UNIVERSITY OF
CAMBRIDGE

ECT1: ECONOMICS TRIPOS PART I

PAPER 3: QUANTITATIVE METHODS IN ECONOMICS

Friday 6 June, 2025   13:30-16:30

This paper is divided into four sections.
Section A answer **four** compulsory questions.
Section B answer **one** question out of two questions.
Section C answer **four** compulsory questions.
Section D answer **one** out of two questions.

Answers from Section A and B (Mathematics) must be written in one booklet; answers from Section C and D (Statistics) must written in a separate booklet. Write the letters of the sections on each cover sheet.

Section A and C will each carry 30% of the marks for this paper. Section B and D each carry 20% of the marks.

Each question within each section carries equal weight.

Write your **Blind Grade Number** (not your name) on your answers.

Candidates are asked to note that there may be a reduction in marks for scripts with illegible handwriting.

If you identify an error in this paper, please alert the Invigilator, who will notify the **Examiner**. A general announcement will be made if the error is validated.

STATIONERY REQUIREMENTS
20 Page booklet $\times 2$
Rough work pads
Tags

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAMINATION
Calculator – students are permitted to bring an approved calculator.
New Cambridge Elementary Statistical Tables

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator.

SECTION A

A1 Let S be a subset of the two-dimensional Euclidean space, that consists of all pairs (x, y), such that both x and y are rational numbers.

(a) Carefully explain whether S is a convex set or not.

(b) Is the set S closed, open, or neither? Justify your answer.

A2 Consider a 100-dimensional Euclidean space $\mathbb{R}^{100}$ with standard basis vectors $e_1, \ldots, e_{100}$. Suppose that $A^{(k)}$, $k = 1, \ldots, 100$ are 100×100 matrices representing the following linear transformations: the transformation corresponding to $A^{(k)}$ sends e_1 to e_k, e_k to e_1, and leaves all the other basis vectors unchanged. In particular, $A^{(1)}$ corresponds to a linear transformation that leaves all the basis vectors unchanged, so $A^{(1)}$ is simply the identity matrix I_{100}.

(a) What is the value of the element in the second row and the fifth column of matrix $A^{(2)}$? Justify your answer.

(b) Let $B = A^{(100)}A^{(99)}...A^{(2)}A^{(1)}$, that is, B is the product of all the matrices $A^{(k)}$, $k = 100, 99, \ldots, 2, 1$. What is the first column of B? Explain.

A3 Consider function $f(x, y) = \sin(2x - y)$, with $x \in \mathbb{R}$ and $y \in \mathbb{R}$.

(a) Find and report all critical points of this function.

(b) Classify these critical points, that is, indicate which of the critical points are local maxima, which of them are local minima, and which of them are saddle points. Justify your answer.

A4 In econometrics, the logistic cumulative distribution function is defined as

$$F(x) = \frac{e^x}{1 + e^x}, \qquad \text{for any } x \in \mathbb{R}.$$

(a) Let $P_1(x) = a + bx$ be the linear Taylor approximation of $F(x)$ around $x = 0$. Find the values of a and b and evaluate $P_1(x)$ at $x = 1$.

(b) Suppose that you know that $\max_{x\in[0,1]} |F''(x)|$ is achieved at $x = 1$. Explain how this fact can be used to find an upper bound on $|F(1) - P_1(1)|$. Find and report such an upper bound.

SECTION B

B5 A consumer derives utility from consuming two goods. Their utility function has the following form

$$U(x, y; a) = -a(x-5)^2 - (y-10)^2,$$

where x is the consumption of the first good, y is the consumption of the second good, and $a > 0$ is a parameter. The unit prices of the first and the second goods are p_1 and p_2, respectively. The consumer's income that they can spend on the two goods is $m > 0$.

(a) Use the Lagrange method to find the optimal values of x and y when $a = p_1 = p_2 = 1$ and $m = 9$. Explain why your solution is indeed delivering the maximum utility.

(b) Interpret the Lagrange multiplier corresponding to the optimal solution you found in (a).

