Answer **all** questions.

1 An analysis of a random sample of 150 urban dwellings for sale showed that 102 are semi-detached.

An analysis of an independent random sample of 80 rural dwellings for sale showed that 36 are semi-detached.

(a) Construct an approximate 99% confidence interval for the difference between the proportion of urban dwellings for sale that are semi-detached and the proportion of rural dwellings for sale that are semi-detached. *(6 marks)*

(b) Hence comment on the claim that there is no difference between these two proportions. *(2 marks)*

2 A hotel chain has hotels in three types of location: city, coastal and country. The percentages of the chain's reservations for each of these locations are 30, 55 and 15 respectively.

Each of the chain's hotels offers three types of reservation: Bed & Breakfast, Half Board and Full Board.

The percentages of these types of reservation for **each** of the three types of location are shown in the table.

		Type of location		
		City	**Coastal**	**Country**
Type of reservation	**Bed & Breakfast**	80	10	30
	Half Board	15	65	50
	Full Board	5	25	20

For example, 80 per cent of reservations for hotels in city locations are for Bed & Breakfast.

(a) For a reservation selected at random:

(i) show that the probability that it is for Bed & Breakfast is 0.34; *(2 marks)*

(ii) calculate the probability that it is for Half Board in a hotel in a coastal location; *(2 marks)*

(iii) calculate the probability that it is for a hotel in a coastal location, given that it is for Half Board. *(4 marks)*

(b) A random sample of 3 reservations for Half Board is selected.

Calculate the probability that these 3 reservations are for hotels in different types of location. *(5 marks)*

3 The proportion, p, of an island's population with blood type A Rh^+ is believed to be approximately 0.35 .

A medical organisation, requiring a more accurate estimate, specifies that a 98% confidence interval for p should have a width of at most 0.1 .

Calculate, to the nearest 10, an estimate of the minimum sample size necessary in order to achieve the organisation's requirement. *(6 marks)*

4 Holly, a horticultural researcher, believes that the mean height of stems on Tahiti daffodils exceeds that on Jetfire daffodils by more than 15 cm.

She measures the heights, x centimetres, of stems on a random sample of 65 Tahiti daffodils and finds that their mean, $\bar{x}$, is 40.7 and that their standard deviation, s_x, is 3.4 .

She also measures the heights, y centimetres, of stems on a random sample of 75 Jetfire daffodils and finds that their mean, $\bar{y}$, is 24.4 and that their standard deviation, s_y, is 2.8 .

Investigate, at the 1% level of significance, Holly's belief. *(8 marks)*

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

(a) Given that

$$\mathrm{E}(X) = np \quad \text{and} \quad \mathrm{E}(X(X-1)) = n(n-1)p^2$$

find an expression for $\mathrm{Var}(X)$. *(3 marks)*

(b) Given that X has a mean of 36 and a standard deviation of 4.8 :

(i) find values for n and p; *(3 marks)*

(ii) use a distributional approximation to estimate $\mathrm{P}(30 < X < 40)$. *(4 marks)*

6 The table shows the probability distribution for the number of weekday (Monday to Friday) morning newspapers, X, purchased by the Reed household per week.

x	0	1	2	3	4	5
P(X=x)	0.16	0.15	0.25	0.25	0.15	0.04

(a) Find values for $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. *(3 marks)*

(b) The number of weekday (Monday to Friday) evening newspapers, Y, purchased by the same household per week is such that

$$\mathrm{E}(Y) = 2.0, \quad \mathrm{Var}(Y) = 1.5 \quad \text{and} \quad \mathrm{Cov}(X, Y) = -0.43$$

Find values for the mean and variance of:

(i) $S = X + Y$;

(ii) $D = X - Y$. *(5 marks)*

(c) The total cost per week, L, of the Reed household's weekday morning and evening newspapers may be assumed to be normally distributed with a mean of £2.31 and a standard deviation of £0.89.

The total cost per week, M, of the household's weekend (Saturday and Sunday) newspapers may be assumed to be independent of L and normally distributed with a mean of £2.04 and a standard deviation of £0.43.

Determine the probability that the total cost per week of the Reed household's newspapers is more than £5. *(5 marks)*

7 The daily number of customers visiting a small arts and crafts shop may be modelled by a Poisson distribution with a mean of 24.

