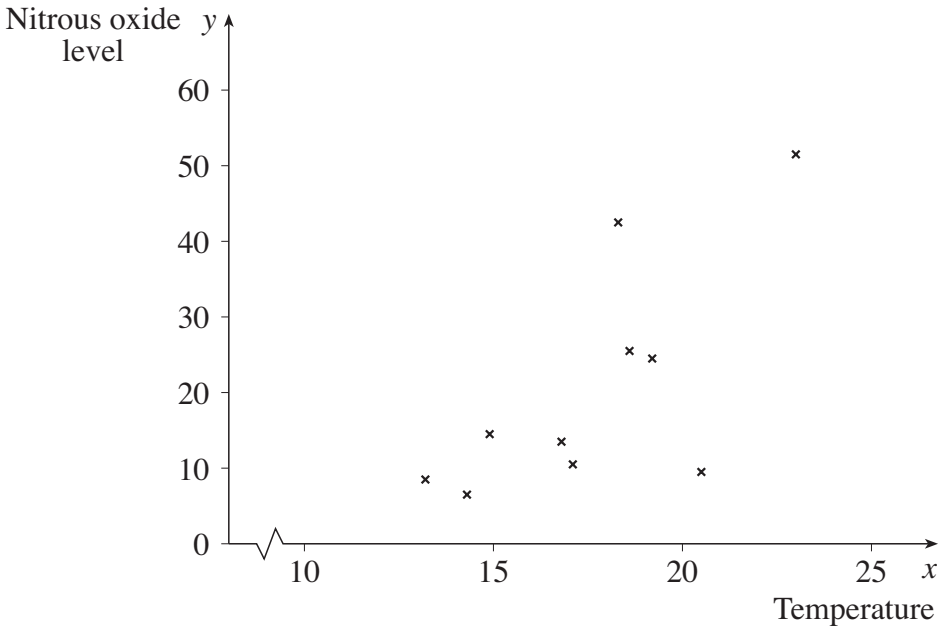
(*B*) Use a suitable approximating distribution to find the probability that at least 10 tests are positive. [4]

(iii) Because of genetic factors, a particular innocent athlete has an abnormally high natural hematocrit level. This athlete's measured level is Normally distributed with mean 48.0 and standard deviation 2.0. The usual limit of 50 for a positive test is to be altered for this athlete to a higher value h. Find the value of h for which this athlete would test positive on average just once in 200 occasions. [4]

[Turn over

3 A researcher is investigating the relationship between temperature and levels of the air pollutant nitrous oxide at a particular site. The researcher believes that there will be a positive correlation between the daily maximum temperature, x, and nitrous oxide level, y. Data are collected for 10 randomly selected days. The data, measured in suitable units, are given in the table and illustrated on the scatter diagram.

x	13.3	17.2	16.9	18.7	18.4	19.3	23.1	15.0	20.6	14.4
y	9	11	14	26	43	25	52	15	10	7



(i) Calculate the value of Spearman's rank correlation coefficient for these data. [5]

(ii) Perform a hypothesis test at the 5% level to check the researcher's belief, stating your hypotheses clearly. [5]

(iii) It is suggested that it would be preferable to carry out a test based on the product moment correlation coefficient. State the distributional assumption required for such a test to be valid. Explain how a scatter diagram can be used to check whether the distributional assumption is likely to be valid and comment on the validity in this case. [3]

(iv) A statistician investigates data over a much longer period and finds that the assumptions for the use of the product moment correlation coefficient are in fact valid. Give the critical region for the test at the 1% level, based on a sample of 60 days. [2]

(v) In a different research project, into the correlation between daily temperature and ozone pollution levels, a positive correlation is found. It is argued that this shows that high temperatures cause increased ozone levels. Comment on this claim. [3]

4 The table summarises the usual method of travelling to school for 200 randomly selected pupils from primary and secondary schools in a city.

		Primary	Secondary
Method of travel	Bus	21	49
	Car	65	15
	Cycle or Walk	34	16

(i) Write down null and alternative hypotheses for a test to examine whether there is any association between method of travel and type of school. [1]

(ii) Calculate the expected frequency for primary school bus users. Calculate also the corresponding contribution to the test statistic for the usual χ^2 test. [4]

(iii) Given that the value of the test statistic for the usual χ^2 test is 42.64, carry out the test at the 5% level of significance, stating your conclusion clearly. [4]

The mean travel time for pupils who travel by bus is known to be 18.3 minutes. A survey is carried out to determine whether the mean travel time to school by car is different from 18.3 minutes. In the survey, 20 pupils who travel by car are selected at random. Their mean travel time is found to be 22.4 minutes.

(iv) Assuming that car travel times are Normally distributed with standard deviation 8.0 minutes, carry out a test at the 10% level, stating your hypotheses and conclusion clearly. [7]

(v) Comment on the suggestion that pupils should use a bus if they want to get to school quickly. [2]

OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 4767

Statistics 2

Monday **22 MAY 2006** Morning 1 hour 30 minutes

Additional materials:
8 page answer booklet
Graph paper
MEI Examination Formulae and Tables (MF2)

TIME 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is 72.

This question paper consists of 3 printed pages and 1 blank page.

Registered Charity 1066969
[Turn over

1 A low-cost airline charges for breakfasts on its early morning flights. On average, 10% of passengers order breakfast.

(i) Find the probability that, out of 8 randomly selected passengers, exactly 1 orders breakfast. [2]

(ii) Use a suitable Poisson approximating distribution to find the probability that the number of breakfasts ordered by 30 randomly selected passengers is

(*A*) exactly 6,

(*B*) at least 8. [5]

(iii) State the conditions under which the use of a Poisson distribution is appropriate as an approximation to a binomial distribution. [2]

(iv) The aircraft carries 120 passengers and the flight is always full. Find the mean μ and variance σ^2 of a Normal approximating distribution suitable for modelling the total number of passengers on the flight who order breakfast. [2]

(v) Use your Normal approximating distribution to calculate the probability that more than 15 breakfasts are ordered on a particular flight. [3]

(vi) The airline wishes to be at least 99% certain that the plane will have sufficient breakfasts for all passengers who order them. Find the minimum number of breakfasts which should be carried on each flight. [4]

2 The head circumference of 3-year-old boys is known to be Normally distributed with mean 49.7 cm and standard deviation 1.6 cm.

(i) Find the probability that the head circumference of a randomly selected 3-year-old boy will be

(*A*) over 51.5 cm,

(*B*) between 48.0 and 51.5 cm. [5]

(ii) Four 3-year-old boys are selected at random. Find the probability that exactly one of them has head circumference between 48.0 and 51.5 cm. [3]

(iii) The head circumference of 3-year-old girls is known to be Normally distributed with mean μ and standard deviation σ. Given that 60% of 3-year-old girls have head circumference below 49.0 cm and 30% have head circumference below 47.5 cm, find the values of μ and σ. [4]

A nutritionist claims that boys who have been fed on a special organic diet will have a larger mean head circumference than other boys. A random sample of ten 3-year-old boys who have been fed on this organic diet is selected. It is found that their mean head circumference is 50.45 cm.

(iv) Using the null and alternative hypotheses H_0: $\mu = 49.7$ cm, H_1: $\mu > 49.7$ cm, carry out a test at the 10% significance level to examine the nutritionist's claim. Explain the meaning of μ in these hypotheses. You may assume that the standard deviation of the head circumference of organically fed 3-year-old boys is 1.6 cm. [6]

3 A student is investigating the relationship between the length x mm and circumference y mm of plums from a large crop. The student measures the dimensions of a random sample of 10 plums from this crop. Summary statistics for these dimensions are as follows.

$$\sum x = 4715 \quad \sum y = 13\,175 \quad \sum x^2 = 2\,237\,725$$

$$\sum y^2 = 17\,455\,825 \quad \sum xy = 6\,235\,575 \quad n = 10$$

(i) Calculate the sample product moment correlation coefficient. [5]

(ii) Carry out a hypothesis test at the 5% significance level to determine whether there is any correlation between length and circumference of plums from this crop. State your hypotheses clearly, defining any symbols which you use. [6]

(iii) (*A*) Explain the meaning of a 5% significance level. [2]

(*B*) State one advantage and one disadvantage of using a 1% significance level rather than a 5% significance level in a hypothesis test. [2]

The student decides to take another random sample of 10 plums. Using the same hypotheses as in part **(ii)**, the correlation coefficient for this second sample is significant at the 5% level. The student decides to ignore the first result and concludes that there is correlation between the length and circumference of plums in the crop.

(iv) Comment on the student's decision to ignore the first result. Suggest a better way in which the student could proceed. [3]

4 A survey of a random sample of 250 people is carried out. Their musical preferences are categorized as pop, classical or jazz. Their ages are categorized as under 25, 25 to 50, or over 50. The results are as follows.

