1 A newspaper article consists of 800 words. For each word, the probability that it is misprinted is 0.005, independently of all other words. Use a suitable approximation to find the probability that the total number of misprinted words in the article is no more than 6. Give a reason to justify your approximation. **[4]**

2 The continuous random variable Y has the distribution $\mathrm{N}(23.0, 5.0^2)$. The mean of n observations of Y is denoted by $\overline{Y}$. It is given that $\mathrm{P}(\overline{Y} > 23.625) = 0.0228$. Find the value of n. **[4]**

3 The number of incidents of radio interference per hour experienced by a certain listener is modelled by a random variable with distribution Po(0.42).

(i) Find the probability that the number of incidents of interference in one randomly chosen hour is

(a) 0,

(b) exactly 1.

[3]

(ii) Find the probability that the number of incidents in a randomly chosen 5-hour period is greater than 3. **[3]**

(iii) One hundred hours of listening are monitored and the numbers of 1-hour periods in which 0, 1, 2, … incidents of interference are experienced are noted. A bar chart is drawn to represent the results. Without any further calculations, sketch the shape that you would expect for the bar chart. (There is no need to use an exact numerical scale on the frequency axis.) **[2]**

4 A television company believes that the proportion of adults who watched a certain programme is 0.14. Out of a random sample of 22 adults, it is found that 2 watched the programme.

(i) Carry out a significance test, at the 10% level, to determine, on the basis of this sample, whether the television company is overestimating the proportion of adults who watched the programme. **[8]**

(ii) The sample was selected randomly. State what properties of this method of sampling are needed to justify the use of the distribution used in your test. **[2]**

5 The continuous random variables S and T have probability density functions as follows.

$$S: \qquad \mathrm{f}(x) = \begin{cases} \frac{1}{4} & -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

$$T: \qquad \mathrm{g}(x) = \begin{cases} \frac{5}{64}x^4 & -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

(i) Sketch, on the same axes, the graphs of f and g. **[3]**

(ii) Describe in everyday terms the difference between the distributions of the random variables S and T. (Answers that comment only on the shapes of the graphs will receive no credit.) **[2]**

(iii) Calculate the variance of T. **[4]**

6 The weight of a plastic box manufactured by a company is W grams, where $W \sim N(\mu, 20.25)$. A significance test of the null hypothesis $H_0 : \mu = 50.0$, against the alternative hypothesis $H_1 : \mu \neq 50.0$, is carried out at the 5% significance level, based on a sample of size n.

(i) Given that $n = 81$,

(a) find the critical region for the test, in terms of the sample mean $\overline{W}$, **[5]**

(b) find the probability that the test results in a Type II error when $\mu = 50.2$. **[5]**

(ii) State how the probability of this Type II error would change if n were greater than 81. **[1]**

7 A motorist records the time taken, T minutes, to drive a particular stretch of road on each of 64 occasions. Her results are summarised by

$$\Sigma t = 876.8, \quad \Sigma t^2 = 12\,657.28.$$

(i) Test, at the 5% significance level, whether the mean time for the motorist to drive the stretch of road is greater than 13.1 minutes. **[11]**

(ii) Explain whether it is necessary to use the Central Limit Theorem in your test. **[1]**

8 A sales office employs 21 representatives. Each day, for each representative, the probability that he or she achieves a sale is 0.7, independently of other representatives. The total number of representatives who achieve a sale on any one day is denoted by K.

(i) Using a suitable approximation (which should be justified), find $P(K \geqslant 16)$. **[7]**

(ii) Using a suitable approximation (which should be justified), find the probability that the mean of 36 observations of K is less than or equal to 14.0. **[7]**

1 A random sample of observations of a random variable X is summarised by

$$n = 100, \quad \Sigma x = 4830.0, \quad \Sigma x^2 = 249\,509.16.$$

(i) Obtain unbiased estimates of the mean and variance of X. [4]

(ii) The sample mean of 100 observations of X is denoted by $\overline{X}$. Explain whether you would need any further information about the distribution of X in order to estimate $\mathrm{P}(\overline{X} > 60)$. [You should not attempt to carry out the calculation.] [2]

2 It is given that on average one car in forty is yellow. Using a suitable approximation, find the probability that, in a random sample of 130 cars, exactly 4 are yellow. [5]

3 The proportion of adults in a large village who support a proposal to build a bypass is denoted by p. A random sample of size 20 is selected from the adults in the village, and the members of the sample are asked whether or not they support the proposal.

