This paper is not to be removed from the Examination Halls

UNIVERSITY OF LONDON

279 004b ZA
990 004b ZA
996 D04b ZA

BSc degrees in Economics, Management, Finance and the Social Sciences, the Diploma in Economics and Access Route for Students in the External Programme

Statistics 2 (half unit)

Monday, 12 May 2003 : 2.30pm to 4.30pm

Candidates should answer **THREE** of the following **FIVE** questions: **QUESTION 1** of Section A (40 marks in total) and **TWO** questions from Section B (30 marks each). **Candidates are strongly advised to divide their time accordingly.**

A list of formulae is given at the end of the paper.

Graph paper is provided. If used, it must be securely fastened inside the answer book.

New Cambridge Statistical Tables (second edition) are provided.

A hand held non-programmable calculator may be used when answering questions on this paper. The make and type of machine must be stated clearly on the front of the answer book.

PLEASE TURN OVER

UL03/405

SECTION A

Answer all **seven** parts of question 1 (40 marks).

1. (a) For each of i to iv below, say whether the statement is true or false and briefly give your reasons:

   i. If $A \cap B = \emptyset$ then $A = \emptyset$ or $B = \emptyset$.

   ii. If the correlation of X, Y is negative, then $\mathrm{var}(X + Y) \leq \mathrm{var}\, X + \mathrm{var}\, Y$.

   iii. 'I have 95% confidence intervals for 100 different parameters, so I expect that 5 of those intervals will not include the true parameter value'.

   iv. An unbiased estimator is one where the statistician has not cheated.

(8 marks)

(b) Write brief notes on each of the topics below. Explain in what way each topic is part of statistics, and why it is important.

   i. The χ^2 distribution.

   ii. Extrapolation.

   iii. Error Sum of Squares.

   iv. Type II error.

(6 marks)

(c) A comparison of the forecasts of changes in the FTSE 100 by an expert with what actually occurred is given below:

		Forecast	
		Up	Down
Actual	Up	35	20
	Down	5	20

   i. What is the proportion of correct forecasts by the expert?

   ii. Give a 95% Confidence Interval for the proportion of correct forecasts by the expert.

(5 marks)

(d) X is a random variable with expected value zero and $P(X = 2) = 0.2$ and $P(X = 3) = 0.3$. X takes just one other value besides 2 and 3.

   i. What is the probability that X is negative?

   ii. What is the other value that X takes?

   iii. What is the variance of X?

(4 marks)

(e) i. From your tables find the probability that there are at least 6 successes for the binomial distribution with 17 trials and probability of success 0.35.

ii. From your tables find the probability that $X > 7$ when X has a Poisson distribution with mean 4.5.

(4 marks)

(f) A university department has 7 academic staff. There are 4 mathematicians and 3 statisticians. At the end of the year, one of the academic staff given a prize as the best researcher for the year. This selection is made at random among the 7 staff. If the winner is a mathematician, then an additional mathematician is hired. If a statistician wins, then he/she leaves the department for a better post and is not replaced. At the end of the second year, another winner is randomly selected. What is the probability that a mathematician won in year 1, given that a mathematician won in year 2?

(5 marks)

(g) 30 observations are thought to be independent realisations x of a Binomial random variable X with $n = 20$ trials and probability of success $\pi = 0.2$. They are tabulated below:

$x \leq 3$	$3 < x \leq 5$	$5 < x$
15	9	6

Test the hypothesis that the observations on X are from a Binomial distribution with number of trials $n = 20$ and probability of success $\pi = 0.2$.

(8 marks)

PLEASE TURN OVER

SECTION B

Answer **two** questions from this section (30 marks each).

2. (a) Four balls are thrown at random into 4 boxes, with each ball having an equal chance of going into every box. Work out the probability that exactly m boxes are empty, where $m = 0, 1, 2, 3$. **(12 marks)**

   (b) X is a random variable with density function

$$f_X(x) = (2 - x)^3/4$$

   over the range $(0, 2)$. Using calculus, find the mean and the variance of X. **(18 marks)**

3. (a) Suppose that have a random sample of size n from a normal distribution with mean μ and variance σ^2. What is the mean squared error of $\bar{X} + 1$ as an estimator of μ?

   (5 marks)

   (b) A box contains 3 balls. Either 2 balls are red and 1 ball is green, or 1 ball is red and 2 balls are green. One ball is taken at random from the box. Estimators S and T of the number of red balls in the box at the start are defined as below:

	Red ball taken	Green ball taken
S	2	1
T	1.5	1.5

   i. What are the mean and variance for each of S and T? (Remember to work with both of the possible initial numbers of red balls.)

   ii. Which of S, T is the best estimator?

   (12 marks)

   (c) An anxious academic times his journey to work from Golders Green on Mondays and Thursdays for ten weeks. His record shows time in minutes for each journey:

	Day	
Week	Monday	Thursday
1	51	48
2	54	52
3	52	50
4	57	51
5	65	70
6	53	55
7	59	52
8	60	49
9	56	50
10	51	53

i. Find an 85% confidence interval for the difference between mean travel time on Mondays and Thursdays.

