EXAMINATIONS

September 1997

Subject C — Statistics

Paper One

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Write your surname in full, the initials of your other names and your Candidate's Number on the front of the answer booklet.*
2. *Begin your answers to Parts One, Two and Three on a separate sheet.*
3. *Mark allocations are shown in brackets.*
4. *Attempt all 17 questions.*

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet and this question paper.

In addition to this paper you should have available Actuarial Tables and an electronic calculator.

PART ONE

For questions 1–8 indicate in your answer booklet which one of the answers A, B, C or D is correct.

1 The number of demands made on a service team each day has a Poisson distribution with mean 2. Under current arrangements the service team can handle, at most, 3 demands per day and no demands are carried forward.

The mean number of demands handled by the service team per day is:

A 1.353
B 1.579
C 1.782
D 2 [3]

2 In a particular road it is estimated that there is a 25% chance that any specified house will be burgled over a period of two years, independently for each house. There are six houses in the road.

The probability that fewer than two houses will be burgled over the period is:

A 0.005
B 0.297
C 0.356
D 0.534 [3]

3 Customer electricity charges C are calculated according to the formula

$$C = 7.00 + 0.0742N$$

where N denotes the number of units used.

In a particular area, N is modelled as a random variable with mean 600 and variance 250.

The mean and variance of the charges C in their respective units are:

	mean	variance
A	44.52	1.38
B	44.52	18.55
C	51.52	1.38
D	51.52	8.38

[2]

4 Independent random samples of size $n_1 = 10$ and $n_2 = 25$, respectively, are taken from normal distributions with the same variance. The sample variances are S_1^2 and S_2^2, respectively.

The value of k such that $P\left(\frac{S_1^2}{S_2^2} > k\right) = 0.01$ is:

A 2.300
B 2.900
C 3.256
D 4.729 [2]

5 A transformation which converts randomly generated numbers y_i, sampled from the uniform distribution on the interval (0, 1), into simulated values x_i from the density function

$$f(x) = 5\exp(-5x),\ 0 < x < \infty$$

is:

A $x_i = 0.2\log(1 - y_i)$

B $x_i = 1 - \exp(-5y_i)$

C $x_i = -0.2\log\left(\frac{y_i}{5}\right)$

D $x_i = 0.2\log\left(\frac{1}{y_i}\right)$ [2]

6 The amounts (£) of pocket-money received by a random sample of 100 13-year-old girls in the UK in 1994 had sample mean £4.56 and sample standard deviation £1.21.

A 95% confidence interval for the mean amount received in the population of all such girls is:

A £4.25 to £4.87
B £4.32 to £4.80
C £4.44 to £4.68
D £4.42 to £4.90 [3]

 PLEASE TURN OVER

7 In part of an internal quality audit of an insurance company the auditors have to inspect a large number of claim files for household contents policies. They intend to select a random sample of files and give these a thorough examination.

A sample of 400 claim files is examined. The auditors record whether or not the claim required a visit by a senior inspector. It was noted that 62 claims did require such a visit.

A 99% confidence interval for the true percentage of claims requiring a visit by a senior inspector is:

A 15.5 ± 1.8%
B 15.5 ± 3.5%
C 15.5 ± 4.2%
D 15.5 ± 4.7% [3]

8 *See question 7.*

A sample of 200 claim files is examined. The auditors record the time elapsed between the original notification of the claim and the final payment made on the claim. They wish to test the hypothesis that the true mean time elapsed is 50 days against the alternative that it is greater than 50 days.

For the sample data, the mean time was calculated as 54.5 days and the standard deviation as 40.3 days. The probability value for the appropriate test is:

A 0.0059
B 0.012
C 0.057
D 0.11 [3]

PART TWO

9 The table below shows a grouped frequency distribution for 100 claim amounts on a certain class of insurance policy.

Claim Amount	*Frequency*
under £100	4
£100–149.99	10
£150–199.99	25
£200–249.99	30
£250–299.99	15
£300–349.99	12
£350–399.99	4
£400 or over	0

Determine approximate values for the median and the interquartile range of these claims amounts. [4]

10 A random variable X follows a gamma distribution with parameters α and λ.

(i) Show that the moment generating function of X is given by

$$M(t) = \left(1 - \frac{t}{l}\right)^{-a}, \quad t < \lambda. \qquad [3]$$

(ii) Use this moment generating function to show that the mean and variance of X are $\frac{a}{l}$ and $\frac{a}{l^2}$, respectively.

[2]

[Total 5]

11 An insurance company issues policies of types A and B. The number of claims per week for type A policies has a Poisson distribution with mean 75. The number of claims per week for type B policies has a Poisson distribution with mean 25. All claims arise independently of one another.

Suppose that the size of a claim for a type A policy is always £1000, and the size of a claim for a type B policy is always £5000.

