Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Wednesday 18 November 2015 (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 **[60 marks]**.

8815–7210

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: 7]

It is known that the standard deviation of the heights of men in a certain country is $15.0\,\text{cm}$.

(a) One hundred men from that country, selected at random, had their heights measured. The mean of this sample was $185\,\text{cm}$. Calculate a $95\,\%$ confidence interval for the mean height of the population. [3]

(b) A second random sample of size n is taken from the same population. Find the minimum value of n needed for the width of a $95\,\%$ confidence interval to be less than $3\,\text{cm}$. [4]

2. [Maximum mark: 11]

The strength of beams compared against the moisture content of the beam is indicated in the following table. You should assume that strength and moisture content are each normally distributed.

Strength	21.1	22.7	23.1	21.5	22.4	22.6	21.1	21.7	21.0	21.4
Moisture content	11.1	8.9	8.8	8.9	8.8	9.9	10.7	10.5	10.5	10.7

(a) Determine the product moment correlation coefficient for these data. [2]

(b) Perform a two-tailed test, at the $5\,\%$ level of significance, of the hypothesis that strength is independent of moisture content. [5]

(c) If the moisture content of a beam is found to be 9.5, use the appropriate regression line to estimate the strength of the beam. [4]

3. [Maximum mark: 9]

Two students are selected at random from a large school with equal numbers of boys and girls. The boys' heights are normally distributed with mean $178\,\text{cm}$ and standard deviation $5.2\,\text{cm}$, and the girls' heights are normally distributed with mean $169\,\text{cm}$ and standard deviation $5.4\,\text{cm}$.

Calculate the probability that the taller of the two students selected is a boy.

4. [Maximum mark: 22]

A discrete random variable U follows a geometric distribution with $p = \frac{1}{4}$.

(a) Find $F(u)$, the cumulative distribution function of U, for $u = 1, 2, 3 \ldots$ [3]

(b) Hence, or otherwise, find the value of $P(U > 20)$. [2]

(c) Prove that the probability generating function of U is given by $G_u(t) = \frac{t}{4 - 3t}$. [4]

(d) Given that $U_i \sim \text{Geo}\left(\frac{1}{4}\right)$, $i = 1, 2, 3$, and that $V = U_1 + U_2 + U_3$, find

(i) $\text{E}(V)$;

(ii) $\text{Var}(V)$;

(iii) $G_v(t)$, the probability generating function of V. [6]

A third random variable W, has probability generating function $G_w(t) = \frac{1}{(4 - 3t)^3}$.

(e) By differentiating $G_w(t)$, find $\text{E}(W)$. [4]

(f) Prove that $V = W + 3$. [3]

5. [Maximum mark: 11]

A biased cubical die has its faces labelled 1, 2, 3, 4, 5 and 6. The probability of rolling a 6 is p, with equal probabilities for the other scores.

The die is rolled once, and the score X_1 is noted.

(a) (i) Find $\mathrm{E}(X_1)$.

(ii) Hence obtain an unbiased estimator for p. [4]

The die is rolled a second time, and the score X_2 is noted.

(b) (i) Show that $k(X_1 - 3) + \left(\frac{1}{3} - k\right)(X_2 - 3)$ is also an unbiased estimator for p for all values of $k \in \mathbb{R}$.

(ii) Find the value for k, which maximizes the efficiency of this estimator. [7]