1. A telephone directory contains 50 000 names. A researcher wishes to select a systematic sample of 100 names from the directory.

(a) Explain in detail how the researcher should obtain such a sample.

(2)

(b) Give one advantage and one disadvantage of

(i) quota sampling,

(ii) systematic sampling.

(4)

Question 1 continued

(Total 6 marks)

Q1

2. The heights of a random sample of 10 imported orchids are measured. The mean height of the sample is found to be 20.1 cm. The heights of the orchids are normally distributed.

Given that the population standard deviation is 0.5 cm,

(a) estimate limits between which 95% of the heights of the orchids lie,

(3)

(b) find a 98% confidence interval for the mean height of the orchids.

(4)

A grower claims that the mean height of this type of orchid is 19.5 cm.

(c) Comment on the grower's claim. Give a reason for your answer.

(2)

Question 2 continued

Q2

(Total 9 marks)

3. A doctor is interested in the relationship between a person's Body Mass Index (BMI) and their level of fitness. She believes that a lower BMI leads to a greater level of fitness. She randomly selects 10 female 18 year-olds and calculates each individual's BMI. The females then run a race and the doctor records their finishing positions. The results are shown in the table.

Individual	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
BMI	17.4	21.4	18.9	24.4	19.4	20.1	22.6	18.4	25.8	28.1
Finishing position	3	5	1	9	6	4	10	2	7	8

(a) Calculate Spearman's rank correlation coefficient for these data.
(5)

(b) Stating your hypotheses clearly and using a one tailed test with a 5% level of significance, interpret your rank correlation coefficient.
(5)

(c) Give a reason to support the use of the rank correlation coefficient rather than the product moment correlation coefficient with these data.
(1)

Question 3 continued

Q3

(Total 11 marks)

4. A sample of size 8 is to be taken from a population that is normally distributed with mean 55 and standard deviation 3. Find the probability that the sample mean will be greater than 57.

(5)

Question 4 continued

Q4

(Total 5 marks)

5. The number of goals scored by a football team is recorded for 100 games. The results are summarised in Table 1 below.

Number of goals	**Frequency**
0	40
1	33
2	14
3	8
4	5

Table 1

(a) Calculate the mean number of goals scored per game.
(2)

The manager claimed that the number of goals scored per match follows a Poisson distribution. He used the answer in part (a) to calculate the expected frequencies given in Table 2.

Number of goals	**Expected Frequency**
0	34.994
1	r
2	s
3	6.752
$\geqslant 4$	2.221

Table 2

(b) Find the value of r and the value of s giving your answers to 3 decimal places.
(3)

(c) Stating your hypotheses clearly, use a 5% level of significance to test the manager's claim.
(7)

Question 5 continued

Question 5 continued

Question 5 continued

Q5

(Total 12 marks)

6. The lengths of a random sample of 120 limpets taken from the upper shore of a beach had a mean of 4.97 cm and a standard deviation of 0.42 cm. The lengths of a second random sample of 150 limpets taken from the lower shore of the same beach had a mean of 5.05 cm and a standard deviation of 0.67 cm.

(a) Test, using a 5% level of significance, whether or not the mean length of limpets from the upper shore is less than the mean length of limpets from the lower shore. State your hypotheses clearly.

(8)

(b) State two assumptions you made in carrying out the test in part (a).

(2)

Question 6 continued

Q6

(Total 10 marks)

7. A company produces climbing ropes. The lengths of the climbing ropes are normally distributed. A random sample of 5 ropes is taken and the length, in metres, of each rope is measured. The results are given below.

120.3 120.1 120.4 120.2 119.9

(a) Calculate unbiased estimates for the mean and the variance of the lengths of the climbing ropes produced by the company.

(5)

The lengths of climbing rope are known to have a standard deviation of 0.2 m. The company wants to make sure that there is a probability of at least 0.90 that the estimate of the population mean, based on a random sample size of n, lies within 0.05 m of its true value.

(b) Find the minimum sample size required.

(6)

Question 7 continued

Q7

(Total 11 marks)

8. The random variable A is defined as

$$A = 4X - 3Y$$

where $X \sim N(30, 3^2)$, $Y \sim N(20, 2^2)$ and X and Y are independent.

Find

(a) E(A),

(2)

(b) Var(A).

(3)

The random variables Y_1, Y_2, Y_3 and Y_4 are independent and each has the same distribution as Y. The random variable B is defined as

$$B = \sum_{i=1}^{4} Y_i$$

(c) Find P($B > A$).

