Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Thursday 21 November 2019 (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]**.

8819 – 7210

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: 7]

Peter, the Principal of a college, believes that there is an association between the score in a Mathematics test, X, and the time taken to run $500\,\text{m}$, Y seconds, of his students. The following paired data are collected.

Mathematics test score X	70	75	76	66	60	61
Time taken to run $500\,\text{m}$ Y	100	105	95	109	89	101

It can be assumed that (X, Y) follow a bivariate normal distribution with product moment correlation coefficient ρ.

(a) (i) State suitable hypotheses H_0 and H_1 to test Peter's claim, using a two-tailed test.

(ii) Carry out a suitable test at the 5% significance level. With reference to the p-value, state your conclusion in the context of Peter's claim. [5]

(b) Peter uses the regression line of y on x as $y = 0.248x + 83.0$ and calculates that a student with a Mathematics test score of 73 will have a running time of 101 seconds. Comment on the validity of his calculation. [2]

2. [Maximum mark: 15]

(a) Three independent random variables X_1, X_2, X_3 are taken from a distribution with mean μ and variance σ^2. Three estimators are proposed for μ.

$$T_1 = \frac{X_1 + X_2 + X_3}{3},\ T_2 = \frac{X_1 + 2X_2 + 3X_3}{3},\ T_3 = \frac{X_1 + 2X_2}{3}$$

(i) Show that one of these estimators for μ is biased and show that the other two are unbiased.

(ii) For the two unbiased estimators, determine, with a reason, which one is more efficient. [9]

(b) Consider the random variable Y, which follows a negative binomial distribution $Y \sim \text{NB}(4, p)$. A random sample is taken from this distribution and the mean is denoted by $\bar{Y}$.

(i) Find $\text{E}(\bar{Y})$.

(ii) Hence suggest an unbiased estimator for $\frac{1}{p}$ in terms of $\bar{Y}$. [2]

(This question continues on the following page)

(Question 2 continued)

(c) A discrete random variable W has a probability distribution given by the following table.

w	1	2
$\mathrm{P}(W = w)$	0.5	0.5

(i) Calculate $\mathrm{E}(W)$.

(ii) Calculate $\mathrm{E}\left(\dfrac{1}{W}\right)$.

(iii) Hence explain why your estimator for $\dfrac{1}{p}$ in (b)(ii) does not directly suggest an unbiased estimator for p. [4]

3. [Maximum mark: 14]

(a) State the central limit theorem as applied to a random sample of size n, taken from a distribution with mean μ and variance σ^2. [2]

A random variable X has a distribution with mean μ and variance 4. A random sample of size 100 is to be taken from the distribution of X.

(b) Jack takes a random sample of size 100 and calculates that $\bar{x} = 60.2$. Find an approximate 90% confidence interval for μ. [2]

(c) Josie takes a different random sample of size 100 to test the null hypothesis that $\mu = 60$ against the alternative hypothesis that $\mu > 60$ at the 5% level.

(i) Find the critical region for Josie's test, giving your answer correct to two decimal places.

(ii) Write down the probability that Josie makes a Type I error.

(iii) Given that the probability that Josie makes a Type II error is 0.25, find the value of μ, giving your answer correct to three significant figures. [10]

4. [Maximum mark: 14]

Consider the random variable X, which follows a negative binomial distribution $X \sim \mathrm{NB}(r, p)$. The probability generating function for X is given by

$$G_X(t) = \frac{p^r t^r}{(1-qt)^r}, \text{ where } q = 1-p.$$

(a) Use this probability generating function to find and simplify $\mathrm{E}(X)$. [5]

Consider another independent random variable Y, where $Y \sim \mathrm{NB}(s, p)$.
Let $W = X + Y$.

(b) (i) Find the probability generating function for W.

(ii) Hence identify the distribution that W follows and state its parameters.

(iii) Given that $r = 2$ and $s = 3$, calculate $\mathrm{P}(X = 3 \mid W = 7)$. [9]

Mathematics
Higher level
Paper 3 – statistics and probability

Wednesday 15 May 2019 (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 **[50 marks]**.

2219 – 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: 16]

The continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} kx & 0 \le x < 1 \\ kx^2 & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}.$$

(a) Show that $k = \frac{6}{17}$. [4]

(b) Find the cumulative distribution function of X. [6]

(c) Find the median, m, of X. [3]

(d) Find $\mathrm{P}(|X - m| < 0.75)$. [3]

2. [Maximum mark: 12]

Employees answer the telephone in a customer relations department. The time taken for an employee to deal with a customer is a random variable which can be modelled by a normal distribution with mean 150 seconds and standard deviation 45 seconds.

(a) Find the probability that the time taken for a randomly chosen customer to be dealt with by an employee is greater than 180 seconds. [2]

(b) Find the probability that the time taken by an employee to deal with a queue of three customers is less than nine minutes. [4]

At the start of the day, one employee, Amanda, has a queue of four customers. A second employee, Brian, has a queue of three customers. You may assume they work independently.

(c) Find the probability that Amanda's queue will be dealt with before Brian's queue. [6]

3. [Maximum mark: 10]

In a large population of hens, the weight of a hen is normally distributed with mean μ kg and standard deviation σ kg. A random sample of 100 hens is taken from the population. The mean weight for the sample is denoted by $\bar{X}$.

(a) State the distribution of $\bar{X}$ giving its mean and variance. [1]

The sample values are summarized by $\sum x = 199.8$ and $\sum x^2 = 407.8$ where x kg is the weight of a hen.

(b) Find an unbiased estimate for μ. [1]

(c) Find an unbiased estimate for σ^2. [2]

(d) Find a 90 % confidence interval for μ. [3]

(e) It is found that $\sigma = 0.27$. It is decided to test, at the 1 % level of significance, the null hypothesis $\mu = 1.95$ against the alternative hypothesis $\mu > 1.95$.

(i) Find the p-value for the test.

(ii) Write down the conclusion reached. [3]

4. [Maximum mark: 12]

It is given that X, Y, Z are random variables and c is a constant.

(a) Show that $\mathrm{Cov}(X + c, Y) = \mathrm{Cov}(X, Y)$. [3]

(b) Show that $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$. [3]

It is given that S and T are two independent normal variables with mean 0 and variance 1.

(c) Using the results from (a) and (b), find the value of $\mathrm{Cov}\left(1 + S, S + ST^2\right)$. [6]

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Thursday 15 November 2018 (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]**.

