This paper is not to be removed from the Examination Halls

UNIVERSITY OF LONDON

279 004b ZB
990 004b ZB
996 D04b ZB

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, B is independent of C and A is independent of C, then the three events are independent.

   ii. If A and B are two independent events, then $\Pr(A|B) = \Pr(A)$.

   iii. The average of two unbiased estimators for the same parameter is also unbiased for the same parameter.

   iv. The average of two unbiased estimators for the same parameter is a better estimator of the parameter than either of them.

   v. Let T be an unbiased estimator for the parameter θ. Then, T^2 is not an unbiased estimator for the parameter θ^2.

   vi. Two fair dice are thrown. The events $A = \{$the dice show a total for 9, 10 or 12$\}$ and $B = \{$the dice show a total score that is odd$\}$ are independent.

(16 marks)

b) Briefly explain the meaning of the following.

   i. Extrapolation

   ii. Type I error

(4 marks)

c) A random variable with expected value 1.2 and $\Pr(X = 0) = 0.1$ and $\Pr(X = 2) = 0.2$ takes one more value besides 0 and 2. Find its variance.

(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 0 and 1. The joint probabilities for each pair are given by the following table.

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	0.1	0.4	0.1
$Y = 1$	0.1	0.1	0.2

Let $Z = \min(X, Y)$ be the smaller of the two variables. Find $E(Y)$, $E(Y|X = 1)$, $E(Z)$ and $\mathrm{Cov}(X, Z)$.

(10 marks)

(question continues on next page)

c) The following observations are supposed to be observations of a Poisson random variable: 3,0,2,0,1,1,4,1,1,2,0,2,0,3,1. Estimate the mean of the Poisson distribution and use a goodness of fit test, where observations of value 3 and above are grouped together, to test whether they are indeed from a Poisson distribution.

(6 marks)

SECTION B

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

2. a) A man suffering from headaches is trying two different pills. Every morning he decides on the pill he will take (if any). The probability that he takes no pill is 0.4, the probability that he will take pill A only is 0.3, the probability that he will take pill B only is 0.2 and the probability that he will take both pills is 0.1. If he gets no pill he will develop a headache with probability 0.5, if he gets pill A only, he will develop a headache with probability 0.4, if he gets B only, he will develop a headache with probability 0.3 and if he gets both pills, he will develop a headache with probability 0.2.

   i. What is the probability that he develops a headache on any day?
   ii. Given that he developed a headache yesterday, what is the probability he took pill A (alone or together with pill B)?
   iii. Are the events "he took pill A alone" and "he did not develop a headache" independent?

(13 marks)

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

$$f(x) = \frac{3x^2 + 1}{2}$$

defined over the region $0 \leq x \leq 1$. Find $\Pr(X < 0.5 | X > 0.25)$, $E(X)$, $\mathrm{Var}(X)$ and $\mathrm{Cov}(X, X^2)$.

(17 marks)

PLEASE TURN OVER

3. Three tests A, B and C are sat by 5 different students. Their grades are given by the following table

	A	B	C
Student 1	52	48	45
Student 2	54	45	30
Student 3	61	60	58
Student 4	43	42	50
Student 5	95	90	99

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

(8 marks)

b) Is there a significant difference between the performances of different students?

(2 marks)

c) Is there a significant difference in difficulty between the tests?

(3 marks)

d) Give a 95% confidence interval for the difference one would expect between the grades of students 3 and 4.

(3 marks)

e) Suppose now that the scores for test C are not available and one has the scores of the 5 students for tests A and B. Construct a two sided 90% confidence interval for the difference in grades between the two tests.

(7 marks)

f) Continuing from (e) use an appropriate hypothesis test to find out if there is significant evidence that the two tests differ in difficulty.

(4 marks)

g) Discuss any differences between your conclusions in parts (c), (e) and (f).

(3 marks)

4. a) Suppose you are given four pairs of observations (x_i, y_i) such that

$$y_1 = \alpha + \beta x_1 + \varepsilon_1,$$

$$y_2 = -\alpha - \beta x_2 + \varepsilon_2$$

and

$$y_i = \beta x_i + \varepsilon_i$$

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

(11 marks)

b) The table below gives the geographical latitude in degrees, the distance from the coast in miles and the average annual rainfall in inches for various weather stations in California.

Station	Latitude	Distance from Coast	Rainfall
Eureka	40.8	1	30.6
Fort Bragg	39.4	1	37.5
San Francisco	37.8	5	21.8
San Jose	37.4	28	14.2
Salinas	36.7	12	13.8
Bakersfield	35.4	75	6
Santa Barbara	34.4	1	18
Los Angeles	34.1	16	15
San Diego	32.7	5	9.9

Fit a straight line through these data using geographical latitude as the dependent variable and rainfall as the independent variable.

(7 marks)

c) A Mexican town is just across the border at a geographical latitude of 32.9 degrees. Give a 95% confidence interval for the annual rainfall there.

(5 marks)

d) How reliable is your answer in (c)?

(3 marks)

e) The full model including the distance from the sea is

$$\text{"Rainfall"} = -72 + 2.57\text{"Latitude"}. - 0.2\text{"Distance from the coast"}$$

Carefully interpret this equation. **(4 marks)**

PLEASE TURN OVER

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

	$X=0$	$X=1$	$X=2$
$Y=0$	0.1	0.15	0.1
$Y=1$	0.1	0.1	0.1
$Y=2$	0.2	0.05	0.1

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 $E(V|U = 2)$

(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

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

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

Chi-squared Statistic

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

END OF PAPER