OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 4768

Statistics 3

Thursday **12 JANUARY 2006** Afternoon 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.

1 A railway company is investigating operations at a junction where delays often occur. Delays (in minutes) are modelled by the random variable T with the following cumulative distribution function.

$$F(t) = \begin{cases} 0 & t \leqslant 0 \\ 1 - e^{-\frac{1}{3}t} & t > 0 \end{cases}$$

(i) Find the median delay and the 90th percentile delay. [5]

(ii) Derive the probability density function of T. Hence use calculus to find the mean delay. [5]

(iii) Find the probability that a delay lasts longer than the mean delay. [2]

You are given that the variance of T is 9.

(iv) Let $\overline{T}$ denote the mean of a random sample of 30 delays. Write down an approximation to the distribution of $\overline{T}$. [3]

(v) A random sample of 30 delays is found to have mean 4.2 minutes. Does this cast any doubt on the modelling? [3]

2 Geoffrey is a university lecturer. He has to prepare five questions for an examination. He knows by experience that it takes about 3 hours to prepare a question, and he models the time (in minutes) taken to prepare one by the Normally distributed random variable X with mean 180 and standard deviation 12, independently for all questions.

(i) One morning, Geoffrey has a gap of 2 hours 50 minutes (170 minutes) between other activities. Find the probability that he can prepare a question in this time. [3]

(ii) One weekend, Geoffrey can devote 14 hours to preparing the complete examination paper. Find the probability that he can prepare all five questions in this time. [3]

A colleague, Helen, has to check the questions.

(iii) She models the time (in minutes) to check a question by the Normally distributed random variable Y with mean 50 and standard deviation 6, independently for all questions and independently of X. Find the probability that the total time for Geoffrey to prepare a question and Helen to check it exceeds 4 hours. [3]

(iv) When working under pressure of deadlines, Helen models the time to check a question in a different way. She uses the Normally distributed random variable $\frac{1}{4}X$, where X is as above. Find the length of time, as given by this model, which Helen needs to ensure that, with probability 0.9, she has time to check a question. [4]

Ian, an educational researcher, suggests that a better model for the time taken to prepare a question would be a constant k representing "thinking time" plus a random variable T representing the time required to write the question itself, independently for all questions.

(v) Taking k as 45 and T as Normally distributed with mean 120 and standard deviation 10 (all units are minutes), find the probability according to Ian's model that a question can be prepared in less than 2 hours 30 minutes. [2]

Juliet, an administrator, proposes that the examination should be reduced in time and shorter questions should be used.

(vi) Juliet suggests that Ian's model should be used for the time taken to prepare such shorter questions but with $k = 30$ and T replaced by $\frac{3}{5}T$. Find the probability as given by this model that a question can be prepared in less than $1\frac{3}{4}$ hours. [3]

[Turn over

3 A production line has two machines, A and B, for delivering liquid soap into bottles. Each machine is set to deliver a nominal amount of 250 ml, but it is not expected that they will work to a high level of accuracy. In particular, it is known that the ambient temperature affects the rate of flow of the liquid and leads to variation in the amounts delivered.

The operators think that machine B tends to deliver a somewhat greater amount than machine A, no matter what the ambient temperature. This is being investigated by an experiment. A random sample of 10 results from the experiment is shown below. Each column of data is for a different ambient temperature.

Ambient temperature	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
Amount delivered by machine A	246.2	251.6	252.0	246.6	258.4	251.0	247.5	247.1	248.1	253.4
Amount delivered by machine B	248.3	252.6	252.8	247.2	258.8	250.0	247.2	247.9	249.0	254.5

(a) Use an appropriate *t* test to examine, at the 5% level of significance, whether the mean amount delivered by machine B may be taken as being greater than that delivered by machine A, stating carefully your null and alternative hypotheses and the required distributional assumption. [11]

(b) Using the data for machine A in the table above, provide a two-sided 95% confidence interval for the mean amount delivered by this machine, stating the required distributional assumption. Explain whether you would conclude that the machine appears to be working correctly in terms of the nominal amount as set. [7]

4 Quality control inspectors in a factory are investigating the lengths of glass tubes that will be used to make laboratory equipment.

(i) Data on the observed lengths of a random sample of 200 glass tubes from one batch are available in the form of a frequency distribution as follows.

Length x (mm)	Observed frequency
$x \leqslant 298$	1
$298 < x \leqslant 300$	30
$300 < x \leqslant 301$	62
$301 < x \leqslant 302$	70
$302 < x \leqslant 304$	34
$x > 304$	3

The sample mean and standard deviation are 301.08 and 1.2655 respectively.

The corresponding expected frequencies for the Normal distribution with parameters estimated by the sample statistics are

Length x (mm)	Expected frequency
$x \leqslant 298$	1.49
$298 < x \leqslant 300$	37.85
$300 < x \leqslant 301$	55.62
$301 < x \leqslant 302$	58.32
$302 < x \leqslant 304$	44.62
$x > 304$	2.10

Examine the goodness of fit of a Normal distribution, using a 5% significance level. [7]

(ii) It is thought that the lengths of tubes in another batch have an underlying distribution similar to that for the batch in part **(i)** but possibly with different location and dispersion parameters. A random sample of 10 tubes from this batch gives the following lengths (in mm).

301.3 301.4 299.6 302.2 300.3 303.2 302.6 301.8 300.9 300.8

(*A*) Discuss briefly whether it would be appropriate to use a t test to examine a hypothesis about the population mean length for this batch. [2]

(*B*) Use a Wilcoxon test to examine at the 10% significance level whether the population median length for this batch is 301 mm. [9]

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 4768

Statistics 3

Wednesday **24 MAY 2006** Afternoon 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.

1 Design engineers are simulating the load on a particular part of a complex structure. They intend that the simulated load, measured in a convenient unit, should be given by the random variable X having probability density function

$$f(x) = 12x^3 - 24x^2 + 12x, \qquad 0 \leqslant x \leqslant 1.$$

(i) Find the mean and the mode of X. [6]

(ii) Find the cumulative distribution function $F(x)$ of X.

Verify that $F\left(\frac{1}{4}\right) = \frac{67}{256}$, $F\left(\frac{1}{2}\right) = \frac{11}{16}$ and $F\left(\frac{3}{4}\right) = \frac{243}{256}$. [3]

The engineers suspect that the process for generating simulated loads might not be working as intended. To investigate this, they generate a random sample of 512 loads. These are recorded in a frequency distribution as follows.

Load x	$0 \leqslant x \leqslant \frac{1}{4}$	$\frac{1}{4} < x \leqslant \frac{1}{2}$	$\frac{1}{2} < x \leqslant \frac{3}{4}$	$\frac{3}{4} < x \leqslant 1$
Frequency	126	209	131	46

(iii) Use a suitable statistical procedure to assess the goodness of fit of X to these data. Discuss your conclusions briefly. [9]

2 A bus route runs from the centre of town A through the town's urban area to a point B on its boundary and then through the country to a small town C. Because of traffic congestion and general road conditions, delays occur on both the urban and the country sections. All delays may be considered independent.

The scheduled time for the journey from A to B is 24 minutes. In fact, journey times over this section are given by the Normally distributed random variable X with mean 26 minutes and standard deviation 3 minutes.

The scheduled time for the journey from B to C is 18 minutes. In fact, journey times over this section are given by the Normally distributed random variable Y with mean 15 minutes and standard deviation 2 minutes.

Journey times on the two sections of route may be considered independent. The timetable published to the public does not show details of times at intermediate points; thus, if a bus is running early, it merely continues on its journey and is not required to wait.

(i) Find the probability that a journey from A to B is completed in less than the scheduled time of 24 minutes. [3]

(ii) Find the probability that a journey from A to C is completed in less than the scheduled time of 42 minutes. [3]

(iii) It is proposed to introduce a system of bus lanes in the urban area. It is believed that this would mean that the journey time from A to B would be given by the random variable $0.85X$. Assuming this to be the case, find the probability that a journey from A to B would be completed in less than the currently scheduled time of 24 minutes. [3]

(iv) An alternative proposal is to introduce an express service. This would leave out some bus stops on both sections of the route and its overall journey time from A to C would be given by the random variable $0.9X + 0.8Y$. The scheduled time from A to C is to be given as a whole number of minutes. Find the least possible scheduled time such that, with probability 0.75, buses would complete the journey on time or early. [6]

(v) A programme of minor road improvements is undertaken on the country section. After their completion, it is thought that the random variable giving the journey time from B to C is still Normally distributed with standard deviation 2 minutes. A random sample of 15 journeys is found to have a sample mean journey time from B to C of 13.4 minutes. Provide a two-sided 95% confidence interval for the population mean journey time from B to C. [3]

3 An employer has commissioned an opinion polling organisation to undertake a survey of the attitudes of staff to proposed changes in the pension scheme. The staff are categorised as management, professional and administrative, and it is thought that there might be considerable differences of opinion between the categories. There are 60, 140 and 300 staff respectively in the categories. The budget for the survey allows for a sample of 40 members of staff to be selected for in-depth interviews.