8818 – 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: 9]

Two independent random variables X and Y follow Poisson distributions.
Given that $\mathrm{E}(X) = 3$ and $\mathrm{E}(Y) = 4$, calculate

(a) $\mathrm{E}(2X + 7Y)$; [2]

(b) $\mathrm{Var}(4X - 3Y)$; [3]

(c) $\mathrm{E}\left(X^2 - Y^2\right)$. [4]

2. [Maximum mark: 9]

The times t, in minutes, taken by a random sample of 75 workers of a company to travel to work can be summarized as follows

$$\sum t = 2165\,, \ \sum t^2 = 76475\,.$$

Let T be the random variable that represents the time taken to travel to work by a worker of this company.

(a) Find unbiased estimates of

(i) the mean of T;

(ii) the variance of T. [3]

(b) Assuming that T is normally distributed, find

(i) the 90% confidence interval for the mean time taken to travel to work by the workers of this company;

(ii) the 95% confidence interval for the mean time taken to travel to work by the workers of this company. [3]

Before seeing these results the managing director believed that the mean time was 26 minutes.

(c) Explain whether your answers to part (b) support her belief. [3]

3. [Maximum mark: 21]

(a) Mr Sailor owns a fish farm and he claims that the weights of the fish in one of his lakes have a mean of 550 grams and standard deviation of 8 grams.

Assume that the weights of the fish are normally distributed and that Mr Sailor's claim is true.

(i) Find the probability that a fish from this lake will have a weight of more than 560 grams.

(ii) The maximum weight a hand net can hold is 6 kg. Find the probability that a catch of 11 fish can be carried in the hand net. [6]

(b) Kathy is suspicious of Mr Sailor's claim about the mean and standard deviation of the weights of the fish. She collects a random sample of fish from this lake whose weights are shown in the following table.

Fish	A	B	C	D	E	F	G	H
Weight (g)	545	554	548	551	558	541	543	549

Using these data, test at the 5% significance level the null hypothesis $H_0: \mu = 550$ against the alternative hypothesis $H_1: \mu < 550$, where μ grams is the population mean weight.

(i) State the distribution of your test statistic, including the parameter.

(ii) Find the p-value for the test.

(iii) State the conclusion of the test, justifying your answer. [6]

(c) Kathy decides to use the same fish sample to test at the 5% significance level whether or not there is a positive association between the weights and the lengths of the fish in the lake. The following table shows the lengths of the fish in the sample. The lengths of the fish can be assumed to be normally distributed.

Fish	A	B	C	D	E	F	G	H
Length (mm)	351	365	355	353	357	349	348	354

(i) State suitable hypotheses for the test.

(ii) Find the product-moment correlation coefficient r.

(iii) State the p-value and interpret it in this context. [6]

(d) Use an appropriate regression line to estimate the weight of a fish with length 360 mm. [3]

4. [Maximum mark: 11]

Let $X \sim \mathrm{Geo}(p)$.

(a) Show that the probability generating function of X is given by

$$G_X(t) = \frac{pt}{1 - qt} \text{ where } q = 1 - p. \qquad [3]$$

(b) Hence prove that $\mathrm{E}(X) = \frac{1}{p}$. [4]

(c) Find the probability generating function of the random variable $Y = 2X + 3$. [4]

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Wednesday 9 May 2018 (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]**.

2218 – 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: 11]

The weights, X kg, of the males of a species of bird may be assumed to be normally distributed with mean 4.8 kg and standard deviation 0.2 kg.

(a) Find the probability that a randomly chosen male bird weighs between 4.75 kg and 4.85 kg. [1]

The weights, Y kg, of female birds of the same species may be assumed to be normally distributed with mean 2.7 kg and standard deviation 0.15 kg.

(b) Find the probability that the weight of a randomly chosen male bird is more than twice the weight of a randomly chosen female bird. [6]

(c) Two randomly chosen male birds and three randomly chosen female birds are placed on a weighing machine that has a weight limit of 18 kg. Find the probability that the total weight of these five birds is greater than the weight limit. [4]

2. [Maximum mark: 8]

Consider an unbiased tetrahedral (four-sided) die with faces labelled 1, 2, 3 and 4 respectively.
The random variable X represents the number of throws required to obtain a 1.

(a) State the distribution of X. [1]

(b) Show that the probability generating function, $G(t)$, for X is given by $G(t) = \frac{t}{4 - 3t}$. [4]

(c) Find $G'(t)$. [2]

(d) Determine the mean number of throws required to obtain a 1. [1]

3. [Maximum mark: 12]

A smartphone's battery life is defined as the number of hours a fully charged battery can be used before the smartphone stops working. A company claims that the battery life of a model of smartphone is, on average, 9.5 hours. To test this claim, an experiment is conducted on a random sample of 20 smartphones of this model. For each smartphone, the battery life, b hours, is measured and the sample mean, $\bar{b}$, calculated. It can be assumed the battery lives are normally distributed with standard deviation 0.4 hours.

(a) State suitable hypotheses for a two-tailed test. [1]

(b) Find the critical region for testing $\bar{b}$ at the 5% significance level. [4]

It is then found that this model of smartphone has an average battery life of 9.8 hours.

(c) Find the probability of making a Type II error. [3]

Another model of smartphone whose battery life may be assumed to be normally distributed with mean μ hours and standard deviation 1.2 hours is tested. A researcher measures the battery life of six of these smartphones and calculates a confidence interval of $[10.2, 11.4]$ for μ.

(d) Calculate the confidence level of this interval. [4]

4. [Maximum mark: 11]

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 a negative linear association exists between X and Y. [1]

A random sample of 11 observations on X, Y was obtained and the value of the sample product moment correlation coefficient, r, was calculated to be -0.708.

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

(ii) State your conclusion at the 1% significance level. [4]

The covariance of the random variables U, V is defined by

$$\mathrm{Cov}(U, V) = \mathrm{E}\big((U - \mathrm{E}(U))(V - \mathrm{E}(V))\big).$$

(c) (i) Show that $\mathrm{Cov}(U, V) = \mathrm{E}(UV) - \mathrm{E}(U)\mathrm{E}(V)$.

(ii) Hence show that if U, V are independent random variables then the population product moment correlation coefficient, ρ, is zero. [6]

5. [Maximum mark: 8]

The random variable X has a binomial distribution with parameters n and p.

(a) Show that $P = \frac{X}{n}$ is an unbiased estimator of p. [2]

Let $U = nP(1 - P)$.

(b) (i) Show that $\mathrm{E}(U) = (n - 1)p(1 - p)$.

