Answer **all** questions.

1 The crossing times for a ferry travelling from Port P on the English coast to Port Q on the French coast were recorded on 8 occasions.

The times, correct to the nearest minute, were

95 92 102 90 81 84 109 108

The crossing times may be assumed to be a random sample from a normal distribution with standard deviation σ.

(a) Calculate a 95% confidence interval for σ^2. *(6 marks)*

(b) Comment on the suggestion that $\sigma = 6$. *(2 marks)*

2 The numbers of girls in 240 families, each having 3 children, are recorded below.

Number of girls	0	1	2	3
Number of families	36	96	83	25

(a) Test, at the 5% level of significance, the hypothesis that these data may be modelled by a binomial distribution with parameter $p = \frac{1}{2}$. *(8 marks)*

(b) (i) Explain how you would estimate p if it was not known to be $\frac{1}{2}$. *(2 marks)*

(ii) State the number of degrees of freedom that there would have been in your test in part (a) if p had needed to be estimated from the data. *(1 mark)*

3 (a) The time, in years, that a taxi driver keeps his taxi, before replacing it with a new one, can be modelled by an exponential distribution with parameter 0.2 .

Find the probability that he keeps his taxi:

(i) for less than 2 years; *(2 marks)*

(ii) for more than 3 years. *(2 marks)*

(b) The continuous random variable X has an exponential distribution with probability density function $\mathrm{f}(x)$, where

$$\mathrm{f}(x) = \lambda \mathrm{e}^{-\lambda x} \quad \text{for } x \geqslant 0$$

Use integration to find:

(i) $\mathrm{E}(X)$; *(4 marks)*

(ii) the median value of X . *(3 marks)*

(c) Breakdowns occur on a stretch of road at a mean rate of 0.3 per day. The number of breakdowns follows a Poisson distribution.

Find, **in hours**:

(i) the mean time between breakdowns; *(1 mark)*

(ii) the median time between breakdowns. *(2 marks)*

4 The numbers of red blood cells, measured in millions per cubic millimetre of blood, for 10 women and 8 men were found to be as follows:

Women	5.05	3.98	4.73	5.36	4.92	5.44	4.04	4.40	4.15	5.33
Men	4.23	4.92	5.53	5.33	5.31	4.86	5.36	4.75		

Assume that these are independent random samples from normal populations.

(a) Show, at the 5% level of significance, that the hypothesis that the population variances are equal is accepted. *(8 marks)*

(b) Investigate, at the 5% level of significance, the hypothesis that the mean number of red blood cells is greater for men than for women. *(9 marks)*

Turn over for the next question

5 A random variable X has mean 2μ and variance 13, and an independent random variable Y has mean μ and variance 3. The random variable $aX + bY$ is an unbiased estimator of μ, where a and b are constants.

(a) Show that $2a + b = 1$. *(2 marks)*

(b) Show that $\text{Var}(aX + bY) = 3 - 12a + 25a^2$. *(3 marks)*

(c) Find values of a and b such that $aX + bY$ has minimum variance. *(3 marks)*

(d) A single observation is made on each of X and Y. The values observed are 15 and 10 respectively.

Obtain an unbiased estimate of μ which has minimum variance. *(2 marks)*

6 (a) Javinder is trying to start his old motorcycle which is known, on average, to start twice in every five attempts. It may be assumed that each attempt is independent of every other attempt, and that the probability of it starting on any attempt remains constant.

Calculate the probability that:

(i) his motorcycle will start on the third attempt; *(2 marks)*

(ii) it will take more than three attempts to start his motorcycle. *(2 marks)*

(b) The discrete random variable X has a geometric distribution with parameter p.

(i) Prove that $\text{E}(X) = \dfrac{1}{p}$. *(3 marks)*

(ii) Given that $\text{E}(X^2) = \dfrac{2-p}{p^2}$, show that $\text{Var}(X) = \dfrac{1-p}{p^2}$. *(2 marks)*

(c) Kylie's old car has a faulty starter motor. The number of attempts, Y, required to start her car may be assumed to follow a geometric distribution with parameter p, such that

$$\text{P}(Y = 1 \text{ or } 2) = 0.36$$

(i) Verify that the value of p is 0.2. *(1 mark)*

(ii) State the values of the mean, μ, and variance, σ^2, of Y. *(2 marks)*

(iii) Hence calculate $\text{P}\left(Y \leqslant \mu - \dfrac{\sigma}{\sqrt{5}}\right)$. *(3 marks)*

END OF QUESTIONS

Practice 2

Leave blank

1. A random sample $X_1, X_2, ..., X_{10}$ is taken from a population with mean μ and variance σ^2.

(a) Determine the bias, if any, of each of the following estimators of μ.

$$\theta_1 = \frac{X_3 + X_4 + X_5}{3},$$

$$\theta_2 = \frac{X_{10} - X_1}{3},$$

$$\theta_3 = \frac{3X_1 + 2X_2 + X_{10}}{6}.$$

(4)

(b) Find the variance of each of these estimators.

