



ADVANCED GCE
MATHEMATICS (MEI)
 Statistics 3

4768

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



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

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

- (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

- 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.

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

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