(ii) Hence write down an unbiased estimator of $\mathrm{Var}(X)$. [6]

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Thursday 16 November 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]**.

8817 – 7210

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: 7]

A continuous random variable T has a probability density function defined by

$$f(t) = \begin{cases} \dfrac{t\left(4-t^2\right)}{4}, & 0 \le t \le 2 \\ 0, & \text{otherwise} \end{cases}.$$

(a) Find the cumulative distribution function $F(t)$, for $0 \le t \le 2$. [3]

(b) (i) Sketch the graph of $F(t)$ for $0 \le t \le 2$, clearly indicating the coordinates of the endpoints.

(ii) Given that $\mathrm{P}(T < a) = 0.75$, find the value of a. [4]

2. [Maximum mark: 8]

Anne is a farmer who grows and sells pumpkins. Interested in the weights of pumpkins produced, she records the weights of eight pumpkins and obtains the following results in kilograms.

$$7.7 \quad 7.5 \quad 8.4 \quad 8.8 \quad 7.3 \quad 9.0 \quad 7.8 \quad 7.6$$

Assume that these weights form a random sample from a $\mathrm{N}\left(\mu, \sigma^2\right)$ distribution.

(a) Determine unbiased estimates for μ and σ^2. [3]

(b) Anne claims that the mean pumpkin weight is 7.5 kilograms. In order to test this claim, she sets up the null hypothesis $\mathrm{H}_0 : \mu = 7.5$.

(i) Use a two-tailed test to determine the p-value for the above results.

(ii) Interpret your p-value at the 5% level of significance, justifying your conclusion. [5]

3. [Maximum mark: 12]

A random variable X is distributed with mean μ and variance σ^2. Two independent random samples of sizes n_1 and n_2 are taken from the distribution of X. The sample means are $\overline{X_1}$ and $\overline{X_2}$ respectively.

(a) Show that $U = a\overline{X_1} + (1-a)\overline{X_2}$, $a \in \mathbb{R}$, is an unbiased estimator of μ. [3]

(b) (i) Show that $\mathrm{Var}(U) = a^2\dfrac{\sigma^2}{n_1} + (1-a)^2\dfrac{\sigma^2}{n_2}$.

(ii) Find, in terms of n_1 and n_2, an expression for a which gives the most efficient estimator of this form.

(iii) Hence find an expression for the most efficient estimator and interpret the result. [9]

4. [Maximum mark: 8]

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

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

A random sample of 12 observations on U, V is obtained to determine whether there is a correlation between U and V. The sample product moment correlation coefficient is denoted by r. A test to determine whether or not U, V are independent is carried out at the 1% level of significance.

(b) Find the least value of $|r|$ for which the test concludes that $\rho \neq 0$. [6]

5. [Maximum mark: 15]

The random variable X follows a Poisson distribution with mean λ. The probability generating function of X is given by $G_X(t) = \mathrm{e}^{\lambda(t-1)}$.

(a) (i) Find expressions for $G'_X(t)$ and $G''_X(t)$.

(ii) Hence show that $\mathrm{Var}(X) = \lambda$. [5]

The random variable Y, independent of X, follows a Poisson distribution with mean μ.

(b) By considering the probability generating function, $G_{X+Y}(t)$, of $X+Y$, show that $X+Y$ follows a Poisson distribution with mean $\lambda+\mu$. [3]

(c) (i) Show that $\mathrm{P}(X = x \mid X+Y = n) = \binom{n}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^{x}\left(1-\frac{\lambda}{\lambda+\mu}\right)^{n-x}$, where n, x are non-negative integers and $n \geq x$.

(ii) Identify the probability distribution given in part (c)(i) and state its parameters. [7]

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

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 $\mathrm{N}\left(\mu,\, \sigma^2\right)$.

(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]

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Friday 18 November 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]**.

8816 – 7210

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: 17]

In this question you may assume that these data are a random sample from a bivariate normal distribution, with population product moment correlation coefficient ρ.
Richard wishes to do some research on two types of exams which are taken by a large number of students. He takes a random sample of the results of 10 students, which are shown in the following table.

Student	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Exam 1	51	70	10	22	99	33	45	8	65	82
Exam 2	52	64	8	25	90	43	50	50	70	50

(a) For these data find the product moment correlation coefficient, r. [2]

Using these data, it is decided to test, at the 1% level, the null hypothesis $H_0 : \rho = 0$ against the alternative hypothesis $H_1 : \rho > 0$.

(b) (i) State the distribution of the test statistic (including any parameters).

(ii) Find the p-value for the test.

(iii) State the conclusion, in the context of the question, with the word "correlation" in your answer. Justify your answer. [6]

Richard decides to take the exams himself. He scored 11 on Exam 1 but his result on Exam 2 was lost.

(c) Using a suitable regression line, find an estimate for his score on Exam 2, giving your answer to the nearest integer. [3]

Caroline believes that the population mean mark on Exam 2 is 6 marks higher than the population mean mark on Exam 1. Using the original data from the 10 students, it is decided to test, at the 5% level, this hypothesis against the alternative hypothesis that the mean of the differences, $d =$ exam 2 mark – exam 1 mark, is less than 6 marks.

(d) (i) State the distribution of your test statistic (including any parameters).

(ii) Find the p-value.

(iii) State the conclusion, justifying the answer. [6]

2. [Maximum mark: 17]

John rings a church bell 120 times. The time interval, T_i , between two successive rings is a random variable with mean of 2 seconds and variance of $\frac{1}{9}$ seconds^2.

Each time interval, T_i, is independent of the other time intervals. Let $X = \sum_{i=1}^{119} T_i$ be the total time between the first ring and the last ring.

(a) Find

(i) $\mathrm{E}(X)$;

(ii) $\mathrm{Var}(X)$. [3]

(b) Explain why a normal distribution can be used to give an approximate model for X. [2]

(c) Use this model to find the values of A and B such that $\mathrm{P}(A < X < B) = 0.9$, where A and B are symmetrical about the mean of X. [7]

The church vicar subsequently becomes suspicious that John has stopped coming to ring the bell and that he is letting his friend Ray do it. When Ray rings the bell the time interval, T_i, has a mean of 2 seconds and variance of $\frac{1}{25}$ seconds^2.

The church vicar makes the following hypotheses:

H_0: Ray is ringing the bell; H_1: John is ringing the bell.

He records four values of X. He decides on the following decision rule:

If $236 \le X \le 240$ for all four values of X he accepts H_0, otherwise he accepts H_1.