(i) Name the probability distribution that would be used in a hypothesis test for the value of p. [1]

(ii) State the properties of a random sample that explain why the distribution in part **(i)** is likely to be a good model. [2]

4 X is a continuous random variable.

(i) State two conditions needed for X to be well modelled by a normal distribution. [2]

(ii) It is given that $X \sim \mathrm{N}(50.0,\ 8^2)$. The mean of 20 random observations of X is denoted by $\overline{X}$. Find $\mathrm{P}(\overline{X} > 47.0)$. [4]

5 The number of system failures per month in a large network is a random variable with the distribution $\mathrm{Po}(\lambda)$. A significance test of the null hypothesis $\mathrm{H}_0 : \lambda = 2.5$ is carried out by counting R, the number of system failures in a period of 6 months. The result of the test is that H_0 is rejected if $R > 23$ but is not rejected if $R \leqslant 23$.

(i) State the alternative hypothesis. [1]

(ii) Find the significance level of the test. [3]

(iii) Given that $\mathrm{P}(R > 23) < 0.1$, use tables to find the largest possible actual value of λ. You should show the values of any relevant probabilities. [3]

6 In a rearrangement code, the letters of a message are rearranged so that the frequency with which any particular letter appears is the same as in the original message. In ordinary German the letter *e* appears 19% of the time. A certain encoded message of 20 letters contains one letter *e*.

(i) Using an exact binomial distribution, test at the 10% significance level whether there is evidence that the proportion of the letter *e* in the language from which this message is a sample is less than in German, i.e., less than 19%. [8]

(ii) Give a reason why a binomial distribution might not be an appropriate model in this context. [1]

7 Two continuous random variables S and T have probability density functions as follows.

$$S: \qquad f(x) = \begin{cases} \frac{1}{2} & -1 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

$$T: \qquad g(x) = \begin{cases} \frac{3}{2}x^2 & -1 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

(i) *Sketch* on the same axes the graphs of $y = f(x)$ and $y = g(x)$. [You should not use graph paper or attempt to plot points exactly.] [3]

(ii) Explain in everyday terms the difference between the two random variables. [2]

(iii) Find the value of t such that $P(T > t) = 0.2$. [5]

8 A random variable Y is normally distributed with mean μ and variance 12.25. Two statisticians carry out significance tests of the hypotheses $H_0 : \mu = 63.0$, $H_1 : \mu > 63.0$.

(i) Statistician A uses the mean $\overline{Y}$ of a sample of size 23, and the critical region for his test is $\overline{Y} > 64.20$. Find the significance level for A's test. [4]

(ii) Statistician B uses the mean of a sample of size 50 and a significance level of 5%.

(a) Find the critical region for B's test. [3]

(b) Given that $\mu = 65.0$, find the probability that B's test results in a Type II error. [4]

(iii) Given that, when $\mu = 65.0$, the probability that A's test results in a Type II error is 0.1365, state with a reason which test is better. [2]

9 **(a)** The random variable G has the distribution $B(n, 0.75)$. Find the set of values of n for which the distribution of G can be well approximated by a normal distribution. [3]

(b) The random variable H has the distribution $B(n, p)$. It is given that, using a normal approximation, $P(H \geqslant 71) = 0.0401$ and $P(H \leqslant 46) = 0.0122$.

(i) Find the mean and standard deviation of the approximating normal distribution. [6]

(ii) Hence find the values of n and p. [4]

Answer **all** questions.

1 The administration department at Hitech College uses two photocopiers, A and B, which operate independently.

The number of breakdowns each week, X, for photocopier A can be modelled by a Poisson distribution with a mean of 0.5 .

