Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Wednesday 18 May 2016 (morning)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- Answer all the questions.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A graphic display calculator is required for this paper.
- A clean copy of the **mathematics HL and further mathematics HL formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[60 marks]**.

2216 – 7210

Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. In particular, solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 12]

Adam does the crossword in the local newspaper every day. The time taken by Adam, X minutes, to complete the crossword is modelled by the normal distribution $\mathrm{N}(22, 5^2)$.

(a) Given that, on a randomly chosen day, the probability that he completes the crossword in less than a minutes is equal to 0.8, find the value of a. [3]

(b) Find the probability that the total time taken for him to complete five randomly chosen crosswords exceeds 120 minutes. [3]

Beatrice also does the crossword in the local newspaper every day. The time taken by Beatrice, Y minutes, to complete the crossword is modelled by the normal distribution $\mathrm{N}(40, 6^2)$.

(c) Find the probability that, on a randomly chosen day, the time taken by Beatrice to complete the crossword is more than twice the time taken by Adam to complete the crossword. Assume that these two times are independent. [6]

2. [Maximum mark: 10]

The random variables X, Y follow a bivariate normal distribution with product moment correlation coefficient ρ.

(a) State suitable hypotheses to investigate whether or not X, Y are independent. [2]

A random sample of 10 observations on X, Y was obtained and the value of r, the sample product moment correlation coefficient, was calculated to be 0.486.

(b) (i) Determine the p-value.

(ii) State your conclusion at the 5% significance level. [7]

(c) Explain why the equation of the regression line of y on x should not be used to predict the value of y corresponding to $x = x_0$, where x_0 lies within the range of values of x in the sample. [1]

3. [Maximum mark: 15]

The continuous random variable X takes values in the interval $[0, \theta]$ and

$$\mathrm{E}(X) = \frac{\theta}{2} \text{ and } \mathrm{Var}(X) = \frac{\theta^2}{24}.$$

To estimate the unknown parameter θ, a random sample of size n is obtained from the distribution of X. The sample mean is denoted by $\bar{X}$ and $U = k\bar{X}$ is an unbiased estimator for θ.

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

(b) (i) Calculate an unbiased estimate for θ, using the random sample, $8.3, 4.2, 6.5, 10.3, 2.7, 1.2, 3.3, 4.3$.

(ii) Explain briefly why this is not a good estimate for θ. [4]

(c) (i) Show that $\mathrm{Var}(U) = \frac{\theta^2}{6n}$.

(ii) Show that U^2 is not an unbiased estimator for θ^2.

(iii) Find an unbiased estimator for θ^2 in terms of U and n. [8]

4. [Maximum mark: 11]

The owner of a factory is asked to produce bricks of weight $2.2\,\mathrm{kg}$. The quality control manager wishes to test whether or not, on a particular day, the mean weight of bricks being produced is $2.2\,\mathrm{kg}$.

(a) State hypotheses to enable the quality control manager to test the mean weight using a two-tailed test. [2]

He therefore collects a random sample of 20 of these bricks and determines the weight, $x\,\mathrm{kg}$, of each brick. He produces the following summary statistics.

$$\sum x = 42.0, \sum x^2 = 89.2$$

(b) (i) Calculate unbiased estimates of the mean and the variance of the weights of the bricks being produced.

(ii) Assuming that the weights of the bricks are normally distributed, determine the p-value of the above results and state the conclusion in context using a 5% significance level. [7]

(c) The owner is more familiar with using confidence intervals. Determine a 95% confidence interval for the mean weight of bricks produced on that particular day. [2]

5. [Maximum mark: 12]

The continuous random variable X has probability density function

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

The discrete random variable Y is defined as the integer part of X, that is the largest integer less than or equal to X.

(a) Show that the probability distribution of Y is given by $\mathrm{P}(Y = y) = e^{-y}\left(1 - e^{-1}\right),\ y \in \mathbb{N}$. [4]

(b) (i) Show that $G(t)$, the probability generating function of Y, is given by $G(t) = \dfrac{1 - e^{-1}}{1 - e^{-1}t}$.

(ii) Hence determine the value of $E(Y)$ correct to three significant figures. [8]