(d) Calculate the probability that he makes a Type II error. [5]

3. [Maximum mark: 15]

Alun answers mathematics questions and checks his answer after doing each one. The probability that he answers any question correctly is always $\frac{6}{7}$, independently of all other questions. He will stop for coffee immediately following a second incorrect answer. Let X be the number of questions Alun answers before he stops for coffee.

(a) (i) State the distribution of X, including its parameters.

(ii) Calculate $\mathrm{E}(X)$.

(iii) Calculate $\mathrm{P}(X = 5)$. [6]

Nic answers mathematics questions and checks his answer after doing each one. The probability that he answers any question correctly is initially $\frac{6}{7}$. After his first incorrect answer, Nic loses confidence in his own ability and from this point onwards, the probability that he answers any question correctly is now only $\frac{4}{7}$.

Both before and after his first incorrect answer, the result of each question is independent of the result of any other question. Nic will also stop for coffee immediately following a second incorrect answer. Let Y be the number of questions Nic answers before he stops for coffee.

(b) (i) Calculate $\mathrm{E}(Y)$.

(ii) Calculate $\mathrm{P}(Y = 5)$. [9]

4. [Maximum mark: 11]

Two independent discrete random variables X and Y have probability generating functions $G(t)$ and $H(t)$ respectively. Let $Z = X + Y$ have probability generating function $J(t)$.

(a) Write down an expression for $J(t)$ in terms of $G(t)$ and $H(t)$. [1]

(b) By differentiating $J(t)$, prove that

(i) $\mathrm{E}(Z) = \mathrm{E}(X) + \mathrm{E}(Y)$;

(ii) $\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. [10]

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

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: 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]

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 $k\left(X_1 - 3\right) + \left(\frac{1}{3} - k\right)\left(X_2 - 3\right)$ 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]

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics
Higher level
Paper 3 – statistics and probability

Thursday 21 May 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]**.

2215 – 7210

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: 13]

Engine oil is sold in cans of two capacities, large and small. The amount, in millilitres, in each can, is normally distributed according to Large $\sim \mathrm{N}(5000, 40)$ and Small $\sim \mathrm{N}(1000, 25)$.

(a) A large can is selected at random. Find the probability that the can contains at least 4995 millilitres of oil. [2]

(b) A large can and a small can are selected at random. Find the probability that the large can contains at least 30 millilitres more than five times the amount contained in the small can. [6]

(c) A large can and five small cans are selected at random. Find the probability that the large can contains at least 30 millilitres less than the total amount contained in the small cans. [5]

2. [Maximum mark: 12]

Eleven students who had under-performed in a philosophy practice examination were given extra tuition before their final examination. The differences between their final examination marks and their practice examination marks were

$$10, -1, 6, 7, -5, -5, 2, -3, 8, 9, -2.$$

Assume that these differences form a random sample from a normal distribution with mean μ and variance σ^2.

(a) Determine unbiased estimates of μ and σ^2. [4]

(b) (i) State suitable hypotheses to test the claim that extra tuition improves examination marks.

(ii) Calculate the p-value of the sample.

(iii) Determine whether or not the above claim is supported at the 5% significance level. [8]

3. [Maximum mark: 9]

A manufacturer of stopwatches employs a large number of people to time the winner of a 100 metre sprint. It is believed that if the true time of the winner is μ seconds, the times recorded are normally distributed with mean μ seconds and standard deviation 0.03 seconds.

The times, in seconds, recorded by six randomly chosen people are

$$9.765\,,\ 9.811\,,\ 9.783\,,\ 9.797\,,\ 9.804\,,\ 9.798\,.$$

(a) Calculate a 99% confidence interval for μ. Give your answer correct to three decimal places. [4]

(b) Interpret the result found in (a). [2]

(c) Find the confidence level of the interval that corresponds to halving the width of the 99% confidence interval. Give your answer as a percentage to the nearest whole number. [3]

4. [Maximum mark: 15]

A random variable X has a population mean μ.

(a) Explain briefly the meaning of

(i) an estimator of μ;

(ii) an unbiased estimator of μ. [3]

(b) A random sample X_1, X_2, X_3 of three independent observations is taken from the distribution of X.

An unbiased estimator of μ, $\mu \neq 0$, is given by $U = \alpha X_1 + \beta X_2 + (\alpha - \beta) X_3$, where $\alpha, \beta \in \mathbb{R}$.

(i) Find the value of α.

(ii) Show that $\mathrm{Var}(U) = \sigma^2\left(2\beta^2 - \beta + \dfrac{1}{2}\right)$ where $\sigma^2 = \mathrm{Var}(X)$.

(iii) Find the value of β which gives the most efficient estimator of μ of this form.

(iv) Write down an expression for this estimator and determine its variance.

(v) Write down a more efficient estimator of μ than the one found in (iv), justifying your answer. [12]

5. [Maximum mark: 11]

(a) Determine the probability generating function for $X \sim \mathrm{B}(1, p)$. [4]

(b) Explain why the probability generating function for $\mathrm{B}(n, p)$ is a polynomial of degree n. [2]

(c) Two independent random variables X_1 and X_2 are such that $X_1 \sim \mathrm{B}(1, p_1)$ and $X_2 \sim \mathrm{B}(1, p_2)$. Prove that if $X_1 + X_2$ has a binomial distribution then $p_1 = p_2$. [5]

88147210

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

MATHEMATICS
HIGHER LEVEL
PAPER 3 – STATISTICS AND PROBABILITY

Thursday 13 November 2014 (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]*.

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: 9]*

A random variable X has probability density function

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \le x < 1 \\ \frac{1}{4} & 1 \le x < 3 \\ 0 & x \ge 3 \end{cases}.$$

(a) Sketch the graph of $y = f(x)$. *[1]*

(b) Find the cumulative distribution function for X. *[5]*

(c) Find the interquartile range for X. *[3]*

2. *[Maximum mark: 9]*

Eric plays a game at a fairground in which he throws darts at a target. Each time he throws a dart, the probability of hitting the target is 0.2. He is allowed to throw as many darts as he likes, but it costs him \$1 a throw. If he hits the target a total of three times he wins \$10.

(a) Find the probability he has his third success of hitting the target on his sixth throw. *[3]*

(b) (i) Find the expected number of throws required for Eric to hit the target three times.

