University of Nottingham Malaysia

BUSINESS SCHOOL

A LEVEL 2 MODULE, SPRING SEMESTER 2023-2024

INTRODUCTORY ECONOMETRICS

Time allowed ONE Hour THIRTY Minutes

Candidates may complete the front cover of their answer book and sign their desk card but must NOT write anything else until the start of the examination period is announced

Answer ALL questions in Section A and TWO questions from Section B.

Section A accounts for 30% of the total marks available for this examination.

Section B questions carry equal weight of 35% each.

Figures following each part indicate the marks available for that part.

Only calculators from Approved Calculators Lists A are permitted in this examination

Approved Calculators Lists A

Basic Models	***Scientific Calculators***
Aurora HC133	*Aurora AX-582*
Casio HS-5D	*Casio FX82 family*
Deli – DL1654	*Casio FX83 family*
Sharp EL-233	*Casio FX85 family*
	Casio FX350 family
	Casio FX570 family
	Casio FX 991 family
	Sharp EL-531 family
	Texas Instruments TI-30 family
	Texas BA II+ family

Dictionaries are not allowed with one exception. Those whose first language is not English may use a standard translation dictionary to translate between that language and English provided that neither language is the subject of this examination. Subject specific translation dictionaries are not permitted.

No electronic devices capable of storing and retrieving text, including electronic dictionaries, may be used.

DO NOT turn examination paper over until instructed to do so

ADDITIONAL MATERIAL: Formula Sheet
Statistical Tables

SECTION A

Answer **all** questions in this section

1. A random variable has *n* observations $(x_1, x_2, \ldots, x_n)$ that are independently and identically distributed with the population mean μ and the population variance σ^2.

   An estimator, $\tau = \frac{\sum_{i=1}^{n} x_i}{n-2}$, is used to estimate the population mean μ.

   (a) Is τ an unbiased estimator for μ? Explain.

   [3 marks]

   (b) Find an expression for the variance of the estimator τ. Comparing it to the variance of a sample mean, which estimator is efficient? Explain your answer.

   [3 marks]

   (c) Is τ a consistent estimator for μ? Explain your answer.

   [4 marks]

2. You invested a proportion of your wealth, ω, in Apple (A) and the remaining proportion $(1-\omega)$ in Boing (B). The return on your two-stock portfolio, P, can be described as:

   $$P = \omega A + (1-\omega)B$$

   where random variables A and B represent the returns from the respective stock.
   The returns on stocks are independent of each other and random, with $E(A) = 0.3$, $E(B) = 0.5$, and the variances of returns are $\text{Var}(A) = 0.5$ and $\text{Var}(Y) = 0.8$.

   (a) Find the fraction of the wealth to be invested in Boing (B) if you want to achieve the expected return of 0.45 from the portfolio.

   [3 marks]

   (b) Find the variance of the return on the portfolio suggested in part (a).

   [2 marks]

   (c) Find the fraction of wealth to be invested in Apple (A) if you want to minimise the variance of the return on the portfolio.

   [5 marks]

3. A quality control test of smart watch batteries indicates that 50% of the batteries lasted 4 hours, 30% lasted 3 hours, and the remaining 20% lasted 2 hours.

   (a) Let X represents the hours lasted. Construct the probability density table and find the population mean and the variance of X.

   [4 marks]

   (b) Random samples of two batteries are drawn with replacement. Construct the probability distribution for the 5 possible values the sample mean.

   [3 marks]

   (c) Using probability distribution for the sample means in (b), calculate the expected value of sample means. What do you observe?

   [3 marks]

SECTION B

Answer any **two** questions from this section

4. In a study on the return to education and the gender gap, a researcher estimated the following regression model (Model 1) using a sample data set of 54 full-time workers.

$$\widehat{AHE} = 16.806 - 2.248FEMALE + 1.624EDU + 0.282(FEMALE \times EDU)$$
$$\text{S.E} \quad (5.6462) \quad (0.1456) \qquad (0.5461) \quad (0.1524)$$

$$R^2 = 0.54$$
$$\text{Residual Sum of Squares (RSS)} = 2307.45$$

where *AHE* is average hourly earnings in US$, *FEMALE* is a dummy variable taking the value 1 for female workers and, 0 otherwise, and *EDU* is years of education. Standard errors are given in parentheses below the corresponding estimates.

(a) Interpret the estimated coefficient of *FEMALE*.

[6 marks]

(b) Explain why the researcher included the interaction term between *FEMALE* and *EDU* in Model 1.