		Musical preference			**Row totals**
		Pop	Classical	Jazz	
Age group	Under 25	57	15	12	**84**
	25 – 50	43	21	21	**85**
	Over 50	22	32	27	**81**
Column totals		**122**	**68**	**60**	**250**

(i) Carry out a test at the 5% significance level to examine whether there is any association between musical preference and age group. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. [12]

(ii) Discuss briefly how musical preferences vary between the age groups, as shown by the contributions to the test statistic. [6]

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE UNIT **4767/01**

MATHEMATICS (MEI)

Statistics 2

FRIDAY 12 JANUARY 2007 Morning

Time: 1 hour 30 minutes

Additional Materials:
Answer booklet (8 pages)
Graph paper
MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- 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 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

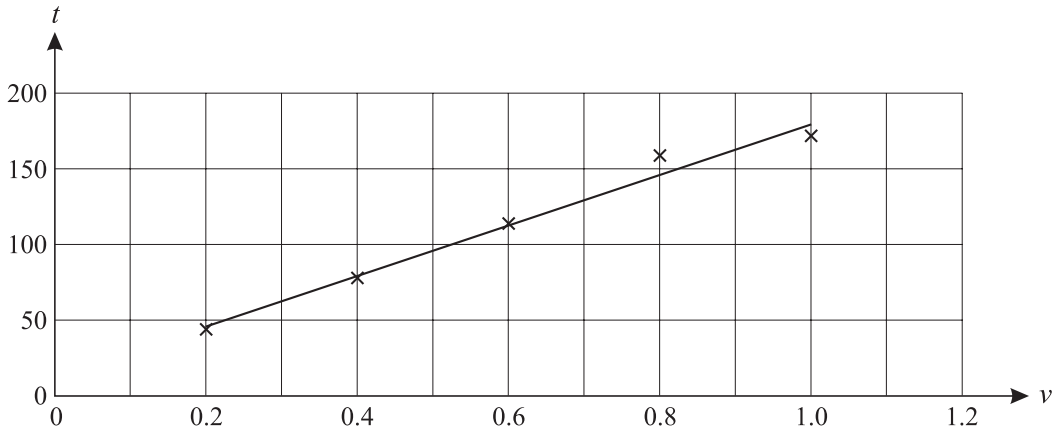
This document consists of **6** printed pages and **2** blank pages.

 OCR is an exempt Charity **[Turn over**

1 In a science investigation into energy conservation in the home, a student is collecting data on the time taken for an electric kettle to boil as the volume of water in the kettle is varied. The student's data are shown in the table below, where v litres is the volume of water in the kettle and t seconds is the time taken for the kettle to boil (starting with the water at room temperature in each case). Also shown are summary statistics and a scatter diagram on which the regression line of t on v is drawn.

v	0.2	0.4	0.6	0.8	1.0
t	44	78	114	156	172

$n = 5,\ \Sigma v = 3.0,\ \Sigma t = 564,\ \Sigma v^2 = 2.20,\ \Sigma vt = 405.2.$



(i) Calculate the equation of the regression line of t on v, giving your answer in the form $t = a + bv$. [5]

(ii) Use this equation to predict the time taken for the kettle to boil when the amount of water which it contains is

(*A*) 0.5 litres,

(*B*) 1.5 litres.

Comment on the reliability of each of these predictions. [4]

(iii) In the equation of the regression line found in part **(i)**, explain the role of the coefficient of v in the relationship between time taken and volume of water. [2]

(iv) Calculate the values of the residuals for $v = 0.8$ and $v = 1.0$. [4]

(v) Explain how, on a scatter diagram with the regression line drawn accurately on it, a residual could be measured and its sign determined. [3]

2 **(a)** A farmer grows Brussels sprouts. The diameter of sprouts in a particular batch, measured in mm, is Normally distributed with mean 28 and variance 16. Sprouts that are between 24 mm and 33 mm in diameter are sold to a supermarket.

(i) Find the probability that the diameter of a randomly selected sprout will be within this range. [4]

(ii) The farmer sells the sprouts in this range to the supermarket for 10 pence per kilogram. The farmer sells sprouts under 24 mm in diameter to a frozen food factory for 5 pence per kilogram. Sprouts over 33 mm in diameter are thrown away. Estimate the total income received by the farmer for the batch, which weighs 25 000 kg. [3]

(iii) By harvesting sprouts earlier, the mean diameter for another batch can be reduced to k mm. Find the value of k for which only 5% of the sprouts will be above 33 mm in diameter. You may assume that the variance is still 16. [3]

(b) The farmer also grows onions. The weight in kilograms of the onions is Normally distributed with mean 0.155 and variance 0.005. He is trying out a new variety, which he hopes will yield a higher mean weight. In order to test this, he takes a random sample of 25 onions of the new variety and finds that their total weight is 4.77 kg. You should assume that the weight in kilograms of the new variety is Normally distributed with variance 0.005.

(i) Write down suitable null and alternative hypotheses for the test in terms of μ. State the meaning of μ in this case. [2]

(ii) Carry out the test at the 1% level. [6]

3 An electrical retailer gives customers extended guarantees on washing machines. Under this guarantee all repairs in the first 3 years are free. The retailer records the numbers of free repairs made to 80 machines.

Number of repairs	0	1	2	3	>3
Frequency	53	20	6	1	0

(i) Show that the sample mean is 0.4375. [1]

(ii) The sample standard deviation s is 0.6907. Explain why this supports a suggestion that a Poisson distribution may be a suitable model for the distribution of the number of free repairs required by a randomly chosen washing machine. [2]

The random variable X denotes the number of free repairs required by a randomly chosen washing machine. For the remainder of this question you should assume that X may be modelled by a Poisson distribution with mean 0.4375.

(iii) Find $P(X = 1)$. Comment on your answer in relation to the data in the table. [4]

(iv) The manager decides to monitor 8 washing machines sold on one day. Find the probability that there are at least 12 free repairs in total on these 8 machines. You may assume that the 8 machines form an independent random sample. [3]

(v) A launderette with 8 washing machines has needed 12 free repairs. Why does your answer to part **(iv)** suggest that the Poisson model with mean 0.4375 is unlikely to be a suitable model for free repairs on the machines in the launderette? Give a reason why the model may not be appropriate for the launderette. [3]

The retailer also sells tumble driers with the same guarantee. The number of free repairs on a tumble drier in three years can be modelled by a Poisson distribution with mean 0.15. A customer buys a tumble drier and a washing machine.

(vi) Assuming that free repairs are required independently, find the probability that

(*A*) the two appliances need a total of 3 free repairs between them,

(*B*) each appliance needs exactly one free repair. [5]

4 Two educational researchers are investigating the relationship between personal ambitions and home location of students. The researchers classify students into those whose main personal ambition is good academic results and those who have some other ambition. A random sample of 480 students is selected.

(i) One researcher summarises the data as follows.

Observed		Home location	
		City	Non-city
Ambition	Good results	102	147
	Other	75	156

Carry out a test at the 5% significance level to examine whether there is any association between home location and ambition. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. [9]

(ii) The other researcher summarises the same data in a different way as follows.

Observed		Home location		
		City	Town	Country
Ambition	Good results	102	83	64
	Other	75	64	92

(*A*) Calculate the expected frequencies for both 'Country' cells. [2]

(*B*) The test statistic for these data is 10.94. Carry out a test at the 5% level based on this table, using the same hypotheses as in part **(i)**. [3]

(*C*) The table below gives the contribution of each cell to the test statistic. Discuss briefly how personal ambitions are related to home location. [2]

Contribution to the test statistic		Home location		
		City	Town	Country
Ambition	Good results	1.129	0.596	3.540
	Other	1.217	0.643	3.816

(iii) Comment briefly on whether the analysis in part **(ii)** means that the conclusion in part **(i)** is invalid. [2]

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE UNIT **4767/01**

MATHEMATICS (MEI)

Statistics 2

MONDAY 21 MAY 2007 Morning

Time: 1 hour 30 minutes

Additional Materials:
Answer booklet (8 pages)
Graph paper
MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- 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 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

This document consists of **4** printed pages.

 OCR is an exempt Charity **[Turn over**

1 The random variable X represents the time taken in minutes for a haircut at a barber's shop. X is Normally distributed with mean 11 and standard deviation 3.

(i) Find $P(X < 10)$. [4]

(ii) Find the probability that exactly 3 out of 8 randomly selected haircuts take less than 10 minutes. [3]

(iii) Use a suitable approximating distribution to find the probability that at least 50 out of 100 randomly selected haircuts take less than 10 minutes. [4]

A new hairdresser joins the shop. The shop manager suspects that she takes longer on average than the other staff to do a haircut. In order to test this, the manager records the time taken for 25 randomly selected cuts by the new hairdresser. The mean time for these cuts is 12.34 minutes. You should assume that the time taken by the new hairdresser is Normally distributed with standard deviation 3 minutes.

(iv) Write down suitable null and alternative hypotheses for the test. [3]

(v) Carry out the test at the 5% level. [5]

2 A medical student is trying to estimate the birth weight of babies using pre-natal scan images. The actual weights, x kg, and the estimated weights, y kg, of ten randomly selected babies are given in the table below.

x	2.61	2.73	2.87	2.96	3.05	3.14	3.17	3.24	3.76	4.10
y	3.2	2.6	3.5	3.1	2.8	2.7	3.4	3.3	4.4	4.1

(i) Calculate the value of Spearman's rank correlation coefficient. [5]

(ii) Carry out a hypothesis test at the 5% level to determine whether there is positive association between the student's estimates and the actual birth weights of babies in the underlying population. [5]

(iii) Calculate the value of the product moment correlation coefficient of the sample. You may use the following summary statistics in your calculations:

$$\Sigma x = 31.63, \quad \Sigma y = 33.1, \quad \Sigma x^2 = 101.92, \quad \Sigma y^2 = 112.61, \quad \Sigma xy = 106.51.$$ [5]

(iv) Explain why, if the underlying population has a bivariate Normal distribution, it would be preferable to carry out a hypothesis test based on the product moment correlation coefficient.

Comment briefly on the significance of the product moment correlation coefficient in relation to that of Spearman's rank correlation coefficient. [4]

3 The number of calls received at an office per 5 minutes is modelled by a Poisson distribution with mean 3.2.

(i) Find the probability of

(*A*) exactly one call in a 5-minute period,

(*B*) at least 6 calls in a 5-minute period. [4]

(ii) Find the probability of

(*A*) exactly one call in a 1-minute period,

(*B*) exactly one call in each of five successive 1-minute periods. [4]

(iii) Use a suitable approximating distribution to find the probability of at most 45 calls in a period of 1 hour. [4]

Two assumptions required for a Poisson distribution to be a suitable model are that calls arrive

- at a uniform average rate,
- independently of each other.