(a) Using a distributional approximation, estimate the probability that there was a total of at most 150 customers visiting the shop during a given 6-day period. *(5 marks)*

(b) The shop offers a picture framing service. The daily number of requests, Y, for this service may be assumed to have a Poisson distribution.

Prior to the shop advertising this service in the local free newspaper, the mean value of Y was 2. Following the advertisement, the shop received a total of 17 requests for the service during a period of 5 days.

(i) Using a Poisson distribution, carry out a test, at the 10% level of significance, to investigate the claim that the advertisement increased the mean daily number of requests for the shop's picture framing service. *(5 marks)*

(ii) Determine the critical value of Y for your test in part (b)(i). *(3 marks)*

(iii) Hence, assuming that the advertisement increased the mean value of Y to 3, determine the power of your test in part (b)(i). *(4 marks)*

END OF QUESTIONS

Answer **all** questions in the spaces provided.

1 Ffion, as part of her research project, measured the stem length and the cap diameter of each of a random sample of 24 matsutake mushrooms. Using these measurements, she calculated the value of the product moment correlation coefficient to be 0.336, correct to three significant figures.

Assuming that her measurements came from a bivariate normal distribution, test, at the 5% level of significance, the hypothesis that there is no correlation between the stem length and the cap diameter of matsutake mushrooms. *(4 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

2 Rodney and Derrick, two independent fruit and vegetable market stallholders, sell punnets of locally-grown raspberries from their stalls during June and July.

The following information, based on independent random samples, was collected as part of an investigation by Trading Standards Officers.

			Weight of raspberries in a punnet (grams)	
		Sample size	**Sample mean**	**Sample standard deviation, s**
Stallholder	**Rodney**	50	225	5
	Derrick	75	219	8

(a) Construct a 99% confidence interval for the difference between the mean weight of raspberries in a punnet sold by Rodney and the mean weight of raspberries in a punnet sold by Derrick. *(5 marks)*

(b) What can be concluded from your confidence interval? *(2 marks)*

(c) In addition to weight, state one other factor that may influence whether customers buy raspberries from Rodney or from Derrick. *(1 mark)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

3 The weekly number of hits, S, on Sam's website may be modelled by a Poisson distribution with parameter λ_S. The weekly number of hits, T, on Tina's website may be modelled by a Poisson distribution with parameter λ_T.

During a period of 40 weeks, the number of hits on Sam's website was 940.

During a period of 60 weeks, the number of hits on Tina's website was 1560.

Assuming that S and T are independent random variables, investigate, at the 2% level of significance, Tina's claim that the mean weekly number of hits on her website is greater than that on Sam's website. *(7 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

4 It is proposed to introduce, for all males at age 60, screening tests, A and B, for a certain disease.

Test B is administered only when the result of Test A is inconclusive.

It is known that 10% of 60-year-old men suffer from the disease.

For those 60-year-old men suffering from the disease:

- Test A is known to give a positive result, indicating a presence of the disease, in 90% of cases, a negative result in 2% of cases and a requirement for the administration of Test B in 8% of cases;
- Test B is known to give a positive result in 98% of cases and a negative result in 2% of cases.

For those 60-year-old men not suffering from the disease:

- Test A is known to give a positive result in 1% of cases, a negative result in 80% of cases and a requirement for the administration of Test B in 19% of cases;
- Test B is known to give a positive result in 1% of cases and a negative result in 99% of cases.

(a) Draw a tree diagram to represent the above information. *(4 marks)*

(b) (i) Hence, or otherwise, determine the probability that:

(A) a 60-year-old man, suffering from the disease, tests negative;

(B) a 60-year-old man, not suffering from the disease, tests positive. *(2 marks)*

(ii) A random sample of ten thousand 60-year-old men is given the screening tests. Calculate, to the nearest 10, the number who you would expect to be given an **incorrect** diagnosis. *(2 marks)*

(c) Determine the probability that:

(i) a 60-year-old man suffers from the disease given that the tests provide a positive result;

(ii) a 60-year-old man does not suffer from the disease given that the tests provide a negative result. *(5 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

QUESTION PART REFERENCE

5 In the manufacture of desk drawer fronts, a machine cuts sheets of veneered chipboard into rectangular pieces of width W millimetres and height H millimetres. The 4 edges of each of these pieces are then covered with matching veneered tape.

