

EXAMINATIONS

September 1999

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.*

Graph paper is required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet and this question paper.

<p><i>In addition to this paper you should have available graph paper, Actuarial Tables and an electronic calculator.</i></p>

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 amounts paid out by a company on a sample of 40 third-party motor policy claims include six in which nothing was actually paid out. Four were in excess of £100,000 and the remaining thirty were as follows (in order, in units of £1,000):

0.41	0.89	1.02	1.27	1.63	1.82	1.96	2.23	2.29	2.41
2.78	3.11	3.64	4.20	4.52	4.94	5.61	5.72	7.14	10.04
12.21	15.67	19.33	25.18	28.82	31.20	35.19	46.88	61.94	85.04

In which of the following intervals does the interquartile range of the full sample of 40 amounts lie?

- A 0–£10,000
B £10,001–£20,000
C £20,001–£25,000
D over £25,000 [3]

- 2 Five students are selected at random, one after the other and without replacement, from a group of twenty students of whom twelve are men and eight are women.

The probability that the fifth student selected is a man is:

- A 0.5
B 0.563
C 0.6
D 0.625 [2]

- 3 Consider the displayed bivariate probability distribution for the discrete random variables X and Y ($x = 0, 1, 2, 3$; $y = 0, 1, 2$).

		X			
		0	1	2	3
Y	0	.01	.03	.05	.02
	1	.10	.12	.15	.08
	2	.05	.16	.20	.03

The conditional expectation $E(Y|X = 0)$ is:

- A 0.20
B 1.25
C 1.33
D 1.73 [3]

- 4 Claim amounts (£) for a certain type of claim are modelled by a gamma distribution with parameters $\alpha = 40$ and $\lambda = 0.1$.

The probability that a claim amount exceeds £500 is approximately:

- A 0.057
 - B 0.21
 - C 0.49
 - D 0.94
- [3]

- 5 A random sample of size 10 is taken from a normal population with standard deviation $\sigma = 15$ and the sample standard deviation, S , is calculated.

The value of “a” such that $P(S > a) = 0.95$ is

- A 6.59
 - B 9.12
 - C 9.92
 - D 20.57
- [3]

- 6 A set of 10 independent sample values of a normal random variable, X , yield a sample mean of 130.2, and a sample standard deviation of 25.0.

A 90% confidence interval for the population mean is

- A 84.38 to 176.02
 - B 112.59 to 147.81
 - C 115.71 to 144.69
 - D 117.20 to 143.20
- [3]

- 7 A goodness-of-fit test of a model to an observed frequency distribution has been conducted using a χ^2 statistic and the calculated value of the statistic is k .

An equal number of additional observations becomes available, and their frequency distribution is identical to the one you already have. The two sets of data are combined, so that the observed frequencies in the cells double and the proportions in the various cells remain the same.

The revised value of the χ^2 statistic is:

- A $k/2$
 - B k
 - C $2k$
 - D $4k$
- [2]

PART TWO

- 8** A set of paired data has correlation coefficient $r = -0.6$. In fitting a linear regression model to the data, the sum of squares of the residuals is

$$\sum (y_i - \hat{y}_i)^2 = 2160.$$

Calculate the variation in the responses explained by the model (the regression sum of squares). [3]

- 9** Two insurance investigators have each collected sample data on claim amounts. John used a sample of size 15 which yielded a mean of £452 and a sample standard deviation of £205. Mary used a sample of size 20 which yielded a mean of £385 and a sample standard deviation of £182.

They combined their data to give a sample of size 35.

Calculate the mean and standard deviation for this combined sample. [4]

- 10** The random variable X is distributed uniformly over $[2, 4]$.

(i) State the value of the population mean $E(X)$. [1]

(ii) Show that $\text{Var}(X) = \frac{1}{3}$. [2]

(iii) Let Y be defined by $Y = \sum_{i=1}^{48} X_i$, where X_1, X_2, \dots, X_{48} are independent variables distributed as X .

Calculate an approximate value for $P(140 \leq Y \leq 152)$. [3]

[Total 6]

- 11** An insurance executive wishes to take a sample of policies from a portfolio in order to estimate the percentage that have been in force for more than 10 years. Each policy is chosen with equal probability.

Determine the minimum sample size required to ensure that 95% confidence limits for the underlying population proportion are of the form “ $\pm 5\%$ ”, justifying any approximations used. [4]

- 12** Two sales representatives deal with the sales of insurance policies in a local office. In a particular section of the market they have three types of policy available for sale; I, II and III.