(6)

Question 8 continued

Question 8 continued

Q8

(Total 11 marks)

TOTAL FOR PAPER: 75 MARKS

END

mock papers 7

1. A report states that employees spend, on average, 80 minutes every working day on personal use of the Internet. A company takes a random sample of 100 employees and finds their mean personal Internet use is 83 minutes with a standard deviation of 15 minutes. The company's managing director claims that his employees spend more time on average on personal use of the Internet than the report states.

Test, at the 5% level of significance, the managing director's claim. State your hypotheses clearly.

(7)

Question 1 continued

(Total 7 marks)

Q1

Turn over

2. Philip and James are racing car drivers. Philip's lap times, in seconds, are normally distributed with mean 90 and variance 9. James' lap times, in seconds, are normally distributed with mean 91 and variance 12. The lap times of Philip and James are independent. Before a race, they each take a qualifying lap.

(a) Find the probability that James' time for the qualifying lap is less than Philip's.
(4)

The race is made up of 60 laps. Assuming that they both start from the same starting line and lap times are independent,

(b) find the probability that Philip beats James in the race by more than 2 minutes.
(5)

Question 2 continued

(Total 9 marks)

Q2

Turn over

3. A woodwork teacher measures the width, w mm, of a board. The measured width, X mm, is normally distributed with mean w mm and standard deviation 0.5 mm.

(a) Find the probability that X is within 0.6 mm of w.
(2)

The same board is measured 16 times and the results are recorded.

(b) Find the probability that the mean of these results is within 0.3 mm of w.
(4)

Given that the mean of these 16 measurements is 35.6 mm,

(c) find a 98% confidence interval for w.
(4)

Question 3 continued

Turn over

Question 3 continued

Question 3 continued

Q3

(Total 10 marks)

Turn over

4. A researcher claims that, at a river bend, the water gradually gets deeper as the distance from the inner bank increases. He measures the distance from the inner bank, b cm, and the depth of a river, s cm, at seven positions. The results are shown in the table below.

Position	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Distance from inner bank b cm	100	200	300	400	500	600	700
Depth s cm	60	75	85	76	110	120	104

(a) Calculate Spearman's rank correlation coefficient between b and s.

(6)

(b) Stating your hypotheses clearly, test whether or not the data provides support for the researcher's claim. Use a 1% level of significance.

(4)

Question 4 continued

Q4

(Total 10 marks)

Turn over

5. A random sample of 100 people were asked if their finances were worse, the same or better than this time last year. The sample was split according to their annual income and the results are shown in the table below.

Annual income \ Finances	Worse	Same	Better
Under £15 000	14	11	9
£15 000 and above	17	20	29

Test, at the 5% level of significance, whether or not the relative state of their finances is independent of their income range. State your hypotheses and show your working clearly.

(10)

Question 5 continued

Turn over

Question 5 continued

Question 5 continued

(Total 10 marks)

Q5

Turn over

6. A total of 228 items are collected from an archaeological site. The distance from the centre of the site is recorded for each item. The results are summarised in the table below.

Distance from the centre of the site (m)	0–1	1–2	2–4	4–6	6–9	9–12
Number of items	22	15	44	37	52	58

Test, at the 5% level of significance, whether or not the data can be modelled by a continuous uniform distribution. State your hypotheses clearly.

(12)

Question 6 continued

Turn over

Question 6 continued

Question 6 continued

(Total 12 marks)

Q6

7. A large company surveyed its staff to investigate the awareness of company policy. The company employs 6000 full time staff and 4000 part time staff.

(a) Describe how a stratified sample of 200 staff could be taken.
(3)

(b) Explain an advantage of using a stratified sample rather than a simple random sample.
(1)

A random sample of 80 full time staff and an independent random sample of 80 part time staff were given a test of policy awareness. The results are summarised in the table below.

	Mean score ($\bar{x}$)	Variance of scores (s^2)
Full time staff	52	21
Part time staff	50	19

(c) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean policy awareness scores for full time and part time staff are different.
(7)

(d) Explain the significance of the Central Limit Theorem to the test in part (c).
(2)

(e) State an assumption you have made in carrying out the test in part (c).
(1)

After all the staff had completed a training course the 80 full time staff and the 80 part time staff were given another test of policy awareness. The value of the test statistic z was 2.53

(f) Comment on the awareness of company policy for the full time and part time staff in light of this result. Use a 1% level of significance.
(2)

(g) Interpret your answers to part (c) and part (f).
(1)

Question 7 continued

Turn over

Question 7 continued

Question 7 continued

Turn over

Question 7 continued

(Total 17 marks)

Q7

TOTAL FOR PAPER: 75 MARKS

END

1. Explain what you understand by the Central Limit Theorem.

(3)

Q1

(Total 3 marks)

2. A county councillor is investigating the level of hardship, h, of a town and the number of calls per 100 people to the emergency services, c. He collects data for 7 randomly selected towns in the county. The results are shown in the table below.