(i) Explain why it would be unwise to select a simple random sample from all the staff. [2]

(ii) Discuss whether it would be sensible to consider systematic sampling. [3]

(iii) What are the advantages of stratified sampling in this situation? [2]

(iv) State the sample sizes in each category if stratified sampling with as nearly as possible proportional allocation is used. [1]

The opinion polling organisation needs to estimate the average wealth of staff in the categories, in terms of property, savings, investments and so on. In a random sample of 11 professional staff, the sample mean is £345 818 and the sample standard deviation is £69 241.

(v) Assuming the underlying population is Normally distributed, test at the 5% level of significance the null hypothesis that the population mean is £300 000 against the alternative hypothesis that it is greater than £300 000. Provide also a two-sided 95% confidence interval for the population mean. [10]

4 A company has many factories. It is concerned about incidents of trespassing and, in the hope of reducing if not eliminating these, has embarked on a programme of installing new fencing.

(i) Records for a random sample of 9 factories of the numbers of trespass incidents in typical weeks before and after installation of the new fencing are as follows.

Factory	A	B	C	D	E	F	G	H	I
Number before installation	8	12	6	4	14	22	4	13	14
Number after installation	6	11	0	1	18	10	11	5	4

Use a Wilcoxon test to examine at the 5% level of significance whether it appears that, on the whole, the number of trespass incidents per week is lower after the installation of the new fencing than before. [9]

(ii) Records are also available of the costs of damage from typical trespass incidents before and after the introduction of the new fencing for a random sample of 7 factories, as follows (in £).

Factory	T	U	V	W	X	Y	Z
Cost before installation	1215	95	546	467	2356	236	550
Cost after installation	1268	110	578	480	2417	318	620

Stating carefully the required distributional assumption, provide a two-sided 99% confidence interval based on a *t* distribution for the population mean difference between costs of damage before and after installation of the new fencing.

Explain why this confidence interval should not be based on the Normal distribution. [9]

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE UNIT **4768/01**

MATHEMATICS (MEI)

Statistics 3

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 **4** printed pages.

1 The continuous random variable X has probability density function

$$f(x) = k(1 - x) \quad \text{for } 0 \leqslant x \leqslant 1$$

where k is a constant.

(i) Show that $k = 2$. Sketch the graph of the probability density function. [4]

(ii) Find $E(X)$ and show that $\text{Var}(X) = \frac{1}{18}$. [5]

(iii) Derive the cumulative distribution function of X. Hence find the probability that X is greater than the mean. [4]

(iv) Verify that the median of X is $1 - \dfrac{1}{\sqrt{2}}$. [2]

(v) $\overline{X}$ is the mean of a random sample of 100 observations of X. Write down the approximate distribution of $\overline{X}$. [3]

2 The manager of a large country estate is preparing to plant an area of woodland. He orders a large number of saplings (young trees) from a nursery. He selects a random sample of 12 of the saplings and measures their heights, which are as follows (in metres).

0.63 0.62 0.58 0.56 0.59 0.62 0.64 0.58 0.55 0.61 0.56 0.52

(i) The manager requires that the mean height of saplings at planting is at least 0.6 metres. Carry out the usual t test to examine this, using a 5% significance level. State your hypotheses and conclusion carefully. What assumption is needed for the test to be valid? [11]

(ii) Find a 95% confidence interval for the true mean height of saplings. Explain carefully what is meant by a 95% confidence interval. [5]

(iii) Suppose the assumption needed in part **(i)** cannot be justified. Identify an alternative test that the manager could carry out in order to check that the saplings meet his requirements, and state the null hypothesis for this test. [2]

3 Bill and Ben run their own gardening company. At regular intervals throughout the summer they come to work on my garden, mowing the lawns, hoeing the flower beds and pruning the bushes. From past experience it is known that the times, in minutes, spent on these tasks can be modelled by independent Normally distributed random variables as follows.

	Mean	Standard deviation
Mowing	44	4.8
Hoeing	32	2.6
Pruning	21	3.7

(i) Find the probability that, on a randomly chosen visit, it takes less than 50 minutes to mow the lawns. [3]

(ii) Find the probability that, on a randomly chosen visit, the total time for hoeing and pruning is less than 50 minutes. [3]

(iii) If Bill mows the lawns while Ben does the hoeing and pruning, find the probability that, on a randomly chosen visit, Ben finishes first. [4]

Bill and Ben do my gardening twice a month and send me an invoice at the end of the month.

(iv) Write down the mean and variance of the **total** time (in minutes) they spend on mowing, hoeing and pruning per month. [2]

(v) The company charges for the **total** time spent at 15 pence per minute. There is also a fixed charge of £10 per month. Find the probability that the total charge for a month does not exceed £40. [6]

4 **(a)** An amateur weather forecaster has been keeping records of air pressure, measured in atmospheres. She takes the measurement at the same time every day using a barometer situated in her garden. A random sample of 100 of her observations is summarised in the table below. The corresponding expected frequencies for a Normal distribution, with its two parameters estimated by sample statistics, are also shown in the table.

Pressure (a atmospheres)	Observed frequency	Frequency as given by Normal model
$a \leqslant 0.98$	4	1.45
$0.98 < a \leqslant 0.99$	6	5.23
$0.99 < a \leqslant 1.00$	9	13.98
$1.00 < a \leqslant 1.01$	15	23.91
$1.01 < a \leqslant 1.02$	37	26.15
$1.02 < a \leqslant 1.03$	21	18.29
$1.03 < a$	8	10.99

Carry out a test at the 5% level of significance of the goodness of fit of the Normal model. State your conclusion carefully and comment on your findings. [9]

(b) The forecaster buys a new digital barometer that can be linked to her computer for easier recording of observations. She decides that she wishes to compare the readings of the new barometer with those of the old one. For a random sample of 10 days, the readings (in atmospheres) of the two barometers are shown below.

Day	A	B	C	D	E	F	G	H	I	J
Old	0.992	1.005	1.001	1.011	1.026	0.980	1.020	1.025	1.042	1.009
New	0.985	1.003	1.002	1.014	1.022	0.988	1.030	1.016	1.047	1.025

Use an appropriate Wilcoxon test to examine at the 10% level of significance whether there is any reason to suppose that, on the whole, readings on the old and new barometers do not agree. [9]

ADVANCED GCE UNIT

MATHEMATICS (MEI)

4768/01

Statistics 3

TUESDAY 5 JUNE 2007

Afternoon

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.

1 A manufacturer of fireworks is investigating the lengths of time for which the fireworks burn. For a particular type of firework this length of time, in minutes, is modelled by the random variable T with probability density function

$$\mathrm{f}(t) = kt^3(2-t) \quad \text{for } 0 < t \leqslant 2$$

where k is a constant.

(i) Show that $k = \frac{5}{8}$. [2]

(ii) Find the modal time. [2]

(iii) Find $\mathrm{E}(T)$ and show that $\mathrm{Var}(T) = \frac{8}{63}$. [5]

(iv) A large random sample of n fireworks of this type is tested. Write down in terms of n the approximate distribution of $\overline{T}$, the sample mean time. [3]

(v) For a random sample of 100 such fireworks the times are summarised as follows.

$$\Sigma t = 145.2 \qquad \Sigma t^2 = 223.41$$

Find a 95% confidence interval for the mean time for this type of firework and hence comment on the appropriateness of the model. [6]

2 The operator of a section of motorway toll road records its weekly takings according to the types of vehicles using the motorway. For purposes of charging, there are three types of vehicle: cars, coaches, lorries. The weekly takings (in thousands of pounds) for each type are assumed to be Normally distributed. These distributions are independent of each other and are summarised in the table.

Vehicle type	Mean	Standard deviation
Cars	60.2	5.2
Coaches	33.9	6.3
Lorries	52.4	4.9

(i) Find the probability that the weekly takings for coaches are less than £40 000. [3]

(ii) Find the probability that the weekly takings for lorries exceed the weekly takings for cars. [4]

(iii) Find the probability that over a 4-week period the total takings for cars exceed £225 000. What assumption must be made about the four weeks? [5]

(iv) Each week the operator allocates part of the takings for repairs. This is determined for each type of vehicle according to estimates of the long-term damage caused. It is calculated as follows: 5% of takings for cars, 10% for coaches and 20% for lorries. Find the probability that in any given week the total amount allocated for repairs will exceed £20 000. [6]

3 The management of a large chain of shops aims to reduce the level of absenteeism among its workforce by means of an incentive bonus scheme. In order to evaluate the effectiveness of the scheme, the management measures the percentage of working days lost before and after its introduction for each of a random sample of 11 shops. The results are shown below.

Shop	A	B	C	D	E	F	G	H	I	J	K
% days lost before	3.5	5.0	3.5	3.2	4.5	4.9	4.1	6.0	6.8	8.1	6.0
% days lost after	1.8	4.3	2.9	4.5	4.4	5.8	3.5	6.7	6.4	5.4	5.1

(a) The management decides to carry out a t test to investigate whether there has been a reduction in absenteeism.

(i) State clearly the hypotheses that should be used together with any necessary assumptions. [4]

(ii) Carry out the test using a 5% significance level. [7]

(b) Find a 95% confidence interval for the true mean percentage of days lost after the introduction of the incentive scheme and state any assumption needed. The management has set a target that the mean percentage should be 3.5. Do you think this has been achieved? Explain your answer. [7]

4 A machine produces plastic strip in a continuous process. Occasionally there is a flaw at some point along the strip. The length of strip (in hundreds of metres) between successive flaws is modelled by a continuous random variable X with probability density function $f(x) = \frac{18}{(3+x)^3}$ for $x > 0$. The table below gives the frequencies for 100 randomly chosen observations of X. It also gives the probabilities for the class intervals using the model.

