OCR
Oxford Cambridge and RSA

Wednesday 13 June 2018 – Morning

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

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

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

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

- The Question Paper will be found inside 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**. If additional space is required, you should use the lined page(s) at the end of this booklet. The question number(s) must be clearly shown.
- 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 barcodes.
- 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

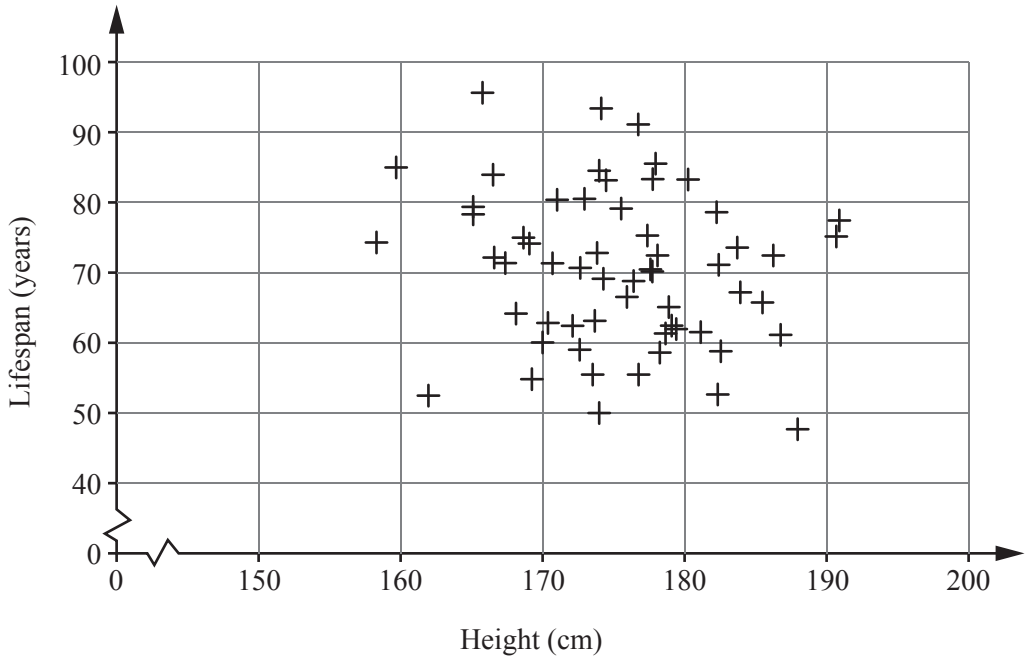
This information is the same on the Printed Answer Book and the Question Paper.

- 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 **8** 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 medical student thinks that there may be some correlation between the heights and lifespans of people in the area in which she lives. She has access to the medical records, going back many years, of deceased people. The scatter diagram below shows the heights x cm and lifespans y years of a random sample of 60 deceased people from the area.



Summary statistics for these data are as follows:

$$n = 60, \quad \Sigma x = 10\,524, \quad \Sigma y = 4219, \quad \Sigma x^2 = 1\,849\,100, \quad \Sigma y^2 = 303\,700, \quad \Sigma xy = 739\,140.$$

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

(ii) The medical student uses these data to test her theory. Carry out a hypothesis test at the 5% significance level to investigate whether her theory may be correct. **[6]**

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

(iv) A friend of the medical student proposes to carry out a similar test using data from a different area. However, a doctor tells the friend that it is known that the correlation coefficient between x and y for the whole population in this other area is –0.134. Explain why it is not sensible for the friend to carry out the proposed test. **[2]**

(v) The student's friend suggests that being tall causes a person to have a shorter life. Comment on this suggestion. **[2]**

2 At a manufacturing plant, work on a production line often has to be stopped due to faults with machinery.

(i) State conditions required for a Poisson distribution to be a suitable model for the number of faults which occur in a day. **[2]**

You may assume that these conditions are satisfied. You are given that the faults occur at an average rate of 8.4 per day.