(iv) Comment briefly on the validity of each of these assumptions if the office is

(*A*) the enquiry department of a bank,

(*B*) a police emergency control room. [4]

4 The sexes and ages of a random sample of 300 runners taking part in marathons are classified as follows.

Observed		Sex		Row totals
		Male	Female	
Age group	Under 40	70	54	124
	40–49	76	36	112
	50 and over	52	12	64
Column totals		198	102	300

(i) Carry out a test at the 5% significance level to examine whether there is any association between age group and sex. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. [10]

(ii) Does your analysis support the suggestion that women are less likely than men to enter marathons as they get older? Justify your answer. [3]

For marathons in general, on average 3% of runners are 'Female, 50 and over'. The random variable X represents the number of 'Female, 50 and over' runners in a random sample of size 300.

(iii) Use a suitable approximating distribution to find $P(X \geqslant 12)$. [5]

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE **4767/01**

MATHEMATICS (MEI)

Statistics 2

TUESDAY 15 JANUARY 2008 Morning

Time: 1 hour 30 minutes

Additional materials: Answer Booklet (8 pages)
Graph paper
MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

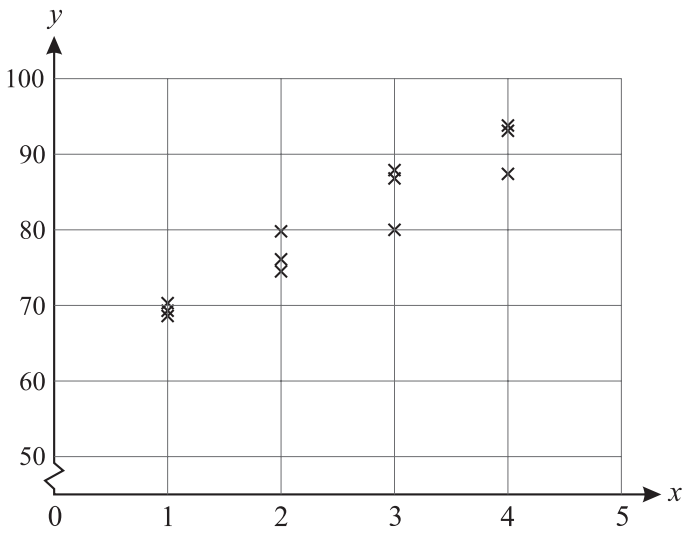
INFORMATION FOR CANDIDATES

- 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 72.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

This document consists of **4** printed pages.

[Turn over

1 A biology student is carrying out an experiment to study the effect of a hormone on the growth of plant shoots. The student applies the hormone at various concentrations to a random sample of twelve shoots and measures the growth of each shoot. The data are illustrated on the scatter diagram below, together with the summary statistics for these data. The variables x and y, measured in suitable units, represent concentration and growth respectively.



$n = 12,\ \Sigma x = 30,\ \Sigma y = 967.6,\ \Sigma x^2 = 90,\ \Sigma y^2 = 78\,926,\ \Sigma xy = 2530.3.$

(i) State which of the two variables x and y is the independent variable and which is the dependent variable. Briefly explain your answers. [3]

(ii) Calculate the equation of the regression line of y on x. [5]

(iii) Use the equation of the regression line to calculate estimates of shoot growth for concentrations of

(*A*) 1.2,

(*B*) 4.3.

Comment on the reliability of each of these estimates. [4]

(iv) Calculate the value of the residual for the data point where $x = 3$ and $y = 80$. [3]

(v) In further experiments, the student finds that using concentration $x = 6$ results in shoot growths of around $y = 20$. In the light of all the available information, what can be said about the relationship between x and y? [3]

2 A large hotel has 90 bedrooms. Sometimes a guest makes a booking for a room, but then does not arrive. This is called a 'no-show'. On average 10% of bookings are no-shows. The hotel manager accepts up to 94 bookings before saying that the hotel is full. If at least 4 of these bookings are no-shows then there will be enough rooms for all of the guests. 94 bookings have been made for each night in August. You should assume that all bookings are independent.

(i) State the distribution of the number of no-shows on one night in August. [2]

(ii) State the conditions under which the use of a Poisson distribution is appropriate as an approximation to a binomial distribution. [2]

(iii) Use a Poisson approximating distribution to find the probability that, on one night in August,

(*A*) there are exactly 4 no-shows,

(*B*) there are enough rooms for all of the guests who do arrive. [5]

(iv) Find the probability that, on all of the 31 nights in August, there are enough rooms for all of the guests who arrive. [2]

(v) (*A*) In August there are $31 \times 94 = 2914$ bookings altogether. State the exact distribution of the total number of no-shows during August. [2]

(*B*) Use a suitable approximating distribution to find the probability that there are at most 300 no-shows altogether during August. [5]

3 In a large population, the diastolic blood pressure (DBP) of 5-year-old children is Normally distributed with mean 56 and standard deviation 6.5.

(i) Find the probability that the DBP of a randomly selected 5-year-old child is between 52.5 and 57.5. [4]

The DBP of young adults is Normally distributed with mean 68 and standard deviation 10.

(ii) A 5-year-old child and a young adult are selected at random. Find the probability that the DBP of one of them is over 62 and the other is under 62. [5]

(iii) Sketch both distributions on a single diagram. [4]

(iv) For another age group, the DBP is Normally distributed with mean 82. The DBP of 12% of people in this age group is below 62. Find the standard deviation for this age group. [4]

[Question 4 is printed overleaf.]

4 **(a)** A researcher believes that there may be some association between a student's sex and choice of certain subjects at A-level. A random sample of 250 A-level students is selected. The table below shows, for each sex, how many study either or both of the two subjects, Mathematics and English.

	Mathematics only	English only	Both	Neither	**Row totals**
Male	38	19	6	32	**95**
Female	42	55	9	49	**155**
Column totals	**80**	**74**	**15**	**81**	**250**

Carry out a test at the 5% significance level to examine whether there is any association between a student's sex and choice of subjects. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. [12]

(b) Over a long period it has been determined that the mean score of students in a particular English module is 67.4 and the standard deviation is 8.9. A new teaching method is introduced with the aim of improving the results. A random sample of 12 students taught by the new method is selected. Their mean score is found to be 68.3. Carry out a test at the 10% level to investigate whether the new method appears to have been successful. State carefully your null and alternative hypotheses. You should assume that the scores are Normally distributed and that the standard deviation is unchanged. [7]

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE **4767/01**

MATHEMATICS (MEI)

Statistics 2

FRIDAY 23 MAY 2008 Morning

Time: 1 hour 30 minutes

Additional materials: Answer Booklet (8 pages)
Graph paper
MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- 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 72.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

This document consists of **4** printed pages.

 [Turn over

1 A researcher believes that there is a negative correlation between money spent by the government on education and population growth in various countries. A random sample of 48 countries is selected to investigate this belief. The level of government spending on education x, measured in suitable units, and the annual percentage population growth rate y, are recorded for these countries. Summary statistics for these data are as follows.

$$\Sigma x = 781.3 \quad \Sigma y = 57.8 \quad \Sigma x^2 = 14\,055 \quad \Sigma y^2 = 106.3 \quad \Sigma xy = 880.1 \quad n = 48$$

(i) Calculate the sample product moment correlation coefficient. [5]

(ii) Carry out a hypothesis test at the 5% significance level to investigate the researcher's belief. State your hypotheses clearly, defining any symbols which you use. [6]

(iii) State the distributional assumption which is necessary for this test to be valid. Explain briefly how a scatter diagram may be used to check whether this assumption is likely to be valid. [2]

(iv) A student suggests that if the variables are negatively correlated then population growth rates can be reduced by increasing spending on education. Explain why the student may be wrong. Discuss an alternative explanation for the correlation. [3]

(v) State briefly one advantage and one disadvantage of using a smaller sample size in this investigation. [2]

2 A public water supply contains bacteria. Each day an analyst checks the water quality by counting the number of bacteria in a random sample of 5 ml of water.

Throughout this question, you should assume that the bacteria occur randomly at a mean rate of 0.37 bacteria per 5 ml of water.

(i) Use a Poisson distribution to

(*A*) find the probability that a 5 ml sample contains exactly 2 bacteria, [2]

(*B*) show that the probability that a 5 ml sample contains more than 2 bacteria is 0.0064. [3]

(ii) The month of September has 30 days. Find the probability that during September there is at most one day when a 5 ml sample contains more than 2 bacteria. [4]

The daily 5 ml sample is the first stage of the quality control process. The remainder of the process is as follows.

- If the 5 ml sample contains more than 2 bacteria, then a 50 ml sample is taken.
- If this 50 ml sample contains more than 8 bacteria, then a sample of 1000 ml is taken.
- If this 1000 ml sample contains more than 90 bacteria, then the supply is declared to be 'questionable'.

(iii) Find the probability that a random sample of 50 ml contains more than 8 bacteria. [3]

(iv) Use a suitable approximating distribution to find the probability that a random sample of 1000 ml contains more than 90 bacteria. [4]

(v) Find the probability that the supply is declared to be questionable. [2]

3 A company has a fleet of identical vans. Company policy is to replace all of the tyres on a van as soon as any one of them is worn out. The random variable X represents the number of miles driven before the tyres on a van are replaced. X is Normally distributed with mean 27 500 and standard deviation 4000.

(i) Find $P(X > 25\,000)$. [4]

(ii) 10 vans in the fleet are selected at random. Find the probability that the tyres on exactly 7 of them last for more than 25 000 miles. [3]

(iii) The tyres of 99% of vans last for more than k miles. Find the value of k. [3]

A tyre supplier claims that a different type of tyre will have a greater mean lifetime. A random sample of 15 vans is fitted with these tyres. For each van, the number of miles driven before the tyres are replaced is recorded. A hypothesis test is carried out to investigate the claim. You may assume that these lifetimes are also Normally distributed with standard deviation 4000.