(5)

(c) State, giving reasons, which of these three estimators for μ is

(i) the best estimator,

(ii) the worst estimator.

(4)

Leave blank

Question 1 continued

Leave blank

Question 1 continued

Leave blank

Question 1 continued

(Total 13 marks)

Q1

Leave blank

2. A large number of students are split into two groups A and B. The students sit the same test but under different conditions. Group A has music playing in the room during the test, and group B has no music playing during the test. Small samples are then taken from each group and their marks recorded. The marks are normally distributed.

The marks are as follows:

Sample from Group A	42	40	35	37	34	43	42	44	49
Sample from Group B	40	44	38	47	38	37	33		

(a) Stating your hypotheses clearly, and using a 10% level of significance, test whether or not there is evidence of a difference between the variances of the marks of the two groups.
(8)

(b) State clearly an assumption you have made to enable you to carry out the test in part (a).
(1)

(c) Use a two tailed test, with a 5% level of significance, to determine if the playing of music during the test has made any difference in the mean marks of the two groups. State your hypotheses clearly.
(7)

(d) Write down what you can conclude about the effect of music on a student's performance during the test.
(1)

Leave blank

Question 2 continued

Turn over

Leave blank

Question 2 continued

Leave blank

Question 2 continued

(Total 17 marks)

Q2

Turn over

Leave blank

3. The weights, in grams, of mice are normally distributed. A biologist takes a random sample of 10 mice. She weighs each mouse and records its weight.

The ten mice are then fed on a special diet. They are weighed again after two weeks.

Their weights in grams are as follows:

Mouse	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Weight before diet	50.0	48.3	47.5	54.0	38.9	42.7	50.1	46.8	40.3	41.2
Weight after diet	52.1	47.6	50.1	52.3	42.2	44.3	51.8	48.0	41.9	43.6

Stating your hypotheses clearly, and using a 1% level of significance, test whether or not the diet causes an increase in the mean weight of the mice.

(8)

Leave blank

Question 3 continued

(Total 8 marks)

Q3

Leave blank

4. A town council is concerned that the mean price of renting two bedroom flats in the town has exceeded £650 per month. A random sample of eight two bedroom flats gave the following results, £x, per month.

705, 640, 560, 680, 800, 620, 580, 760

[You may assume $\sum x = 5345$ $\sum x^2 = 3621025$]

(a) Find a 90% confidence interval for the mean price of renting a two bedroom flat.
(6)

(b) State an assumption that is required for the validity of your interval in part (a).
(1)

(c) Comment on whether or not the town council is justified in being concerned. Give a reason for your answer.
(2)

Leave blank

Question 4 continued

(Total 9 marks)

Q4

Leave blank

5. A machine is filling bottles of milk. A random sample of 16 bottles was taken and the volume of milk in each bottle was measured and recorded. The volume of milk in a bottle is normally distributed and the unbiased estimate of the variance, s^2, of the volume of milk in a bottle is 0.003

(a) Find a 95% confidence interval for the variance of the population of volumes of milk from which the sample was taken.

(5)

The machine should fill bottles so that the standard deviation of the volumes is equal to 0.07

(b) Comment on this with reference to your 95% confidence interval.

(3)

Leave blank

Question 5 continued

Q5

(Total 8 marks)

Leave blank

6. A drug is claimed to produce a cure to a certain disease in 35% of people who have the disease. To test this claim a sample of 20 people having this disease is chosen at random and given the drug. If the number of people cured is between 4 and 10 inclusive the claim will be accepted. Otherwise the claim will not be accepted.

(a) Write down suitable hypotheses to carry out this test.
(2)

(b) Find the probability of making a Type I error.
(3)

The table below gives the value of the probability of the Type II error, to 4 decimal places, for different values of p where p is the probability of the drug curing a person with the disease.

P(cure)	0.2	0.3	0.4	0.5
P(Type II error)	0.5880	r	0.8565	s

(c) Calculate the value of r and the value of s.
(3)

(d) Calculate the power of the test for $p = 0.2$ and $p = 0.4$
(2)

(e) Comment, giving your reasons, on the suitability of this test procedure.
(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 12 marks)

Turn over

Leave blank

7. An engineering firm buys steel rods. The steel rods from its present supplier are known to have a mean tensile strength of 230 N/mm^2.