The distributions of W and H are such that

$$\mathrm{E}(W) = 350 \qquad \mathrm{Var}(W) = 5 \qquad \mathrm{E}(H) = 210 \qquad \mathrm{Var}(H) = 4 \qquad \rho_{WH} = 0.75$$

(a) Calculate the mean and the variance of the length of tape, $T = 2W + 2H$, needed for the edges of a drawer front. *(5 marks)*

(b) A desk has 4 such drawers whose sizes may be assumed to be independent.

Given that T may be assumed to be normally distributed, determine the probability that the total length of tape needed for the edges of the desk's 4 drawer fronts does not exceed 4.5 metres. *(5 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

6 (a) A district council claimed that more than 80 per cent of the complaints that it received about the delivery of its services were answered to the satisfaction of complainants before reaching formal status.

An analysis of a random sample of 175 complaints revealed that 28 reached formal status.

(i) Construct an approximate 95% confidence interval for the proportion of complaints that reach formal status. *(5 marks)*

(ii) Hence comment on the council's claim. *(2 marks)*

(b) The district council also claimed that less than 40 per cent of all formal complaints were due to a failing in the delivery of its services.

An analysis of the 50 formal complaints received during 2007/08 showed that 16 were due to a failing in the delivery of its services.

(i) Using an exact test, investigate the council's claim at the 10% level of significance. The 50 formal complaints received during 2007/08 may be assumed to be a random sample. *(5 marks)*

(ii) Determine the critical value for your test in part **(b)(i)**. *(2 marks)*

(iii) In fact, only 25 per cent of all formal complaints were due to a failing in the delivery of the council's services.

Determine the probability of a Type II error for a test of the council's claim at the 10% level of significance and based on the analysis of a random sample of 50 formal complaints. *(4 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

QUESTION PART REFERENCE

QUESTION PART REFERENCE

7 The random variable X has a Poisson distribution with parameter λ.

(a) (i) Prove, from first principles, that $E(X) = \lambda$. *(3 marks)*

(ii) Hence, given that $E(X(X-1)) = \lambda^2$, find, in terms of λ, an expression for $\text{Var}(X)$. *(2 marks)*

(b) The mode, m, of X is such that

$$P(X = m) \geqslant P(X = m - 1) \quad \text{and} \quad P(X = m) \geqslant P(X = m + 1)$$

(i) Show that $\lambda - 1 \leqslant m \leqslant \lambda$. *(3 marks)*

(ii) Given that $\lambda = 4.9$, determine $P(X = m)$. *(2 marks)*

(c) The random variable Y has a Poisson distribution with mode d and standard deviation 15.5.

Use a distributional approximation to estimate $P(Y \geqslant d)$. *(5 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Turn over ▶

QUESTION PART REFERENCE

END OF QUESTIONS

Answer **all** questions in the spaces provided.

1 A consumer report claimed that more than 25 per cent of visitors to a theme park were dissatisfied with the catering facilities provided.

In a survey, 375 visitors who had used the catering facilities were interviewed independently, and 108 of them stated that they were dissatisfied with the catering facilities provided.

(a) Test, at the 2% level of significance, the consumer report's claim. *(6 marks)*

(b) State an assumption about the 375 visitors that was necessary in order for the hypothesis test in part **(a)** to be valid. *(1 mark)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Turn over ▶

2 The number of emergency calls received by a fire station may be modelled by a Poisson distribution.

During a given period of 13 weeks, the station received a total of 108 emergency calls.

(a) Construct an approximate 98% confidence interval for the average **weekly** number of emergency calls received by the station. *(5 marks)*

(b) Hence comment on the station officer's claim that the station receives an average of one emergency call **per day**. *(2 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

3 An IT help desk has three telephone stations: Alpha, Beta and Gamma. Each of these stations deals only with telephone enquiries.

The probability that an enquiry is received at Alpha is 0.60.
The probability that an enquiry is received at Beta is 0.25.
The probability that an enquiry is received at Gamma is 0.15.

Each enquiry is resolved at the station that receives the enquiry. The **percentages** of enquiries resolved within various times at **each** station are shown in the table.

		Time		
		$\leqslant$ 1 hour	**$\leqslant$ 24 hours**	**$\leqslant$ 72 hours**
Station	**Alpha**	55	80	100
	Beta	60	85	100
	Gamma	40	75	100

For example:
80 per cent of enquiries received at Alpha are resolved within 24 hours;
25 per cent of enquiries received at Alpha take between 1 hour and 24 hours to resolve.