(iv) Write down suitable null and alternative hypotheses for the test. [3]

(v) For the 15 vans, it is found that the mean lifetime of the tyres is 28 630 miles. Carry out the test at the 5% level. [5]

[Question 4 is printed overleaf.]

4 A student is investigating whether there is any association between the species of shellfish that occur on a rocky shore and where they are located. A random sample of 160 shellfish is selected and the numbers of shellfish in each category are summarised in the table below.

		Location		
		Exposed	Sheltered	Pool
Species	Limpet	24	32	16
	Mussel	24	11	3
	Other	5	22	23

(i) Write down null and alternative hypotheses for a test to examine whether there is any association between species and location. [1]

The contributions to the test statistic for the usual χ^2 test are shown in the table below.

Contribution		Location		
		Exposed	Sheltered	Pool
Species	Limpet	0.0009	0.2585	0.4450
	Mussel	10.3472	1.2756	4.8773
	Other	8.0719	0.1402	7.4298

The sum of these contributions is 32.85.

(ii) Calculate the expected frequency for mussels in pools. Verify the corresponding contribution 4.8773 to the test statistic. [4]

(iii) Carry out the test at the 5% level of significance, stating your conclusion clearly. [5]

(iv) For each species, comment briefly on how its distribution compares with what would be expected if there were no association. [5]

(v) If 3 of the 160 shellfish are selected at random, one from each of the 3 types of location, find the probability that all 3 of them are limpets. [3]

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI)

4767

Statistics 2

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
None

Monday 19 January 2009
Afternoon

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- This document consists of **4** pages. Any blank pages are indicated.

1 A researcher is investigating whether there is a relationship between the population size of cities and the average walking speed of pedestrians in the city centres. Data for the population size, x thousands, and the average walking speed of pedestrians, $y\,\mathrm{m\,s^{-1}}$, of eight randomly selected cities are given in the table below.

x	18	43	52	94	98	206	784	1530
y	1.15	0.97	1.26	1.35	1.28	1.42	1.32	1.64

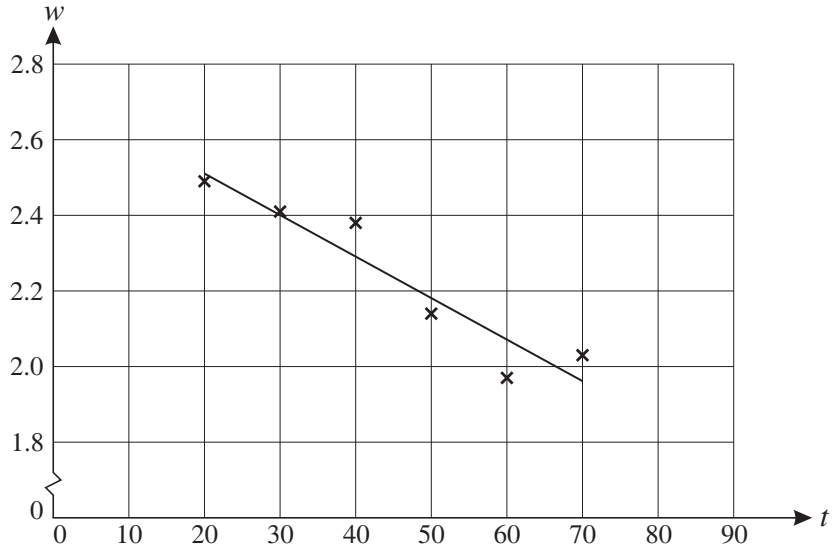
(i) Calculate the value of Spearman's rank correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 5% significance level to determine whether there is any association between population size and average walking speed. **[6]**

In another investigation, the researcher selects a random sample of six adult males of particular ages and measures their maximum walking speeds. The data are shown in the table below, where t years is the age of the adult and $w\,\mathrm{m\,s^{-1}}$ is the maximum walking speed. Also shown are summary statistics and a scatter diagram on which the regression line of w on t is drawn.

t	20	30	40	50	60	70
w	2.49	2.41	2.38	2.14	1.97	2.03

$n = 6 \quad \Sigma t = 270 \quad \Sigma w = 13.42 \quad \Sigma t^2 = 13\,900 \quad \Sigma w^2 = 30.254 \quad \Sigma tw = 584.6$



(iii) Calculate the equation of the regression line of w on t. **[5]**

(iv) (*A*) Use this equation to calculate an estimate of maximum walking speed of an 80-year-old male. **[2]**

(*B*) Explain why it might not be appropriate to use the equation to calculate an estimate of maximum walking speed of a 10-year-old male. **[2]**

2 Clover stems usually have three leaves. Occasionally a clover stem has four leaves. This is considered by some to be lucky and is known as a four-leaf clover. On average 1 in 10 000 clover stems is a four-leaf clover. You may assume that four-leaf clovers occur randomly and independently.

A random sample of 5000 clover stems is selected.

(i) State the exact distribution of X, the number of four-leaf clovers in the sample. **[2]**

(ii) Explain why X may be approximated by a Poisson distribution. Write down the mean of this Poisson distribution. **[3]**

(iii) Use this Poisson distribution to find the probability that the sample contains at least one four-leaf clover. **[2]**

(iv) Find the probability that in 20 samples, each of 5000 clover stems, there are exactly 9 samples which contain at least one four-leaf clover. **[3]**

(v) Find the expected number of these 20 samples which contain at least one four-leaf clover. **[2]**

The table shows the numbers of four-leaf clovers in these 20 samples.

Number of four-leaf clovers	0	1	2	>2
Number of samples	11	7	2	0

(vi) Calculate the mean and variance of the data in the table. **[3]**

(vii) Briefly comment on whether your answers to parts **(v)** and **(vi)** support the use of the Poisson approximating distribution in part **(iii)**. **[3]**

3 The number of minutes, X, for which a particular model of laptop computer will run on battery power is Normally distributed with mean 115.3 and standard deviation 21.9.

(i) (*A*) Find $P(X < 120)$. **[3]**

(*B*) Find $P(100 < X < 110)$. **[3]**

(*C*) Find the value of k for which $P(X > k) = 0.9$. **[3]**

The number of minutes, Y, for which a different model of laptop computer will run on battery power is known to be Normally distributed with mean μ and standard deviation σ.

(ii) Given that $P(Y < 180) = 0.7$ and $P(Y < 140) = 0.15$, find the values of μ and σ. **[4]**

(iii) Find values of a and b for which $P(a < Y < b) = 0.95$. **[4]**

4 A gardening research organisation is running a trial to examine the growth and the size of flowers of various plants.

(i) In the trial, seeds of three types of plant are sown. The growth of each plant is classified as good, average or poor. The results are shown in the table.

		Growth			Row totals
		Good	Average	Poor	
Type of plant	Coriander	12	28	15	55
	Aster	7	18	23	48
	Fennel	14	22	11	47
Column totals		33	68	49	150

Carry out a test at the 5% significance level to examine whether there is any association between growth and type of plant. State carefully your null and alternative hypotheses. Include a table of the contributions of each cell to the test statistic. **[12]**

(ii) It is known that the diameter of marigold flowers is Normally distributed with mean 47 mm and standard deviation 8.5 mm. A certain fertiliser is expected to cause flowers to have a larger mean diameter, but without affecting the standard deviation. A large number of marigolds are grown using this fertiliser. The diameters of a random sample of 50 of the flowers are measured and the mean diameter is found to be 49.2 mm. Carry out a hypothesis test at the 1% significance level to check whether flowers grown with this fertiliser appear to be larger on average. Use hypotheses $H_0 : \mu = 47$, $H_1 : \mu > 47$, where μ mm represents the mean diameter of all marigold flowers grown with this fertiliser. **[5]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI)

4767

Statistics 2

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
None

Monday 15 June 2009
Afternoon

Duration: 1 hour 30 minutes

4767

INSTRUCTIONS TO CANDIDATES

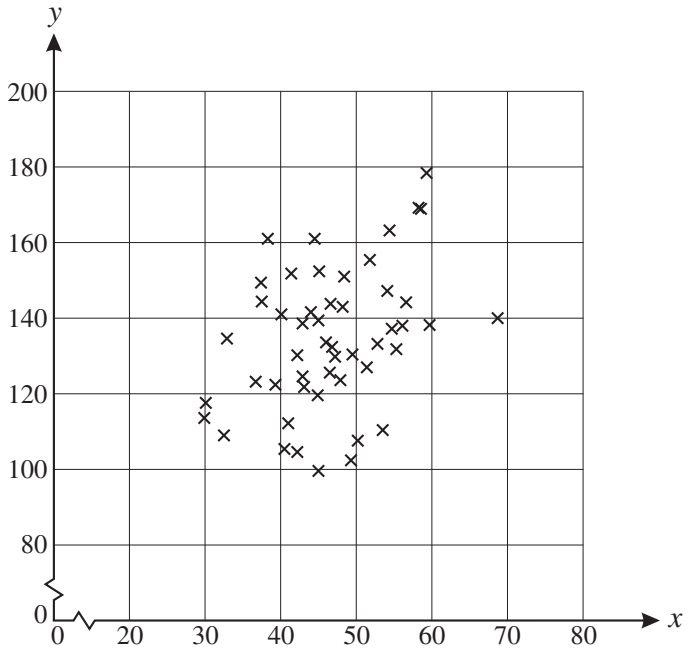
- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- This document consists of **4** pages. Any blank pages are indicated.