Find the mean and standard deviation of the total amount of claims in a week for both types of policy together. [5]

12 The probabilities of death within one year for the respective partners in a marriage relationship are denoted by q_1 and q_2 and the deaths are assumed to be independent events.

(i) Write down expressions, in terms of q_1 and q_2, for the probabilities of neither, one only, and both partners dying within one year. [2]

(ii) If, in a random sample of 1000 such couples taken from a population for which $q_1 = 0.1$ and $q_2 = 0.09$, neither partner dies within one year in 825 cases and exactly one partner dies within one year in 159 cases, test the goodness of fit of the above model and comment briefly on the result. [5]

[Total 7]

13 In two series of hauls to determine the number of plankton organisms inhabiting the waters of a lake, the following counts per unit volume of water were observed:

Series I: 80 96 102 77 97 110 99 88 103 108 $\sum x = 960$ $\Sigma x^2 = 93{,}276$

Series II: 74 122 92 81 104 92 90 $\sum y = 655$ $\Sigma y^2 = 62{,}765$

In Series I the hauls were made at different times at the same place in the lake; in Series II they were made at the same time but at different points scattered over the lake. Investigate whether the variability is greater between different places than between different times at the same place on the evidence of these data. State all assumptions you make.

[5]

14 A sample of 12 paired observations (x_i, y_i) yielded a sample correlation coefficient $r = 0.91$. Use Fisher's transformation to perform the following test on the population correlation coefficient ρ.

$$H_0 : \rho = 0.8 \quad \text{v.} \quad H_1 : \rho > 0.8$$ [5]

PART THREE

15 It is known that a reasonable model for the occurrence of claims on many types of insurance policy is the Poisson model. Such a model has one parameter λ, being the claim rate per year (or other unit of time). In order to estimate the claim rate, a random sample of *n* policies is taken and the number of claims in the last year is recorded for each one, giving the data $x_1, x_2, \ldots, x_n$.

(i) (a) Show that the maximum likelihood estimator of λ is $\overline{X}$, the sample mean.

(b) Determine the Cramer-Rao lower bound for this situation and hence state the asymptotic distribution of $\overline{X}$. [7]

(ii) (a) By standardising $\overline{X}$ (using its distribution from part (i)(b)), use $\dfrac{\overline{X} - l}{\sqrt{\dfrac{l}{n}}}$ as a pivotal quantity to develop approximate 95% confidence limits for λ.

(b) If a random sample of 100 policies yields a total of 145 claims, calculate the 95% confidence limits for λ as found in part (ii)(a). [6]

(iii) (a) It is more usual in practice to use $\dfrac{\overline{X} - l}{\sqrt{\dfrac{\overline{X}}{n}}}$ as a pivotal quantity. Develop approximate 95% confidence limits for λ in this way.

(b) For the data in part (ii)(b), calculate these 95% confidence limits for λ and comment briefly on their comparison with those from part (ii)(b). [4]

[Total 17]

16 It is thought that a suitable model for a plumber's charges when called out for a job is a linear one based on a fixed call-out charge and an hourly rate. A random sample of 10 of his invoices gave the following results:

Duration of job (hours) x	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.5	5.0	5.5
Cost of job (£) y	40	55	45	65	80	75	95	100	120	130

$\Sigma x = 29 \quad \Sigma x^2 = 110.5 \quad \Sigma y = 805 \quad \Sigma y^2 = 73{,}225 \quad \Sigma xy = 2{,}795$

(i) (a) Plot these data and comment on the suitability of the proposed model.

(b) Calculate the least squares estimates of:

(1) the plumber's call-out charge, and,

(2) the plumber's hourly rate charge. [6]

(ii) (a) Determine a 90% confidence interval for the plumber's hourly rate charge.

(b) Calculate 95% confidence intervals for the expected cost of single jobs lasting

(1) three hours, and

(2) six hours

and comment briefly on their different widths. [9]

[Total 15]

17 (i) A random sample of size 20 is taken from the normal distribution $N(m, s^2)$ and the following values recorded:

–0.31	–0.90	–0.22	1.00	0.12	–0.01	–0.16	–1.31	0.38	–0.38
–1.07	1.65	–1.02	–0.06	–0.47	–0.10	0.71	0.94	–0.29	–0.57

$$\Sigma x = -2.07 \qquad \Sigma x^2 = 10.9385$$

(a) Test the hypothesis $H_0 : s^2 = 1$ against the alternative $H_1 : s^2 \neq 1$.

(b) Derive the maximum likelihood estimates for μ and s^2.

(c) Determine 95% confidence intervals for μ:

(1) assuming $s^2 = 1$, and

(2) making no assumption about the value of s^2. [11]

(ii) A random sample of size 20 is taken from the normal distribution $N(\mu, 1)$ but instead of recording the actual values, as in part (i) above, it is observed that 14 of the 20 observations are negative and the remaining 6 observations are positive. Using only this information, determine the maximum likelihood estimate of μ. [5]

[Total 16]