(c) Suppose that the value of parameter a changes from 1 to 1.01. Then, by how much, approximately, will the optimal utility's value change? Justify your answer.

(d) Suppose now that parameter a is a function of t. Specifically, let $a = a(t)$ so that $a(0) = 1$ and

$$\frac{\mathrm{d}}{\mathrm{d}t}a(t) = -ta(t).$$

What are the optimal values of x and y if $p_1 = p_2 = 1$, $m = 9$ as above, but $a = a(6)$? Justify your answer. (Hint: you may use the approximation $e^{-18} \approx 0$.)

(TURN OVER)

B6 Suppose that the marginal utility of income for a person is given by

$$\frac{\mathrm{d}}{\mathrm{d}y}U(y) = (1+y)e^{-y},$$

where U is the utility function and y is the income level. Suppose further that $U(0) = 0$.

(a) Find the value of $U(1)$. Show your work.

(b) Let $f(y) = \frac{\mathrm{d}}{\mathrm{d}y}U(y)$. Show that $\frac{\mathrm{d}}{\mathrm{d}y}f(y) = e^{-y} - f(y)$. Using this identity, demonstrate that

$$f^{(k)}(y) = (-1)^{k-1}e^{-y} - f^{(k-1)}(y) \quad \text{for } k = 1, 2, 3, \ldots,$$

where $f^{(k)}(y)$ denotes the k-th order derivative of $f(y)$, and $f^{(0)}(y) = f(y)$.

(c) Using the identity established in (b) or otherwise, derive the Taylor series for $f(y)$ around $y = 0$. In particular, derive an explicit form of the n-th term of the Taylor series (the one involving the n-th derivative of $f(y)$) for a general positive integer n.

(d) Suppose that you would like to approximate the marginal utility at $y = 5$ by the value of the second order Taylor polynomial around $y = 0$. Using the Taylor theorem, provide an upper bound on the absolute value of the error of such an approximation.

SECTION C

C7 Two players A and B play a "best of three sets" tennis match. The match ends when a player has won two sets (sets cannot be drawn). Suppose that $\Pr(A \text{ wins a set}) = p$. Which of the following two outcomes is more likely: (a) the match will last two sets, (b) the match will last three sets. State clearly any further assumptions you make and the (mathematical) reasons for your answer.

C8 Independent variables X_1 and X_2 satisfy $E(X_1) = \mu$, $E(X_2) = 2\mu$, $Var(X_1) = \sigma^2$, $Var(X_2) = 2\sigma^2$. You have one observation of X_1 and one of X_2. An estimator of μ is defined as $\tilde{\mu} = w_1 X_1 + w_2 X_2$, where w_1 and w_2 are some constants. What values of w_1 and w_2 would minimise the variance of $\tilde{\mu}$ whilst making $\tilde{\mu}$ unbiased?

C9 Suppose that $x_i, i = 1, \ldots, n$, are i.i.d. random variables with mean μ and variance σ^2. Further, suppose that $n = 64$ and that you observe the following summary statistics:

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} = 0.6,$$

$$\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 = s^2 = 14.2.$$

You are employed to test the null hypothesis $H_0 : \mu = 0$ against the alternative $H_1 : \mu = 1$ using a one-tailed test. Your employer cares mainly about Type II errors and they are willing to incur a 10% probability of such an error. Explain how you would design a test of H_0 vs H_1 that meets this requirement and calculate the size of this test.

C10 An analyst writes down the following linear regression equation

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \tag{1}$$

where Y_i denotes the random variable of interest, X_i is a binary variable, and ε_i is an error term.

(a) Show that α and β in equation (1) can be written, respectively, as $E[Y_i|X_i = 0]$ and $E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$. Clearly state all assumptions that you are making.

(b) If the objective is to test the null hypothesis

$$H_0 : \beta = 0 \qquad \text{against} \qquad H_1 : \beta \neq 0,$$

comment on the observation that a regression equation may not be needed. What is the advantage of utilising a regression model?

