OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) **4767**

Statistics 2

4767

INSTRUCTIONS TO CANDIDATES

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

INFORMATION FOR CANDIDATES

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

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

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

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

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

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

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

$$n = 6 \quad \Sigma t = 270 \quad \Sigma w = 13.42 \quad \Sigma t^2 = 13\,900 \quad \Sigma w^2 = 30.254 \quad \Sigma tw = 584.6$$
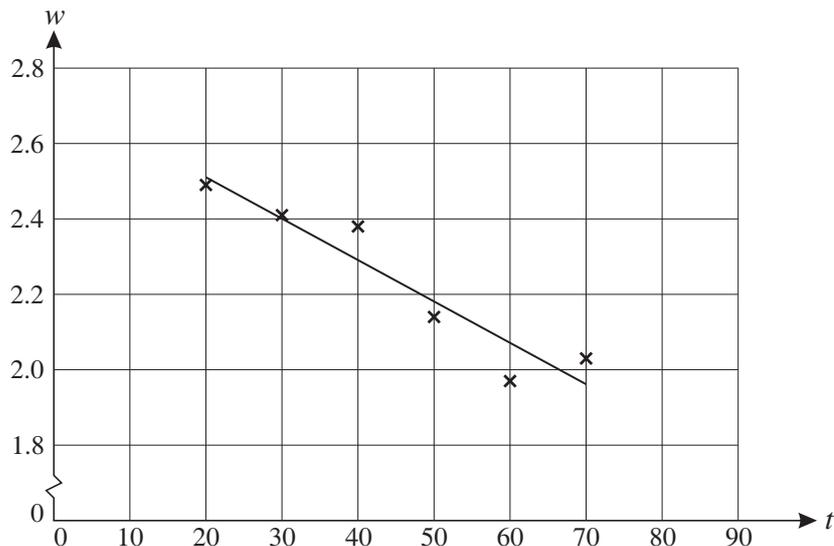
 (iii) Calculate the equation of the regression line of w on t. **[5]**

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

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

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

A random sample of 5000 clover stems is selected.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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