OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE

MEI STATISTICS   **G242**

Statistics 2 (Z2)

Monday 1 June 2009
Morning

Duration: 1 hour 30 minutes

G242

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 **8** pages. Any blank pages are indicated.

OCR is an exempt Charity

Turn over

1 A wine producer wishes to determine whether the quality of the grapes harvested on his estate is associated with the part of the estate on which they are grown. To investigate this, the estate is divided into three different areas. A random sample of 300 bunches of grapes is taken from across the entire estate. The area from which each bunch is taken is noted and the quality of each bunch is classified as excellent, good or satisfactory. The results are as follows.

	Area A	Area B	Area C
Excellent	36	18	30
Good	36	50	42
Satisfactory	21	31	36

(i) State null and alternative hypotheses for a test to examine whether these data provide any evidence of an association between these classification factors. **[1]**

The following tables show some of the expected frequencies and contributions to the test statistic.

Expected frequencies

	Area A	Area B	Area C
Excellent	26.04	27.72	30.24
Good	39.68		
Satisfactory	27.28		

Contributions to the test statistic

	Area A	Area B	Area C
Excellent	3.8096	3.4083	0.0019
Good	0.3413		
Satisfactory	1.4457		

(ii) Calculate the remaining expected frequencies and contributions to the statistic for the χ^2 test, and carry out the test at the 5% significance level. **[10]**

(iii) With reference to the values in the above tables, discuss briefly how the quality of grapes harvested differs between areas. **[3]**

2 **(i)** A coal merchant supplies smokeless fuel in bags labelled as containing 25 kg. A machine is used to fill the bags. The quantity delivered to each bag, X kg, may be modelled using a Normal distribution with mean 25.2 kg and standard deviation 0.1 kg.

    (*A*) Find the probability that the machine delivers less than 25 kg of smokeless fuel to a bag. **[3]**

    (*B*) Find the probability that, in a customer's order of five bags, there is at least one bag containing less than 25 kg. **[3]**

  (ii) The coal merchant also sells house coal in bags labelled as containing 25 kg. The coal merchant investigates whether or not the machine used to fill bags with house coal is delivering the nominal amount of 25 kg on average. The amount, y kg, in each of a random sample of 50 bags is measured. The results are summarised as follows.

$$\Sigma y = 1295 \qquad \Sigma y^2 = 33\,544$$

    (*A*) Use these data to show that the sample variance is $0.071\,43$ kg^2, correct to 4 significant figures. **[2]**

    (*B*) Find a two-sided 95% confidence interval for the mean amount delivered by this machine. **[4]**

    (*C*) With reference to the confidence interval found in part (*B*), describe what the coal merchant could conclude about the amount delivered by this machine. **[3]**

3 A practice manager is monitoring the length of time, during evening surgeries, that patients wait to see a doctor. She has found the median waiting time to be 23 minutes. She decides to change the system of allocating appointment times.

Following the introduction of a new system, she measures the waiting times, during evening surgeries, of a random sample of 12 patients. The results, in minutes, for the sample are as follows.

    14   33   12   11   6   16   27   18   29   8   9   20

  (i) Use a Wilcoxon test to examine, at the 5% significance level, whether the new appointment booking system has been successful in reducing the median waiting time. State your null and alternative hypotheses clearly. **[12]**

  (ii) Suppose that it could be assumed that the underlying distribution of these waiting times is Normal. State, with a reason, what the most appropriate test procedure would be. **[2]**

4   A botanist is investigating how the seeds of the Black Nightshade plant spread. He believes that seeds spread randomly, and that *X*, the number of seedlings in plots of a given area, may be modelled using a Poisson distribution. To test his belief, the botanist divides a region into plots of equal area, and then counts the number of Black Nightshade seedlings in each of a random sample of 120 plots.

The botanist's results are as follows.

Number of seedlings, *x*	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	16	25	24	18	12	13	12	0

The sample standard deviation is 1.880 correct to 3 decimal places.

(i) (*A*) Verify that the sample mean is 2.6. **[2]**

(*B*) Do the sample statistics provide any reason to doubt the appropriateness of the Poisson model? Justify your answer. **[2]**

(*C*) Does the pattern of observed frequencies provide any reason to doubt the appropriateness of the Poisson model? Justify your answer. **[2]**

The botanist wishes to carry out a test of the goodness of fit of the Poisson model. He uses 2.6 as an estimate for the mean of the underlying population. The following tables show some of the expected frequencies and corresponding contributions to the test statistic.

Expected frequencies

Number of seedlings, *x*	0	1	2	3	4	5	$\geqslant 6$
Expected frequency	8.916	23.172	30.120	26.112	16.968		

Contributions to the test statistic

Number of seedlings, *x*	0	1	2	3	4	5	$\geqslant 6$
Contribution	5.6284	0.1442	1.2435	2.5201			6.3698

(ii) Use the appropriate cumulative probability tables to find P(*X* = 5) and P(*X* $\geqslant$ 6) and hence calculate the remaining expected frequencies and contributions. Carry out the test at the 5% level of significance. **[10]**

5 A golf club manufacturer is testing clubs made with a new alloy to find out if the average ball striking distance is greater than for clubs made with the standard alloy. A random sample of 10 observations from a test using the new alloy is as follows. All distances are in yards.

$$269 \quad 275 \quad 279 \quad 266 \quad 273 \quad 274 \quad 276 \quad 275 \quad 267 \quad 276$$

   (i) Use these data to estimate the mean and variance of the underlying population. **[2]**

In previous tests it has been found that, when using clubs made with the standard alloy, the ball striking distance is Normally distributed with a mean of 268 yards.

  (ii) Use a *t* test at the 5% level of significance to examine whether this sample provides evidence that the mean striking distance for clubs made with the new alloy is greater than 268 yards. State your null and alternative hypotheses clearly. **[9]**

 (iii) When testing golf clubs, a robot arm is used to ensure that all balls are struck with a club head speed of 100 mph. Tests are carried out at an outdoor golf driving range. Describe briefly two other factors that the golf club manufacturer could consider to help ensure fairness in the testing procedure. **[2]**

BLANK PAGE

BLANK PAGE