Length x (hundreds of metres)	Observed frequency	Probability
$0 < x \leqslant 0.5$	21	0.2653
$0.5 < x \leqslant 1$	24	0.1722
$1 < x \leqslant 2$	12	0.2025
$2 < x \leqslant 3$	15	0.1100
$3 < x \leqslant 5$	13	0.1094
$5 < x \leqslant 10$	9	0.0874
$x > 10$	6	0.0532

(i) Examine the fit of this model to the data at the 5% level of significance. [9]

You are given that the median length between successive flaws is 124 metres. At a later date the following random sample of ten lengths (in metres) between flaws is obtained.

239 77 179 221 100 312 52 129 236 42

(ii) Test at the 10% level of significance whether the median length may still be assumed to be 124 metres. [9]

OCR RECOGNISING ACHIEVEMENT

ADVANCED GCE **4768/01**

MATHEMATICS (MEI)

Statistics 3

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)** The time (in milliseconds) taken by my computer to perform a particular task is modelled by the random variable T. The probability that it takes more than t milliseconds to perform this task is given by the expression $\mathrm{P}(T > t) = \dfrac{k}{t^2}$ for $t \geqslant 1$, where k is a constant.

(i) Write down the cumulative distribution function of T and hence show that $k = 1$. [3]

(ii) Find the probability density function of T. [2]

(iii) Find the mean time for the task. [3]

(b) For a different task, the times (in milliseconds) taken by my computer on 10 randomly chosen occasions were as follows.

6.4 5.9 5.0 6.2 6.8 6.0 5.2 6.5 5.7 5.3

From past experience it is thought that the median time for this task is 5.4 milliseconds. Carry out a test at the 5% level of significance to investigate this, stating your hypotheses carefully. [10]

2 In the vegetable section of a local supermarket, leeks are on sale either loose (and unprepared) or prepared in packs of 4.

The weights of unprepared leeks are modelled by the random variable X which has the Normal distribution with mean 260 grams and standard deviation 24 grams. The prepared leeks have had 40% of their weight removed, so that their weights, Y, are modelled by $Y = 0.6X$.

(i) Find the probability that a randomly chosen unprepared leek weighs less than 300 grams. [3]

(ii) Find the probability that a randomly chosen prepared leek weighs more than 175 grams. [3]

(iii) Find the probability that the total weight of 4 randomly chosen prepared leeks in a pack is less than 600 grams. [3]

(iv) What total weight of prepared leeks in a randomly chosen pack of 4 is exceeded with probability 0.975? [3]

(v) Sandie is making soup. She uses 3 unprepared leeks and 2 onions. The weights of onions are modelled by the Normal distribution with mean 150 grams and standard deviation 18 grams. Find the probability that the total weight of her ingredients is more than 1000 grams. [3]

(vi) A large consignment of unprepared leeks is delivered to the supermarket. A random sample of 100 of them is taken. Their weights have sample mean 252.4 grams and sample standard deviation 24.6 grams. Find a 99% confidence interval for the true mean weight of the leeks in this consignment. [3]

3 Engineers in charge of a chemical plant need to monitor the temperature inside a reaction chamber. Past experience has shown that when functioning correctly the temperature inside the chamber can be modelled by a Normal distribution with mean 380 °C. The engineers are concerned that the mean operating temperature may have fallen. They decide to test the mean using the following random sample of 12 recent temperature readings.

374.0	378.1	363.0	357.0	377.9	388.4
379.6	372.4	362.4	377.3	385.2	370.6

(i) Give three reasons why a *t* test would be appropriate. [3]

(ii) Carry out the test using a 5% significance level. State your hypotheses and conclusion carefully. [9]

(iii) Find a 95% confidence interval for the true mean temperature in the reaction chamber. [4]

(iv) Describe briefly one advantage and one disadvantage of having a 99% confidence interval instead of a 95% confidence interval. [2]

4 **(a)** In Germany, towards the end of the nineteenth century, a study was undertaken into the distribution of the sexes in families of various sizes. The table shows some data about the numbers of girls in 500 families, each with 5 children. It is thought that the binomial distribution B(5, *p*) should model these data.

Number of girls	Number of families
0	32
1	110
2	154
3	125
4	63
5	16

(i) Use this information to calculate an estimate for the mean number of girls per family of 5 children. Hence show that 0.45 can be taken as an estimate of *p*. [3]

(ii) Investigate at a 5% significance level whether the binomial model with *p* estimated as 0.45 fits the data. Comment on your findings and also on the extent to which the conditions for a binomial model are likely to be met. [12]

(b) A researcher wishes to select 50 families from the 500 in part **(a)** for further study. Suggest what sort of sample she might choose and describe how she should go about choosing it. [3]

ADVANCED GCE

4768/01

MATHEMATICS (MEI)

Statistics 3

WEDNESDAY 21 MAY 2008

Afternoon

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.

1 **(a)** Sarah travels home from work each evening by bus; there is a bus every 20 minutes. The time at which Sarah arrives at the bus stop varies randomly in such a way that the probability density function of X, the length of time in minutes she has to wait for the next bus, is given by

$$\mathrm{f}(x) = k(20 - x) \text{ for } 0 \leqslant x \leqslant 20, \text{ where } k \text{ is a constant.}$$

(i) Find k. Sketch the graph of $\mathrm{f}(x)$ and use its shape to explain what can be deduced about how long Sarah has to wait. [5]

(ii) Find the cumulative distribution function of X and hence, or otherwise, find the probability that Sarah has to wait more than 10 minutes for the bus. [4]

(iii) Find the median length of time that Sarah has to wait. [3]

(b) **(i)** Define the term 'simple random sample'. [2]

(ii) Explain briefly how to carry out cluster sampling. [3]

(iii) A researcher wishes to investigate the attitudes of secondary school pupils to pollution. Explain why he might prefer to collect his data using a cluster sample rather than a simple random sample. [2]

2 An electronics company purchases two types of resistor from a manufacturer. The resistances of the resistors (in ohms) are known to be Normally distributed. Type A have a mean of 100 ohms and standard deviation of 1.9 ohms. Type B have a mean of 50 ohms and standard deviation of 1.3 ohms.

(i) Find the probability that the resistance of a randomly chosen resistor of type A is less than 103 ohms. [3]

(ii) Three resistors of type A are chosen at random. Find the probability that their total resistance is more than 306 ohms. [3]

(iii) One resistor of type A and one resistor of type B are chosen at random. Find the probability that their total resistance is more than 147 ohms. [3]

(iv) Find the probability that the total resistance of two randomly chosen type B resistors is within 3 ohms of one randomly chosen type A resistor. [5]

(v) The manufacturer now offers type C resistors which are specified as having a mean resistance of 300 ohms. The resistances of a random sample of 100 resistors from the first batch supplied have sample mean 302.3 ohms and sample standard deviation 3.7 ohms. Find a 95% confidence interval for the true mean resistance of the resistors in the batch. Hence explain whether the batch appears to be as specified. [4]

3 **(a)** A tea grower is testing two types of plant for the weight of tea they produce. A trial is set up in which each type of plant is grown at each of 8 sites. The total weight, in grams, of tea leaves harvested from each plant is measured and shown below.

Site	A	B	C	D	E	F	G	H
Type I	225.2	268.9	303.6	244.1	230.6	202.7	242.1	247.5
Type II	215.2	242.1	260.9	241.7	245.5	204.7	225.8	236.0

(i) The grower intends to perform a *t* test to examine whether there is any difference in the mean yield of the two types of plant. State the hypotheses he should use and also any necessary assumption. [3]

(ii) Carry out the test using a 5% significance level. [7]

(b) The tea grower deals with many types of tea and employs tasters to rate them. The tasters do this by giving each tea a score out of 100. The tea grower wishes to compare the scores given by two of the tasters. Their scores for a random selection of 10 teas are as follows.

Tea	Q	R	S	T	U	V	W	X	Y	Z
Taster 1	69	79	85	63	81	65	85	86	89	77
Taster 2	74	75	99	66	75	64	96	94	96	86

Use a Wilcoxon test to examine, at the 5% level of significance, whether it appears that, on the whole, the scores given to teas by these two tasters differ. [8]

4 **(a)** A researcher is investigating the feeding habits of bees. She sets up a feeding station some distance from a beehive and, over a long period of time, records the numbers of bees arriving each minute. For a random sample of 100 one-minute intervals she obtains the following results.

Number of bees	0	1	2	3	4	5	6	7	$\geqslant 8$
Number of intervals	6	16	19	18	17	14	6	4	0

(i) Show that the sample mean is 3.1 and find the sample variance. Do these values support the possibility of a Poisson model for the number of bees arriving each minute? Explain your answer. [3]

(ii) Use the mean in part **(i)** to carry out a test of the goodness of fit of a Poisson model to the data. [10]

(b) The researcher notes the length of time, in minutes, that each bee spends at the feeding station. The times spent are assumed to be Normally distributed. For a random sample of 10 bees, the mean is found to be 1.465 minutes and the standard deviation is 0.3288 minutes. Find a 95% confidence interval for the overall mean time. [4]

OCR RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) **4768**

Statistics 3

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

Thursday 15 January 2009
Morning

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)** A continuous random variable X has probability density function

$$f(x) = \lambda x^c, \quad 0 \leqslant x \leqslant 1,$$

where c is a constant and the parameter λ is greater than 1.