(c) Show that β in equation (1) can be written as $\frac{\mathrm{Cov}(Y_i, X_i)}{\mathrm{Var}(X_i)}$.

(d) We now consider an extension of equation (1) by including the variable Z_i.

$$Y_i = \alpha + \beta X_i + \theta Z_i + \varepsilon_i$$

If the analyst was unaware of this change and proceeds to estimate β based on equation (1), evaluate the statement that the ordinary least squares estimator $\widehat{\beta}$ will be biased unless $\mathrm{Cov}(X_i, Z_i) = 0$.

SECTION D

D11 A discrete random variable X has the following probability mass function

x	$\Pr(X = x)$
-1	p
0	$1 - 2p$
$+1$	p

where $0 \leq p \leq \frac{1}{2}$ is a parameter.

(a) Calculate $E(X)$ and $Var(X)$.

(b) Calculate $E(|X|)$ and $E(I(X > 0))$, where $|X|$ denotes absolute value and $I(X > 0)$ is an indicator function, that is, $I(X > 0) = 0$ if $X \leq 0$ and $I(X > 0) = 1$ if $X > 0$.

You have an i.i.d. sample $X_1, X_2, \ldots, X_N$ of size N drawn from this distribution and wish to estimate p. Two estimators are proposed

$$\hat{p} = k_1 \frac{\sum_{i=1}^{N} |X_i|}{N}$$

and

$$\tilde{p} = k_2 \frac{\sum_{i=1}^{N} I(X_i > 0)}{N},$$

where k_1 and k_2 are constants.

(c) Find the values of k_1 and k_2 that ensure that estimators $\hat{p}$ and $\tilde{p}$ are unbiased.

(d) For the values of k_1 and k_2 found in (c), calculate the mean squared error of the corresponding estimators $\hat{p}$ and $\tilde{p}$.

(e) Which estimator is preferred? Justify your answer.

D12 A new drug is being tested on a group of 80 people (40 men and 40 women). Of those, half are randomly selected and given the drug, $D = 1$, and the other half are given a placebo, $D = 0$. The table below reports recovery rates for men and women. $R = 1$ means that the individual recovered, and $R = 0$ means that the individual did not recover.

Table: Men and Women

Men ($M = 1$)			Women ($M = 0$)		
	$R = 1$	$R = 0$		$R = 1$	$R = 0$
$D = 1$	18	12	$D = 1$	2	8
$D = 0$	7	3	$D = 0$	9	21

(a) Based on the statistics in the table, estimate $\Pr(R = 1|D = a, M = b)$ for all possible combinations of a and b, that is, $(a, b) \in \{(1, 1), (0, 1), (1, 0), (0, 0)\}$. Interpret the results.

(b) Conduct a test of the equivalence of recovery rates:
$H_0 : \pi_1 = \pi_0 \qquad H_1 : \pi_1 \neq \pi_0,$ where
π_1 = population probability of recovery for men given drug,
π_0 = population probability of recovery for men given placebo.

(c) Use your results from (a) to estimate $\Pr(R = 1|D = 1)$ and $\Pr(R = 1|D = 0)$. Interpret the results.

(d) For each combination of gender and treatment we write a conditional probability of recovery:

$$\begin{aligned}
\Pr(R_i &= 1|M_i = 0, D_i = 0) = \alpha \\
\Pr(R_i &= 1|M_i = 0, D_i = 1) = \alpha + \beta_1 \\
\Pr(R_i &= 1|M_i = 1, D_i = 0) = \alpha + \beta_2 \\
\Pr(R_i &= 1|M_i = 1, D_i = 1) = \alpha + \beta_1 + \beta_2 + \beta_3
\end{aligned}$$

where R_i is an indicator of recovery for individual i, D_i denotes whether a drug has been taken by individual i, and M_i denotes whether individual i is a male.

i) Demonstrate how these four conditional probabilities can be combined into a single regression equation.

ii) Formulate the hypothesis that women who take the drug are equally likely to recover as men who take the drug in terms of the coefficients in your regression equation from i).

(END OF PAPER)