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 and Diplomas for Graduates 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)

Tuesday, 16 May 2006 : 2.30pm to 4.30pm

Candidates should answer **THREE** of the following **FIVE** questions: **QUESTION 1** of Section A (40 marks) 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 fastened securely inside the answer book.

New Cambridge Statistical Tables (second edition) are provided.

A hand held calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics, text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

PLEASE TURN OVER

SECTION A

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

1. a) For each one of the statements below say whether the statement is true or false explaining your answer.

   i. For three events A, $B$and C, if A is independent of B and $C \subset A$, then C is independent of B.

   ii. If A and B are two events such that $\Pr(B|A) > \Pr(B)$, then $\Pr(A|B) > \Pr(A)$.

   iii. It is possible to find two independent events A and B such that $\Pr(B|A) > \Pr(B)$.

   iv. Let T_1 and T_2 be two independent unbiased estimators for the same parameter with the same variance. Their average $\frac{T_1+T_2}{2}$ is an unbiased estimator for the same parameter with a smaller variance than either of them.

   v. The random variable X has a binomial distribution with number of trials 2 and probability of success $\sqrt{\theta}$. Let T be such that $T = 1$ if $X = 2$ and $T = 0$ otherwise. Then T is an unbiased estimator for the parameter θ.

   vi. Two fair dice are thrown. The events $A = \{$the dice show a total for 2, 4 or 5$\}$ and $B = \{$the dice show a total score that is even$\}$ are independent.

(16 marks)

b) Briefly explain the meaning of the following.

   i. Correlation

   ii. The power of a test

(4 marks)

c) A random variable can only take the values 0 and 2. Its variance is 1. Find its mean.

(4 marks)

d) Consider two random variables X and Y. X can take the values 0, 1 and 2 and Y can take the values 1 and 2. The joint probabilities for each pair are given by the following table.

	$X=0$	$X=1$	$X=2$
$Y=1$	0.1	0.2	0.3
$Y=2$	0.1	0.1	0.2

Let $Z = \max(X,Y)$ be the larger of the two variables. Find $E(Y)$, $E(Y|X=1)$, $E(Z)$ and $E(Z|X=1)$.

(10 marks)

(question continues on next page)

e) The following observations are supposed to be observations of a Binomial random variable with number of trials 10 and the same probability of success: 2,3,1,0,4,4,1,2,3,0. Estimate the probability of success and use a goodness of fit test, where observations of value 3 and above are grouped together, to test whether they are indeed from the same Binomial distribution.

(6 marks)

SECTION B

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

2. a) The proportion of pregnant women amongst women that take a particular pregnancy test is 75%. Two thirds of the pregnant women are at an early stage of pregnancy. If a woman that takes the test is pregnant at an early stage the test will be positive (indicate that she is pregnant) with probability 0.88. If she is pregnant at a later stage it will be positive with probability 0.96. If she is not pregnant it will be positive with probability 0.2.

   i. What is the probability the test will be positive?
   ii. Given that a woman took the test and it was positive, what is the probability that she is pregnant (at any stage)?
   iii. Given that a woman took the test and it was negative, what is the probability that she is pregnant (at any stage)?
   iv. What is the probability the test will show a false result?

(14 marks)

b) The random variable X has a density function given by

$$f(x) = \frac{3x^2 + 2x}{2}$$

defined over the region $0 \leq x \leq 1$. Find $\Pr(X > 0.8 | X > 0.6)$, $E(X)$, $\mathrm{Var}(X)$ and $\mathrm{Cov}\left(X, \frac{1}{X}\right)$.

(16 marks)

PLEASE TURN OVER

3. Three lubricants 1, 2 and 3 are tested using by using 15 different cars, 3 from each one of the brands A, B, C, D and E. Each one of the three cars of each brand used a different lubricant. Testing took place on the same day and the top speeds in kilometres per hour achieved by each car without overheating are given in the following table

	Lubricant 1	Lubricant 2	Lubricant 3
Brand A	176	174	172
Brand B	177	172	166
Brand C	181	180	177
Brand D	171	170	187
Brand E	197	191	202

a) Give the table for a two-way analysis of variance of the above data.

(8 marks)

b) Is there a significant difference between top speeds achieved by cars of different brands?

(2 marks)

c) Is there a significant difference between the top speeds achieved using different lubricants?

