Cambridge Assessment International Education

Cambridge International AS & A Level

CANDIDATE NAME

CENTRE NUMBER

CANDIDATE NUMBER

6928940735

FURTHER MATHEMATICS **9231/42**

Paper 4 Further Probability & Statistics **October/November 2021**

1 hour 30 minutes

You must answer on the question paper.

You will need: List of formulae (MF19)

INSTRUCTIONS

- Answer **all** questions.
- Use a black or dark blue pen. You may use an HB pencil for any diagrams or graphs.
- Write your name, centre number and candidate number in the boxes at the top of the page.
- Write your answer to each question in the space provided.
- Do **not** use an erasable pen or correction fluid.
- Do **not** write on any bar codes.
- If additional space is needed, you should use the lined page at the end of this booklet; the question number or numbers must be clearly shown.
- You should use a calculator where appropriate.
- You must show all necessary working clearly; no marks will be given for unsupported answers from a calculator.
- Give non-exact numerical answers correct to 3 significant figures, or 1 decimal place for angles in degrees, unless a different level of accuracy is specified in the question.

INFORMATION

- The total mark for this paper is 50.
- The number of marks for each question or part question is shown in brackets [ ].

This document has **16** pages. Any blank pages are indicated.

[Turn over

BLANK PAGE

1 The number, x, of pine trees was counted in each of 40 randomly chosen regions of equal size in country A. The number, y, of pine trees was counted in each of 60 randomly chosen regions of the same equal size in country B. The results are summarised as follows.

$$\Sigma x = 752 \qquad \Sigma x^2 = 14320 \qquad \Sigma y = 1548 \qquad \Sigma y^2 = 40200$$

Find a 95% confidence interval for the difference between the mean number of pine trees in regions of this size in countries A and B. [7]

2 It is claimed that the heights of a particular age group of boys follow a normal distribution with mean 125 cm and standard deviation 12 cm. Observations for a randomly chosen group of 60 boys in this age group are summarised in the following table. The table also gives the expected frequencies, correct to 2 decimal places, based on the normal distribution with mean 125 cm and standard deviation 12 cm.

Height, x cm	$x < 100$	$100 \leqslant x < 110$	$110 \leqslant x < 120$	$120 \leqslant x < 130$	$130 \leqslant x < 140$	$x \geqslant 140$
Observed frequency	0	3	15	23	11	8
Expected frequency	1.12	5.22	13.97	19.38	13.97	6.34

(a) Show how the expected frequency for $130 \leqslant x < 140$ is obtained. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(b) Carry out a goodness of fit test, at the 5% significance level, to determine whether the claim is supported by the data. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

3 The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} a + \frac{1}{5}x & 0 \leqslant x < 1, \\ 2a - \frac{1}{5}x & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a constant.

(a) Find the value of a. [3]

(b) Find $\mathrm{E}(X^2)$. [2]

(c) Find the cumulative distribution function of X. [3]

4 Applicants for a particular college take a written test when they attend for interview. There are two different written tests, *A* and *B*, and each applicant takes one or the other. The interviewer wants to determine whether the medians of the distribution of marks obtained in the two tests are equal. The marks obtained by a random sample of 8 applicants who took test *A* and a random sample of 8 applicants who took test *B* are as follows.

Test *A*	46	32	29	12	33	18	25	40
Test *B*	36	28	49	37	48	35	41	31

(a) Carry out a Wilcoxon rank-sum test at the 5% significance level to determine whether there is a difference in the population median marks obtained in the two tests. [6]

The interviewer considers using the given information to carry out a paired sample t-test to determine whether there is a difference in the population means for the two tests.

(b) Give two reasons why it is not appropriate to use this test. [2]

5 The random variable X is such that $\mathrm{P}(X = r) = kr^2$ for $r = 1, 2, 3, 4$, where k is a constant.

(a) Find the value of k. [1]

.....

.....

.....

.....

.....

.....

(b) Find the probability generating function $\mathrm{G}_X(t)$ of X. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable Y has probability generating function $\mathrm{G}_Y(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$.

The random variable Z is the sum of X and Y.

(c) Assuming that X and Y are independent, find the probability generating function $\mathrm{G}_Z(t)$ of Z as a polynomial in t. [3]

.....

.....

.....

.....

.....

.....

.....

(d) Given that $\mathrm{E}(Z) = \frac{13}{3}$, use $\mathrm{G}_Z(t)$ to find $\mathrm{Var}(Z)$. [3]

6 A scientist is investigating the masses of a particular type of fish found in lakes A and B. He chooses a random sample of 10 fish of this type from lake A and records their masses, x kg, as follows.

2.1 1.8 0.9 3.0 2.4 2.6 1.8 2.2 1.9 2.5

The scientist also chooses a random sample of 12 fish of this type from lake B, but he only has a summary of their masses, y kg, as follows.

$$\Sigma y = 24.48 \qquad \Sigma y^2 = 53.75$$

Test at the 10% significance level whether the mean mass of fish of this type in lake A is greater than the mean mass of fish of this type in lake B. You should state any assumptions that you need to make for the test to be valid. [10]

Additional Page

If you use the following lined page to complete the answer(s) to any question(s), the question number(s) must be clearly shown.

BLANK PAGE

BLANK PAGE