The number of breakdowns each week, Y, for photocopier B can be modelled by a Poisson distribution with a mean of 2.5 .

(a) Calculate:

(i) $\mathrm{P}(X \leqslant 1)$; *(2 marks)*

(ii) $\mathrm{P}(Y = 2)$. *(2 marks)*

(b) (i) In a given week, show that the probability of a total of exactly 2 photocopier breakdowns is 0.224, correct to three decimal places. *(3 marks)*

(ii) Hence find the probability that there will be exactly 2 photocopier breakdowns in each of four consecutive weeks. *(2 marks)*

(c) The total number of photocopier breakdowns, T, in a 4-week period can be modelled by a Poisson distribution with mean μ.

(i) Find the value of μ. *(1 mark)*

(ii) Hence calculate the probability that, in a given 4-week period, there will be a total of at least 18 photocopier breakdowns. *(2 marks)*

2 Year 12 students at Newstatus School choose to participate in one of four sports during the Spring term.

The students' choices are summarised in the table.

	Squash	**Badminton**	**Archery**	**Hockey**	**Total**
Male	5	16	30	19	70
Female	4	20	33	53	110
Total	9	36	63	72	180

(a) Use a χ^2 test, at the 5% level of significance, to determine whether the choice of sport is independent of gender. *(10 marks)*

(b) Interpret your result in part (a) as it relates to students choosing hockey. *(2 marks)*

3 The weight, W grams, of rice contained in a packet can be modelled by a normal distribution with mean μ and variance σ^2.

The weights, in grams, of rice contained in each of a random sample of 10 packets were

301.2	305.6	298.4	297.7	292.6
306.1	296.8	299.2	304.0	308.4

(a) Calculate unbiased estimates for μ and σ^2. *(2 marks)*

(b) Hence construct a 99% confidence interval for μ, giving the limits to one decimal place. *(5 marks)*

(c) The packets are claimed to contain, on average, at least 290 grams of rice.

Comment on this claim. *(2 marks)*

4 (a) A random variable X has probability density function defined by

$$f(x) = \begin{cases} k & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

(i) Show that $k = \dfrac{1}{b-a}$. *(1 mark)*

(ii) Prove, using integration, that $E(X) = \frac{1}{2}(a+b)$. *(4 marks)*

(b) The error, X grams, made when a shopkeeper weighs out loose sweets can be modelled by a rectangular distribution with the following probability density function:

$$f(x) = \begin{cases} k & -2 < x < 4 \\ 0 & \text{otherwise} \end{cases}$$

(i) Write down the value of the mean, μ, of X. *(1 mark)*

(ii) Evaluate the standard deviation, σ, of X. *(2 marks)*

(iii) Hence find $P\left(X < \dfrac{2-\mu}{\sigma}\right)$. *(3 marks)*

Turn over for the next question

5 The Globe Express agency organises trips to the theatre. The cost, £X, of these trips can be modelled by the following probability distribution:

x	40	45	55	74
$\mathrm{P}(X=x)$	0.30	0.24	0.36	0.10

(a) Calculate the mean and standard deviation of X. *(4 marks)*

(b) For special celebrity charity performances, Globe Express increases the cost of the trips to £Y, where

$$Y = 10X + 250$$

Determine the mean and standard deviation of Y. *(2 marks)*

6 Bottles of sherry nominally contain 1000 millilitres. After the introduction of a new method of filling the bottles, there is a suspicion that the mean volume of sherry in a bottle has changed.

In order to investigate this suspicion, a random sample of 12 bottles of sherry is taken and the volume of sherry in each bottle is measured.

The volumes, in millilitres, of sherry in these bottles are found to be

996	1006	1009	999	1007	1003
998	1010	997	996	1008	1007

Assuming that the volume of sherry in a bottle is normally distributed, investigate, at the 5% level of significance, whether the mean volume of sherry in a bottle differs from 1000 millilitres. *(10 marks)*

END OF QUESTIONS

Answer **all** questions.

1 The number of accidents, X, occurring during one week at Joanne's place of work can be modelled by a Poisson distribution with a mean of 0.7.