A new supplier of steel rods offers to supply rods at a cheaper price than the present supplier. A random sample of ten rods from this new supplier gave tensile strengths, x N/mm^2, which are summarised below.

Sample size	Σx	Σx^2
10	2283	524079

(a) Stating your hypotheses clearly, and using a 5% level of significance, test whether or not the rods from the new supplier have a tensile strength lower than the present supplier. (You may assume that the tensile strength is normally distributed).
(7)

(b) In the light of your conclusion to part (a) write down what you would recommend the engineering firm to do.
(1)

Leave blank

Question 7 continued

Turn over

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 8 marks)

TOTAL FOR PAPER: 75 MARKS

END

Answer **all** questions.

1 The headteacher of a school believes that the standard deviation of the annual number of new pupils joining the school is 10.

A statistician on the staff collects the following data on the number of new pupils joining the school during each of a sample of ten years.

124 123 139 136 128 125 128 133 131 133

Investigate, at the 5% level of significance, the headteacher's belief. Assume that these data may be regarded as a random sample from a normal distribution. *(8 marks)*

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

(a) Given that the value of the mean is 4 times that of the variance, find the value of p. *(3 marks)*

(b) Hence determine $\mathrm{P}(X > 4 \mid X > 2)$. *(4 marks)*

3 The assessment of a physics course has two components: a written examination and a practical test. Each component has a maximum mark of 75. The marks achieved by 10 students in each component are shown in the table.

Student	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Written Mark	35	47	54	55	43	48	41	59	47	31
Practical Mark	57	63	47	72	73	27	39	60	53	22

(a) Investigate, using a paired t-test and the 5% level of significance, whether the mean mark in the written examination is less than that in the practical test. *(10 marks)*

(b) State **two** assumptions that were necessary in order to carry out the test in part (a). *(2 marks)*

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

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geqslant 0$$

Prove that the mean value of X is $\frac{1}{\lambda}$. *(4 marks)*

(b) The lifetime of a component in a machine is T hours, where T has probability density function

$$f(t) = \frac{1}{a} e^{-\frac{t}{a}} \quad \text{for } t \geqslant 0$$

The mean lifetime of these components is known to be 62.5 hours.

(i) Find the value of $\frac{1}{a}$. *(2 marks)*

(ii) Calculate the probability that a component will last for at least 80 hours. *(4 marks)*

(iii) Given that a component has lasted for 80 hours, find the probability that it will last for a further 20 hours. *(3 marks)*

5 One hundred 1-millilitre samples of water were taken at random and the number of bacteria in each sample was counted. The results are shown in the table.

Number of bacteria	0	1	2	3	4	5	6	7
Frequency	7	15	27	25	11	10	3	2

(a) For these data, show that the mean number of bacteria per 1-millilitre of water is 2.7. *(2 marks)*

(b) Hence, using a χ^2 goodness of fit test with the 10% level of significance, investigate whether the number of bacteria per 1-millilitre of water can be modelled by a Poisson distribution. *(11 marks)*

Turn over for the next question

6 A random variable X is distributed with mean μ and variance σ^2. Three independent observations, X_1, X_2 and X_3, are taken on X.

The combined statistic

$$T = aX_1 + bX_2 + cX_3$$

where a, b and c are constants, is used as an estimator for μ.

(a) Show that, if T is an unbiased estimator for μ, then $a + b + c = 1$. *(3 marks)*

(b) Two unbiased estimators for μ are T_1 and T_2, defined by

$$T_1 = \tfrac{1}{3}X_1 + \tfrac{1}{2}X_2 + \tfrac{1}{6}X_3$$

$$T_2 = \tfrac{2}{3}X_1 + \tfrac{3}{4}X_2 - \tfrac{5}{12}X_3$$

(i) Calculate the relative efficiency of T_1 with respect to T_2. *(5 marks)*

(ii) With reference to your answer to part (b)(i), state, with a reason, which of T_1 and T_2 is the better unbiased estimator for μ. *(2 marks)*

7 A student at an agricultural college was asked to compare the variability of the weight, X grams, of eggs laid by free-range hens with the weight, Y grams, of eggs laid by battery hens.

The variables X and Y may be assumed to be normally distributed with variances ${\sigma_X}^2$ and ${\sigma_Y}^2$ respectively.

A random sample of 12 values of X resulted in $\sum(x - \bar{x})^2 = 761.2$, where $\bar{x}$ denotes the sample mean.

A random sample of 10 values of Y resulted in $\sum(y - \bar{y})^2 = 386.1$, where $\bar{y}$ denotes the sample mean.