(i) Find c in terms of λ. **[3]**

(ii) Find $E(X)$ in terms of λ. **[3]**

(iii) Show that $\text{Var}(X) = \dfrac{\lambda}{(\lambda + 2)(\lambda + 1)^2}$. **[4]**

(b) Every day, Godfrey does a puzzle from the newspaper and records the time taken in minutes. Last year, his median time was 32 minutes. His times for a random sample of 12 puzzles this year are as follows.

40 20 18 11 47 36 38 35 22 14 12 21

Use an appropriate test, with a 5% significance level, to examine whether Godfrey's times this year have decreased on the whole. **[8]**

2 A factory manufactures paperweights consisting of glass mounted on a wooden base. The volume of glass, in cm^3, in a paperweight has a Normal distribution with mean 56.5 and standard deviation 2.9. The volume of wood, in cm^3, also has a Normal distribution with mean 38.4 and standard deviation 1.1. These volumes are independent of each other. For the purpose of quality control, paperweights for testing are chosen at random from the factory's output.

(i) Find the probability that the volume of glass in a randomly chosen paperweight is less than $60\,cm^3$. **[3]**

(ii) Find the probability that the total volume of a randomly chosen paperweight is more than $100\,cm^3$. **[3]**

The glass has a mass of 3.1 grams per cm^3 and the wood has a mass of 0.8 grams per cm^3.

(iii) Find the probability that the total mass of a randomly chosen paperweight is between 200 and 220 grams. **[6]**

(iv) The factory manager introduces some modifications intended to reduce the mean mass of the paperweights to 200 grams or less. The variance is also affected but not the Normality. Subsequently, for a random sample of 10 paperweights, the sample mean mass is 205.6 grams and the sample standard deviation is 8.51 grams. Is there evidence, at the 5% level of significance, that the intended reduction of the mean mass has not been achieved? **[6]**

3 Pathology departments in hospitals routinely analyse blood specimens. Ideally the analysis should be done while the specimens are fresh to avoid any deterioration, but this is not always possible. A researcher decides to study the effect of freezing specimens for later analysis by measuring the concentrations of a particular hormone before and after freezing. He collects and divides a sample of 15 specimens. One half of each specimen is analysed immediately, the other half is frozen and analysed a month later. The concentrations of the particular hormone (in suitable units) are as follows.

Immediately	15.21	13.36	15.97	21.07	12.82	10.80	11.50	12.05
After freezing	15.96	10.65	13.38	15.00	12.11	12.65	12.48	8.49

Immediately	10.90	18.48	13.43	13.16	16.62	14.91	17.08
After freezing	9.13	15.53	11.84	8.99	16.24	14.03	16.13

A t test is to be used in order to see if, on average, there is a reduction in hormone concentration as a result of being frozen.

(i) Explain why a paired test is appropriate in this situation. **[2]**

(ii) State the hypotheses that should be used, together with any necessary assumptions. **[4]**

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

(iv) A p% confidence interval for the true mean reduction in hormone concentration is found to be (0.4869, 2.8131). Determine the value of p. **[5]**

4 **(i)** Explain the meaning of 'opportunity sampling'. Give one reason why it might be used and state one disadvantage of using it. **[3]**

A market researcher is conducting an 'on-street' survey in a busy city centre, for which he needs to stop and interview 100 people. For each interview the researcher counts the number of people he has to ask until one agrees to be interviewed. The data collected are as follows.

No. of people asked	1	2	3	4	5	6	7 or more
Frequency	26	19	17	13	11	8	6

A model for these data is proposed as follows, where p (assumed constant throughout) is the probability that a person asked agrees to be interviewed, and $q = 1 - p$.

No. of people asked	1	2	3	4	5	6	7 or more
Probability	p	pq	pq^2	pq^3	pq^4	pq^5	q^6

(ii) Verify that these probabilities add to 1 whatever the value of p. **[2]**

(iii) Initially it is thought that on average 1 in 4 people asked agree to be interviewed. Test at the 10% level of significance whether it is reasonable to suppose that the model applies with $p = 0.25$. **[9]**

(iv) Later an estimate of p obtained from the data is used in the analysis. The value of the test statistic (with no combining of cells) is found to be 9.124. What is the outcome of this new test? Comment on your answer in relation to the outcome of the test in part **(iii)**. **[4]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) **4768**

Statistics 3

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

Wednesday 17 June 2009
Morning

Duration: 1 hour 30 minutes

* 4 7 6 8 *

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.

3R–8H16

1 Andy, a carpenter, constructs wooden shelf units for storing CDs. The wood used for the shelves has a thickness which is Normally distributed with mean 14 mm and standard deviation 0.55 mm. Andy works to a design which allows a gap of 145 mm between the shelves, but past experience has shown that the gap is Normally distributed with mean 144 mm and standard deviation 0.9 mm. Dimensions of shelves and gaps are assumed to be independent of each other.

(i) Find the probability that a randomly chosen gap is less than 145 mm. **[3]**

(ii) Find the probability that the combined height of a gap and a shelf is more than 160 mm. **[3]**

A complete unit has 7 shelves and 6 gaps.

(iii) Find the probability that the overall height of a unit lies between 960 mm and 965 mm. Hence find the probability that at least 3 out of 4 randomly chosen units are between 960 mm and 965 mm high. **[7]**

(iv) I buy two randomly chosen CD units made by Andy. The probability that the difference in their heights is less than h mm is 0.95. Find h. **[5]**

2 Pat makes and sells fruit cakes at a local market. On her stall a sign states that the average weight of the cakes is 1 kg. A trading standards officer carries out a routine check of a random sample of 8 of Pat's cakes to ensure that they are not underweight, on average. The weights, in kg, that he records are as follows.

0.957 1.055 0.983 0.917 1.015 0.865 1.013 0.854

(i) On behalf of the trading standards officer, carry out a suitable test at a 5% level of significance, stating your hypotheses clearly. Assume that the weights of Pat's fruit cakes are Normally distributed. **[9]**

(ii) Find a 95% confidence interval for the true mean weight of Pat's fruit cakes. **[4]**

Pat's husband, Tony, is the owner of a factory which makes and supplies fruit cakes to a large supermarket chain. A large random sample of n of these cakes has mean weight $\bar{x}$ kg and variance 0.006 kg^2.

(iii) Write down, in terms of n and $\bar{x}$, a 95% confidence interval for the true mean weight of cakes produced in Tony's factory. **[3]**

(iv) What is the size of the smallest sample that should be taken if the width of the confidence interval in part **(iii)** is to be 0.025 kg at most? **[3]**

3 A company which employs 600 staff wishes to improve its image by introducing new uniforms for the staff to wear. The human resources manager would like to obtain the views of the staff. She decides to do this by means of a systematic sample of 10% of the staff.

(i) How should she go about obtaining such a sample, ensuring that all members of staff are equally likely to be selected? Explain whether this constitutes a simple random sample. **[5]**

At a later stage in the process, the choice of uniform has been reduced to two possibilities. Twelve members of staff are selected to take part in deciding which of the two uniforms to adopt. Each of the twelve assesses each uniform for comfort, appearance and practicality, giving it a total score out of 10. The scores are as follows.

Staff member	1	2	3	4	5	6	7	8	9	10	11	12
Uniform A	4.2	2.6	10.0	9.0	8.2	2.8	5.0	7.4	2.8	6.8	10.0	9.8
Uniform B	5.0	5.2	1.4	2.8	2.2	6.4	7.4	7.8	6.8	1.2	3.4	7.6

A Wilcoxon signed rank test is to be used to decide whether there is any evidence of a preference for one of the uniforms.

(ii) Explain why this test is appropriate in these circumstances and state the hypotheses that should be used. **[4]**

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

4 A random variable X has probability density function $f(x) = \dfrac{2x}{\lambda^2}$ for $0 < x < \lambda$, where λ is a positive constant.

(i) Show that, for any value of λ, $f(x)$ is a valid probability density function. **[3]**

(ii) Find μ, the mean value of X, in terms of λ and show that $P(X < \mu)$ does not depend on λ. **[4]**

(iii) Given that $E(X^2) = \dfrac{\lambda^2}{2}$, find σ^2, the variance of X, in terms of λ. **[2]**

The random variable X is used to model the depth of the space left by the filling machine at the top of a jar of jam. The model gives the following probabilities for X (whatever the value of λ).

$0 < X \leqslant \mu - \sigma$	$\mu - \sigma < X \leqslant \mu$	$\mu < X \leqslant \mu + \sigma$	$\mu + \sigma < X < \lambda$
0.18573	0.25871	0.36983	0.18573

A sample of 50 random observations of X, classified in the same way, is summarised by the following frequencies.