(ii) State the variance of the distribution of the number of faults per day. **[1]**

(iii) Find the probability of at least 8 faults in a randomly chosen day. **[2]**

(iv) Use a suitable approximating distribution to find the probability that there are at least 40 faults in a 5-day week. Explain briefly whether any additional assumption needs to be made in order to calculate this probability. **[6]**

One of the machines on the production line produces steel rods. The length of a rod is denoted by X mm, where X has a Normal distribution with mean 103.2 and variance 0.36.

(v) (*A*) Find $\mathrm{P}(103 < X < 104)$. **[3]**

(*B*) Rods which are shorter than 102 mm have to be rejected. It is proposed, by reducing the variance, to reduce the proportion of rods which are rejected to 0.5%. Assuming that the mean is unchanged, find the new value of the variance. **[4]**

3 A machine manufactures analgesic (pain-killing) tablets. The tablets are sold in packets. On the packet it states that each tablet contains 500 milligrams of active ingredient. The random variable X represents the amount, in milligrams, of the active ingredient in each tablet. It is known that X is Normally distributed with mean 504.7 and variance 16.

(i) Show that $\mathrm{P}(X > 500) = 0.8800$, correct to 4 decimal places. **[2]**

(ii) Calculate the probability that in a random sample of 10 tablets, at least 9 contain more than the amount of active ingredient stated on the packet. **[2]**

(iii) Use a suitable approximation to find the probability that in a random sample of 100 tablets, at least 90 contain more than the amount stated on the packet. **[5]**

The machine which manufactures the tablets is serviced. In order to check if the mean amount of active ingredient is still 504.7 milligrams, a random sample of 25 tablets is selected. The total amount of active ingredient in the 25 tablets is 12 580 milligrams. You should assume that the variance for individual tablets is still 16.

(iv) Carry out a hypothesis test at the 5% significance level to check whether the mean amount is still 504.7 milligrams. **[9]**

(v) Given that in fact the population mean is now 503.8 milligrams, comment briefly on the result of the test. What change could have been made to the test procedure, without changing the significance level, to make it more likely to detect this change in the mean? **[2]**

4 The year in which a car was registered can often be identified from its number plate. A motoring correspondent is investigating whether there is any association between the age of cars and their location. She chooses three locations, a motorway, a supermarket car park and a housing estate, and selects a random sample of cars from each location. The correspondent classifies the ages according to year of registration, 2001–2007, 2008–2012, 2013–2017 and also 'unknown' for those cars for which the number plate does not identify the year of registration.

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

(ii) You are given that the value of the test statistic for the usual χ^2 test for the motoring correspondent's data is 8.752 correct to 3 decimal places. Carry out the test at the 10% significance level. **[4]**

The correspondent thinks that she should carry out another test, but this time taking a new sample which excludes the cars of unknown age. The numbers of cars for the three other age categories are given in the table below.

	2001–2007	2008–2012	2013–2017
Motorway	41	29	16
Supermarket car park	35	22	14
Housing estate	12	24	7

(iii) Calculate the expected frequency for cars on the Housing estate registered in 2001–2007. Verify the corresponding contribution, 2.5310, to the test statistic. **[3]**

The contributions to the test statistic are shown in the table below. The figures are rounded to 4 decimal places.

	2001–2007	2008–2012	2013–2017
Motorway	0.2639	0.3275	0.0005
Supermarket car park	0.4525	0.8034	0.0570
Housing estate	2.5310	3.8459	0.1146

(iv) Using the same hypotheses as in part **(i)**, complete this new test at the 10% significance level. **[3]**

(v) For each location, comment briefly on how the ages of the cars compare with what would be expected if there were no association. You should calculate any expected frequencies that you need in order to make these comments. **[4]**

(vi) A colleague suggests that the motoring correspondent should have subdivided the ages of the cars into more than three categories. Give one advantage and one disadvantage of doing that. **[2]**

END OF QUESTION PAPER

BLANK PAGE

BLANK PAGE

BLANK PAGE