Town	*A*	*B*	*C*	*D*	*E*	*F*	*G*
h	14	20	16	18	37	19	24
c	52	45	43	42	61	82	55

(a) Calculate the Spearman's rank correlation coefficient between h and c.

(6)

After collecting the data, the councillor thinks there is no correlation between hardship and the number of calls to the emergency services.

(b) Test, at the 5% level of significance, the councillor's claim. State your hypotheses clearly.

(4)

Question 2 continued

Q2

(Total 10 marks)

3. A factory manufactures batches of an electronic component. Each component is manufactured in one of three shifts. A component may have one of two types of defect, D_1 or D_2, at the end of the manufacturing process. A production manager believes that the type of defect is dependent upon the shift that manufactured the component. He examines 200 randomly selected defective components and classifies them by defect type and shift. The results are shown in the table below.

Shift \ Defect type	D_1	D_2
First shift	45	18
Second shift	55	20
Third shift	50	12

Stating your hypotheses, test, at the 10% level of significance, whether or not there is evidence to support the manager's belief. Show your working clearly.

(10)

Question 3 continued

Question 3 continued

Question 3 continued

Q3

(Total 10 marks)

4. A shop manager wants to find out if customers spend more money when music is playing in the shop. The amount of money spent by a customer in the shop is £x. A random sample of 80 customers, who were shopping without music playing, and an independent random sample of 60 customers, who were shopping with music playing, were surveyed. The results of both samples are summarised in the table below.

	$\sum x$	$\sum x^2$	Unbiased estimate of mean	Unbiased estimate of variance
Customers shopping **without** music	5 320	392 000	$\bar{x}$	s^2
Customers shopping **with** music	4 140	312 000	69.0	446.44

(a) Find the values of $\bar{x}$ and s^2.

(5)

(b) Test, at the 5% level of significance, whether or not the mean money spent is greater when music is playing in the shop. State your hypotheses clearly.

(8)

Question 4 continued

Question 4 continued

Question 4 continued

Q4

(Total 13 marks)

5. The number of hurricanes per year in a particular region was recorded over 80 years. The results are summarised in Table 1 below.

No of hurricanes, *h*	0	1	2	3	4	5	6	7
Frequency	0	2	5	17	20	12	12	12

Table 1

(a) Write down two assumptions that will support modelling the number of hurricanes per year by a Poisson distribution.
(2)

(b) Show that the mean number of hurricanes per year from Table 1 is 4.4875
(2)

(c) Use the answer in part (b) to calculate the expected frequencies r and s given in Table 2 below to 2 decimal places.
(3)

h	0	1	2	3	4	5	6	7 or more
Expected frequency	0.90	4.04	r	13.55	s	13.65	10.21	13.39

Table 2

(d) Test, at the 5% level of significance, whether or not the data can be modelled by a Poisson distribution. State your hypotheses clearly.
(6)

Question 5 continued

Question 5 continued

Question 5 continued

Q5

(Total 13 marks)

6. The lifetimes of batteries from manufacturer A are normally distributed with mean 20 hours and standard deviation 5 hours when used in a camera.

(a) Find the mean and standard deviation of the total lifetime of a pack of 6 batteries from manufacturer A.

(2)

Judy uses a camera that takes one battery at a time. She takes a pack of 6 batteries from manufacturer A to use in her camera on holiday.

(b) Find the probability that the batteries will last for more than 110 hours on her holiday.

(2)

The lifetimes of batteries from manufacturer B are normally distributed with mean 35 hours and standard deviation 8 hours when used in a camera.

(c) Find the probability that the total lifetime of a pack of 6 batteries from manufacturer A is more than 4 times the lifetime of a single battery from manufacturer B when used in a camera.

(6)

Question 6 continued

Q6

(Total 10 marks)

7. Roastie's Coffee is sold in packets with a stated weight of 250 g. A supermarket manager claims that the mean weight of the packets is less than the stated weight. She weighs a random sample of 90 packets from their stock and finds that their weights have a mean of 248 g and a standard deviation of 5.4 g.

(a) Using a 5% level of significance, test whether or not the manager's claim is justified. State your hypotheses clearly.

(5)

(b) Find the 98% confidence interval for the mean weight of a packet of coffee in the supermarket's stock.

(4)

(c) State, with a reason, the action you would recommend the manager to take over the weight of a packet of Roastie's Coffee.

(2)

Roastie's Coffee company increase the mean weight of their packets to μ g and reduce the standard deviation to 3 g. The manager takes a sample of size n from these new packets. She uses the sample mean $\overline{X}$ as an estimator of μ.