The number of accidents, Y, occurring during one week at Pete's place of work can be modelled by a Poisson distribution with a mean of 1.3.

(a) (i) Determine $\mathrm{P}(X < 3)$. *(1 mark)*

(ii) Calculate $\mathrm{P}(Y = 2)$. *(2 marks)*

(b) Find the probability that, during a particular week, there are at least 4 accidents in total at these two places of work. *(3 marks)*

2 The marks achieved by Pat in her homework assignments may be assumed to be normally distributed with mean μ.

The marks achieved by Pat in a random sample of 8 assignments were recorded as follows:

60 65 62 67 69 71 63 66

Construct a 99% confidence interval for μ. *(5 marks)*

3 The handicap committee of a golf club has indicated that the mean score achieved by the club's members in the past was 85.9.

A group of members believes that recent changes to the golf course have led to a change in the mean score achieved by the club's members and decides to investigate this belief.

A random sample of the scores, x, of 100 club members was taken and is summarised by

$$\sum x = 8350 \quad \text{and} \quad \sum (x - \bar{x})^2 = 15\,321$$

where $\bar{x}$ denotes the sample mean.

Test, at the 5% level of significance, the group's belief that the mean score of 85.9 has changed. *(8 marks)*

4 The number of mistakes, X, that Holli makes as a learner driver when she drives from Ampthill to Bedford can be modelled by the following discrete probability distribution:

x	$\leqslant 1$	2	3	4	5	6	$\geqslant 7$
$\mathrm{P}(X = x)$	0	0.40	0.25	0.18	0.12	k	0

(a) Find the value of k. *(1 mark)*

(b) Find:

(i) $\mathrm{E}(X)$; *(1 mark)*

(ii) $\mathrm{Var}(X)$. *(3 marks)*

(c) When Holli makes the return journey by the same route, the number of mistakes, Y, that she makes can be approximated by

$$Y = 2X - 3$$

Find:

(i) $\mathrm{E}(Y)$; *(1 mark)*

(ii) the standard deviation of Y. *(3 marks)*

5 Jasmine's French teacher states that a homework assignment should take, on average, 30 minutes to complete.

Jasmine believes that he is understating the mean time that the assignment takes to complete and so decides to investigate. She records the times, in minutes, that it takes for a random sample of 10 students to complete the French assignment, with the following results:

29 33 36 42 30 28 31 34 37 35

(a) Test, at the 1% level of significance, Jasmine's belief that her French teacher has understated the mean time that it should take to complete the homework assignment. *(7 marks)*

(b) State an assumption that you must make in order for the test used in part (a) to be valid. *(1 mark)*

Turn over for the next question

6 The waiting time, T minutes, before being served at a local newsagents can be modelled by a continuous random variable with probability density function

$$f(t) = \begin{cases} \frac{3}{8}t^2 & 0 \leqslant t < 1 \\ \frac{1}{16}(t+5) & 1 \leqslant t \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch the graph of f. *(3 marks)*

(b) For a customer selected at random, calculate $P(T \geqslant 1)$. *(2 marks)*

(c) (i) Show that the cumulative distribution function for $1 \leqslant t \leqslant 3$ is given by

$$F(t) = \frac{1}{32}(t^2 + 10t - 7)$$ *(5 marks)*

(ii) Hence find the median waiting time. *(4 marks)*

7 A statistics unit is required to determine whether or not there is an association between students' performances in mathematics at Key Stage 3 and at GCE.

A survey of the results of 500 students showed the following information:

		GCE Grade				
		A	**B**	**C**	**Below C**	**Total**
Key Stage 3 Level	**8**	60	55	47	43	205
	7	55	32	31	26	144
	6	40	38	35	38	151
	Total	155	125	113	107	500

(a) Use a χ^2 test at the 10% level of significance to determine whether there is an association between students' performances in mathematics at Key Stage 3 and at GCE. *(9 marks)*

(b) Comment on the number of students who gained a grade A at GCE having gained a level 7 at Key Stage 3. *(1 mark)*

END OF QUESTIONS

Answer **all** questions.