(ii) Write down his expected profit or loss if he plays until he wins the \$10. *[3]*

(c) If he has just \$8, find the probability he will lose all his money before he hits the target three times. *[3]*

3. *[Maximum mark: 11]*

(a) If X and Y are two random variables such that $\mathrm{E}(X) = \mu_X$ and $\mathrm{E}(Y) = \mu_Y$ then $\mathrm{Cov}(X, Y) = \mathrm{E}\big((X - \mu_X)(Y - \mu_Y)\big)$.

Prove that if X and Y are independent then $\mathrm{Cov}(X, Y) = 0$. *[3]*

(b) In a particular company, it is claimed that the distance travelled by employees to work is independent of their salary. To test this, 20 randomly selected employees are asked about the distance they travel to work and the size of their salaries. It is found that the product moment correlation coefficient, r, for the sample is -0.35.

You may assume that both salary and distance travelled to work follow normal distributions.

Perform a one-tailed test at the 5% significance level to test whether or not the distance travelled to work and the salaries of the employees are independent. *[8]*

4. *[Maximum mark: 21]*

If X is a random variable that follows a Poisson distribution with mean $\lambda > 0$ then the probability generating function of X is $G(t) = e^{\lambda(t-1)}$.

(a) (i) Prove that $\mathrm{E}(X) = \lambda$.

(ii) Prove that $\mathrm{Var}(X) = \lambda$. *[6]*

Y is a random variable, independent of X, that also follows a Poisson distribution with mean λ.

(b) If $S = 2X - Y$ find

(i) $\mathrm{E}(S)$;

(ii) $\mathrm{Var}(S)$. *[3]*

Let $T = \frac{X}{2} + \frac{Y}{2}$.

(c) (i) Show that T is an unbiased estimator for λ.

(ii) Show that T is a more efficient unbiased estimator of λ than S. *[3]*

(d) Could either S or T model a Poisson distribution? Justify your answer. *[1]*

(e) By consideration of the probability generating function, $G_{X+Y}(t)$, of $X+Y$, prove that $X+Y$ follows a Poisson distribution with mean 2λ. *[3]*

(f) Find

(i) $G_{X+Y}(1)$;

(ii) $G_{X+Y}(-1)$. *[2]*

(g) Hence find the probability that $X+Y$ is an even number. *[3]*

5. *[Maximum mark: 10]*

Two species of plant, A and B, are identical in appearance though it is known that the mean length of leaves from a plant of species A is 5.2 cm, whereas the mean length of leaves from a plant of species B is 4.6 cm. Both lengths can be modelled by normal distributions with standard deviation 1.2 cm.

In order to test whether a particular plant is from species A or species B, 16 leaves are collected at random from the plant. The length, x, of each leaf is measured and the mean length evaluated. A one-tailed test of the sample mean, $\bar{X}$, is then performed at the 5 % level, with the hypotheses: $H_0 : \mu = 5.2$ and $H_1 : \mu < 5.2$.

(a) Find the critical region for this test. *[3]*

(b) Find the probability of a Type II error if the leaves are in fact from a plant of species B. *[2]*

It is now known that in the area in which the plant was found 90 % of all the plants are of species A and 10 % are of species B.

(c) Find the probability that $\bar{X}$ will fall within the critical region of the test. *[2]*

(d) If, having done the test, the sample mean is found to lie within the critical region, find the probability that the leaves came from a plant of species A. *[3]*

M14/5/MATHL/HP3/ENG/TZ0/SP

22147210

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

MATHEMATICS
HIGHER LEVEL
PAPER 3 – STATISTICS AND PROBABILITY

Thursday 15 May 2014 (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]*.

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: 14]*

The random variable X has probability distribution $\mathrm{Po}(8)$.

(a) (i) Find $\mathrm{P}(X = 6)$.

(ii) Find $\mathrm{P}(X = 6 \mid 5 \le X \le 8)$. *[5]*

(b) $\bar{X}$ denotes the sample mean of $n > 1$ independent observations from X.

(i) Write down $\mathrm{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$.

(ii) Hence, give a reason why $\bar{X}$ is not a Poisson distribution. *[3]*

(c) A random sample of 40 observations is taken from the distribution for X.

(i) Find $\mathrm{P}(7.1 < \bar{X} < 8.5)$.

(ii) Given that $\mathrm{P}(|\bar{X} - 8| \le k) = 0.95$, find the value of k. *[6]*

2. *[Maximum mark: 16]*

The following table gives the average yield of olives per tree, in kg, and the rainfall, in cm, for nine separate regions of Greece. You may assume that these data are a random sample from a bivariate normal distribution, with correlation coefficient ρ .

Rainfall (x)	11	10	15	13	7	18	22	20	28
Yield (y)	56	53	67	61	54	78	86	88	78

A scientist wishes to use these data to determine whether there is a positive correlation between rainfall and yield.

(a) State suitable hypotheses. *[2]*

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

(c) Determine the associated p-value and comment on this value in the context of the question. *[2]*

(d) Find the equation of the regression line of y on x. *[2]*

(e) Hence, estimate the yield per tree in a tenth region where the rainfall was 19 cm. *[2]*

(f) Determine the angle between the regression line of y on x and that of x on y. Give your answer to the nearest degree. *[6]*

3. *[Maximum mark: 14]*

(a) Consider the random variable X for which $\mathrm{E}(X) = a\lambda + b$, where a and b are constants and λ is a parameter.

Show that $\frac{X-b}{a}$ is an unbiased estimator for λ. *[3]*

(b) The continuous random variable Y has probability density function

$$f(y) = \begin{cases} \frac{2}{9}(3+y-\lambda), & \text{for } \lambda - 3 \le y \le \lambda \\ 0, & \text{otherwise} \end{cases}$$

where λ is a parameter.

(i) Verify that $f(y)$ is a probability density function for all values of λ.

(ii) Determine $\mathrm{E}(Y)$.

(iii) Write down an unbiased estimator for λ. *[11]*

4. *[Maximum mark: 16]*

Consider the random variable $X \sim \text{Geo}(p)$.

(a) State $\text{P}(X < 4)$. *[2]*

(b) Show that the probability generating function for X is given by $G_X(t) = \frac{pt}{1-qt}$, where $q = 1 - p$. *[3]*

Let the random variable $Y = 2X$.

(c) (i) Show that the probability generating function for Y is given by $G_Y(t) = G_X\left(t^2\right)$.

(ii) By considering $G_Y'(1)$, show that $\text{E}(Y) = 2\text{E}(X)$. *[6]*

Let the random variable $W = 2X + 1$.

(d) (i) Find the probability generating function for W in terms of the probability generating function of Y.