(d) Find the minimum value of n such that $\mathrm{P}(|\overline{X} - \mu| < 1) \geqslant 0.98$

(5)

Question 7 continued

Question 7 continued

Question 7 continued

Question 7 continued

Q7

(Total 16 marks)

TOTAL FOR PAPER: 75 MARKS

END

1 X_1 and X_2 are independent random variables with distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively. Assuming that the moment generating function of a normal variable with mean μ and variance σ^2 is $e^{\mu t+\frac{1}{2}\sigma^2 t^2}$, find the moment generating function of $X_1 + X_2$. Hence identify the distribution of $X_1 + X_2$, stating the value(s) of any parameter(s). **[5]**

2 A company wishes to buy a new lathe for making chair legs. Two models of lathe, 'Allegro' and 'Vivace', were trialled. The company asked 12 randomly selected employees to make a particular type of chair leg on each machine. The times, in seconds, for each employee are shown in the table.

Employee	1	2	3	4	5	6	7	8	9	10	11	12
Time on Allegro	162	111	194	159	202	210	183	168	165	150	185	160
Time on Vivace	182	130	193	181	192	205	186	184	192	180	178	189

The company wishes to test whether there is any difference in average times for the two machines.

(i) State the circumstances under which a non-parametric test should be used. **[1]**

(ii) Use two different non-parametric tests and show that they lead to different conclusions at the 5% significance level. **[9]**

(iii) State, with a reason, which conclusion is to be preferred. **[1]**

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

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

(i) Show that the moment generating function of X is $\frac{4}{4-t^2}$, where $|t| < 2$, and explain why the condition $|t| < 2$ is necessary. **[5]**

(ii) Find $\text{Var}(X)$. **[4]**

4 The probability generating function of the discrete random variable Y is given by

$$G_Y(t) = \frac{a + bt^3}{t},$$

where a and b are constants.

(i) Given that $E(Y) = -0.7$, find the values of a and b. **[4]**

(ii) Find $\text{Var}(Y)$. **[2]**

(iii) Find the probability that the sum of 10 random observations of Y is -7. **[4]**

5 Alana and Ben work for an estate agent. The joint probability distribution of the number of houses they sell in a randomly chosen week, X_A and X_B respectively, is shown in the table.

		X_A			
		0	1	2	3
X_B	0	0.02	0.13	0.07	0.03
	1	0.16	0.22	0.03	0.04
	2	0.09	0.06	0.03	0.02
	3	0.03	0.04	0.02	0.01

(i) Find $\mathrm{E}(X_A)$ and $\mathrm{Var}(X_A)$. **[3]**

(ii) Determine whether X_A and X_B are independent. **[2]**

(iii) Given that $\mathrm{E}(X_B) = 1.15$, $\mathrm{Var}(X_B) = 0.8275$ and $\mathrm{E}(X_A X_B) = 1.09$, find $\mathrm{Cov}(X_A, X_B)$ and $\mathrm{Var}(X_A - X_B)$. **[4]**

(iv) During a particular week only one house was sold by Alana and Ben. Find the probability that it was sold by Alana. **[4]**

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

$$\mathrm{f}(x) = \begin{cases} 0 & x < a, \\ \mathrm{e}^{-(x-a)} & x \geqslant a, \end{cases}$$

where a is a constant. $X_1, X_2, \ldots, X_n$ are n independent observations of X, where $n \geqslant 4$.

(i) Show that $\mathrm{E}(X) = a + 1$. **[3]**

T_1 and T_2 are proposed estimators of a, where

$$T_1 = X_1 + 2X_2 - X_3 - X_4 - 1 \quad \text{and} \quad T_2 = \frac{X_1 + X_2}{4} + \frac{X_3 + X_4 + \ldots + X_n}{2(n-2)} - 1.$$

(ii) Show that T_1 and T_2 are unbiased estimators of a. **[4]**

(iii) Determine which is the more efficient estimator. **[4]**

(iv) Suggest another unbiased estimator of a using all of the n observations. **[2]**

[Question 7 is printed overleaf.]

Turn over

7 A particular disease occurs in a proportion p of the population of a town. A diagnostic test has been developed, in which a positive result indicates the presence of the disease. It has a probability 0.98 of giving a true positive result, i.e. of indicating the presence of the disease when it is actually present. The test will give a false positive result with probability 0.08 when the disease is not present. A randomly chosen person is given the test.

(i) Find, in terms of p, the probability that

(a) the person has the disease when the result is positive, **[3]**

(b) the test will lead to a wrong conclusion. **[2]**

It is decided that if the result of the test on someone is positive, that person is tested again. The result of the second test is independent of the result of the first test.