1 An investment analyst thinks that there may be correlation between the cost of oil, x dollars per barrel, and the price of a particular share, y pence. The analyst selects 50 days at random and records the values of x and y. Summary statistics for these data are shown below, together with a scatter diagram.

$$\Sigma x = 2331.3 \quad \Sigma y = 6724.3 \quad \Sigma x^2 = 111\,984 \quad \Sigma y^2 = 921\,361 \quad \Sigma xy = 316\,345 \quad n = 50$$



(i) Calculate the sample product moment correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 5% significance level to investigate the analyst's belief. State your hypotheses clearly, defining any symbols which you use. **[6]**

(iii) An assumption that there is a bivariate Normal distribution is required for this test to be valid. State whether it is the sample or the population which is required to have such a distribution. State, with a reason, whether in this case the assumption appears to be justified. **[3]**

(iv) Explain why a 2-tail test is appropriate even though it is clear from the scatter diagram that the sample has a positive correlation coefficient. **[2]**

2 Jess is watching a shower of meteors (shooting stars). During the shower, she sees meteors at an average rate of 1.3 per minute.

(i) State conditions required for a Poisson distribution to be a suitable model for the number of meteors which Jess sees during a randomly selected minute. **[2]**

You may assume that these conditions are satisfied.

(ii) Find the probability that, during one minute, Jess sees

(*A*) exactly one meteor,

(*B*) at least 4 meteors. **[4]**

(iii) Find the probability that, in a period of 10 minutes, Jess sees exactly 10 meteors. **[3]**

(iv) Use a suitable approximating distribution to find the probability that Jess sees a total of at least 100 meteors during a period of one hour. **[5]**

(v) Jess watches the shower for t minutes. She wishes to be at least 99% certain that she will see one or more meteors. Find the smallest possible integer value of t. **[5]**

3 Intensity of light is measured in lumens. The random variable X represents the intensity of the light from a standard 100 watt light bulb. X is Normally distributed with mean 1720 and standard deviation 90. You may assume that the intensities for different bulbs are independent.

(i) Show that $P(X < 1700) = 0.4121$. **[4]**

(ii) These bulbs are sold in packs of 4. Find the probability that the intensities of exactly 2 of the 4 bulbs in a randomly chosen pack are below 1700 lumens. **[3]**

(iii) Use a suitable approximating distribution to find the probability that the intensities of at least 20 out of 40 randomly selected bulbs are below 1700 lumens. **[5]**

A manufacturer claims that the average intensity of its 25 watt low energy light bulbs is 1720 lumens. A consumer organisation suspects that the true figure may be lower than this. The intensities of a random sample of 20 of these bulbs are measured. A hypothesis test is then carried out to check the claim.

(iv) Write down a suitable null hypothesis and explain briefly why the alternative hypothesis should be $H_1 : \mu < 1720$. State the meaning of μ. **[3]**

(v) Given that the standard deviation of the intensity of such bulbs is 90 lumens and that the mean intensity of the sample of 20 bulbs is 1703 lumens, carry out the test at the 5% significance level. **[5]**

4 In a traffic survey a random sample of 400 cars passing a particular location during the rush hour is selected. The type of car and the sex of the driver are classified as follows.

		Sex		Row totals
		Male	Female	
Type of car	Hatchback	96	36	132
	Saloon	77	35	112
	People carrier	38	44	82
	4WD	19	8	27
	Sports car	22	25	47
Column totals		252	148	400

(i) Carry out a test at the 5% significance level to examine whether there is any association between type of car and sex of driver. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. **[12]**

(ii) For each type of car, comment briefly on how the number of drivers of each sex compares with what would be expected if there were no association. **[5]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI)

4767

Statistics 2

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
None

Monday 25 January 2010
Morning

Duration: 1 hour 30 minutes

* 4 7 6 7 *

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- This document consists of **4** pages. Any blank pages are indicated.

1 A pilot records the take-off distance for his light aircraft on runways at various altitudes. The data are shown in the table below, where a metres is the altitude and t metres is the take-off distance. Also shown are summary statistics for these data.

a	0	300	600	900	1200	1500	1800
t	635	704	776	836	923	1008	1105

$n = 7 \quad \Sigma a = 6300 \quad \Sigma t = 5987 \quad \Sigma a^2 = 8\,190\,000 \quad \Sigma t^2 = 5\,288\,931 \quad \Sigma at = 6\,037\,800$

(i) Draw a scatter diagram to illustrate these data. **[3]**

(ii) State which of the two variables a and t is the independent variable and which is the dependent variable. Briefly explain your answer. **[3]**

(iii) Calculate the equation of the regression line of t on a. **[5]**

(iv) Use the equation of the regression line to calculate estimates of the take-off distance for altitudes

(A) 800 metres,

(B) 2500 metres.

Comment on the reliability of each of these estimates. **[4]**

(v) Calculate the value of the residual for the data point where $a = 1200$ and $t = 923$, and comment on its sign. **[4]**

2 On average 2% of a particular model of laptop computer are faulty. Faults occur independently and randomly.

(i) Find the probability that exactly 1 of a batch of 10 laptops is faulty. **[3]**

(ii) State the conditions under which the use of a Poisson distribution is appropriate as an approximation to a binomial distribution. **[2]**

(iii) A school buys a batch of 150 of these laptops. Use a Poisson approximating distribution to find the probability that

(A) there are no faulty laptops in the batch, **[3]**

(B) there are more than the expected number of faulty laptops in the batch. **[3]**

(iv) A large company buys a batch of 2000 of these laptops for its staff.

(A) State the exact distribution of the number of faulty laptops in this batch. **[2]**

(B) Use a suitable approximating distribution to find the probability that there are at most 50 faulty laptops in this batch. **[5]**

3 In an English language test for 12-year-old children, the raw scores, X, are Normally distributed with mean 45.3 and standard deviation 11.5.

(i) Find

(A) $P(X < 50)$, **[3]**

(B) $P(45.3 < X < 50)$. **[2]**

(ii) Find the least raw score which would be obtained by the highest scoring 10% of children. **[3]**

(iii) The raw score is then scaled so that the scaled score is Normally distributed with mean 100 and standard deviation 15. This scaled score is then rounded to the nearest integer. Find the probability that a randomly selected child gets a rounded score of exactly 111. **[4]**

(iv) In a Mathematics test for 12-year-old children, the raw scores, Y, are Normally distributed with mean μ and standard deviation σ. Given that $P(Y < 15) = 0.3$ and $P(Y < 22) = 0.8$, find the values of μ and σ. **[5]**

[Question 4 is printed overleaf.]

4 A council provides waste paper recycling services for local businesses. Some businesses use the standard service for recycling paper, others use a special service for dealing with confidential documents, and others use both. Businesses are classified as small or large. A survey of a random sample of 285 businesses gives the following data for size of business and recycling service.

		Recycling Service		
		Standard	Special	Both
Size of business	Small	35	26	44
	Large	55	52	73

(i) Write down null and alternative hypotheses for a test to examine whether there is any association between size of business and recycling service used. **[1]**

The contributions to the test statistic for the usual χ^2 test are shown in the table below.

		Recycling Service		
		Standard	Special	Both
Size of business	Small	0.1023	0.2607	0.0186
	Large	0.0597	0.1520	0.0108

The sum of these contributions is 0.6041.

(ii) Calculate the expected frequency for large businesses using the special service. Verify the corresponding contribution 0.1520 to the test statistic. **[4]**

(iii) Carry out the test at the 5% level of significance, stating your conclusion clearly. **[5]**

The council is also investigating the weight of rubbish in domestic dustbins. In 2008 the average weight of rubbish in bins was 32.8 kg. The council has now started a recycling initiative and wishes to determine whether there has been a reduction in the weight of rubbish in bins. A random sample of 50 domestic dustbins is selected and it is found that the mean weight of rubbish per bin is now 30.9 kg, and the standard deviation is 3.4 kg.

(iv) Carry out a test at the 5% level to investigate whether the mean weight of rubbish has been reduced in comparison with 2008. State carefully your null and alternative hypotheses. **[8]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI)

4767

Statistics 2

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
- Scientific or graphical calculator

Friday 18 June 2010
Afternoon

Duration: 1 hour 30 minutes

* 4 7 6 7 *

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- This document consists of **4** pages. Any blank pages are indicated.

1 Two celebrities judge a talent contest. Each celebrity gives a score out of 20 to each of a random sample of 8 contestants. The scores, x and y, given by the celebrities to each contestant are shown below.

Contestant	A	B	C	D	E	F	G	H
x	6	17	9	20	13	15	11	14
y	6	13	10	11	9	7	12	15

(i) Calculate the value of Spearman's rank correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 5% significance level to determine whether there is positive association between the scores allocated by the two celebrities. **[6]**

(iii) State the distributional assumption required for a test based on the product moment correlation coefficient. Sketch a scatter diagram of the scores above, and discuss whether it appears that the assumption is likely to be valid. **[5]**

2 A radioactive source is decaying at a mean rate of 3.4 counts per 5 seconds.

(i) State conditions for a Poisson distribution to be a suitable model for the rate of decay of the source. **[2]**

You may assume that a Poisson distribution with a mean rate of 3.4 counts per 5 seconds is a suitable model.

(ii) State the variance of this Poisson distribution. **[1]**

(iii) Find the probability of

(*A*) exactly 3 counts in a 5-second period,

(*B*) at least 3 counts in a 5-second period. **[5]**

(iv) Find the probability of exactly 40 counts in a period of 60 seconds. **[3]**

(v) Use a suitable approximating distribution to find the probability of at least 40 counts in a period of 60 seconds. **[5]**

(vi) The background radiation rate also, independently, follows a Poisson distribution and produces a mean count of 1.4 per 5 seconds. Find the probability that the radiation source together with the background radiation give a total count of at least 8 in a 5-second period. **[3]**

3 In a men's cycling time trial, the times are modelled by the random variable X minutes which is Normally distributed with mean 63 and standard deviation 5.2.