1 The number of A-grades, X, achieved in total by students at Lowkey School in their Mathematics examinations each year can be modelled by a Poisson distribution with a mean of 3.

(a) Determine the probability that, during a 5-year period, students at Lowkey School achieve a total of more than 18 A-grades in their Mathematics examinations. *(3 marks)*

(b) The number of A-grades, Y, achieved in total by students at Lowkey School in their English examinations each year can be modelled by a Poisson distribution with a mean of 7.

(i) Determine the probability that, during a year, students at Lowkey School achieve a total of fewer than 15 A-grades in their Mathematics and English examinations. *(3 marks)*

(ii) What assumption did you make in answering part (b)(i)? *(1 mark)*

2 It is claimed that the area within which a school is situated affects the age profile of the staff employed at that school. In order to investigate this claim, the age profiles of staff employed at two schools with similar academic achievements are compared.

Academia High School, situated in a rural community, employs 120 staff whilst Best Manor Grammar School, situated in an inner-city community, employs 80 staff.

The **percentage** of staff within each age group, for each school, is given in the table.

Age	**Academia High School**	**Best Manor Grammar School**
22–34	17.5	40.0
35–39	60.0	45.0
40–59	22.5	15.0

(a) (i) Form the data into a contingency table suitable for analysis using a χ^2 distribution. *(2 marks)*

(ii) Use a χ^2 test, at the 1% level of significance, to determine whether there is an association between the age profile of the staff employed and the area within which the school is situated. *(9 marks)*

(b) Interpret your result in part (a)(ii) as it relates to the 22–34 age group. *(2 marks)*

3 The number of aces, X, served by a tennis player during a set can be modelled by the following probability distribution.

x	0	1	2	3	4	5
$\mathrm{P}(X=x)$	0.01	0.09	0.13	0.50	0.15	0.12

(a) Calculate the mean, μ, and the standard deviation, σ, of X. *(4 marks)*

(b) For a set selected at random, calculate $\mathrm{P}(\mu - \sigma < X < \mu + \sigma)$. *(3 marks)*

4 The error, R centilitres, made in measuring the contents of bottles of wine can be modelled by the rectangular distribution

$$\mathrm{f}(r) = \begin{cases} \alpha & -0.4 < r < 0.4 \\ 0 & \text{otherwise} \end{cases}$$

where α is a positive constant.

(a) Find the value of α. *(1 mark)*

(b) Calculate the mean and the standard deviation of R. *(3 marks)*

(c) For a bottle of wine selected at random, calculate the probability that the **magnitude** of the error made in measuring its contents is less than 0.3 cl. *(2 marks)*

Turn over for the next question

5 The lifetime, X hours, of Everwhite camera batteries is normally distributed. The manufacturer claims that the mean lifetime of these batteries is 100 hours.

(a) The members of a photography club suspect that the batteries do not last as long as is claimed by the manufacturer. In order to investigate their suspicion, the members test a random sample of five of these batteries and find the lifetimes, in hours, to be as follows:

85 92 100 95 99

Test the members' suspicion at the 5% level of significance. *(9 marks)*

(b) The manufacturer, believing that the mean lifetime of these batteries has not changed from 100 hours, decides to determine the lifetime, x hours, of each of a random sample of 80 Everwhite camera batteries. The manufacturer obtains the following results, where $\bar{x}$ denotes the sample mean:

$$\sum x = 8080 \quad \text{and} \quad \sum (x - \bar{x})^2 = 6399$$

Test the manufacturer's belief at the 5% level of significance. *(8 marks)*

6 The continuous random variable X has probability density function defined by

$$f(x) = \begin{cases} \frac{1}{5}(2x+1) & 0 \leqslant x \leqslant 1 \\ \frac{1}{15}(4-x)^2 & 1 < x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch the graph of f. *(2 marks)*

(b) (i) Show that the cumulative distribution function, $F(x)$, for $0 \leqslant x \leqslant 1$ is

$$F(x) = \frac{1}{5}x(x+1)$$ *(3 marks)*

(ii) Hence find $P(X \geqslant 0.5)$. *(2 marks)*

(iii) Verify that the lower quartile, q_1, satisfies

$$0.5 < q_1 < 0.75$$ *(3 marks)*

END OF QUESTIONS

Answer **all** questions.