(ii) Find the probability that the person has the disease when the result of the second test is positive. **[2]**

(iii) The town has 24 000 children and plans to test all of them at a cost of £5 per test. Assuming that $p = 0.001$, calculate the expected total cost of carrying out these tests. **[4]**

1 For the mutually exclusive events A and B, $\mathrm{P}(A) = \mathrm{P}(B) = x$, where $x \neq 0$.

(i) Show that $x \leqslant \frac{1}{2}$. [1]

(ii) Show that A and B are not independent. [2]

The event C is independent of A and also independent of B, and $\mathrm{P}(C) = 2x$.

(iii) Show that $\mathrm{P}(A \cup B \cup C) = 4x(1 - x)$. [4]

2 Part of Helen's psychology dissertation involved the reaction times to a certain stimulus. She measured the reaction times of 30 randomly selected students, in seconds correct to 2 decimal places. The results are shown in the following stem-and-leaf diagram.

14	1 2
15	2 4
16	0 3 6
17	1 5 7
18	3 4 5 7 9
19	2 4 6 7 8 9
20	0 1 3 4 5 7 8 9
21	7

Key: 18 | 3 means 1.83 seconds

Helen wishes to test whether the population median time exceeds 1.80 seconds.

(i) Give a reason why the Wilcoxon signed-rank test should not be used. [1]

(ii) Carry out a suitable non-parametric test at the 5% significance level. [7]

3 From the records of Mulcaster United Football Club the following distribution was suggested as a probability model for future matches. X and Y denoted the numbers of goals scored by the home team and the away team respectively.

		X			
		0	1	2	3
Y	0	0.11	0.04	0.06	0.08
	1	0.08	0.05	0.12	0.05
	2	0.05	0.08	0.07	0.03
	3	0.03	0.06	0.07	0.02

Use the model to find

(i) $E(X)$, [3]

(ii) the probability that the away team wins a randomly chosen match, [2]

(iii) the probability that the away team wins a randomly chosen match, given that the home team scores. [4]

One of the directors, an amateur statistician, finds that $\text{Cov}(X, Y) = 0.007$. He states that, as this value is very close to zero, X and Y may be considered to be independent.

(iv) Comment on the director's statement. [2]

4 William takes a bus regularly on the same journey, sometimes in the morning and sometimes in the afternoon. He wishes to compare morning and afternoon journey times. He records the journey times on 7 randomly chosen mornings and 8 randomly chosen afternoons. The results, each correct to the nearest minute, are as follows, where M denotes a morning time and A denotes an afternoon time.

M	A	A	M	M	M	M	M	M	A	A	A	A	A	A
19	20	22	24	25	26	28	30	31	33	35	37	38	39	42

William wishes to test for a difference between the average times of morning and afternoon journeys.

(i) Given that $s_M^2 = 16.5$ and $s_A^2 = 64.5$, with the usual notation, explain why a t-test is not appropriate in this case. [1]

(ii) William chooses a non-parametric test at the 5% significance level. Carry out the test, stating the rejection region. [6]

5 The discrete random variable X has moment generating function $\frac{1}{4}e^{2t} + ae^{3t} + be^{4t}$, where a and b are constants. It is given that $E(X) = 3\frac{3}{8}$.

(i) Show that $a = \frac{1}{8}$, and find the value of b. [6]

(ii) Find $\text{Var}(X)$. [4]

(iii) State the possible values of X. [1]

6 The continuous random variable Y has cumulative distribution function given by

$$\mathrm{F}(y) = \begin{cases} 0 & y < a, \\ 1 - \dfrac{a^3}{y^3} & y \geqslant a, \end{cases}$$

where a is a positive constant. A random sample of 3 observations, Y_1, Y_2, Y_3, is taken, and the smallest is denoted by S.

(i) Show that $\mathrm{P}(S > s) = \left(\frac{a}{s}\right)^9$ and hence obtain the probability density function of S. [5]

(ii) Show that S is not an unbiased estimator of a, and construct an unbiased estimator, T_1, based on S. [4]

It is given that T_2, where $T_2 = \frac{2}{9}(Y_1 + Y_2 + Y_3)$, is another unbiased estimator of a.

(iii) Given that $\mathrm{Var}(Y) = \frac{3}{4}a^2$ and $\mathrm{Var}(S) = \frac{9}{448}a^2$, determine which of T_1 and T_2 is the more efficient estimator. [3]

(iv) The values of Y for a particular sample are 12.8, 4.5 and 7.0. Find the values of T_1 and T_2 for this sample, and give a reason, unrelated to efficiency, why T_1 gives a better estimate of a than T_2 in this case. [3]

7 The probability generating function of the random variable X is given by

$$\mathrm{G}(t) = \frac{1 + at}{4 - t},$$

where a is a constant.