(i) Find

(A) $P(X < 65)$,

(B) $P(60 < X < 65)$. **[6]**

(ii) Find the probability that 5 riders selected at random all record times between 60 and 65 minutes. **[2]**

(iii) A competitor aims to be in the fastest 5% of entrants (i.e. those with the lowest times). Find the maximum time that he can take. **[3]**

It is suggested that holding the time trial on a new course may result in lower times. To investigate this, a random sample of 15 competitors is selected. These 15 competitors do the time trial on the new course. The mean time taken by these riders is 61.7 minutes. You may assume that times are Normally distributed and the standard deviation is still 5.2 minutes. A hypothesis test is carried out to investigate whether times on the new course are lower.

(iv) Write down suitable null and alternative hypotheses for the test. Carry out the test at the 5% significance level. **[8]**

4 In a survey a random sample of 63 runners is selected. The category of runner and the type of running are classified as follows.

		Category of runner			Row totals
		Junior	Senior	Veteran	
Type of running	Track	9	8	2	19
	Road	4	8	12	24
	Both	4	10	6	20
Column totals		17	26	20	63

(i) Carry out a test at the 5% significance level to examine whether there is any association between category of runner and the type of running. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. **[12]**

(ii) For each category of runner, comment briefly on how the type of running compares with what would be expected if there were no association. **[6]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI)

4767

Statistics 2

QUESTION PAPER

Candidates answer on the printed answer book.

OCR supplied materials:
- Printed answer book 4767
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Monday 24 January 2011
Morning

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES
These instructions are the same on the printed answer book and the question paper.

- The question paper will be found in the centre of the printed answer book.
- Write your name, centre number and candidate number in the spaces provided on the printed answer book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the printed answer book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES
This information is the same on the printed answer book and the question paper.

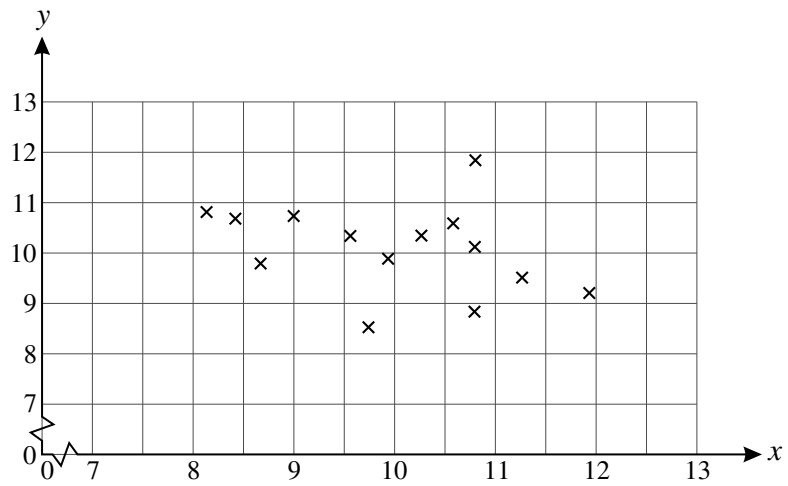
- The number of marks is given in brackets **[ ]** at the end of each question or part question on the question paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The printed answer book consists of **12** pages. The question paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this question paper for marking; it should be retained in the centre or destroyed.

1 The scatter diagram below shows the birth rates x, and death rates y, measured in standard units, in a random sample of 14 countries in a particular year. Summary statistics for the data are as follows.

$$\Sigma x = 139.8 \quad \Sigma y = 140.4 \quad \Sigma x^2 = 1411.66 \quad \Sigma y^2 = 1417.88 \quad \Sigma xy = 1398.56 \quad n = 14$$



(i) Calculate the sample product moment correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 5% significance level to determine whether there is any correlation between birth rates and death rates. **[6]**

(iii) State the distributional assumption which is necessary for this test to be valid. Explain briefly in the light of the scatter diagram why it appears that the assumption may be valid. **[2]**

(iv) The values of x and y for another country in that year are 14.4 and 7.8 respectively. If these values are included, the value of the sample product moment correlation coefficient is –0.5694. Explain why this one observation causes such a large change to the value of the sample product moment correlation coefficient. Discuss whether this brings the validity of the test into question. **[4]**

2 A student is investigating the numbers of sultanas in a particular brand of biscuit. The data in the table show the numbers of sultanas in a random sample of 50 of these biscuits.

Number of sultanas	0	1	2	3	4	5	>5
Frequency	8	15	12	9	4	2	0

(i) Show that the sample mean is 1.84 and calculate the sample variance. **[3]**

(ii) Explain why these results support a suggestion that a Poisson distribution may be a suitable model for the distribution of the numbers of sultanas in this brand of biscuit. **[1]**

For the remainder of the question you should assume that a Poisson distribution with mean 1.84 is a suitable model for the distribution of the numbers of sultanas in these biscuits.

(iii) Find the probability of

(*A*) no sultanas in a biscuit,

(*B*) at least two sultanas in a biscuit. **[5]**

(iv) Show that the probability that there are at least 10 sultanas in total in a packet containing 5 biscuits is 0.4389. **[3]**

(v) Six packets each containing 5 biscuits are selected at random. Find the probability that exactly 2 of the six packets contain at least 10 sultanas. **[3]**

(vi) Sixty packets each containing 5 biscuits are selected at random. Use a suitable approximating distribution to find the probability that more than half of the sixty packets contain at least 10 sultanas. **[5]**

3 The random variable X represents the reaction times, in milliseconds, of men in a driving simulator. X is Normally distributed with mean 355 and standard deviation 52.

(i) Find

(*A*) $P(X < 325)$,

(*B*) $P(300 < X < 400)$. **[6]**

(ii) Find the value of k for which $P(X < k) = 0.2$. **[3]**

It is thought that women may have a different mean reaction time from men. In order to test this, a random sample of 25 women is selected. The mean reaction time of these women in the driving simulator is 344 milliseconds. You may assume that women's reaction times are also Normally distributed with standard deviation 52 milliseconds. A hypothesis test is carried out to investigate whether women have a different mean reaction time from men.

(iii) Carry out the test at the 5% significance level. **[8]**

4 A researcher is investigating the sizes of pebbles at various locations in a river. Three sites in the river are chosen and each pebble sampled at each site is classified as large, medium or small. The results are as follows.

		Site			Row totals
		A	B	C	
Pebble size	Large	15	12	10	37
	Medium	28	17	45	90
	Small	47	33	36	116
Column totals		90	62	91	243

(i) Carry out a test at the 5% significance level to examine whether there is any association between pebble size and site. Your working should include a table of the contributions of each cell to the test statistic. **[12]**

(ii) By referring to each site, comment briefly on how the size of the pebbles compares with what would be expected if there were no association. You should support your answers by referring to your table of contributions. **[6]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI)

4767

Statistics 2

QUESTION PAPER

Candidates answer on the printed answer book.

OCR supplied materials:
- Printed answer book 4767
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Wednesday 22 June 2011
Morning

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES
These instructions are the same on the printed answer book and the question paper.

- The question paper will be found in the centre of the printed answer book.
- Write your name, centre number and candidate number in the spaces provided on the printed answer book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the printed answer book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES
This information is the same on the printed answer book and the question paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the question paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The printed answer book consists of **12** pages. The question paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this question paper for marking; it should be retained in the centre or destroyed.

1 An experiment is performed to determine the response of maize to nitrogen fertilizer. Data for the amount of nitrogen fertilizer applied, x kg/hectare, and the average yield of maize, y tonnes/hectare, in 5 experimental plots are given in the table below.

x	0	30	60	90	120
y	0.5	2.5	4.7	6.2	7.4

(i) Draw a scatter diagram to illustrate these data. **[3]**

(ii) Calculate the equation of the regression line of y on x. **[5]**

(iii) Draw your regression line on your scatter diagram and comment briefly on its fit. **[3]**

(iv) Calculate the value of the residual for the data point where $x = 30$ and $y = 2.5$. **[3]**

(v) Use the equation of the regression line to calculate estimates of average yield with nitrogen fertilizer applications of

(A) 45 kg/hectare,

(B) 150 kg/hectare. **[2]**

(vi) In a plot where 150 kg/hectare of nitrogen fertilizer is applied, the average yield of maize is 8.7 tonnes/hectare. Comment on this result. **[2]**

2 At a drive-through fast food takeaway, cars arrive independently, randomly and at a uniform average rate. The numbers of cars arriving per minute may be modelled by a Poisson distribution with mean 0.62.

(i) Briefly explain the meaning of each of the three terms 'independently', 'randomly' and 'at a uniform average rate', in the context of cars arriving at a fast food takeaway. **[3]**

(ii) Find the probability of at most 1 car arriving in a period of 1 minute. **[3]**

(iii) Find the probability of more than 5 cars arriving in a period of 10 minutes. **[3]**

(iv) State the exact distribution of the number of cars arriving in a period of 1 hour. **[2]**

(v) Use a suitable approximating distribution to find the probability that at least 40 cars arrive in a period of 1 hour. **[5]**

3 The weights of Braeburn apples on display in a supermarket, measured in grams, are Normally distributed with mean 210.5 and standard deviation 15.2.

(i) Find the probability that a randomly selected apple weighs at least 220 grams. **[3]**

(ii) These apples are sold in packs of 3. You may assume that the weights of apples in each pack are independent. Find the probability that all 3 of the apples in a randomly selected pack weigh at least 220 grams. **[2]**

(iii) 100 packs are selected at random.

(*A*) State the exact distribution of the number of these 100 packs in which all 3 apples weigh at least 220 grams. **[2]**

(*B*) Use a suitable approximating distribution to find the probability that in at most one of these packs all 3 apples weigh at least 220 grams. **[5]**

(*C*) Explain why this approximating distribution is suitable. **[1]**

(iv) The supermarket also sells Cox's Orange Pippin apples. The weights of these apples, measured in grams, are Normally distributed with mean 185 and standard deviation σ.