(ii) Hence, show that $\text{E}(W) = 2\text{E}(X) + 1$. *[5]*

N13/5/MATHL/HP3/ENG/TZ0/SP

88137210

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

MATHEMATICS
HIGHER LEVEL
PAPER 3 – STATISTICS AND PROBABILITY

Tuesday 19 November 2013 (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 SL*** *information booklet* is required for this paper.
- The maximum mark for this examination paper is *[60 marks]*.

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: 8]*

A traffic radar records the speed, v kilometres per hour (km h^{-1}), of cars on a section of a road. The following table shows a summary of the results for a random sample of 1000 cars whose speeds were recorded on a given day.

Speed	**Number of cars**
$50 \le v < 60$	5
$60 \le v < 70$	13
$70 \le v < 80$	52
$80 \le v < 90$	68
$90 \le v < 100$	98
$100 \le v < 110$	105
$110 \le v < 120$	289
$120 \le v < 130$	142
$130 \le v < 140$	197
$140 \le v < 150$	31

(a) Using the data in the table,

(i) show that an estimate of the mean speed of the sample is $113.21\ \text{km h}^{-1}$;

(ii) find an estimate of the variance of the speed of the cars on this section of the road. *[4]*

(b) Find the 95% confidence interval, I, for the mean speed. *[2]*

(c) Let J be the 90% confidence interval for the mean speed.

Without calculating J, explain why $J \subset I$. *[2]*

2. *[Maximum mark: 14]*

A farmer is selling apples and oranges. The weights X and Y, in grams, of the apples and oranges respectively are normally distributed with $X \sim N(180, 14^2)$ and $Y \sim N(150, 12^2)$.

(a) Find the probability that the weight of a randomly chosen apple is more than 1.5 times the weight of a randomly chosen orange. *[7]*

(b) Katharina buys 4 apples and 6 oranges. Find the probability that the total weight is greater than 1.5 kilograms. *[7]*

3. *[Maximum mark: 14]*

Jenny tosses seven coins simultaneously and counts the number of tails obtained. She repeats the experiment 750 times. The following frequency table shows her results.

Number of tails	Frequency
0	6
1	19
2	141
3	218
4	203
5	117
6	38
7	8

(a) It is claimed that all of these seven coins are fair and it is decided to test this claim using a suitable χ^2 test.

(i) State the null and alternative hypotheses.

(ii) State a decision rule at the 5% level of significance.

(iii) Find the value of the test statistic.

(iv) Write down your conclusion. *[10]*

(b) Explain what can be done with this data to decrease the probability of making a type I error. *[2]*

(c) (i) State the meaning of a type II error.

(ii) Write down how to proceed if it is required to decrease the probability of making both a type I and type II error. *[2]*

4. *[Maximum mark: 10]*

Francisco and his friends want to test whether performance in running 400 metres improves if they follow a particular training schedule. The competitors are tested before and after the training schedule.

The times taken to run 400 metres, in seconds, before and after training are shown in the following table.

Competitor	**A**	**B**	**C**	**D**	**E**
Time before training	75	74	60	69	69
Time after training	73	69	55	72	65

Apply an appropriate test at the 1 % significance level to decide whether the training schedule improves competitors' times, stating clearly the null and alternative hypotheses. (It may be assumed that the distributions of the times before and after training are normal.)

5. *[Maximum mark: 14]*

Let X and Y be independent random variables with $X \sim P_o(3)$ and $Y \sim P_o(2)$.

Let $S = 2X + 3Y$.

(a) Find the mean and variance of S. *[2]*

(b) Hence state with a reason whether or not S follows a Poisson distribution. *[2]*

Let $T = X + Y$.

(c) Find $\mathrm{P}(T = 3)$. *[4]*

(d) Show that $\mathrm{P}(T = t) = \sum_{r=0}^{t} \mathrm{P}(X = r)\mathrm{P}(Y = t - r)$. *[2]*

(e) Hence show that T follows a Poisson distribution with mean 5. *[4]*

22137210

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

MATHEMATICS
HIGHER LEVEL
PAPER 3 – STATISTICS AND PROBABILITY

Tuesday 21 May 2013 (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 SL*** *information booklet* is required for this paper.
- The maximum mark for this examination paper is *[60 marks]*.

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]*

The random variable X is normally distributed with unknown mean μ and unknown variance σ^2. A random sample of 20 observations on X gave the following results.

$$\sum x = 280,\ \sum x^2 = 3977.57$$

(a) Find unbiased estimates of μ and σ^2. *[3 marks]*

(b) Determine a 95 % confidence interval for μ. *[3 marks]*

(c) Given the hypotheses

$$H_0 : \mu = 15;\ H_1 : \mu \neq 15,$$

find the p-value of the above results and state your conclusion at the 1 % significance level. *[4 marks]*

2. *[Maximum mark: 12]*

A hockey team played 60 matches last season. The manager believes that the number of goals scored by the team in a match could be modelled by a Poisson distribution and he produces the following table based on the season's results.

Number of goals	0	1	2	3	4	5
Frequency	8	9	17	14	7	5

(a) State suitable hypotheses to test the manager's belief. *[1 mark]*

(b) The manager decides to carry out an appropriate χ^2 goodness of fit test.

(i) Construct a table of appropriate expected frequencies correct to **four decimal places**.

(ii) Determine the value of χ^2_{calc} and the corresponding p-value.

(iii) State whether or not your analysis supports the manager's belief. *[11 marks]*

3. *[Maximum mark: 9]*

The number of machine breakdowns occurring in a day in a certain factory may be assumed to follow a Poisson distribution with mean μ. The value of μ is known, from past experience, to be 1.2. In an attempt to reduce the value of μ, all the machines are fitted with new control units. To investigate whether or not this reduces the value of μ, the total number of breakdowns, x, occurring during a 30-day period following the installation of these new units is recorded.

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

(b) It is decided to define the critical region by $x \le 25$.

(i) Calculate the significance level.

(ii) Assuming that the value of μ was actually reduced to 0.75, determine the probability of a Type II error. *[8 marks]*

4. *[Maximum mark: 14]*

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \dfrac{3x^2 + 2x}{10}, & \text{for } 1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}.$$

(a) (i) Determine an expression for $F(x)$, valid for $1 \le x \le 2$, where F denotes the cumulative distribution function of X.

(ii) Hence, or otherwise, determine the median of X. *[6 marks]*

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

(ii) A random sample of 150 observations is taken from the distribution of X and $\bar{X}$ denotes the sample mean. Use the central limit theorem to find, approximately, the probability that $\bar{X}$ is greater than 1.6. *[8 marks]*

5. *[Maximum mark: 15]*

When Ben shoots an arrow, he hits the target with probability 0.4. Successive shots are independent.