(a) Find the probability that an enquiry, selected at random, is:

(i) resolved at Gamma; *(1 mark)*

(ii) resolved at Alpha within 1 hour; *(1 mark)*

(iii) resolved within 24 hours; *(2 marks)*

(iv) received at Beta, given that it is resolved within 24 hours. *(3 marks)*

(b) A random sample of 3 enquiries was selected.

Given that all 3 enquiries were resolved within 24 hours, calculate the probability that they were all received at:

(i) Beta; *(2 marks)*

(ii) the same station. *(4 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Turn over ▶

4 The waiting time at a hospital's A&E department may be modelled by a normal distribution with mean μ and standard deviation $\dfrac{\mu}{2}$.

The department's manager wishes a 95% confidence interval for μ to be constructed such that it has a width of at most 0.2μ.

Calculate, to the nearest 10, an estimate of the minimum sample size necessary in order to achieve the manager's wish. *(5 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

5 An examination of 160 e-mails received by Gopal showed that 72 had attachments.

An examination of 250 e-mails received by Haley showed that 102 had attachments.

Stating **two** necessary assumptions about the selection of e-mails, construct an approximate 99% confidence interval for the difference between the proportion of e-mails received by Gopal that have attachments and the proportion of e-mails received by Haley that have attachments. *(8 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

6 The weight, X grams, of a dressed pheasant may be modelled by a normal random variable with a mean of 1000 and a standard deviation of 120.

Pairs of dressed pheasants are selected for packing into boxes. The total weight of a pair, $Y = X_1 + X_2$ grams, may be modelled by a normal distribution with a mean of 2000 and a standard deviation of 140.

(a) (i) Show that $\text{Cov}(X_1, X_2) = -4600$. *(3 marks)*

(ii) Given that $X_1 - X_2$ may be assumed to be normally distributed, determine the probability that the difference between the weights of a selected pair of dressed pheasants exceeds 250 grams. *(5 marks)*

(b) The weight of a box is independent of the total weight of a pair of dressed pheasants, and is normally distributed with a mean of 500 grams and a standard deviation of 40 grams.

Determine the probability that a box containing a pair of dressed pheasants weighs less than 2750 grams. *(5 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Turn over ▶

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Turn over ▶

7 (a) The random variable X has a Poisson distribution with $\mathrm{E}(X) = \lambda$.

(i) Prove, from first principles, that $\mathrm{E}(X(X-1)) = \lambda^2$. *(3 marks)*

(ii) Hence deduce that $\mathrm{Var}(X) = \mathrm{E}(X)$. *(2 marks)*

(b) The random variable Y has a Poisson distribution with $\mathrm{E}(Y) = 2.5$.

Given that $Z = 4Y + 30$:

(i) show that $\mathrm{Var}(Z) = \mathrm{E}(Z)$; *(3 marks)*

(ii) give a reason why the distribution of Z is **not** Poisson. *(1 mark)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Turn over ▶

8 The tensile strength of rope is measured in kilograms.

The standard deviation of the tensile strength of a particular design of 10 mm diameter rope is known to be 285 kilograms. A retail organisation, which buys such rope from two manufacturers, A and B, wishes to compare their ropes for mean tensile strength.

The mean tensile strength, $\bar{x}$, of a random sample of 80 lengths from manufacturer A was 3770 kilograms.

The mean tensile strength, $\bar{y}$, of a random sample of 120 lengths from manufacturer B was 3695 kilograms.

(a) (i) Test, at the 5% level of significance, the hypothesis that there is no difference between the mean tensile strength of rope from manufacturer A and that of rope from manufacturer B. *(6 marks)*

(ii) Why was it **not** necessary to know the distributions of tensile strength in order for your test in part **(a)(i)** to be valid? *(1 mark)*

(b) (i) Deduce that, for your test in part **(a)(i)**, the critical values of $(\bar{x} - \bar{y})$ are ± 80.63, correct to two decimal places. *(2 marks)*

(ii) In fact, the mean tensile strength of rope from manufacturer A exceeds that of rope from manufacturer B by 125 kilograms.