(*A*) Given that 10% of randomly selected Cox's Orange Pippin apples weigh less than 170 grams, calculate the value of σ. **[3]**

(*B*) Sketch the distributions of the weights of both types of apple on a single diagram. **[4]**

[Question 4 is printed overleaf.]

4 **(a)** In a survey on internet usage, a random sample of 200 people is selected. The people are asked how much they have spent on internet shopping during the last three months. The results, classified by amount spent and sex, are shown in the table.

		Sex		Row totals
		Male	Female	
Amount spent	Nothing	28	34	62
	Less than £50	17	21	38
	£50 up to £200	22	26	48
	£200 up to £1000	23	16	39
	£1000 or more	8	5	13
Column totals		98	102	200

(i) Write down null and alternative hypotheses for a test to examine whether there is any association between amount spent and sex of person. **[1]**

The contributions to the test statistic for the usual χ^2 test are shown in the table below.

		Sex	
		Male	Female
Amount spent	Nothing	0.1865	0.1791
	Less than £50	0.1409	0.1354
	£50 up to £200	0.0982	0.0944
	£200 up to £1000	0.7918	0.7608
	£1000 or more	0.4171	0.4007

The sum of these contributions, correct to 3 decimal places, is 3.205.

(ii) Calculate the expected frequency for females spending nothing. Verify the corresponding contribution, 0.1791, to the test statistic. **[3]**

(iii) Carry out the test at the 5% level of significance, stating your conclusion clearly. **[5]**

(b) A bakery sells loaves specified as having a mean weight of 400 grams. It is known that the weights of these loaves are Normally distributed and that the standard deviation is 5.7 grams. An inspector suspects that the true mean weight may be less than 400 grams. In order to test this, the inspector takes a random sample of 6 loaves. Carry out a suitable test at the 5% level, given that the weights, in grams, of the 6 loaves are as follows. **[9]**

392.1 405.8 401.3 387.4 391.8 400.6

OCR
RECOGNISING ACHIEVEMENT

Monday 23 January 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4767 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 Nine long-distance runners are starting an exercise programme to improve their strength. During the first session, each of them has to do a 100 metre run and to do as many push-ups as possible in one minute. The times taken for the run, together with the number of push-ups each runner achieves, are shown in the table.

Runner	A	B	C	D	E	F	G	H	I
100 metre time (seconds)	13.2	11.6	10.9	12.3	14.7	13.1	11.7	13.6	12.4
Push-ups achieved	32	42	22	36	41	27	37	38	33

(i) Draw a scatter diagram to illustrate the data. **[3]**

(ii) Calculate the value of Spearman's rank correlation coefficient. **[5]**

(iii) Carry out a hypothesis test at the 5% significance level to examine whether there is any association between time taken for the run and number of push-ups achieved. **[6]**

(iv) Under what circumstances is it appropriate to carry out a hypothesis test based on the product moment correlation coefficient? State, with a reason, which test is more appropriate for these data. **[3]**

2 The number of printing errors per page in a book is modelled by a Poisson distribution with a mean of 0.85.

(i) State conditions for a Poisson distribution to be a suitable model for the number of printing errors per page. **[2]**

(ii) A page is chosen at random. Find the probability of

(*A*) exactly 1 error on this page,

(*B*) at least 2 errors on this page. **[5]**

10 pages are chosen at random.

(iii) Find the probability of exactly 10 errors in these 10 pages. **[3]**

(iv) Find the least integer k such that the probability of there being k or more errors in these 10 pages is less than 1%. **[4]**

30 pages are chosen at random.

(v) Use a suitable approximating distribution to find the probability of no more than 30 errors in these 30 pages. **[5]**

3 The lifetime of a particular type of light bulb is X hours, where X is Normally distributed with mean 1100 and variance 2000.

(i) Find $P(1100 < X < 1200)$. **[3]**

(ii) Use a suitable approximating distribution to find the probability that, in a random sample of 100 of these light bulbs, no more than 40 have a lifetime between 1100 and 1200 hours. **[5]**

(iii) A factory has a large number of these light bulbs installed. As soon as 1% of the bulbs have come to the end of their lifetimes, it is company policy to replace all of the bulbs. After how many hours should the bulbs need to be replaced? **[3]**

(iv) The bulbs are to be replaced by low-energy bulbs. The lifetime of these bulbs is Normally distributed and the mean is claimed by the manufacturer to be 7000 hours. The standard deviation is known to be 100 hours. A random sample of 25 low-energy bulbs is selected. Their mean lifetime is found to be 6972 hours. Carry out a 2-tail test at the 10% level to investigate the claim. **[8]**

[Question 4 is printed overleaf.]

4 Birds are observed at feeding stations in three different places – woodland, farm and garden. The numbers of finches, thrushes and tits observed at each site are summarised in the table. The birds observed are regarded as a random sample from the population of birds of these species that use these feeding stations.

Observed Frequency		Place			
		Farm	Garden	Woodland	Totals
Species	Thrushes	11	74	7	92
	Tits	70	26	88	184
	Finches	17	2	10	29
	Totals	98	102	105	305

The expected frequencies under the null hypothesis for the usual χ^2 test are shown in the table below.

Expected Frequency		Place		
		Farm	Garden	Woodland
Species	Thrushes	29.5607	30.7672	31.6721
	Tits	59.1213	61.5344	63.3443
	Finches	9.3180	9.6984	9.9836

(i) Verify that the entry 9.3180 is correct. **[2]**

The corresponding contributions to the test statistic are shown in the table below.

Contribution		Place		
		Farm	Garden	Woodland
Species	Thrushes	11.6539	60.7489	19.2192
	Tits	2.0017	20.5201	9.5969
	Finches	6.3332	6.1108	0.0000

(ii) Verify that the entry 6.3332 is correct. **[2]**

(iii) Carry out the test at the 1% level of significance. **[7]**

(iv) For each place, use the table of contributions to comment briefly on the differences between the observed and expected distributions of species. **[6]**

OCR
RECOGNISING ACHIEVEMENT

Thursday 31 May 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4767 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4715850612

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 The times, in seconds, taken by ten randomly selected competitors for the first and last sections of an Olympic bobsleigh run are denoted by x and y respectively. Summary statistics for these data are as follows.

$$\Sigma x = 113.69 \quad \Sigma y = 52.81 \quad \Sigma x^2 = 1292.56 \quad \Sigma y^2 = 278.91 \quad \Sigma xy = 600.41 \quad n = 10$$

(i) Calculate the sample product moment correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 10% significance level to investigate whether there is any correlation between times taken for the first and last sections of the bobsleigh run. **[6]**

(iii) State the distributional assumption which is necessary for this test to be valid. Explain briefly how a scatter diagram may be used to check whether this assumption is likely to be valid. **[2]**

(iv) A commentator says that in order to have a fast time on the last section, you must have a fast time on the first section. Comment briefly on this suggestion. **[2]**

(v) (*A*) Would your conclusion in part **(ii)** have been different if you had carried out the hypothesis test at the 1% level rather than the 10% level? Explain your answer. **[2]**

(*B*) State one advantage and one disadvantage of using a 1% significance level rather than a 10% significance level in a hypothesis test. **[2]**

2 A particular genetic mutation occurs in one in every 300 births on average. A random sample of 1200 births is selected.

(i) State the exact distribution of X, the number of births in the sample which have the mutation. **[2]**

(ii) Explain why X has, approximately, a Poisson distribution. **[2]**

(iii) Use a Poisson approximating distribution to find

(*A*) $P(X = 1)$,

(*B*) $P(X > 4)$. **[5]**

(iv) Twenty independent samples, each of 1200 births, are selected. State the mean and variance of a Normal approximating distribution suitable for modelling the total number of births with the mutation in the twenty samples. **[2]**

(v) Use this Normal approximating distribution to

(*A*) find the probability that there are at least 90 births which have the mutation, **[3]**

(*B*) find the least value of k such that the probability that there are at most k births with this mutation is greater than 5%. **[4]**

3 At a vineyard, the process used to fill bottles with wine is subject to variation. The contents of bottles are independently Normally distributed with mean $\mu = 751.4$ ml and standard deviation $\sigma = 2.5$ ml.

(i) Find the probability that a randomly selected bottle contains at least 750 ml. **[3]**

(ii) A case of wine consists of 6 bottles. Find the probability that all 6 bottles in a case contain at least 750 ml. **[2]**

(iii) Find the probability that, in a random sample of 25 cases, there are at least 2 cases in which all 6 bottles contain at least 750 ml. **[4]**

It is decided to increase the proportion of bottles which contain at least 750 ml to 98%.

(iv) This can be done by changing the value of μ, but retaining the original value of σ. Find the required value of μ. **[4]**

(v) An alternative is to change the value of σ, but retain the original value of μ. Find the required value of σ. **[3]**

(vi) Comment briefly on which method might be easier to implement and which might be preferable to the vineyard owners. **[2]**

[Question 4 is printed overleaf.]

4 **(a)** Mary is opening a cake shop. As part of her market research, she carries out a survey into which type of cake people like best. She offers people 4 types of cake to taste: chocolate, carrot, lemon and ginger. She selects a random sample of 150 people and she classifies the people as children and adults. The results are as follows.

		Classification of person		Row totals
		Child	Adult	
Type of cake	Chocolate	34	23	57
	Carrot	16	18	34
	Lemon	4	18	22
	Ginger	13	24	37
Column totals		67	83	150

The contributions to the test statistic for the usual χ^2 test are shown in the table below.

		Classification of person	
		Child	Adult
Type of cake	Chocolate	2.8646	2.3124
	Carrot	0.0436	0.0352
	Lemon	3.4549	2.7889
	Ginger	0.7526	0.6075

The sum of these contributions, correct to 2 decimal places, is 12.86.