4	11	20	15

(iv) Carry out a suitable test at the 5% level of significance to assess the goodness of fit of X to these data. Explain briefly how your conclusion may be affected by the choice of significance level. **[9]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) **4768**

Statistics 3

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

Wednesday 20 January 2010
Afternoon

Duration: 1 hour 30 minutes

4768

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 Coastal wildlife wardens are monitoring populations of herring gulls. Herring gulls usually lay 3 eggs per nest and the wardens wish to model the number of eggs per nest that hatch. They assume that the situation can be modelled by the binomial distribution $B(3, p)$ where p is the probability that an egg hatches. A random sample of 80 nests each containing 3 eggs has been observed with the following results.

Number of eggs hatched	0	1	2	3
Number of nests	7	23	29	21

(i) Initially it is assumed that the value of p is $\frac{1}{2}$. Test at the 5% level of significance whether it is reasonable to suppose that the model applies with $p = \frac{1}{2}$. **[10]**

(ii) The model is refined by estimating p from the data. Find the mean of the observed data and hence an estimate of p. **[2]**

(iii) Using the estimated value of p, the value of the test statistic X^2 turns out to be 2.3857. Is it reasonable to suppose, at the 5% level of significance, that this refined model applies? **[3]**

(iv) Discuss the reasons for the different outcomes of the tests in parts **(i)** and **(iii)**. **[2]**

2 **(a)** A continuous random variable, X, has probability density function

$$f(x) = \begin{cases} \frac{1}{72}(8x - x^2) & 2 \leqslant x \leqslant 8, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $F(x)$, the cumulative distribution function of X. **[3]**

(ii) Sketch $F(x)$. **[3]**

(iii) The median of X is m. Show that m satisfies the equation $m^3 - 12m^2 + 148 = 0$. Verify that $m \approx 4.42$. **[3]**

(b) The random variable in part **(a)** is thought to model the weights, in kilograms, of lambs at birth. The birth weights, in kilograms, of a random sample of 12 lambs, given in ascending order, are as follows.

3.16 3.62 3.80 3.90 4.02 4.72 5.14 6.36 6.50 6.58 6.68 6.78

Test at the 5% level of significance whether a median of 4.42 is consistent with these data. **[10]**

3 Cholesterol is a lipid (fat) which is manufactured by the liver from the fatty foods that we eat. It plays a vital part in allowing the body to function normally. However, when high levels of cholesterol are present in the blood there is a risk of arterial disease. Among the factors believed to assist with achieving and maintaining low cholesterol levels are weight loss and exercise.

A doctor wishes to test the effectiveness of exercise in lowering cholesterol levels. For a random sample of 12 of her patients, she measures their cholesterol levels before and after they have followed a programme of exercise. The measurements obtained are as follows.

Patient	A	B	C	D	E	F	G	H	I	J	K	L
Before	5.7	5.7	4.0	6.8	7.4	5.5	6.7	6.4	7.2	7.2	7.1	4.4
After	5.8	4.0	5.2	5.7	6.0	5.0	5.8	4.2	7.3	5.2	6.4	4.1

(i) A t test is to be used in order to see if, on average, the exercise programme seems to be effective in lowering cholesterol levels. State the distributional assumption necessary for the test, and carry out the test using a 1% significance level. **[11]**

(ii) A second random sample of 12 patients gives a 95% confidence interval of (−0.5380, 1.4046) for the true mean reduction (before – after) in cholesterol level. Find the mean and standard deviation for this sample. How might the doctor interpret this interval in relation to the exercise programme? **[7]**

4 The weights of a particular variety (A) of tomato are known to be Normally distributed with mean 80 grams and standard deviation 11 grams.

(i) Find the probability that a randomly chosen tomato of variety A weighs less than 90 grams. **[3]**

The weights of another variety (B) of tomato are known to be Normally distributed with mean 70 grams. These tomatoes are packed in sixes using packaging that weighs 15 grams.

(ii) The probability that a randomly chosen pack of 6 tomatoes of variety B, including packaging, weighs less than 450 grams is 0.8463. Show that the standard deviation of the weight of single tomatoes of variety B is 6 grams, to the nearest gram. **[5]**

(iii) Tomatoes of variety A are packed in fives using packaging that weighs 25 grams. Find the probability that the total weight of a randomly chosen pack of variety A is greater than the total weight of a randomly chosen pack of variety B. **[5]**

(iv) A new variety (C) of tomato is introduced. The weights, c grams, of a random sample of 60 of these tomatoes are measured giving the following results.

$$\Sigma c = 3126.0 \qquad \Sigma c^2 = 164\,223.96$$

Find a 95% confidence interval for the true mean weight of these tomatoes. **[5]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) **4768**

Statistics 3

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

Tuesday 22 June 2010
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 **(i)** The manager of a company that employs 250 travelling sales representatives wishes to carry out a detailed analysis of the expenses claimed by the representatives. He has an alphabetical (by surname) list of the representatives. He chooses a sample of representatives by selecting the 10th, 20th, 30th and so on. Name the type of sampling the manager is attempting to use. Describe a weakness in his method of using it, and explain how he might overcome this weakness. **[3]**

The representatives each use their own cars to drive to meetings with customers. The total distance, in miles, travelled by a representative in a month is Normally distributed with mean 2018 and standard deviation 96.

(ii) Find the probability that, in a randomly chosen month, a randomly chosen representative travels more than 2100 miles. **[3]**

(iii) Find the probability that, in a randomly chosen 3-month period, a randomly chosen representative travels less than 6000 miles. What assumption is needed here? Give a reason why it may not be realistic. **[5]**

(iv) Each month every representative submits a claim for travelling expenses plus commission. Travelling expenses are paid at the rate of 45 pence per mile. The commission is 10% of the value of sales in that month. The value, in £, of the monthly sales has the distribution $N(21\,200,\ 1100^2)$. Find the probability that a randomly chosen claim lies between £3000 and £3300. **[7]**

2 William Sealy, a biochemistry student, is doing work experience at a brewery. One of his tasks is to monitor the specific gravity of the brewing mixture during the brewing process. For one particular recipe, an initial specific gravity of 1.040 is required. A random sample of 9 measurements of the specific gravity at the start of the process gave the following results.

1.046 1.048 1.039 1.055 1.038 1.054 1.038 1.051 1.038

(i) William has to test whether the specific gravity of the mixture meets the requirement. Why might a t test be used for these data and what assumption must be made? **[3]**

(ii) Carry out the test using a significance level of 10%. **[9]**

(iii) Find a 95% confidence interval for the true mean specific gravity of the mixture and explain what is meant by a 95% confidence interval. **[6]**

3 **(a)** In order to prevent and/or control the spread of infectious diseases, the Government has various vaccination programmes. One such programme requires people to receive a booster injection at the age of 18. It is felt that the proportion of people receiving this booster could be increased and a publicity campaign is undertaken for this purpose. In order to assess the effectiveness of this campaign, health authorities across the country are asked to report the percentage of 18-year-olds receiving the booster before and after the campaign. The results for a randomly chosen sample of 9 authorities are as follows.

Authority	A	B	C	D	E	F	G	H	I
Before	76	98	88	81	86	84	83	93	80
After	82	97	93	77	83	95	91	95	89

This sample is to be tested to see whether the campaign appears to have been successful in raising the percentage receiving the booster.

(i) Explain why the use of paired data is appropriate in this context. **[1]**

(ii) Carry out an appropriate Wilcoxon signed rank test using these data, at the 5% significance level. **[10]**

(b) Benford's Law predicts the following probability distribution for the first significant digit in some large data sets.

Digit	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

On one particular day, the first significant digits of the stock market prices of the shares of a random sample of 200 companies gave the following results.

Digit	1	2	3	4	5	6	7	8	9
Frequency	55	34	27	16	15	17	12	15	9

Test at the 10% level of significance whether Benford's Law provides a reasonable model in the context of share prices. **[7]**

[Question 4 is printed overleaf.]

4 A random variable X has an exponential distribution with probability density function $\mathrm{f}(x) = \lambda \mathrm{e}^{-\lambda x}$ for $x \geqslant 0$, where λ is a positive constant.

(i) Verify that $\int_0^\infty \mathrm{f}(x)\,\mathrm{d}x = 1$ and sketch $\mathrm{f}(x)$. **[5]**

(ii) In this part of the question you may use the following result.

$$\int_0^\infty x^r \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{r!}{\lambda^{r+1}} \quad \text{for } r = 0,\ 1,\ 2,\ \ldots\ .$$

Derive the mean and variance of X in terms of λ. **[6]**

The random variable X is used to model the lifetime, in years, of a particular type of domestic appliance. The manufacturer of the appliance states that, based on past experience, the mean lifetime is 6 years.

(iii) Let $\overline{X}$ denote the mean lifetime, in years, of a random sample of 50 appliances. Write down an approximate distribution for $\overline{X}$. **[4]**

(iv) A random sample of 50 appliances is found to have a mean lifetime of 7.8 years. Does this cast any doubt on the model? **[3]**

ADVANCED GCE

MATHEMATICS (MEI) **4768**

Statistics 3

Candidates answer on the answer booklet.

OCR supplied materials:
- 8 page answer booklet (sent with general stationery)
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Wednesday 19 January 2011
Afternoon

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet. Please write clearly and in capital letters.
- 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

- 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 Each month the amount of electricity, measured in kilowatt-hours (kWh), used by a particular household is Normally distributed with mean 406 and standard deviation 12.