(i) Find the value of a. [2]

(ii) Find $\mathrm{P}(X = 3)$. [4]

The sum of 3 independent observations of X is denoted by Y. The probability generating function of Y is denoted by $\mathrm{H}(t)$.

(iii) Use $\mathrm{H}(t)$ to find $\mathrm{E}(Y)$. [5]

(iv) By considering $\mathrm{H}(-1) + \mathrm{H}(1)$, show that $\mathrm{P}(Y \text{ is an even number}) = \frac{62}{125}$. [2]

1 The random variable X has the distribution $B(n, p)$.

(i) Show, from the definition, that the probability generating function of X is $(q + pt)^n$, where $q = 1 - p$. **[2]**

(ii) The independent random variable Y has the distribution $B(2n, p)$ and $T = X + Y$. Use probability generating functions to determine the distribution of T, giving its parameters. **[4]**

2 A botanist believes that some species of plants produce more flowers at high altitudes than at low altitudes. In order to investigate this belief the botanist randomly samples 11 species of plants each of which occurs at both altitudes. The numbers of flowers on the plants are shown in the table.

Species	1	2	3	4	5	6	7	8	9	10	11
Number of flowers at low altitude	5	3	4	7	2	9	6	5	4	11	2
Number of flowers at high altitude	1	6	10	8	14	16	20	21	15	2	12

(i) Use the Wilcoxon signed rank test at the 5% significance level to test the botanist's belief. **[7]**

(ii) Explain why the Wilcoxon rank sum test should not be used for this test. **[1]**

3 For the events A and B, $P(A) = P(B) = \frac{3}{4}$ and $P(A \mid B') = \frac{1}{2}$.

(i) Find $P(A \cap B)$. **[4]**

For a third event C, $P(C) = \frac{1}{4}$ and C is independent of the event $A \cap B$.

(ii) Find $P(A \cap B \cap C)$. **[1]**

(iii) Given that $P(C \mid A) = \lambda$ and $P(B \mid C) = 3\lambda$, and that no event occurs outside $A \cup B \cup C$, find the value of λ. **[5]**

4 The discrete random variable X has moment generating function $\left(\frac{1}{4} + \frac{3}{4}e^t\right)^3$.

(i) Find $E(X)$. **[3]**

(ii) Find $P(X = 2)$. **[3]**

(iii) Show that X can be expressed as a sum of 3 independent observations of a random variable Y. Obtain the probability distribution of Y, and the variance of Y. **[4]**

5 A test was carried out to compare the breaking strengths of two brands of elastic band, A and B, of the same size. Random samples of 6 were selected from each brand and the breaking strengths were measured. The results, in suitable units and arranged in ascending order for each brand, are as follows.

Brand A:	5.6	8.7	9.2	10.7	11.2	12.6
Brand B:	10.1	11.6	12.0	12.2	12.9	13.5

(i) Give one advantage that a non-parametric test might have over a parametric test in this context. **[1]**

(ii) Carry out a suitable Wilcoxon test at the 5% significance level of whether there is a difference between the average breaking strengths of the two brands. **[7]**

(iii) An extra elastic band of brand B was tested and found to have a breaking strength exceeding all of the other 12 bands. Determine whether this information alters the conclusion of your test. **[3]**

6 A City Council comprises 16 Labour members, 14 Conservative members and 6 members of Other parties. A sample of two members was chosen at random to represent the Council at an event. The number of Labour members and the number of Conservative members in this sample are denoted by L and C respectively. The joint probability distribution of L and C is given in the following table.

		L = 0	L = 1	L = 2
C	0	$\frac{1}{42}$	$\frac{16}{105}$	$\frac{4}{21}$
	1	$\frac{2}{15}$	$\frac{16}{45}$	0
	2	$\frac{13}{90}$	0	0

(i) Verify the two non-zero probabilities in the table for which $C = 1$. **[4]**

(ii) Find the expected number of Conservatives in the sample. **[3]**

(iii) Find the expected number of Other members in the sample. **[3]**

(iv) Explain why L and C are not independent, and state what can be deduced about $\text{Cov}(L, C)$. **[3]**

[Question 7 is printed overleaf.]

7 The continuous random variable U has unknown mean μ and known variance σ^2. In order to estimate μ, two random samples, one of 4 observations of U and the other of 6 observations of U, are taken. The sample means are denoted by $\overline{U}_4$ and $\overline{U}_6$ respectively. One estimator S, given by $S = \frac{1}{2}(\overline{U}_4 + \overline{U}_6)$, is proposed.

(i) Show that S is unbiased and find $\text{Var}(S)$ in terms of σ^2. **[4]**

A second estimator T of the form $a\overline{U}_4 + b\overline{U}_6$ is proposed, where a and b are chosen such that T is an unbiased estimator for μ with the smallest possible variance.