Determine the probability of a Type II error for a test of the hypothesis in part **(a)(i)** at the 5% level of significance, based upon a random sample of 80 lengths from manufacturer A and a random sample of 120 lengths from manufacturer B. *(4 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

QUESTION PART REFERENCE

END OF QUESTIONS

1 The lengths of rivets produced by a certain factory are checked each day by measuring a random sample of 100 rivets. For a particular day's sample the lengths, x mm, are summarised by $\Sigma x = 761.2$ and $\Sigma x^2 = 6115.04$. The mean and standard deviation of the lengths of all rivets produced that day are denoted by μ mm and σ mm respectively.

(i) Find an unbiased estimate of σ^2. [2]

(ii) Calculate a 95% confidence interval for μ. [3]

(iii) Explain what distributional assumptions (if any) are required for the validity of your calculated confidence interval. [1]

2 The saturated fat content of a particular brand of olive spread is monitored regularly in order to maintain a mean percentage content of 12.6. This is carried out by measuring the saturated fat content in random samples of 10 cartons. For a particular sample, the sample mean and an unbiased estimate of the population variance are calculated. The unbiased estimate of the population variance is 0.1195. It may be assumed that percentage fat content has a normal distribution. Find the critical region for a test at the 10% significance level of whether the population mean percentage fat content exceeds 12.6. [5]

3 A large sample of people were surveyed and classified by 4 levels of income and by which of 3 newspapers they read. The results were arranged in a contingency table consisting of 4 columns and 3 rows. In a χ^2 test of independence between income and choice of newspaper, it was found necessary to combine two of the columns. The value of the test statistic was 12.32.

(i) State a suitable null hypothesis for the test. [1]

(ii) Determine the largest significance level, obtained from tables or calculator, for which independence would be accepted. [3]

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

$$f(x) = \begin{cases} 0 & x < 0, \\ \frac{4}{3}x^3 & 0 \leqslant x \leqslant 1, \\ \frac{4}{3x^3} & x > 1. \end{cases}$$

(i) Find $P(X < 2)$. [3]

(ii) Show that the median of X exceeds 1. [3]

(iii) Find $E(X)$. [3]

(iv) Show that $\mathrm{Var}(X)$ is not finite. [3]

5 The proportion of syringes of brand A that are faulty is 2.2%. The corresponding proportion for brand B is 2.5%. Random samples of 75 brand A and 90 brand B syringes are taken and the total number of faulty syringes is denoted by X.

(i) Show that the distribution of X can be approximated by a Poisson distribution, and state its mean. [5]

(ii) Find $P(X > 5)$. [2]

6 The proportion of teapots with faulty spouts produced in a factory is denoted by p. In a random sample of 50 teapots, the number with faulty spouts was found to be 6.

(i) Find a 98% confidence interval for p. [4]

(ii) Find an estimate of the sample size for which the sample proportion would differ from p by less than 0.05 with 98% confidence. [5]

7 A psychologist believed that teenage boys worry more than teenage girls and he devised a questionnaire to examine his belief. He gave the questionnaire to a random sample of 24 girls and a random sample of 18 boys. The scores, x_G and x_B for the girls and boys, are summarised by $\Sigma x_G = 1526.8$ and $\Sigma x_B = 1238.4$. Unbiased estimates of the respective population variances, obtained from the samples, are $s_G^2 = 86.79$ and $s_B^2 = 93.01$. Larger scores indicate greater levels of worry.

(i) State two assumptions required for the validity of a t-test to examine the psychologist's belief. [2]

(ii) Comment on one of these assumptions in the light of the data. [1]

(iii) Carry out the test at the 5% significance level. [9]

[Question 8 is printed overleaf.]

[Turn over

8 The numbers of goals scored by my local football team in 80 matches are summarised in the following table.

Number of goals	0	1	2	3	4	5	$\geqslant 6$
Number of matches	11	15	33	16	2	3	0

(i) Show that the mean of the distribution is 1.9, and find the variance of the distribution. [3]

(ii) Without carrying out a test, explain whether the values of the mean and variance indicate that a Poisson distribution could be a suitable model for the number of goals scored in a match. [2]

The table below gives the expected frequencies, correct to 2 decimal places, for a χ^2 goodness of fit test of a Poisson distribution.