(i) Find the probability that, in a randomly chosen month, less than 420 kWh is used. **[3]**

The charge for electricity used is 14.6 pence per kWh.

(ii) Write down the distribution of the total charge for the amount of electricity used in any one month. Hence find the probability that, in a randomly chosen month, the total charge is more than £60. **[3]**

(iii) The household receives a bill every three months. Assume that successive months may be regarded as independent of each other.

Find the value of b such that the probability that a randomly chosen bill is less than £b is 0.99. **[4]**

In a different household, the amount of electricity used per month was Normally distributed with mean 432 kWh. This household buys a new washing machine that is claimed to be cheaper to run than the old one. Over the next six months the amounts of electricity used, in kWh, are as follows.

404 433 420 423 413 440

(iv) Treating this as a random sample, carry out an appropriate test, with a 5% significance level, to see if there is any evidence to suggest that the amount of electricity used per month by this household has decreased on average. **[9]**

2 **(a)** **(i)** What is stratified sampling? Why would it be used? **[4]**

(ii) A local authority official wishes to conduct a survey of households in the borough. He decides to select a stratified sample of 2000 households using Council Tax property bands as the strata. At the time of the survey there are 79 368 households in the borough. The table shows the numbers of households in the different tax bands.

Tax band	A – B	C – D	E – F	G – H
Number of households	32 298	33 211	9739	4120

Calculate the number of households that the official should choose from each stratum in order to obtain his sample of 2000 households so that each stratum is represented proportionally. **[2]**

(b) **(i)** What assumption needs to be made when using a Wilcoxon single sample test? **[2]**

(ii) As part of an investigation into trends in local authority spending, one of the categories of expenditure considered was 'Highways and the Environment'. For a random sample of 10 local authorities, the percentages of their total expenditure spent on Highways and the Environment in 1999 and then in 2009 are shown in the table.

Local authority	A	B	C	D	E	F	G	H	I	J
1999	9.60	8.40	8.67	9.32	9.89	9.35	7.91	8.08	9.61	8.55
2009	8.94	8.42	7.87	8.41	10.17	10.11	8.31	9.76	9.54	9.67

Use a Wilcoxon test, with a significance level of 10%, to determine whether there appears to be any change to the average percentage of total expenditure spent on Highways and the Environment between 1999 and 2009. **[10]**

3 The masses, in kilograms, of a random sample of 100 chickens on sale in a large supermarket were recorded as follows.

Mass (m kg)	$m < 1.6$	$1.6 \leqslant m < 1.8$	$1.8 \leqslant m < 2.0$	$2.0 \leqslant m < 2.2$	$2.2 \leqslant m < 2.4$	$2.4 \leqslant m < 2.6$	$2.6 \leqslant m$
Frequency	2	8	30	42	11	5	2

(i) Assuming that the first and last classes are the same width as the other classes, calculate an estimate of the sample mean and show that the corresponding estimate of the sample standard deviation is 0.2227 kg. **[3]**

A Normal distribution using the mean and standard deviation found in part **(i)** is to be fitted to these data. The expected frequencies for the classes are as follows.

Mass (m kg)	$m < 1.6$	$1.6 \leqslant m < 1.8$	$1.8 \leqslant m < 2.0$	$2.0 \leqslant m < 2.2$	$2.2 \leqslant m < 2.4$	$2.4 \leqslant m < 2.6$	$2.6 \leqslant m$
Expected frequency	2.17	10.92	f	33.85	19.22	5.13	0.68

(ii) Use the Normal distribution to find f. **[3]**

(iii) Carry out a goodness of fit test of this Normal model using a significance level of 5%. **[9]**

(iv) Discuss the outcome of the test with reference to the contributions to the test statistic and to the possibility of other significance levels. **[3]**

4 A timber supplier cuts wooden fence posts from felled trees. The posts are of length $(k + X)$ cm where k is a constant and X is a random variable which has probability density function

$$f(x) = \begin{cases} 1 + x & -1 \leqslant x < 0, \\ 1 - x & 0 \leqslant x \leqslant 1, \\ 0 & \text{elsewhere.} \end{cases}$$

(i) Sketch $f(x)$. **[3]**

(ii) Write down the value of $E(X)$ and find $Var(X)$. **[5]**

(iii) Write down, in terms of k, the approximate distribution of $\bar{L}$, the mean length of a random sample of 50 fence posts. Justify your choice of distribution. **[4]**

(iv) In a particular sample of 50 posts, the mean length is 90.06 cm. Find a 95% confidence interval for the true mean length of the fence posts. **[4]**

(v) Explain whether it is reasonable to suppose that $k = 90$. **[1]**

OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) **4768**

Statistics 3

Candidates answer on the answer booklet.

OCR supplied materials:
- 8 page answer booklet
  (sent with general stationery)
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Thursday 23 June 2011
Morning

Duration: 1 hour 30 minutes

* 4 7 6 8 *

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet. Please write clearly and in capital letters.
- 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

- 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 Gerry runs 5000-metre races for his local athletics club. His coach has been monitoring his practice times for several months and he believes that they can be modelled using a Normal distribution with mean 15.3 minutes. The coach suggests that Gerry should try running with a pacemaker in order to see if this can improve his times. Subsequently a random sample of 10 of Gerry's times with the pacemaker is collected to see if any reduction has been achieved. The sample of times (in minutes) is as follows.

14.86 15.00 15.62 14.44 15.27 15.64 14.58 14.30 15.08 15.08

(i) Why might a t test be used for these data? **[2]**

(ii) Using a 5% significance level, carry out the test to see whether, on average, Gerry's times have been reduced. **[9]**

(iii) What is meant by 'a 5% significance level'? What would be the consequence of decreasing the significance level? **[3]**

(iv) Find a 95% confidence interval for the true mean of Gerry's times using a pacemaker. **[4]**

2 Scientists researching into the chemical composition of dust in space collect specimens using a specially designed spacecraft. The craft collects the particles of dust in trays that are made up of a large array of cells containing aerogel. The aerogel traps the particles that penetrate into the cells.

(i) For a random sample of 100 cells, the number of particles of dust in each cell was counted, giving the following results.

Number of particles	0	1	2	3	4	5	6	7	8	9	10+
Frequency	4	7	10	20	17	15	10	9	5	3	0

It is thought that the number of particles collected in each cell can be modelled using the distribution Poisson(4.2) since 4.2 is the sample mean for these data.

Some of the calculations for a χ^2 test are shown below. The cells for 8, 9 and 10+ particles have been combined.

Number of particles	...	5	6	7	8+
Observed frequency	...	15	10	9	8
Expected frequency	...	16.33	11.44	6.86	6.39
Contribution to X^2	...	0.1083	0.1813	0.6676	0.4056

Complete the calculations and carry out the test using a 10% significance level to see whether the number of particles per cell may be modelled in this way. **[12]**

(ii) The diameters of the dust particles are believed to be distributed symmetrically about a median of 15 micrometres (μm). For a random sample of 20 particles, the sum of the signed ranks of the diameters of the particles smaller than 15 μm (W_-) is found to be 53. Test at the 5% level of significance whether the median diameter appears to be more than 15 μm. **[6]**

3 The time, in hours, until an electronic component fails is represented by the random variable X. In this question two models for X are proposed.

(i) In one model, X has cumulative distribution function

$$\mathrm{G}(x) = \begin{cases} 0 & x \leqslant 0, \\ 1 - \left(1 + \dfrac{x}{200}\right)^{-2} & x > 0. \end{cases}$$

(A) Sketch $\mathrm{G}(x)$. **[3]**

(B) Find the interquartile range for this model. Hence show that a lifetime of more than 454 hours (to the nearest hour) would be classed as an outlier. **[6]**

(ii) In the alternative model, X has probability density function

$$\mathrm{f}(x) = \begin{cases} \frac{1}{200}\mathrm{e}^{-\frac{1}{200}x} & x > 0, \\ 0 & \text{elsewhere.} \end{cases}$$

(A) For this model show that the cumulative distribution function of X is

$$\mathrm{F}(x) = \begin{cases} 0 & x \leqslant 0, \\ 1 - \mathrm{e}^{-\frac{1}{200}x} & x > 0. \end{cases}$$ **[3]**

(B) Show that $\mathrm{P}(X > 50) = \mathrm{e}^{-0.25}$. **[2]**

(C) It is observed that a particular component is still working after 400 hours. Find the conditional probability that it will still be working after a further 50 hours (i.e. after a total of 450 hours) given that it is still working after 400 hours. **[4]**

4 The weights of Avonley Blue cheeses made by a small producer are found to be Normally distributed with mean 10 kg and standard deviation 0.4 kg.

(i) Find the probability that a randomly chosen cheese weighs less than 9.5 kg. **[3]**

One particular shop orders four Avonley Blue cheeses each week from the producer. From experience, the shopkeeper knows that the weekly demand from customers for Avonley Blue cheese is Normally distributed with mean 35 kg and standard deviation 3.5 kg. In the interests of food hygiene, no cheese is kept by the shopkeeper from one week to the next.

(ii) Find the probability that, in a randomly chosen week, demand from customers for Avonley Blue will exceed the supply. **[4]**

Following a campaign to promote Avonley Blue cheese, the shopkeeper finds that the weekly demand for it has increased by 30% (i.e. the mean and standard deviation are both increased by 30%). Therefore the shopkeeper increases his weekly order by one cheese.

