OCR
RECOGNISING ACHIEVEMENT

# Thursday 31 May 2012 – Morning

## AS GCE MEI STATISTICS

**G242** Statistics 2 (Z2)

**QUESTION PAPER**

*G215990612*

Candidates answer on the Printed Answer Book.

**OCR supplied materials:**
- Printed Answer Book G242
- 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 in the centre of 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.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- 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 bar codes.
- 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**.
- This Printed Answer Book consists of **12** pages. The Question Paper consists of **4** 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 manufacturer produces capacitors for the computer industry. The manufacturer's quality control department has found that 0.8% of capacitors produced are faulty. A sample of 50 capacitors is tested and the number of faulty capacitors is noted.

**(i)** State the conditions required for the binomial distribution B(50, 0.008) to be a suitable model for the number of faulty capacitors in the sample. **[1]**

**(ii)** Explain why the probability distribution in part **(i)** may be approximated by a Poisson distribution. State the parameter of this distribution. **[2]**

**(iii)** Use the approximating distribution in part **(ii)** to find the probability that a sample of 50 capacitors contains at least 3 which are faulty. **[2]**

The manufacturer also produces diodes. As part of the quality control process, diodes are tested in batches of 50. The number of faulty diodes in each batch is recorded. The following table summarises the results from 200 batches.

| Number of faulty diodes | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Observed frequency | 40 | 61 | 64 | 29 | 6 |

The manufacturer wishes to test the goodness of fit of a Poisson model. A mean of 1.5, calculated from the data, is used as an estimate for the mean of the underlying population. The following table shows the corresponding expected frequencies.

| Number of faulty diodes | 0 | 1 | 2 | 3 | $\geqslant 4$ |
|---|---|---|---|---|---|
| Expected frequency | 44.63 | 66.94 | 50.20 | 25.10 | 13.13 |

**(iv)** Show how the expected frequency of 66.94 for 1 faulty diode is obtained. **[2]**

**(v)** Given that the value of the chi-squared test statistic is 9.279, carry out the test of the goodness of fit of the Poisson model at the 5% level of significance. **[6]**

**2** The mass of a tomato harvested from a particular variety of tomato plant is $M$ grams, where $M$ is Normally distributed with mean $\mu$ and variance 20.

**(i)** Given that $P(M < 72) = P(M > 98)$, explain why $\mu = 85$. **[1]**

**(ii)** Calculate the probability that a randomly selected tomato has a mass less than 80 g. **[3]**

A randomly selected sample of 4 tomatoes is chosen and the sample mean is calculated.

**(iii)** Explain the difference between a population mean and a sample mean. **[2]**

**(iv)** Calculate the standard error of the sample mean for the 4 tomatoes. **[2]**

**(v)** Calculate the probability that the sample mean exceeds 90 g. **[3]**

**(vi)** Explain whether or not it was necessary to apply the Central Limit Theorem to calculate the probability in part **(v)**. **[2]**

**3** A motorcycle racer wishes to improve his lap times for a particular racing circuit. He makes an adjustment to his motorcycle's gears which he hopes will reduce his average lap time. Before the adjustment was made, his median lap time for this circuit was 258 seconds. Following the adjustment, he recorded his lap times for a sample of 10 laps. The results, in seconds, are as follows.

253 255 251 246 250 271 264 256 248 267

**(i)** Stating any necessary assumptions, use a Wilcoxon test to examine, at the 5% significance level, whether the motorcyclist's adjustment has been successful in reducing the average lap time. **[13]**

**(ii)** State the advantage of using a Wilcoxon test rather than using a hypothesis test based on the *t* distribution. **[1]**

**(iii)** Discuss whether a test based on the *t* distribution would be suitable in this case. **[2]**

**4** A car manufacturer is developing a battery-powered car. The manufacturer requires a battery which, when fully charged, is capable of powering the car for an average distance of 240 km. A particular type of battery is chosen for testing. In each test, a car with a fully charged battery is driven around a test track until the battery fails; the distance travelled is measured. The distances, in kilometres, for a random sample of 12 tests are as follows.

239 241 238 239 237 242 238 242 238 240 239 235

**(i)** Use these data to show that the sample mean is 239 and calculate the sample standard deviation. **[3]**

**(ii)** Explain, with reference to the sample, whether or not you think that these data could have an underlying Normal distribution. **[2]**

**(iii)** Assuming that these distances do have an underlying Normal distribution, obtain a 95% confidence interval, based on the *t* distribution, for the mean distance travelled. **[5]**

**(iv)** Discuss the confidence interval found in part **(iii)** in relation to the manufacturer's requirement. **[3]**

**(v)** Explain why, in this case, a confidence interval based on the *t* distribution is more suitable than a confidence interval based on the Normal distribution. **[2]**

**[Question 5 is printed overleaf.]**

**5** A beekeepers' organisation is concerned about the continuing reduction in the number of bees. It funds a variety of research projects to investigate the reasons for this reduction. One such project aims to discover if there is an association between the change in size of a bee colony over the course of a year and the intensity of pesticide use over the area in which the colony is located. 120 colonies, regarded as a random sample, are selected and the results are summarised in the table below.

| | | Pesticide use | | |
|---|---|---|---|---|
| | | High | Medium | Low |
| Change in bee population | Minimal change | 3 | 16 | 20 |
| | Decrease of 10% to 20% | 11 | 16 | 15 |
| | Over 20% decrease | 15 | 13 | 11 |

A test to examine whether these data provide any evidence of an association between these classification factors is to be carried out. The following tables show some of the expected frequencies and contributions to the test statistic.

| **Expected frequencies** | | Pesticide use | | |
|---|---|---|---|---|
| | | High | Medium | Low |
| Change in bee population | Minimal change | 9.425 | 14.625 | 14.95 |
| | Decrease of 10% to 20% | 10.15 | | |
| | Over 20% decrease | 9.425 | | |

| **Contributions to the test statistic** | | Pesticide use | | |
|---|---|---|---|---|
| | | High | Medium | Low |
| Change in bee population | Minimal change | 4.380 | 0.129 | 1.706 |
| | Decrease of 10% to 20% | 0.071 | | |
| | Over 20% decrease | 3.298 | | |

**(i)** Calculate the remaining expected frequencies and contributions. Carry out the test at the 5% level of significance. **[11]**

**(ii)** The cell corresponding to high pesticide use and minimal change in population provides the largest single contribution to the test statistic. Explain how this can be interpreted in relation to your hypotheses. **[2]**

**(iii)** Which of the three levels of pesticide use shows the least association with change in population? Explain your answer. **[2]**