Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Monday 8 May 2017 (afternoon)

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 **[50 marks]**.

2217 – 7210

5 pages

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

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: 10]

A farmer sells bags of potatoes which he states have a mean weight of $7\,\text{kg}$. An inspector, however, claims that the mean weight is less than $7\,\text{kg}$. In order to test this claim, the inspector takes a random sample of 12 of these bags and determines the weight, $x\,\text{kg}$, of each bag. He finds that

$$\sum x = 83.64;\ \sum x^2 = 583.05\,.$$

You may assume that the weights of the bags of potatoes can be modelled by the normal distribution $\text{N}\left(\mu,\, \sigma^2\right)$.

(a) State suitable hypotheses to test the inspector's claim. [1]

(b) Find unbiased estimates of μ and σ^2. [3]

(c) (i) Carry out an appropriate test and state the p-value obtained.

(ii) Using a 10% significance level and justifying your answer, state your conclusion in context. [6]

2. [Maximum mark: 14]

The continuous random variable X has cumulative distribution function F given by

$$F(x) = \begin{cases} 0, & x < 0 \\ x\mathrm{e}^{x-1}, & 0 \le x \le 1. \\ 1, & x > 1 \end{cases}$$

(a) Determine

(i) $P(0.25 \le X \le 0.75)$;

(ii) the median of X. [4]

(b) (i) Show that the probability density function f of X is given, for $0 \le x \le 1$, by

$$f(x) = (x+1)\mathrm{e}^{x-1}.$$

(ii) Hence determine the mean and the variance of X. [6]

(c) (i) State the central limit theorem.

(ii) A random sample of 100 observations is obtained from the distribution of X. If $\overline{X}$ denotes the sample mean, use the central limit theorem to find an approximate value of $P(\overline{X} > 0.65)$. Give your answer correct to two decimal places. [4]

3. [Maximum mark: 9]

The discrete random variable X has the following probability distribution.

$$\mathrm{P}(X = x) = \begin{cases} pq^{\frac{x}{2}} & \text{for } \quad x = 0, 2, 4, 6 \ldots \text{ where } p + q = 1, 0 < p < 1. \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that the probability generating function for X is given by $G(t) = \dfrac{p}{1 - qt^2}$. [2]

(b) Hence determine $\mathrm{E}(X)$ in terms of p and q. [4]

(c) The random variable Y is given by $Y = 2X + 1$. Find the probability generating function for Y. [3]

4. [Maximum mark: 10]

The random variables X_1 and X_2 are a random sample from $\mathrm{N}(\mu, 2\sigma^2)$. The random variables Y_1, Y_2 and Y_3 are a random sample from $\mathrm{N}(2\mu, \sigma^2)$.
The estimator U is used to estimate μ where $U = a(X_1 + X_2) + b(Y_1 + Y_2 + Y_3)$ and a, b are constants.

(a) Given that U is unbiased, show that $2a + 6b = 1$. [3]

(b) Show that $\mathrm{Var}(U) = (39b^2 - 12b + 1)\sigma^2$. [3]

(c) Hence find

(i) the value of a and the value of b which give the best unbiased estimator of this form, giving your answers as fractions.

(ii) the variance of this best unbiased estimator. [4]

5. [Maximum mark: 7]

A teacher decides to use the marks obtained by a random sample of 12 students in Geography and History examinations to investigate whether or not there is a positive association between marks obtained by students in these two subjects. You may assume that the distribution of marks in the two subjects is bivariate normal.

(a) State suitable hypotheses for this investigation. [1]

He gives the marks to Anne, one of his students, and asks her to use a calculator to carry out an appropriate test at the 5% significance level. Anne reports that the p-value is 0.177.

(b) State, in context, what conclusion should be drawn from this p-value. [1]

(c) The teacher then asks Anne for the values of the t-statistic and the product moment correlation coefficient r produced by the calculator but she has deleted these. Starting with the p-value, calculate these values of t and r. [5]