(iii) Find the probability that, in a randomly chosen week, demand will now exceed supply. **[4]**

(iv) Following complaints, the cheese producer decides to check the mean weight of the Avonley Blue cheeses. For a random sample of 12 cheeses, she finds that the mean weight is 9.73 kg. Assuming that the population standard deviation of the weights is still 0.4 kg, find a 95% confidence interval for the true mean weight of the cheeses and comment on the result. Explain what is meant by a 95% confidence interval. **[7]**

THERE ARE NO QUESTIONS PRINTED ON THIS PAGE

OCR
RECOGNISING ACHIEVEMENT

Friday 20 January 2012 – Afternoon

A2 GCE MATHEMATICS (MEI)

4768 Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4768
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4732190112

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.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- 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)** Define simple random sampling. Describe briefly one difficulty associated with simple random sampling. **[4]**

(b) Freeze-drying is an economically important process used in the production of coffee. It improves the retention of the volatile aroma compounds. In order to maintain the quality of the coffee, technologists need to monitor the drying rate, measured in suitable units, at regular intervals. It is known that, for best results, the mean drying rate should be 70.3 units and anything substantially less than this would be detrimental to the coffee. Recently, a random sample of 12 observations of the drying rate was as follows.

66.0 66.1 59.8 64.0 70.9 71.4 66.9 76.2 65.2 67.9 69.2 68.5

(i) Carry out a test to investigate at the 5% level of significance whether the mean drying rate appears to be less than 70.3. State the distributional assumption that is required for this test. **[10]**

(ii) Find a 95% confidence interval for the true mean drying rate. **[4]**

2 In a particular chain of supermarkets, one brand of pasta shapes is sold in small packets and large packets. Small packets have a mean weight of 505 g and a standard deviation of 11 g. Large packets have a mean weight of 1005 g and a standard deviation of 17 g. It is assumed that the weights of packets are Normally distributed and are independent of each other.

(i) Find the probability that a randomly chosen large packet weighs between 995 g and 1020 g. **[3]**

(ii) Find the probability that the weights of two randomly chosen small packets differ by less than 25 g. **[3]**

(iii) Find the probability that the total weight of two randomly chosen small packets exceeds the weight of a randomly chosen large packet. **[4]**

(iv) Find the probability that the weight of one randomly chosen small packet exceeds half the weight of a randomly chosen large packet by at least 5 g. **[4]**

(v) A different brand of pasta shapes is sold in packets of which the weights are assumed to be Normally distributed with standard deviation 14 g. A random sample of 20 packets of this pasta is found to have a mean weight of 246 g. Find a 95% confidence interval for the population mean weight of these packets. **[4]**

3 **(a)** A medical researcher is looking into the delay, in years, between first and second myocardial infarctions (heart attacks). The following table shows the results for a random sample of 225 patients.

Delay (years)	0 –	1 –	2 –	3 –	4 – 10
Number of patients	160	40	13	9	3

The mean of this sample is used to construct a model which gives the following expected frequencies.

Delay (years)	0 –	1 –	2 –	3 –	4 – 10
Number of patients	142.23	52.32	19.25	7.08	4.12

Carry out a test, using a 2.5% level of significance, of the goodness of fit of the model to the data. **[9]**

(b) A further piece of research compares the incidence of myocardial infarction in men aged 55 to 70 with that in women aged 55 to 70. Incidence is measured by the number of infarctions per 10 000 of the population. For a random sample of 8 health authorities across the UK, the following results for the year 2010 were obtained.

Health authority	A	B	C	D	E	F	G	H
Incidence in men	47	56	15	51	45	54	50	32
Incidence in women	36	30	30	47	54	55	27	27

A Wilcoxon paired sample test, using the hypotheses H_0: $m = 0$ and H_1: $m \neq 0$ where m is the population median difference, is to be carried out to investigate whether there is any difference between men and women on the whole.

(i) Explain why a paired test is being used in this context. **[1]**

(ii) Carry out the test using a 10% level of significance. **[8]**

[Question 4 is printed overleaf.]

4 At the school summer fair, one of the games involves throwing darts at a circular dartboard of radius a lying on the ground some distance away. Only darts that land on the board are counted. The distance from the centre of the board to the point where a dart lands is modelled by the random variable R. It is assumed that the probability that a dart lands inside a circle of radius r is proportional to the area of the circle.

(i) By considering $P(R < r)$ show that $F(r)$, the cumulative distribution function of R, is given by

$$F(r) = \begin{cases} 0 & r < 0, \\ \dfrac{r^2}{a^2} & 0 \leqslant r \leqslant a, \\ 1 & r > a. \end{cases}$$ **[3]**

(ii) Find $f(r)$, the probability density function of R. **[2]**

(iii) Find $E(R)$ and show that $\text{Var}(R) = \dfrac{a^2}{18}$. **[7]**

The radius a of the dartboard is 22.5 cm.

(iv) Let $\bar{R}$ denote the mean distance from the centre of the board of a random sample of 100 darts. Write down an approximation to the distribution of $\bar{R}$. **[3]**

(v) A random sample of 100 darts is found to give a mean distance of 13.87 cm. Does this cast any doubt on the modelling? **[3]**

OCR RECOGNISING ACHIEVEMENT

Friday 22 June 2012 – Afternoon

A2 GCE MATHEMATICS (MEI)

4768 Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4768
- 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.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- 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 Technologists at a company that manufactures paint are trying to develop a new type of gloss paint with a shorter drying time than the current product. In order to test whether the drying time has been reduced, the technologists paint a square metre of each of the new and old paints on each of 10 different surfaces. The lengths of time, in hours, that each square metre takes to dry are as follows.

Surface	A	B	C	D	E	F	G	H	I	J
Old paint	16.6	17.0	16.5	15.6	16.3	16.5	16.4	15.9	16.3	16.1
New paint	15.9	16.3	16.3	15.9	15.5	16.6	16.1	16.0	16.2	15.6

(i) Explain why a paired sample is used in this context. **[1]**

(ii) The mean reduction in drying time is to be investigated. Why might a t test be appropriate in this context and what assumption needs to be made? **[4]**

(iii) Using a significance level of 5%, carry out a test to see if there appears to be any reduction in mean drying time. **[9]**

(iv) Find a 95% confidence interval for the true mean reduction in drying time. **[4]**

2 **(a)** **(i)** Give two reasons why an investigator might need to take a sample in order to obtain information about a population. **[2]**

(ii) State two requirements of a sample. **[2]**

(iii) Discuss briefly the advantage of the sampling being random. **[2]**

(b) **(i)** Under what circumstances might one use a Wilcoxon single sample test in order to test a hypothesis about the median of a population? What distributional assumption is needed for the test? **[2]**

(ii) On a stretch of road leading out of the centre of a town, highways officials have been monitoring the speed of the traffic in case it has increased. Previously it was known that the median speed on this stretch was 28.7 miles per hour. For a random sample of 12 vehicles on the stretch, the following speeds were recorded.

32.0 29.1 26.1 35.2 34.4 28.6 32.3 28.5 27.0 33.3 28.2 31.9

Carry out a test, with a 5% significance level, to see whether the speed of the traffic on this stretch of road seems to have increased on the whole. **[10]**

3 The triathlon is a sports event in which competitors take part in three stages, swimming, cycling and running, one straight after the other. The winner is the competitor with the shortest overall time. In this question the times for the separate stages are assumed to be Normally distributed and independent of each other.

For a particular triathlon event in which there was a very large number of competitors, the mean and standard deviation of the times, measured in minutes, for each stage were as follows.

	Mean	Standard deviation
Swimming	11.07	2.36
Cycling	57.33	8.76
Running	24.23	3.75

(i) For a randomly chosen competitor, find the probability that the swimming time is between 10 and 13 minutes. **[3]**

(ii) For a randomly chosen competitor, find the probability that the running time exceeds the swimming time by more than 10 minutes. **[4]**

(iii) For a randomly chosen competitor, find the probability that the swimming and running times combined exceed $\frac{2}{3}$ of the cycling time. **[4]**

(iv) In a different triathlon event the total times, in minutes, for a random sample of 12 competitors were as follows.

103.59 99.04 85.03 81.34 106.79 89.14 98.55 98.22 108.87 116.29 102.51 92.44

Find a 95% confidence interval for the mean time of all competitors in this event. **[5]**

(v) Discuss briefly whether the assumptions of Normality and independence for the stages of triathlon events are reasonable. **[2]**

[Question 4 is printed overleaf.]

4 The numbers of call-outs per day received by a fire station for a random sample of 255 weekdays were recorded as follows.

Number of call-outs	0	1	2	3	4	5 or more
Frequency (days)	145	79	22	6	3	0

The mean number of call-outs per day for these data is 0.6. A Poisson model, using this sample mean of 0.6, is fitted to the data, and gives the following expected frequencies (correct to 3 decimal places).

Number of call-outs	0	1	2	3	4	5 or more
Expected frequency	139.947	83.968	25.190	5.038	0.756	0.101

(i) Using a 5% significance level, carry out a test to examine the goodness of fit of the model to the data. **[9]**

The time T, measured in days, that elapses between successive call-outs can be modelled using the exponential distribution for which f(t), the probability density function, is

$$\mathrm{f}(t) = \begin{cases} 0 & t < 0, \\ \lambda \mathrm{e}^{-\lambda t} & t \geqslant 0, \end{cases}$$

where λ is a positive constant.