Number of goals	0	1	2	3	4	$\geqslant 5$
Expected frequency	11.97	22.73	21.60	13.68	6.50	3.52

(iii) Show how the value 13.68 for 3 goals is obtained. [2]

(iv) Stating a required assumption regarding the data, carry out the test at the 5% significance level. Does the outcome of the test confirm your answer to part **(ii)**? [8]

(v) Without further calculation, state two ways in which the test would be different if it were a goodness of fit test of the distribution Po(2), also at the 5% significance level. [2]

1 At a particular hospital, admissions of patients as a result of visits to the Accident and Emergency Department occur randomly at a uniform average rate of 0.75 per day. Independently, admissions that result from G.P. referrals occur randomly at a uniform average rate of 6.4 per *week*. The total number of admissions from these two causes over a randomly chosen period of four weeks is denoted by T. State the distribution of T and obtain its expectation and variance. **[4]**

2 The continuous random variable U has (cumulative) distribution function given by

$$F(u) = \begin{cases} \frac{1}{5}e^{u} & u < 0, \\ 1 - \frac{4}{5}e^{-\frac{1}{4}u} & u \geqslant 0. \end{cases}$$

(i) Find the upper quartile of U. **[3]**

(ii) Find the probability density function of U. **[2]**

3 In a random sample of credit card holders, it was found that 28% of them used their card for internet purchases.

(i) Given that the sample size is 1200, find a 98% confidence interval for the percentage of all credit card holders who use their card for internet purchases. **[4]**

(ii) Estimate the smallest sample size for which a 98% confidence interval would have a width of at most 5%, and state why the value found is only an estimate. **[4]**

4 The weekly sales of petrol, X thousand litres, at a garage may be modelled by a continuous random variable with probability density function given by

$$f(x) = \begin{cases} c & 25 \leqslant x \leqslant 45, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. The weekly profit, in £, is given by $(400\sqrt{X} - 240)$.

(i) Obtain the value of c. **[1]**

(ii) Find the expected weekly profit. **[3]**

(iii) Find the probability that the weekly profit exceeds £2000. **[3]**

5 The concentration level of mercury in a large lake is known to have a normal distribution with standard deviation 0.24 in suitable units. At the beginning of June 2008, the mercury level was measured at five randomly chosen places on the lake, and the sample mean is denoted by $\bar{x}_1$. Towards the end of June 2008 there was a spillage in the lake which may have caused the mercury level to rise. Because of this the level was then measured at six randomly chosen points of the lake, and the mean of this sample is denoted by $\bar{x}_2$.

(i) State hypotheses for a test based on the two samples for whether, on average, the level of mercury had increased. Define any parameters that you use. **[2]**

(ii) Find the set of values of $\bar{x}_2 - \bar{x}_1$ for which there would be evidence at the 5% significance level that, on average, the level of mercury had increased. **[4]**

(iii) Given that the average level had actually increased by 0.3 units, find the probability of making a Type II error in your test, and comment on its value. **[4]**

6 A mathematics examination is taken by 29 boys and 26 girls. Experience has shown that the probability that any boy forgets to bring a calculator to the examination is 0.3, and that any girl forgets is 0.2. Whether or not any student forgets to bring a calculator is independent of all other students. The numbers of boys and girls who forget to bring a calculator are denoted by B and G respectively, and $F = B + G$.

(i) Find $\mathrm{E}(F)$ and $\mathrm{Var}(F)$. **[5]**

(ii) Using suitable approximations to the distributions of B and G, which should be justified, find the smallest number of spare calculators that should be available in order to be at least 99% certain that all 55 students will have a calculator. **[8]**

7 A tutor gives a randomly selected group of 8 students an English Literature test, and after a term's further teaching, she gives the group a similar test. The marks for the two tests are given in the table.

Student	A	B	C	D	E	F	G	H
First test	38	27	55	43	32	24	51	46
Second test	37	26	57	43	30	26	54	48

(i) Stating a necessary condition, show by carrying out a suitable t-test, at the 1% significance level, that the marks do not give evidence of an improvement. **[8]**

(ii) The tutor later found that she had marked the second test too severely, and she decided to add a constant amount k to each mark. Find the least integer value of k for which the increased marks would give evidence of improvement at the 1% significance level. **[3]**

[Question 8 is printed overleaf.]

8 A soft drinks factory produces lemonade which is sold in packs of 6 bottles. As part of the factory's quality control, random samples of 75 packs are examined at regular intervals. The number of underfilled bottles in a pack of 6 bottles is denoted by the random variable X. The results of one quality control check are shown in the following table.

Number of underfilled bottles	0	1	2	3
Number of packs	44	20	8	3

A researcher assumes that $X \sim B(3, p)$.