The table below shows for a particular month the number of each type of policy each representative sells.

		<i>Policy Type</i>		
		I	II	III
<i>Representative</i>	A	30	35	40
	B	26	20	13

Investigate the claim that the proportion of policies of each type sold is independent of the representative. [6]

- 13** In a particular linear regression situation there are reasons for assuming that an appropriate model is a straight line through the origin, that is,

$$Y_i = \beta x_i + e_i \text{ where } e_i \text{ has mean } 0 \text{ and variance } \sigma^2.$$

(i) Derive the least squares estimate of the slope parameter β . [3]

(ii) Discuss whether the fitted line will pass through the point (\bar{x}, \bar{y}) . [2]

[Total 5]

- 14** A random sample of 50 pairs of observations is taken from a bivariate normal distribution with correlation coefficient ρ . The sample correlation coefficient is 0.58.

By using Fisher's transformation, calculate a 95% confidence interval for ρ . [6]

PART THREE

- 15** (i) Show that the exponential random variable, X , with probability density function

$$f(x) = \lambda e^{-\lambda x}; \quad \lambda > 0, x \geq 0.$$

has moment generating function

$$\frac{\lambda}{\lambda - t}; \quad t < \lambda \quad [2]$$

- (ii) (a) The random variable Z has probability density function

$$\begin{cases} \frac{\alpha\beta}{\alpha + \beta} e^{-\alpha z} & ; \quad z \geq 0 \\ \frac{\alpha\beta}{\alpha + \beta} e^{\beta z} & ; \quad z \leq 0 \end{cases} \quad \alpha > 0, \beta > 0$$

By using the results of (i) or otherwise, show that the moment generating function of Z is

$$\frac{\alpha\beta}{(\beta + t)(\alpha - t)}; \quad -\beta < t < \alpha$$

- (b) By taking logs, or otherwise, derive $E(Z)$ and $\text{Var}(Z)$ from the moment generating function.
- (c) By considering the values of the mean and mode, discuss briefly the skewness of the distribution. [11]
- (iii) The two independent random variables X and Y are both exponentially distributed with respective parameters α and β {i.e. in (i) $\lambda = \alpha$, β respectively}.

Use moment generating function techniques to determine the probability density function of $R = X - Y$ [3]

[Total 16]

- 16** Let X_1, X_2, \dots, X_n be a random sample of size n from the normal distribution $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. In particular it is required to estimate the variance σ^2 .

- (i) Show that the maximum likelihood estimators are

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2. \quad [5]$$

- (ii) Use the result $\frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$ to show that $\hat{\sigma}^2$ is biased as an estimator of σ^2 and determine expressions for its bias and its variance. [3]
- (iii) Show that the sample variance $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ is unbiased for σ^2 and determine an expression for its variance. [2]
- (iv) Show that the mean square error of $\hat{\sigma}^2$ is less than the mean square error of S^2 . [4]
- (v) An alternative estimator for σ^2 is $\frac{1}{n+1} \sum (X_i - \bar{X})^2$. Given that its mean square error is $\frac{2}{n+1} \sigma^4$, discuss whether this is a better estimator of σ^2 than $\hat{\sigma}^2$ or S^2 . [2]
- [Total 16]

17 A market analyst is investigating the experience of two insurance companies, A and B, with regard to a particular class of insurance business (which is written by both companies). The analyst has access to a limited amount of information - a random sample of 10 recent claim amounts from company A and a random sample of 8 recent claim amounts from company B.

The data follow (amounts in units of £100), with summaries:

A: 27.6 43.0 32.7 22.1 31.5 55.4 32.2 10.7 38.7 29.9
 B: 31.2 26.8 11.2 22.8 36.5 28.9 13.2 25.1

Company A: $n = 10$, $\Sigma x = 323.8$, $\Sigma x^2 = 11772.90$

Company B: $n = 8$, $\Sigma x = 195.7$, $\Sigma x^2 = 5308.67$

- (i) (a) Plot the two sets of data on a simple diagram for comparison purposes, and comment on the assumption that these samples may be regarded as coming from normal distributions.
- (b) The analyst is quite happy to assume that the variation in claim amounts is the same for both companies.
- Perform a test to justify this assumption. [7]
- (ii) (a) Calculate a two-sided 95% confidence interval for the difference between the mean claim amounts for all such business for the two companies, and comment briefly on the result.
- (b) Perform a test to investigate whether the mean claim amount for all such business for company A exceeds that for company B. (State your approximate probability-value and your conclusion clearly). [8]

[Total 15]