(ii) For the distribution above, it can be shown that $\mathrm{E}(T) = \frac{1}{\lambda}$. Given that the mean time between successive call-outs is $\frac{5}{3}$ days, write down the value of λ. **[1]**

(iii) Find F(t), the cumulative distribution function. **[3]**

(iv) Find the probability that the time between successive call-outs is more than 1 day. **[2]**

(v) Find the median time that elapses between successive call-outs. **[3]**

OCR RECOGNISING ACHIEVEMENT

Wednesday 23 January 2013 – Morning

A2 GCE MATHEMATICS (MEI)

4768/01 Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4768/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4733020113

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 certain industrial process requires a supply of water. It has been found that, for best results, the mean water pressure in suitable units should be 7.8. The water pressure is monitored by taking measurements at regular intervals. On a particular day, a random sample of the measurements is as follows.

7.50 7.64 7.68 7.51 7.70 7.85 7.34 7.72 7.74

These data are to be used to carry out a hypothesis test concerning the mean water pressure.

(i) Why is a test based on the Normal distribution not appropriate in this case? **[2]**

(ii) What distributional assumption is needed for a test based on the *t* distribution? **[1]**

(iii) Carry out a *t* test, with a 2% level of significance, to see whether it is reasonable to assume that the mean pressure is 7.8. **[9]**

(iv) Explain what is meant by a 95% confidence interval. **[2]**

(v) Find a 95% confidence interval for the actual mean water pressure. **[4]**

2 A particular species of reed that grows up to 2 metres in length is used for thatching. The lengths in metres of the reeds when harvested are modelled by the random variable X which has the following probability density function, $\mathrm{f}(x)$.

$$\mathrm{f}(x) = \begin{cases} \frac{3}{16}(4x - x^2) & \text{for } 0 \leqslant x \leqslant 2 \\ 0 & \text{elsewhere} \end{cases}$$

(i) Sketch $\mathrm{f}(x)$. **[3]**

(ii) Show that $\mathrm{E}(X) = \frac{5}{4}$ and find the standard deviation of the lengths of the harvested reeds. **[8]**

(iii) Find the standard error of the mean length for a random sample of 100 reeds. **[2]**

Once the harvested reeds have been collected, any that are shorter than 1 metre are discarded.

(iv) Find the proportion of reeds that should be discarded according to the model. **[2]**

(v) Reeds are harvested from a large area which is divided into several reed beds. A sample of the harvested reeds is required for quality control. How might the method of cluster sampling be used to obtain it? **[3]**

3 In the manufacture of child car seats, a resin made up of three ingredients is used. The ingredients are two polymers and an impact modifier. The resin is prepared in batches. Each ingredient is supplied by a separate feeder and the amount supplied to each batch, in kg, is assumed to be Normally distributed with mean and standard deviation as shown in the table below. The three feeders are also assumed to operate independently of each other.

	Mean	Standard deviation
Polymer 1	2025	44.6
Polymer 2	1565	21.8
Impact modifier	1410	33.8

(i) Find the probability that, in a randomly chosen batch of resin, there is no more than 2100 kg of polymer 1. **[3]**

(ii) Find the probability that, in a randomly chosen batch of resin, the amount of polymer 1 exceeds the amount of polymer 2 by at least 400 kg. **[4]**

(iii) Find the value of b such that the total amount of the ingredients in a randomly chosen batch exceeds b kg 95% of the time. **[4]**

(iv) Polymer 1 costs £1.20 per kg, polymer 2 costs £1.30 per kg and the impact modifier costs £0.80 per kg. Find the mean and variance of the total cost of a batch of resin. **[3]**

(v) Each batch of resin is used to make a large number of car seats from which a random sample of 50 seats is selected in order that the tensile strength (in suitable units) of the resin can be measured. From one such sample, the 99% confidence interval for the true mean tensile strength of the resin in that batch was calculated as (123.72, 127.38). Find the mean and standard deviation of the sample. **[4]**

[Question 4 is printed overleaf.]

4 **(a)** At a college, two examiners are responsible for marking, independently, the students' projects. Each examiner awards a mark out of 100 to each project. There is some concern that the examiners' marks do not agree, on average. Consequently a random sample of 12 projects is selected and the marks awarded to them are compared.

(i) Describe how a random sample of projects should be chosen. **[2]**

(ii) The marks given for the projects in the sample are as follows.

Project	1	2	3	4	5	6	7	8	9	10	11	12
Examiner A	58	37	72	78	67	77	62	41	80	60	65	70
Examiner B	73	47	74	71	78	96	54	27	97	73	60	66

Carry out a test at the 10% level of significance of the hypotheses H_0: $m = 0$, H_1: $m \neq 0$, where m is the population median difference. **[8]**

(b) A calculator has a built-in random number function which can be used to generate a list of random digits. If it functions correctly then each digit is equally likely to be generated. When it was used to generate 100 random digits, the frequencies of the digits were as follows.

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	6	8	11	14	12	9	15	5	14	6

Use a goodness of fit test, with a significance level of 10%, to investigate whether the random number function is generating digits with equal probability. **[8]**

OCR RECOGNISING ACHIEVEMENT

Friday 21 June 2013 – Morning

A2 GCE MATHEMATICS (MEI)

4768/01 Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4768/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4715870613

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.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- 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 In the past, the times for workers in a factory to complete a particular task had a known median of 7.4 minutes. Following a review, managers at the factory wish to know if the median time to complete the task has been reduced.

(i) A random sample of 12 times, in minutes, gives the following results.

6.90 7.23 6.54 7.62 7.04 7.33 6.74 6.45 7.81 7.71 7.50 6.32

Carry out an appropriate test using a 5% level of significance. **[10]**

(ii) Some time later, a much larger random sample of times gives the following results.

$$n = 80 \qquad \Sigma x = 555.20 \qquad \Sigma x^2 = 3863.9031$$

Find a 95% confidence interval for the true mean time for the task. Justify your choice of which distribution to use. **[6]**

(iii) Describe briefly one advantage and one disadvantage of having a 99% confidence interval instead of a 95% confidence interval. **[2]**

2 A company supplying cattle feed to dairy farmers claims that its new brand of feed will increase average milk yields by 10 litres per cow per week. A farmer thinks the increase will be less than this and decides to carry out a statistical investigation using a paired *t* test. A random sample of 10 dairy cows are given the new feed and then their milk yields are compared with their yields when on the old feed. The yields, in litres per week, for the 10 cows are as follows.

Cow	A	B	C	D	E	F	G	H	I	J
Old feed	144	130	132	146	137	140	140	149	138	133
New feed	148	139	138	159	138	148	146	156	147	145

(i) Why is it sensible to use a paired test? **[1]**

(ii) State the condition necessary for a paired *t* test. **[2]**

(iii) Assuming the condition stated in part **(ii)** is met, carry out the test, using a significance level of 5%, to see whether it appears that the company's claim is justified. **[10]**

(iv) Find a 95% confidence interval for the mean increase in the milk yield using the new feed. **[4]**

3 The random variable X has the following probability density function, $\mathrm{f}(x)$.

$$\mathrm{f}(x) = \begin{cases} kx(x-5)^2 & 0 \leqslant x < 5 \\ 0 & \text{elsewhere} \end{cases}$$

(i) Sketch $\mathrm{f}(x)$. **[3]**

(ii) Find, in terms of k, the cumulative distribution function, $\mathrm{F}(x)$. **[3]**

(iii) Hence show that $k = \dfrac{12}{625}$. **[2]**

The random variable X is proposed as a model for the amount of time, in minutes, lost due to stoppages during a football match. The times lost in a random sample of 60 matches are summarised in the table. The table also shows some of the corresponding expected frequencies given by the model.

Time (minutes)	$0 \leqslant x < 1$	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$
Observed frequency	5	15	23	11	6
Expected frequency			17.76	9.12	1.632

(iv) Find the remaining expected frequencies. **[3]**

(v) Carry out a goodness of fit test, using a significance level of 2.5%, to see if the model might be suitable in this context. **[8]**

4 A company that makes meat pies includes a "small" size in its product range. These pies consist of a pastry case and meat filling, the weights of which are independent of each other. The weight of the pastry case, C, is Normally distributed with mean 96 g and variance 21 g^2. The weight of the meat filling, M, is Normally distributed with mean 57 g and variance 14 g^2.

(i) Find the probability that, in a randomly chosen pie, the weight of the pastry case is between 90 and 100 g. **[4]**

(ii) The wrappers on the pies state that the weight is 145 g. Find the proportion of pies that are underweight. **[3]**

(iii) The pies are sold in packs of 4. Find the value of w such that, in 95% of packs, the total weight of the 4 pies in a randomly chosen pack exceeds w g. **[5]**

(iv) It is required that the weight of the meat filling in a pie should be at least 35% of the total weight. Show that this means that $0.65M - 0.35C \geqslant 0$. Hence find the probability that, in a randomly chosen pie, this requirement is met. **[6]**

THERE ARE NO QUESTIONS PRINTED ON THIS PAGE.