(i) Calculate the expected frequency for children preferring chocolate cake. Verify the corresponding contribution, 2.8646, to the test statistic. **[3]**

(ii) Carry out the test at the 1% level of significance. **[5]**

(b) Mary buys flour in bags which are labelled as containing 5 kg. She suspects that the average contents of these bags may be less than 5 kg. In order to test this, she selects a random sample of 8 bags and weighs their contents. Assuming that weights are Normally distributed with standard deviation 0.0072 kg, carry out a test at the 5% level, given that the weights of the 8 bags in kg are as follows.

4.992 4.981 5.006 4.982 4.996 5.009 4.991 5.003 **[9]**

OCR
RECOGNISING ACHIEVEMENT

Friday 25 January 2013 – Afternoon

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 A manufacturer of playground safety tiles is testing a new type of tile. Tiles of various thicknesses are tested to estimate the maximum height at which people would be unlikely to sustain injury if they fell onto a tile. The results of the test are as follows.

Thickness (t mm)	20	40	60	80	100
Maximum height (h m)	0.72	1.09	1.62	1.97	2.34

(i) Draw a scatter diagram to illustrate these data. **[3]**

(ii) State which of the two variables is the independent variable, giving a reason for your answer. **[1]**

(iii) Calculate the equation of the regression line of maximum height on thickness. **[5]**

(iv) Use the equation of the regression line to calculate estimates of the maximum height for thicknesses of

(A) 70 mm,

(B) 120 mm.

Comment on the reliability of each of these estimates. **[4]**

(v) Calculate the value of the residual for the data point at which $t = 40$. **[3]**

(vi) In a further experiment, the manufacturer tests a tile with a thickness of 200 mm and finds that the corresponding maximum height is 2.96 m. What can be said about the relationship between tile thickness and maximum height? **[3]**

2 John is observing butterflies being blown across a fence in a strong wind. He uses the Poisson distribution with mean 2.1 to model the number of butterflies he observes in one minute.

(i) Find the probability that John observes

(A) no butterflies in a minute, **[2]**

(B) at least 2 butterflies in a minute, **[2]**

(C) between 5 and 10 butterflies inclusive in a period of 5 minutes. **[3]**

(ii) Use a suitable approximating distribution to find the probability that John observes at least 130 butterflies in a period of 1 hour. **[5]**

In fact some of the butterflies John observes being blown across the fence are being blown in pairs.

(iii) Explain why this invalidates one of the assumptions required for a Poisson distribution to be a suitable model. **[1]**

John decides to revise his model for the number of butterflies he observes in one minute. In this new model, the number of pairs of butterflies is modelled by the Poisson distribution with mean 0.2, and the number of single butterflies is modelled by an independent Poisson distribution with mean 1.7.

(iv) Find the probability that John observes no more than 3 butterflies altogether in a period of one minute. **[5]**

3 The amount of data, X megabytes, arriving at an internet server per second during the afternoon is modelled by the Normal distribution with mean 435 and standard deviation 30.

(i) Find

(A) $\mathrm{P}(X < 450)$, **[3]**

(B) $\mathrm{P}(400 < X < 450)$. **[3]**

(ii) Find the probability that, during 5 randomly selected seconds, the amounts of data arriving are all between 400 and 450 megabytes. **[2]**

The amount of data, Y megabytes, arriving at the server during the evening is modelled by the Normal distribution with mean μ and standard deviation σ.

(iii) Given that $\mathrm{P}(Y < 350) = 0.2$ and $\mathrm{P}(Y > 390) = 0.1$, find the values of μ and σ. **[5]**

(iv) Find values of a and b for which $\mathrm{P}(a < Y < b) = 0.95$. **[4]**

4 **(a)** A random sample of 60 students studying mathematics was selected. Their grades in the Core 1 module are summarised in the table below, classified according to whether they worked less than 5 hours per week or at least 5 hours per week. Test, at the 5% significance level, whether there is any association between grade and hours worked.

		Hours worked	
		Less than 5	At least 5
Grade	A or B	20	11
	C or lower	13	16

[9]

(b) At a canning factory, cans are filled with tomato purée. The machine which fills the cans is set so that the volume of tomato purée in a can, measured in millilitres, is Normally distributed with mean 420 and standard deviation 3.5. After the machine is recalibrated, a quality control officer wishes to check whether the mean is still 420 millilitres. A random sample of 10 cans of tomato purée is selected and the volumes, measured in millilitres, are as follows.

417.2 422.6 414.3 419.6 420.4 410.0 418.3 416.9 418.9 419.7

Carry out a test at the 1% significance level to investigate whether the mean is still 420 millilitres. You should assume that the volumes are Normally distributed with unchanged standard deviation. **[9]**

THERE ARE NO QUESTIONS PRINTED ON THIS PAGE.

OCR
RECOGNISING ACHIEVEMENT

Thursday 6 June 2013 – Morning

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

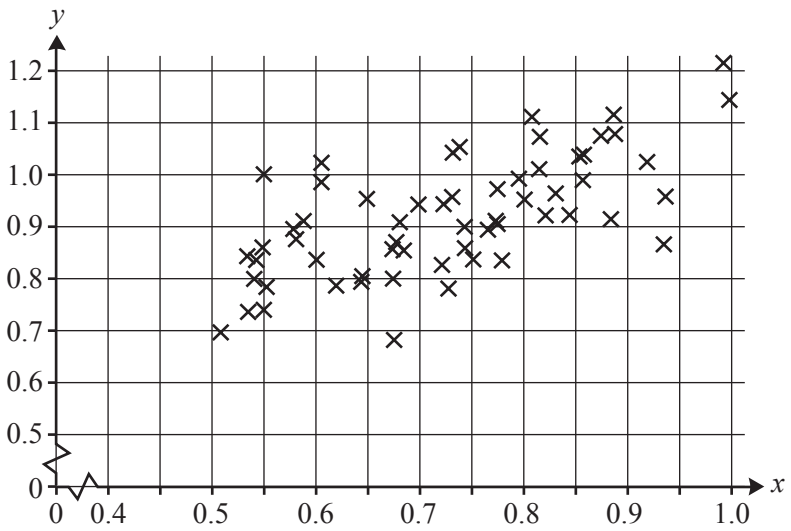
This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 Salbutamol is a drug used to improve lung function. In a medical trial, a random sample of 60 people with impaired lung function was selected. The forced expiratory volume in one second (FEV1) was measured for each person, both before being given salbutamol and again after a two-week course of the drug. The variables x and y, measured in suitable units, represent FEV1 before and after the two-week course respectively. The data are illustrated in the scatter diagram below, together with the summary statistics for these data.



Summary statistics:

$$n = 60, \quad \Sigma x = 43.62, \quad \Sigma y = 55.15, \quad \Sigma x^2 = 32.68, \quad \Sigma y^2 = 51.44, \quad \Sigma xy = 40.66$$

(i) Calculate the sample product moment correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 5% significance level to investigate whether there is positive correlation between FEV1 before and after the course. **[6]**

(iii) State the distributional assumption which is necessary for this test to be valid. State, with a reason, whether the assumption appears to be valid. **[2]**

(iv) Explain the meaning of the term 'significance level'. **[2]**

(v) Calculate the values of the summary statistics if the data point $x = 0.55$, $y = 1.00$ had been incorrectly recorded as $x = 1.00$, $y = 0.55$. **[3]**

2 Suppose that 3% of the population of a large city have red hair.

(i) A random sample of 10 people from the city is selected. Find the probability that there is at least one person with red hair in this sample. **[2]**

A random sample of 60 people from the city is selected. The random variable X represents the number of people in this sample who have red hair.

(ii) Explain why the distribution of X may be approximated by a Poisson distribution. Write down the mean of this Poisson distribution. **[3]**

(iii) Hence find

(A) $\mathrm{P}(X = 2)$, **[2]**

(B) $\mathrm{P}(X > 2)$. **[2]**

(iv) Discuss whether or not it would be appropriate to model X using a Normal approximating distribution. **[2]**

A random sample of 5000 people from the city is selected.

(v) State the exact distribution of the number of people with red hair in the sample. **[2]**

(vi) Use a suitable Normal approximating distribution to find the probability that there are at least 160 people with red hair in the sample. **[5]**

3 The scores, X, in Paper 1 of an English examination have an underlying Normal distribution with mean 76 and standard deviation 12. The scores are reported as integer marks. So, for example, a score for which $75.5 \leqslant X < 76.5$ is reported as 76 marks.

(i) Find the probability that a candidate's reported mark is 76. **[4]**

(ii) Find the probability that a candidate's reported mark is at least 80. **[3]**

(iii) Three candidates are chosen at random. Find the probability that exactly one of these three candidates' reported marks is at least 80. **[2]**

The proportion of candidates who receive an A* grade (the highest grade) must not exceed 10% but should be as close as possible to 10%.

(iv) Find the lowest reported mark that should be awarded an A* grade. **[5]**

The scores in Paper 2 of the examination have an underlying Normal distribution with mean μ and standard deviation 12.

(v) Given that 20% of candidates receive a reported mark of 50 or less, find the value of μ. **[4]**

4 An art gallery is holding an exhibition. A random sample of 150 visitors to the exhibition is selected. The visitors are asked which of four artists they prefer. Their preferences, classified according to whether the visitor is female or male, are given in the table.

		Artist preferred			
		Monet	Renoir	Degas	Cézanne
Sex	Male	8	25	18	19
	Female	18	35	10	17

(i) Carry out a test at the 10% significance level to examine whether there is any association between artist preferred and sex of visitor. Your working should include a table showing the contributions of each cell to the test statistic. **[12]**

(ii) For each artist, comment briefly on how the preferences of each sex compare with what would be expected if there were no association. **[6]**