1 Two groups of patients, suffering from the same medical condition, took part in a clinical trial of a new drug. One of the groups was given the drug whilst the other group was given a placebo, a drug that has no physical effect on their medical condition.

The table shows the number of patients in each group and whether or not their condition improved.

	Placebo	**Drug**
Condition improved	20	46
Condition did not improve	55	29

Conduct a χ^2 test, at the 5% level of significance, to determine whether the condition of the patients at the conclusion of the trial is associated with the treatment that they were given.
(10 marks)

2 The number of telephone calls per day, X, received by Candice may be modelled by a Poisson distribution with mean 3.5 .

The number of e-mails per day, Y, received by Candice may be modelled by a Poisson distribution with mean 6.0 .

(a) For any particular day, find:

(i) $\mathrm{P}(X = 3)$; *(2 marks)*

(ii) $\mathrm{P}(Y \geqslant 5)$. *(2 marks)*

(b) (i) Write down the distribution of T, the total number of telephone calls and e-mails per day received by Candice. *(1 mark)*

(ii) Determine $\mathrm{P}(7 \leqslant T \leqslant 10)$. *(3 marks)*

(iii) Hence calculate the probability that, on each of three consecutive days, Candice will receive a total of at least 7 but at most 10 telephone calls and e-mails.
(2 marks)

3 David is the professional coach at the golf club where Becki is a member. He claims that, after having a series of lessons with him, the mean number of putts that Becki takes per round of golf will reduce from her present mean of 36.

After having the series of lessons with David, Becki decides to investigate his claim.

She therefore records, for each of a random sample of 50 rounds of golf, the number of putts, x, that she takes to complete the round. Her results are summarised below, where $\bar{x}$ denotes the sample mean.

$$\sum x = 1730 \qquad \text{and} \qquad \sum (x - \bar{x})^2 = 784$$

Using a z-test and the 1% level of significance, investigate David's claim. *(8 marks)*

4 Ten students each independently carried out the same experiment in order to measure, in $\mathrm{m\,s^{-2}}$, the value of g, the acceleration due to gravity, with the following results:

9.75	9.72	9.71	9.69	9.66	9.70	9.72	9.71	9.69	9.65

(a) Assuming that values from the experiment are normally distributed, with mean g, construct a 95% confidence interval for g. *(6 marks)*

(b) It was subsequently discovered that the equipment used in the experiment was faulty. As a consequence, each of the values above is $0.10\,\mathrm{m\,s^{-2}}$ less than the actual value.

Use this additional information to write down a revised 95% confidence interval for g. *(2 marks)*

Turn over for the next question

5 A discrete random variable X has probability distribution as given in the table.

$\boldsymbol{x}$	1	2	3	4
$\mathbf{P}(\boldsymbol{X=x})$	p	p	p	$1-3p$

(a) Show that, for this to be a valid distribution, $0 \leqslant p \leqslant \frac{1}{3}$. *(3 marks)*

(b) (i) Find an expression, in terms of p, for $\mathrm{E}(X)$. *(1 mark)*

(ii) Show that $\mathrm{Var}(X) = 2p(7-18p)$. *(3 marks)*

(c) (i) Find the value of p for which $\mathrm{Var}(X)$ is a maximum. *(2 marks)*

(ii) Find the maximum value of the standard deviation of X. *(3 marks)*

6 The continuous random variable X has the probability density function given by

$$\mathrm{f}(x) = \begin{cases} 3x^2 & 0 < x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine:

(i) $\mathrm{E}\left(\dfrac{1}{X}\right)$; *(3 marks)*

(ii) $\mathrm{Var}\left(\dfrac{1}{X}\right)$. *(4 marks)*

(b) Hence, or otherwise, find the mean and the variance of $\left(\dfrac{5+2X}{X}\right)$. *(5 marks)*

END OF QUESTIONS