[5 marks]

(c) A researcher comments that the above model is mis-specified because it omits an additional dummy variable that represents male workers. Do you agree with this comment? Explain.

[5 marks]

(d) The researcher has also estimated the following regression model (Model 2) using the same sample data set. Test the null hypothesis that the gender has no effect (both direct and moderating effects) on earnings, using a 1% level of significance. State the null and alternative hypotheses, and your conclusion clearly.

$$\widehat{AHE} = 18.026 + 1.846EDU$$
$$\text{S.E} \quad (5.6628) \quad (0.3612)$$

$$R^2 = 0.32$$
$$\text{Residual Sum of Squares (RSS)} = 2891.88$$

[7 marks]

(e) Based on the conclusion of the hypothesis test conducted in part (d) and the sizes of the estimated coefficients given in Model 1, does it appear that the return to education (estimated coefficient of *EDU*) is similar for both male and female workers? Explain.

[6 marks]

(f) Based on the conclusion of the hypothesis test conducted in part (d), what are the consequences of estimating the regression model (Model 2) provided in part (d)?

[6 marks]

5. The following table shows the results of a regression model that predicts the sales of new residential properties in Malaysia using annual time-series data for the years 1993 to 2022 inclusive (30 observations).

Model 1: OLS, using observations 1993-2022 (T =30)
Dependent variable: *resid*

Variables	***Coefficient***	***p-value***
const	1033.44	0.1486
income	12.76	0.0031
interest	- 10.63	0.1153
inflation	- 4.15	0.1953

Mean dependent var	1894.157	S.D. dependent var	311.345
Sum squared resid	2698682	S.E. of regression	219.523
R-squared	0.828139	Adjusted R-squared	0.5028
rho	0.195388	Durbin-Watson	1.0952

The variables used in the regression model are defined as:

resid = Number of new residential properties sold (in thousands)

income = Real disposable income per capita in Ringgit (in thousands)

interest = interest rate in percent

inflation = inflation rate in percent

(a) Interpret the estimated coefficient of *inflation*.

[6 marks]

(b) Test, at the 10% significance level, whether the interest rate had a negative impact on the sales of new residential properties. State your hypotheses and conclusion clearly.

[6 marks]

(c) Conduct the test of overall significance of the model at the 5% significance level. State your hypotheses and conclusion clearly.

[7 marks]

(d) Another researcher suggests that the variables '*interest*' and '*inflation*' are likely to be highly correlated. Identify the potential issue, its consequences and discuss how it can be solved.

[6 marks]

(e) Test, at the 5% significance level, whether the estimated regression model satisfies the OLS assumption of no autocorrelation of error terms. State the null and alternative hypotheses clearly.

[5 marks]

(f) Another researcher suggested that COVID-19 pandemic in 2020 and 2021 might also have an impact on the sales of new residential properties in Malaysia. Explain how you would control the impact of COVID-19 pandemic in a regression model that estimates the sales of new residential properties.

[5 marks]

6. Using a sample of 200 families. the following regression model was used to estimate the travelling distance of families in Malaysia when they take vacation.

$$\ln km_i = \beta_1 + \beta_2 \ln income_i + \beta_3 \ln age_i + kids_i + \varepsilon_i$$

The variables are defined as:

$\ln km$ = logarithm of travelling distance of a family in kilometres

$\ln income$ =logarithm of a family's annual income (in Malaysian Ringgit 1000)

$\ln age$ =logarithm of the average age of adult members of household

$kids$ =number of children in household

The estimation results are given in the following table.

Variables	Coefficient	Std. Error
const	0.0786	0.6949
$\ln income$	0.9595	0.1215
$\ln age$	0.7949	0.1727
$kids$	-0.0942	0.0315

(a) Interpret the estimated coefficient of the variable '*kids*'.

[6 marks]

(b) Interpret the estimated coefficient of the variable 'ln *income*' (logarithm of family's annual income).

[6 marks]

(c) Test, at the 1% significance level, whether a smaller family (lower number of *kids*) travel more distance when they take vacation. State the null and alternative hypotheses clearly.

[6 marks]

(d) Define heteroscedasticity and explain why the OLS assumption of homoscedasticity might be violated in the model estimated above.

[5 marks]

(e) The researcher suspects that the assumption of homoscedasticity was violated. Explain how the White's general test for heteroscedasticity can be conducted. Describe all the steps including the null and alternative hypotheses of the test clearly.

[7 marks]

(f) Explain the rationales for transforming distance, income, and age variables into logarithmic format.

[5 marks]