(a) Find the probability that

(i) he hits the target exactly 4 times in his first 8 shots;

(ii) he hits the target for the 4^{th} time with his 8^{th} shot. *[6 marks]*

(b) Ben hits the target for the 10^{th} time with his X^{th} shot.

(i) Determine the expected value of the random variable X.

(ii) Write down an expression for $\mathrm{P}(X = x)$ and show that

$$\frac{\mathrm{P}(X = x)}{\mathrm{P}(X = x - 1)} = \frac{3(x-1)}{5(x-10)}.$$

(iii) Hence, or otherwise, find the most likely value of X. *[9 marks]*

88127210

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

MATHEMATICS
HIGHER LEVEL
PAPER 3 – STATISTICS AND PROBABILITY

Thursday 8 November 2012 (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 SL*** *information booklet* is required for this paper.
- The maximum mark for this examination paper is *[60 marks]*.

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, e.g. 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: 16]*

Anna has a box with 10 biscuits in it. 4 biscuits are chocolate and 6 are plain. Anna takes a biscuit from her box at random and eats it. She repeats this process until she has eaten 5 biscuits in total.

Let A be the number of chocolate biscuits that Anna eats.

(a) State the distribution of A. *[1 mark]*

(b) Find $\mathrm{P}(A=3)$. *[2 marks]*

(c) Find $\mathrm{P}(A=5)$. *[1 mark]*

Bill also has a box with 10 biscuits in it. 4 biscuits are chocolate and 6 are plain. Bill takes a biscuit from his box at random, looks at it and replaces it in the box. He repeats this process until he has looked at 5 biscuits in total. Let B be the number of chocolate biscuits that Bill takes and looks at.

(d) State the distribution of B. *[1 mark]*

(e) Find $\mathrm{P}(B=3)$. *[2 marks]*

(f) Find $\mathrm{P}(B=5)$. *[2 marks]*

Let $D=B-A$.

(g) Calculate $\mathrm{E}(D)$. *[2 marks]*

(h) Calculate $\mathrm{Var}(D)$, justifying the validity of your method. *[5 marks]*

2. *[Maximum mark: 11]*

The n independent random variables $X_1, X_2, \ldots, X_n$ all have the distribution $\mathrm{N}(\mu, \sigma^2)$.

(a) Find the mean and the variance of

(i) $X_1 + X_2$;

(ii) $3X_1$;

(iii) $X_1 + X_2 - X_3$;

(iv) $\overline{X} = \dfrac{(X_1 + X_2 + \quad + X_n)}{n}$. *[8 marks]*

(b) Find $\mathrm{E}(X_1^2)$ in terms of μ and σ. *[3 marks]*

3. *[Maximum mark: 19]*

(a) The random variable X represents the height of a wave on a particular surf beach. It is known that X is normally distributed with unknown mean μ (metres) and known variance $\sigma^2 = \frac{1}{4}$ (metres^2). Sally wishes to test the claim made in a surf guide that $\mu = 3$ against the alternative that $\mu < 3$. She measures the heights of 36 waves and calculates their sample mean $\bar{x}$. She uses this value to test the claim at the 5 % level.

(i) Find a simple inequality, of the form $\bar{x} < A$, where A is a number to be determined to 4 significant figures, so that Sally will reject the null hypothesis, that $\mu = 3$, if and only if this inequality is satisfied.

(ii) Define a Type I error.

(iii) Define a Type II error.

(iv) Write down the probability that Sally makes a Type I error.

(v) The true value of μ is 2.75. Calculate the probability that Sally makes a Type II error. *[11 marks]*

(b) The random variable Y represents the height of a wave on another surf beach. It is known that Y is normally distributed with unknown mean μ (metres) and unknown variance σ^2 (metres^2). David wishes to test the claim made in a surf guide that $\mu = 3$ against the alternative that $\mu < 3$. He is also going to perform this test at the 5 % level. He measures the heights of 36 waves and finds that the sample mean, $\bar{y} = 2.860$ and the unbiased estimate of the population variance, $s_{n-1}^2 = 0.25$.

(i) State the name of the test that David should perform.

(ii) State the conclusion of David's test, justifying your answer by giving the p-value.

(iii) Using David's results, calculate the 90 % confidence interval for μ, giving your answers to 4 significant figures. *[8 marks]*

4. *[Maximum mark: 14]*

Jenny and her Dad frequently play a board game. Before she can start Jenny has to throw a "six" on an ordinary six-sided dice. Let the random variable X denote the number of times Jenny has to throw the dice in total until she obtains her first "six".

(a) If the dice is fair, write down the distribution of X, including the value of any parameter(s). *[1 mark]*

(b) Write down $\mathrm{E}(X)$ for the distribution in part (a). *[1 mark]*

Jenny has played the game with her Dad 216 times and the table below gives the recorded values of X.

Value of X	1	2	3	4	5	6	7	8	9	10	≥ 11
Frequency	40	34	26	24	16	14	12	10	6	4	30

(c) Use this data to test, at the 10 % significance level, the claim that the probability that the dice lands with a "six" uppermost is $\frac{1}{6}$. Justify your conclusion. *[8 marks]*

Before Jenny's Dad can start, he has to throw two "sixes" using a fair, ordinary six-sided dice. Let the random variable Y denote the total number of times Jenny's Dad has to throw the dice until he obtains his second "six".

(d) Write down the distribution of Y, including the value of any parameter(s). *[1 mark]*

(e) Find the value of y such that $\mathrm{P}(Y = y) = \frac{1}{36}$. *[1 mark]*

(f) Find $\mathrm{P}(Y \leq 6)$. *[2 marks]*

22127210

International Baccalaureate® Baccalauréat International Bachillerato Internacional

MATHEMATICS
HIGHER LEVEL
PAPER 3 – STATISTICS AND PROBABILITY

Monday 7 May 2012 (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 SL*** *information booklet* is required for this paper.
- The maximum mark for this examination paper is *[60 marks]*.

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, e.g. 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: 14]*

A baker produces loaves of bread that he claims weigh on average 800 g each. Many customers believe the average weight of his loaves is less than this. A food inspector visits the bakery and weighs a random sample of 10 loaves, with the following results, in grams:

783, 802, 804, 785, 810, 805, 789, 781, 800, 791.

Assume that these results are taken from a normal distribution.