(a) Calculate unbiased estimates of ${\sigma_X}^2$ and ${\sigma_Y}^2$. *(2 marks)*

(b) (i) Hence determine a 90% confidence interval for the ratio $\dfrac{{\sigma_X}^2}{{\sigma_Y}^2}$. *(8 marks)*

(ii) Comment on the suggestion that the weights of eggs laid by free-range hens are more variable than the weights of eggs laid by battery hens. *(2 marks)*

END OF QUESTIONS

Leave blank

1. A company manufactures bolts with a mean diameter of 5 mm. The company wishes to check that the diameter of the bolts has not decreased. A random sample of 10 bolts is taken and the diameters, x mm, of the bolts are measured. The results are summarised below.

$$\sum x = 49.1 \quad \sum x^2 = 241.2$$

Using a 1% level of significance, test whether or not the mean diameter of the bolts is less than 5 mm.

(You may assume that the diameter of the bolts follows a normal distribution.)

(8)

Leave blank

Question 1 continued

Q1

(Total 8 marks)

Leave blank

2. An emission-control device is tested to see if it reduces CO_2 emissions from cars. The emissions from 6 randomly selected cars are measured with and without the device. The results are as follows.

Car	*A*	*B*	*C*	*D*	*E*	*F*
Emissions without device	151.4	164.3	168.5	148.2	139.4	151.2
Emissions with device	148.9	162.7	166.9	150.1	140.0	146.7

(a) State an assumption that needs to be made in order to carry out a *t*-test in this case.
(1)

(b) State why a paired *t*-test is suitable for use with these data.
(1)

(c) Using a 5% level of significance, test whether or not there is evidence that the device reduces CO_2 emissions from cars.
(8)

(d) Explain, in context, what a type II error would be in this case.
(2)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Q2

(Total 12 marks)

Leave blank

3. Define, in terms of H_0 and/or H_1,

(a) the size of a hypothesis test,

(1)

(b) the power of a hypothesis test.

(1)

The probability of getting a head when a coin is tossed is denoted by p.

This coin is tossed 12 times in order to test the hypotheses H_0: $p = 0.5$ against H_1: $p \neq 0.5$, using a 5% level of significance.

(c) Find the largest critical region for this test, such that the probability in each tail is less than 2.5%.

(4)

(d) Given that $p = 0.4$

(i) find the probability of a type II error when using this test,

(ii) find the power of this test.

(4)

(e) Suggest two ways in which the power of the test can be increased.

(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 12 marks)

Q3

Leave blank

4. A farmer set up a trial to assess whether adding water to dry feed increases the milk yield of his cows. He randomly selected 22 cows. Thirteen of the cows were given dry feed and the other 9 cows were given the feed with water added. The milk yields, in litres per day, were recorded with the following results.

	Sample size	Mean	s^2
Dry feed	13	25.54	2.45
Feed with water added	9	27.94	1.02

You may assume that the milk yield from cows given the dry feed and the milk yield from cows given the feed with water added are from independent normal distributions.

(a) Test, at the 10% level of significance, whether or not the variances of the populations from which the samples are drawn are the same. State your hypotheses clearly.
(5)

(b) Calculate a 95% confidence interval for the difference between the two mean milk yields.
(7)

(c) Explain the importance of the test in part (a) to the calculation in part (b).
(2)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 14 marks)

Leave blank

5. A machine fills jars with jam. The weight of jam in each jar is normally distributed. To check the machine is working properly the contents of a random sample of 15 jars are weighed in grams. Unbiased estimates of the mean and variance are obtained as

$$\hat{\mu} = 560 \quad s^2 = 25.2$$

Calculate a 95% confidence interval for,

(a) the mean weight of jam,

(4)

(b) the variance of the weight of jam.

(5)

A weight of more than 565 g is regarded as too high and suggests the machine is not working properly.

(c) Use appropriate confidence limits from parts (a) and (b) to find the highest estimate of the proportion of jars that weigh too much.

(5)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Q5

(Total 14 marks)

Leave blank

6. A continuous uniform distribution on the interval [0, k] has mean $\frac{k}{2}$ and variance $\frac{k^2}{12}$.

A random sample of three independent variables X_1, X_2 and X_3 is taken from this distribution.

(a) Show that $\frac{2}{3}X_1 + \frac{1}{2}X_2 + \frac{5}{6}X_3$ is an unbiased estimator for k.

(3)

An unbiased estimator for k is given by $\hat{k} = aX_1 + bX_2$ where a and b are constants.

(b) Show that Var ($\hat{k}$) = $(a^2 - 2a + 2)\ \frac{k^2}{6}$

(6)

(c) Hence determine the value of a and the value of b for which $\hat{k}$ has minimum variance, and calculate this minimum variance.