(3 marks)

d) Give a 95% confidence interval for the difference one would expect between the top speeds one would expect between brand E and brand A cars.

(3 marks)

e) Lubricant 3 was eventually dismissed as too expensive. In subsequent days the two brand E cars used to test lubricants 1 and 2 continued using the same lubricant they used originally and the top speeds achieved without overheating are given by the following table

Day	Lubricant 1	Lubricant 2
1	197	191
2	200	188
3	202	180
4	200	180
5	195	178
6	184	185
7	180	183
8	178	179
9	177	178
10	175	179

(question continues on next page)

(The speeds achieved on day 1 are included). Construct a two sided 90% confidence interval for the difference in top speeds achieved by the two cars using the two different lubricants.

(7 marks)

f) By using an appropriate hypothesis test find out if there is significant evidence that higher top speeds are achieved using lubricant 1 rather than lubricant 2.

(5 marks)

g) By referring to your data discuss whether there are any problems with the analysis in parts e) and f).

(2 marks)

4. a) Suppose that you are given observations y_1, y_2, y_3 and y_4 that are such that

$$y_1 = \alpha + \beta + \varepsilon_1,$$
$$y_2 = -\alpha + \beta + \varepsilon_2$$
$$y_3 = \alpha - \beta + \varepsilon_3$$
$$y_4 = -\alpha - \beta + \varepsilon_4.$$

The variables ε_i, $i = 1,2,3,4$ are normally distributed with mean 0 and variance σ^2. Find the least squares estimators for the parameters α and β, verify that they are unbiased and find the variance of the estimator for the parameter α.

(12 marks)

b) The table below gives the annual expenditure in marketing, the annual expenditure in quality control and the annual profits of a company for the last 9 years measured in thousands of pounds.

Year	Marketing	Quality Control	Profits
1	730	170	100
2	800	45	60
3	840	130	140
4	820	150	140
5	770	180	180
6	760	140	150
7	850	170	220
8	920	180	310
9	890	180	380

Fit a straight line through these data using profits as the dependent variable and marketing expenditure as the independent variable.

(7 marks)

(question continues on next page) PLEASE TURN OVER

c) Is there significant evidence that an increase in marketing expenditure will result to an increase in profits?

(4 marks)

d) The full model including quality control expenditure is

$$''\text{Profits}'' = -875 + 1.08''\text{Marketing}'' + 1.18''\text{Quality Control}''$$

Carefully interpret this equation.

(4 marks)

e) Does this show that money spent on quality control is better spent than money spent on marketing?

(3 marks)

5. Consider two random variables X and Y. X can take the values -1, 0 and 1 and Y can take the values 0, 1 and 2. The joint probabilities for each pair are given by the following table.

	$X=-1$	$X=0$	$X=1$
$Y=0$	0.1	0.2	0.1
$Y=1$	0.1	0.05	0.1
$Y=2$	0.1	0.05	0.2

a) Calculate the marginal distributions and expected values of X and Y.

(9 marks)

b) Calculate the covariance of the random variables U and V, where $U = X+Y$ and $V = X-Y$.

(7 marks)

c) Calculate $E(V|U=1)$

(7 marks)

d) The random variable W has the same distribution as X and the random variable Z has the same distribution as Y. The random variables W and Z are independent. Write down the table for the joint probabilities of W and Z and calculate their covariance.

(7 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	$\sum(x_i - \bar{x})(y_i - \bar{y})/(n-1) = (\sum x_i y_i - n\bar{x}\bar{y})/(n-1)$
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

$$\frac{(\bar{x_1}-\bar{x_2})-(\mu_1-\mu_2)}{\sqrt{[1/n_1+1/n_2]\{[(n_1-1)s_1^2+(n_2-1)s_2^2]/(n_1+n_2-2)\}}}$$

Variances for differences of binomial proportions

Pooled

$$\left[\frac{(n_1p_1+n_2p_2)}{(n_1+n_2)}\right]\left[1-\frac{(n_1p_1+n_2p_2)}{(n_1+n_2)}\right]\left[\frac{1}{n_1}+\frac{1}{n_2}\right]$$

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 $\sigma^2/\Sigma(x_i-\bar{x})^2$

Chi-squared Statistic

Σ(Observed − Expected)^2/Expected, with degrees of freedom depending on the hypothesis tested.

END OF PAPER