(a) Determine unbiased estimates for the mean and variance of the distribution. *[3 marks]*

In spite of these results the baker insists that his claim is correct.

(b) Stating appropriate hypotheses, test the baker's claim at the 10 % level of significance. *[7 marks]*

The inspector informs the baker that he must improve his quality control and reject any loaf that weighs less than 790 g. The baker changes his production methods and asserts that he has reduced the number of low weight loaves. On a subsequent visit to the bakery the inspector tests a random sample of loaves for sale. Of the 40 loaves tested, 5 should have been rejected.

(c) Calculate a 95 % confidence interval for the proportion of loaves for sale that should be rejected. *[4 marks]*

2. *[Maximum mark: 6]*

The random variable X has a geometric distribution with parameter p.

(a) Show that $\mathrm{P}(X \le n) = 1-(1-p)^n$, $n \in \mathbb{Z}^+$. *[3 marks]*

(b) Deduce an expression for $\mathrm{P}(m < X \le n)$, $m, n \in \mathbb{Z}^+$ and $m < n$. *[1 mark]*

(c) Given that $p = 0.2$, find the least value of n for which $\mathrm{P}(1 < X \le n) > 0.5$, $n \in \mathbb{Z}^+$. *[2 marks]*

3. *[Maximum mark: 13]*

Each week the management of a football club recorded the number of injuries suffered by their playing staff in that week. The results for a 52-week period were as follows:

Number of injuries per week	0	1	2	3	4	5	6
Number of weeks	6	14	15	9	5	2	1

(a) Calculate the mean and variance of the number of injuries per week. *[2 marks]*

(b) Explain why these values provide supporting evidence for using a Poisson distribution model. *[1 mark]*

(c) Stating your hypotheses, test whether a Poisson distribution is a suitable model for the number of injuries per week at the 5 % level of significance using a χ^2 test. *[10 marks]*

4. *[Maximum mark: 19]*

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} 2x, & 0 \le x \le 0.5, \\ \dfrac{4}{3} - \dfrac{2}{3}x, & 0.5 \le x \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Sketch the function f and show that the lower quartile is 0.5. *[3 marks]*

(b) (i) Determine $\mathrm{E}(X)$.

(ii) Determine $\mathrm{E}(X^2)$. *[4 marks]*

Two independent observations are made from X and the values are added. The resulting random variable is denoted Y.

(c) (i) Determine $\mathrm{E}(Y - 2X)$.

(ii) Determine $\mathrm{Var}(Y - 2X)$. *[5 marks]*

(d) (i) Find the cumulative distribution function for X.

(ii) Hence, or otherwise, find the median of the distribution. *[7 marks]*

5. *[Maximum mark: 8]*

The random variable $X \sim \text{Po}(m)$. Given that $\text{P}(X = k-1) = \text{P}(X = k+1)$, where k is a positive integer,

(a) show that $m^2 = k(k+1)$; *[2 marks]*

(b) hence show that the mode of X is k. *[6 marks]*

88117210

MATHEMATICS
HIGHER LEVEL
PAPER 3 – STATISTICS AND PROBABILITY

Friday 4 November 2011 (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.

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, e.g. 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: 15]*

The weight of tea in *Supermug* tea bags has a normal distribution with mean 4.2 g and standard deviation 0.15 g. The weight of tea in *Megamug* tea bags has a normal distribution with mean 5.6 g and standard deviation 0.17 g.

(a) Find the probability that a randomly chosen *Supermug* tea bag contains more than 3.9 g of tea. *[2 marks]*

(b) Find the probability that, of two randomly chosen *Megamug* tea bags, one contains more than 5.4 g of tea and one contains less than 5.4 g of tea. *[4 marks]*

(c) Find the probability that five randomly chosen *Supermug* tea bags contain a total of less than 20.5 g of tea. *[4 marks]*

(d) Find the probability that the total weight of tea in seven randomly chosen *Supermug* tea bags is more than the total weight in five randomly chosen *Megamug* tea bags. *[5 marks]*

2. *[Maximum mark: 7]*

Neil wants the opinion of teachers on a proposal to change the Mathematics HL curriculum. A questionnaire is sent to a large number of teachers asking for their opinions on the proposal. Of the 200 replies he receives, 160 are in favour of the proposal. Assume that these teachers are a random sample from the population.

(a) Test, at the 5 % level, the hypothesis that the proportion of the population in favour of the proposal is 0.75 against the alternative that it is more than 0.75. *[4 marks]*

(b) Find a 95 % confidence interval for the proportion of the population in favour of the proposal. *[3 marks]*

3. *[Maximum mark: 11]*

The random variable X represents the lifetime in hours of a battery. The lifetime may be assumed to be a continuous random variable X with a probability density function given by $f(x) = \lambda e^{-\lambda x}$, where $x \geq 0$.

(a) Name this distribution and state its mean. *[1 mark]*

(b) Find the cumulative distribution function, $F(x)$, of X. *[3 marks]*

(c) Find the probability that the lifetime of a particular battery is more than twice the mean. *[2 marks]*

(d) Find the median of X in terms of λ. *[3 marks]*

(e) Find the probability that the lifetime of a particular battery lies between the median and the mean. *[2 marks]*

4. *[Maximum mark: 16]*

The random variable X is believed to be modelled by $B(5, 0.5)$. A random sample of size 100 is taken and the observed frequencies are given in the table below.

Value of X	0	1	2	3	4	5
Observed frequency	2	15	s	$69 - s$	12	2

A χ^2 goodness of fit test is carried out on these data.

(a) State the null and alternative hypotheses. *[1 mark]*

(b) Evaluate the χ^2 statistic in the form $as^2 + bs + c$. *[8 marks]*

(c) Find the range of values of s that would result in the null hypothesis being accepted at the 10 % level. *[7 marks]*

5. *[Maximum mark: 11]*

The continuous random variable U has a uniform distribution on $[0, 1]$. The random variable X is defined as follows:

$$X = 2U \text{ when } U \leq \frac{3}{4}$$

$$X = 4U \text{ when } U > \frac{3}{4}.$$

(a) (i) Explain why X cannot take values in the interval $\frac{3}{2} < X \leq 3$.

(ii) Find $\mathrm{P}\left(0 \leq X \leq \frac{3}{2}\right)$.

(iii) Find $\mathrm{P}(3 < X \leq 4)$. *[6 marks]*

(b) Find the lower quartile of X. *[3 marks]*

(c) Find $\mathrm{E}(X)$. *[2 marks]*