(ii) Find the values of a and b and the corresponding variance of T. **[7]**

(iii) State, giving a reason, which of S and T is the better estimator. **[1]**

(iv) Compare the efficiencies of this preferred estimator and the mean of all 10 observations. **[2]**

1 For the events A and B, $P(A) = 0.3$, $P(B) = 0.6$ and $P(A' \cap B') = c$, where $c \neq 0$.

(i) Find $P(A \cap B)$ in terms of c. [3]

(ii) Find $P(B \mid A)$ and deduce that $0.1 \leqslant c \leqslant 0.4$. [3]

2 Of 9 randomly chosen students attending a lecture, 4 were found to be smokers and 5 were non-smokers. During the lecture their pulse-rates were measured, with the following results in beats per minute.

Smokers	77	85	90	98	
Non-smokers	59	64	68	80	88

It may be assumed that these two groups of students were random samples from the student populations of smokers and non-smokers. Using a suitable Wilcoxon test at the 10% significance level, test whether there is a difference in the median pulse-rates of the two populations. [7]

3 The discrete random variables X and Y have the joint probability distribution given in the following table.

		X		
		-1	0	1
Y	1	0.24	0.22	0.04
	2	0.26	0.18	0.06

(i) Show that $\text{Cov}(X, Y) = 0$. [5]

(ii) Find the conditional distribution of X given that $Y = 2$. [2]

4 The levels of impurity in a particular alloy were measured using a random sample of 20 specimens. The results, in suitable units, were as follows.

3.00	2.05	3.15	2.65	3.50	3.25	2.85	3.35	2.65	2.75
2.90	2.20	2.95	3.05	3.65	3.45	2.55	2.15	2.80	2.60

(i) Use the sign test, at the 5% significance level, to decide if there is evidence that the population median level of impurity is greater than 2.70. [7]

(ii) State what other test might have been used, and give one advantage and one disadvantage this other test has over the sign test. [3]

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

$$f(x) = \begin{cases} \frac{1}{(\alpha - 1)!} x^{\alpha-1} e^{-x} & x \geqslant 0, \\ 0 & x < 0, \end{cases}$$

where α is a positive integer.

(i) Explain how you can deduce that $\int_0^\infty x^{\alpha-1} e^{-x}\,dx = (\alpha - 1)!$. [1]

(ii) Write down an integral for the moment generating function $M_X(t)$ of X and show, by using the substitution $x = \frac{u}{1-t}$, that $M_X(t) = (1-t)^{-\alpha}$. [5]

(iii) Use the moment generating function to find, in terms of α,

(a) $E(X)$, [3]

(b) $Var(X)$. [3]

6 The discrete random variable X takes the values 0 and 1 with $P(X = 0) = q$ and $P(X = 1) = p$, where $p + q = 1$.

(i) Write down the probability generating function of X. [1]

The sum of n independent observations of X is denoted by S.

(ii) Write down the probability generating function of S, and name the distribution of S. [2]

(iii) Use the probability generating function of S to find $E(S)$ and $Var(S)$. [6]

(iv) The independent random variables Y and Z are such that Y has the distribution $B(10, \frac{1}{2})$, and Z has probability generating function $e^{-(1-t)}$. Find the probability that the sum of one random observation of Y and one random observation of Z is equal to 2. [6]

[Question 7 is printed overleaf.]

[Turn over

7 The continuous random variable X has a uniform distribution over the interval $[0, \theta]$ so that the probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\theta} & 0 \leqslant x \leqslant \theta, \\ 0 & \text{otherwise,} \end{cases}$$

where θ is a positive constant. A sample of n independent observations of X is taken and the sample mean is denoted by $\overline{X}$.

(i) The estimator T_1 is defined by $T_1 = 2\overline{X}$. Show that T_1 is an unbiased estimator of θ. [2]

It is given that the probability density function of the largest value, U, in the sample is

$$g(u) = \begin{cases} \dfrac{nu^{n-1}}{\theta^n} & 0 \leqslant u \leqslant \theta, \\ 0 & \text{otherwise.} \end{cases}$$

(ii) Find $E(U)$ and show that $\text{Var}(U) = \dfrac{n\theta^2}{(n+1)^2(n+2)}$. [6]

(iii) The estimator T_2 is defined by $T_2 = \dfrac{n+1}{n}U$. Given that T_2 is also an unbiased estimator of θ, show that T_2 is a more efficient estimator than T_1 for $n > 1$. [7]

mock papers 13

1. A medical student is investigating two methods of taking a person's blood pressure. He takes a random sample of 10 people and measures their blood pressure using an arm cuff and a finger monitor. The table below shows the blood pressure for each person, measured by each method.