(9 marks)

ii. Test the null hypothesis that the average travel time on Monday is 2 minutes longer than on Thursday.

(4 marks)

4. (a) Derive the sum of squares identity for a two-way analysis of variance.

(8 marks)

(b) Why might one take logarithms of the numbers in a two-way table before carrying out an analysis of variance?

(6 marks)

(c) The table below shows the Irish population in millions over several years, for some age groups.

	Year			
Age	1991	1996	1997	2000
0-14	26.7	23.7	23.1	21.8
15-24	17.1	17.5	17.4	17.4
25-34	14.1	14.3	14.6	15.0
35-39	18.5	19.9	20.0	20.2
50-64	12.3	13.2	13.5	14.4
65+	11.4	11.4	11.4	11.2

i. Give a two-way analysis of variance table for these data.

(9 marks)

ii. Is there significant evidence that Year effects are different from zero?

(4 marks)

iii. Give several reasons why this table does not look appropriate for analysis of variance.

(3 marks)

PLEASE TURN OVER

5. (a) i. Suppose that one has the estimated regression slopes of Y on X and also of X on Y. Show that the product of those estimators is the square of the correlation coefficient between X and Y.

(6 marks)

ii. With the usual assumptions, find the variance of the estimate of the regression slope of Y on X.

(6 marks)

(b) The table below shows the prices in a supplies catalogue of printer toner cartridges in pounds, with the reduced rates for multiple purchases.

Cartridge	Price for 1	Price for 1 of 5	Price for 1 of 10
1	31.99	31.45	31.15
2	38.99	38.45	38.15
3	29.99	29.45	29.15
4	53.99	53.45	53.15
5	39.99	39.45	39.15
6	84.99	84.65	84.15

i. Fit a straight line to these data using Price of 1 of 5 as the response variable and Price of 1 as the explanatory variable.

(6 marks)

ii. Give an 85% confidence interval for the slope of the regression.

(6 marks)

iii. Explain carefully the results you have obtained. What is the pricing policy used here? Are the usual assumptions of regression models valid? If the additional explanatory variable Price of 1 of 10 is put into the regression model for Price of 1 of 5, what would the fitted model be?

(6 marks)

Formulae for Statistics

Discrete Distributions

Binomial Distribution

The probability of x successes in n trials is

$$\binom{n}{x} \pi^x (1-\pi)^{n-x}$$

for $x = 0, 1, \ldots, n$ The mean number of successes is $n\pi$ and the variance is $n\pi(1-\pi)$.

Poisson Distribution

The probability of x is

$$e^{-\mu}\frac{\mu^x}{x!}.$$

The mean number of successes is μ and the variance is μ.

Hypergeometric Distribution

The probability of x successes in a sample of size n from a population of size N with M successes is

$$\binom{M}{x}\binom{N-M}{n-x}\Big/\binom{N}{n}.$$

The mean number of successes is nM/N and the variance is
$n(M/N)(1-M/N)(N-n)/(N-1)$.

PLEASE TURN OVER

Sample Quantities

Sample Variance	$s^2 = \sum(x_i - \bar{x})^2/(n-1) = (\sum x_i^2 - n\bar{x}^2)/(n-1)$
Sample Covariance	$\sum(x_i - \bar{x})(y_i - \bar{y})/n = (\sum x_i y_i - n\bar{x}\bar{y})/n$
Sample Correlation	$(\sum x_i y_i - n\bar{x}\bar{y})/\sqrt{(\sum y_i^2 - n\bar{y}^2)(\sum x_i^2 - n\bar{x}^2)}$

Inference

Variance of Sample Mean	σ^2/n
One-sample t statistic	$\sqrt{n}(\bar{x}-\mu)/s$ with $(n-1)$ degrees of freedom
Two-sample t statistic	$\dfrac{(\bar{x_1}-\bar{x_2})-(\mu_1-\mu_2)}{\sqrt{[1/n_1+1/n_2]\{[(n_1-1)s_1^2+(n_2-1)s_2^2]/(n_1+n_2-2)\}}}$
Variances for differences of binomial proportions	
Pooled	$\left[\dfrac{(n_1p_1+n_2p_2)}{(n_1+n_2)}\right]\left[1-\dfrac{(n_1p_1+n_2p_2)}{(n_1+n_2)}\right]\left[\dfrac{1}{n_1}+\dfrac{1}{n_2}\right]$
Separate	$p_1(1-p_1)/n_1+p_2(1-p_2)/n_2$
Least Squares	Estimates for $y=\alpha+\beta x$ fitted to (y_i,x_i) for $i=1,2,\ldots,n$ are $a=\bar{y}-b\bar{x}$ and $b=\sum(x_i-\bar{x})(y_i-\bar{y})/\sum(x_i-\bar{x})^2$. The estimate of variance is $[\sum(y_i-\bar{y})^2-b^2\sum(x_i-\bar{x})^2]/(n-2)$. The variance of b is $\sigma^2/\sum(x_i-\bar{x})^2$
Chi-squared Statistic	$\sum$(Observed − Expected)^2/Expected, with degrees of freedom depending on the hypothesis tested.

END OF PAPER