(i) By finding the sample mean, show that an estimate of p is 0.2. **[3]**

(ii) Show that, at the 5% significance level, there is evidence that this binomial distribution does not fit the data. **[10]**

(iii) Another researcher suggests that the goodness of fit test should be for $B(6, p)$. She finds that the corresponding value of χ^2 is 2.74, correct to 3 significant figures. Given that the number of degrees of freedom is the same as in part **(ii)**, state the conclusion of the test at the same significance level. **[1]**

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

$$f(x) = \begin{cases} \frac{2}{5} & -a \leqslant x < 0, \\ \frac{2}{5}e^{-2x} & x \geqslant 0. \end{cases}$$

Find

(i) the value of the constant a, **[3]**

(ii) $E(X)$. **[5]**

2 The amount of tomato juice, X ml, dispensed into cartons of a particular brand has a normal distribution with mean 504 and standard deviation 3. The juice is sold in packs of 4 cartons, filled independently. The total amount of juice in one pack is Y ml.

(i) Find $P(Y < 2000)$. **[4]**

The random variable V is defined as $Y - 4X$.

(ii) Find $E(V)$ and $Var(V)$. **[3]**

(iii) What is the probability that the amount of juice in a randomly chosen pack is more than 4 times the amount of juice in a randomly chosen carton? **[1]**

3 It is given that X_1 and X_2 are independent random variables with $X_1 \sim N(\mu_1, 2.47)$ and $X_2 \sim N(\mu_2, 4.23)$. Random samples of n_1 observations of X_1 and n_2 observations of X_2 are taken. The sample means are denoted by $\overline{X}_1$ and $\overline{X}_2$.

(i) State the distribution of $\overline{X}_1 - \overline{X}_2$, giving its parameters. **[3]**

For two particular samples, $n_1 = 5$, $\Sigma x_1 = 48.25$, $n_2 = 10$ and $\Sigma x_2 = 72.30$.

(ii) Test at the 2% significance level whether μ_1 differs from μ_2. **[6]**

A student stated that because of the Central Limit Theorem the sample means will have normal distributions so it is unnecessary for X_1 and X_2 to have normal distributions.

(iii) Comment on the student's statement. **[1]**

4 The continuous random variable V has (cumulative) distribution function given by

$$F(v) = \begin{cases} 0 & v < 1, \\ 1 - \dfrac{8}{(1+v)^3} & v \geqslant 1. \end{cases}$$

The random variable Y is given by $Y = \dfrac{1}{1+V}$.

(i) Show that the (cumulative) distribution function of Y is $8y^3$, over an interval to be stated, and find the probability density function of Y. **[7]**

(ii) Find $E\left(\dfrac{1}{Y^2}\right)$. **[2]**

5 Each of a random sample of 200 steel bars taken from a production line was examined and 27 were found to be faulty.

(i) Find an approximate 90% confidence interval for the proportion of faulty bars produced. **[4]**

A change in the production method was introduced which, it was claimed, would reduce the proportion of faulty bars. After the change, each of a further random sample of 100 bars was examined and 8 were found to be faulty.

(ii) Test the claim, at the 10% significance level. **[7]**

6 The deterioration of a certain drug over time was investigated as follows. The drug strength was measured in each of a random sample of 8 bottles containing the drug. These were stored for two years and the strengths were then re-measured. The original and final strengths, in suitable units, are shown in the following table.

Bottle	1	2	3	4	5	6	7	8
Original strength	8.7	9.4	9.2	8.9	9.6	8.2	9.9	8.8
Final strength	8.1	9.0	9.0	8.8	9.3	8.0	9.5	8.5

(i) Stating any required assumption, test at the 5% significance level whether the mean strength has decreased by more than 0.2 over the two years. **[9]**

(ii) Calculate a 95% confidence interval for the mean reduction in strength over the two years. **[3]**

7 A chef wished to ascertain her customers' preference for certain vegetables. She asked a random sample of 120 customers for their preferred vegetable from asparagus, broad beans and cauliflower. The responses, classified according to the gender of the customer, are shown in the table.

	Asparagus	Broad beans	Cauliflower
Female preference	31	9	25
Male preference	17	21	17

(i) Test, at the 5% significance level, whether vegetable preference and gender are independent. **[8]**

(ii) Determine whether, at the 10% significance level, the vegetables are equally preferred. **[6]**