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Arm cuff	140	110	138	127	142	112	122	128	132	160
Finger monitor	154	112	156	152	142	104	126	132	144	180

(a) Use a paired t-test to determine, at the 10% level of significance, whether or not there is a difference in the mean blood pressure measured using the two methods. State your hypotheses clearly.

(8)

(b) State an assumption about the underlying distribution of measured blood pressure required for this test.

(1)

Question 1 continued

Q1

(Total 9 marks)

2. The value of orders, in £, made to a firm over the internet has distribution $N(\mu, \sigma^2)$. A random sample of n orders is taken and $\overline{X}$ denotes the sample mean.

(a) Write down the mean and variance of $\overline{X}$ in terms of μ and σ^2.

(2)

A second sample of m orders is taken and $\overline{Y}$ denotes the mean of this sample.
An estimator of the population mean is given by

$$U = \frac{n\overline{X} + m\overline{Y}}{n + m}.$$

(b) Show that U is an unbiased estimator for μ.

(3)

(c) Show that the variance of U is $\frac{\sigma^2}{n+m}$.

(4)

(d) State which of $\overline{X}$ or U is a better estimator for μ. Give a reason for your answer.

(2)

Question 2 continued

Turn over

Question 2 continued

Question 2 continued

Q2

(Total 11 marks)

3. The lengths, x mm, of the forewings of a random sample of male and female adult butterflies are measured. The following statistics are obtained from the data.

	No. of butterflies	Sample mean $\bar{x}$	$\sum x^2$
Females	7	50.6	17 956.5
Males	10	53.2	28 335.1

(a) Assuming the lengths of the forewings are normally distributed test, at the 10% level of significance, whether or not the variances of the two distributions are the same. State your hypotheses clearly.

(7)

(b) Stating your hypotheses clearly test, at the 5% level of significance, whether the mean length of the forewings of the female butterflies is less than the mean length of the forewings of the male butterflies.

(6)

Question 3 continued

Turn over

Question 3 continued

Question 3 continued

(Total 13 marks)

Q3

4. The length X mm of a spring made by a machine is normally distributed N (μ, σ^2).
A random sample of 20 springs is selected and their lengths measured in mm. Using this sample the unbiased estimates of μ and σ^2 are

$$\bar{x} = 100.6 \qquad s^2 = 1.5$$

Stating your hypotheses clearly test, at the 10% level of significance,

(a) whether or not the variance of the lengths of springs is different from 0.9,

(6)

(b) whether or not the mean length of the springs is greater than 100 mm.

(6)

Question 4 continued

Turn over

Question 4 continued

Question 4 continued

(Total 12 marks)

Q4

Turn over

5. The number of tornadoes per year to hit a particular town follows a Poisson distribution with mean λ. A weatherman claims that due to climate changes the mean number of tornadoes per year has decreased. He records the number of tornadoes x to hit the town last year.

To test the hypotheses H_0: $\lambda = 7$ and H_1: $\lambda < 7$, a critical region of $x \leqslant 3$ is used.

(a) Find, in terms λ the power function of this test.

(3)

(b) Find the size of this test.

(2)

(c) Find the probability of a Type II error when $\lambda = 4$.

(2)

Question 5 continued

Turn over

Question 5 continued

Question 5 continued

Q5

(Total 7 marks)

Turn over

6. A butter packing machine cuts butter into blocks. The weight of a block of butter is normally distributed with a mean weight of 250 g and a standard deviation of 4 g. A random sample of 15 blocks is taken to monitor any change in the mean weight of the blocks of butter.

(a) Find the critical region of a suitable test using a 2% level of significance.
(3)

(b) Assuming the mean weight of a block of butter has increased to 254 g, find the probability of a Type II error.
(5)

Question 6 continued

Q6

(Total 8 marks)

Turn over

7. A doctor wishes to study the level of blood glucose in males. The level of blood glucose is normally distributed. The doctor measured the blood glucose of 10 randomly selected male students from a school. The results, in mmol/litre, are given below.

4.7 3.6 3.8 4.7 4.1 2.2 3.6 4.0 4.4 5.0

(a) Calculate a 95% confidence interval for the mean.
(7)

(b) Calculate a 95% confidence interval for the variance.
(4)

A blood glucose reading of more than 7 mmol/litre is counted as high.

(c) Use appropriate confidence limits from parts (a) and (b) to find the highest estimate of the proportion of male students in the school with a high blood glucose level.
(4)

Question 7 continued

Turn over

Question 7 continued

(Total 15 marks)

Q7

TOTAL FOR PAPER: 75 MARKS

END