(6)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 15 marks)

TOTAL FOR PAPER: 75 MARKS

END

Answer **all** questions.

1 The volume of fuel consumed by an aircraft making an east–west transatlantic flight was recorded on 10 occasions with the following results, correct to the nearest litre.

68 860	71 266	69 476	68 973	69 318
70 467	71 231	68 977	70 956	69 465

These volumes of fuel may be assumed to be a random sample from a normal distribution with standard deviation σ.

(a) Construct a 99% confidence interval for σ. *(6 marks)*

(b) State one factor that may cause the volume of fuel consumed to vary. *(1 mark)*

2 (a) The discrete random variable X follows a geometric distribution with parameter p.

Prove that $\mathrm{E}(X) = \dfrac{1}{p}$. *(3 marks)*

(b) A fair six-sided die is thrown repeatedly until a six occurs.

(i) State the expected number of throws required to obtain a six. *(1 mark)*

(ii) Calculate the probability that the number of throws required to obtain a six is greater than the expected value. *(3 marks)*

(iii) Find the least value of r such that, when the die is thrown repeatedly, there is more than a 90% chance of obtaining a six on or before the rth throw. *(4 marks)*

3 A geologist is studying the effect of exposure to weather on the radioactivity of granite. He collects, at random, 9 samples of freshly exposed granite and 8 samples of weathered granite. For each sample, he measures the radioactivity, in counts per minute. The results are shown in the table.

	Counts per minute								
Freshly exposed granite	226	189	166	212	179	172	200	203	181
Weathered granite	178	171	141	133	169	173	171	160	

(a) Assuming that these measurements come from two independent normal distributions with a common variance, construct a 95% confidence interval for the difference between the mean radioactivity of freshly exposed granite and that of weathered granite. *(9 marks)*

(b) Comment on a claim that the difference between the mean radioactivity of freshly exposed granite and that of weathered granite is 10 counts per minute. *(2 marks)*

4 The lifetimes of electrical components follow an exponential distribution with mean 200 hours.

(a) Calculate the probability that the lifetime of a randomly selected component is:

(i) less than 120 hours; *(2 marks)*

(ii) more than 160 hours; *(2 marks)*

(iii) less than 160 hours, given that it has lasted more than 120 hours. *(3 marks)*

(b) Determine the median lifetime of these electrical components. *(3 marks)*

5 It is thought that the marks in an examination may be modelled by a triangular distribution with probability density function

$$\mathrm{f}(x) = \begin{cases} \dfrac{1}{1875}x & 0 \leqslant x < 50 \\ \dfrac{6}{75} - \dfrac{2}{1875}x & 50 \leqslant x \leqslant 75 \\ 0 & \text{otherwise} \end{cases}$$

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

(b) A school enters 60 candidates for the examination. The results are summarised in the table.

Marks	0–	25–	50–75
Number of candidates	7	28	25

(i) Investigate, at the 5% level of significance, whether the triangular distribution in part (a) is an appropriate model for these data. *(9 marks)*

(ii) Describe, with a reason, how the test procedure in part (b)(i) would differ for a school entering 15 candidates, assuming that its results are summarised using the same mark ranges as in the table above. *(2 marks)*

Turn over for the next question

6 (a) The IQs of a random sample of 15 students have a standard deviation of 9.1 .

Test, at the 5% level of significance, whether this sample may be regarded as coming from a population with a variance of 225 . Assume that the population is normally distributed. *(6 marks)*

(b) The weights, in kilograms, of 6 boys and 4 girls were found to be as follows.

Boys	53	37	41	50	57	57
Girls	40	46	37	40		

Assume that these data are independent random samples from normal populations.

Show that, at the 5% level of significance, the hypothesis that the population variances are equal is accepted. *(7 marks)*

7 (a) The random variable X has a distribution with unknown mean μ and unknown variance σ^2.

A random sample of size n, denoted by $X_1, X_2, X_3, ..., X_n$, has mean $\overline{X}$ and variance V, where

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad V = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \overline{X}^2$$

(i) Show that

$$\mathrm{E}(X_i^2) = \sigma^2 + \mu^2 \quad \text{and} \quad \mathrm{E}(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$$ *(3 marks)*

(ii) Hence show that $\dfrac{nV}{n-1}$ is an unbiased estimator for σ^2. *(3 marks)*

(b) A random sample of size 2, denoted by X_1 and X_2, is taken from the distribution in part (a).

Show that $\frac{1}{2}(X_1 - X_2)^2$ is an unbiased estimator for σ^2. *(4 marks)*

END OF QUESTIONS