University of Nottingham Malaysia

BUSINESS SCHOOL

A LEVEL 2 MODULE, SPRING SEMESTER 2023-2024

INTRODUCTORY ECONOMETRICS

Time allowed ONE Hour THIRTY Minutes

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

ADDITIONAL MATERIAL: Formula Sheet
Statistical Tables

SECTION A

Answer **all** questions in this section

1.	A random variable has n observations $(x_1, x_2, \ldots, x_n)$ 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]	
		$E(\tau) = E\left(\frac{\sum_{i=1}^{n} x_i}{n-2}\right) = \frac{1}{n-2}[E(x_1) + E(x_2) + \ldots. + E(x_n)]$ Since $E(x_i) = \mu$ $E(\tau) = \frac{1}{n-2} n\mu = \left(\frac{n}{n-2}\right)\mu \neq \mu$ An estimator is an unbiased estimator for the population parameter if the expected value of estimates is equal to the population parameter. Therefore, τ is not an unbiased estimator for μ.	0.5 mark 0.5 mark 1 mark 1 mark
	(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]	
		$Var(\tau) = Var\left(\frac{\sum_{i=1}^{n} x_i}{n-2}\right) = \left(\frac{1}{n-2}\right)^2 [var(X_1) + var(X_2) + \ldots. + var(X_n)]$ Since $Var(x_i) = \sigma^2$ because x_i is i.i.d. $Var(\tau) = \left(\frac{1}{n-2}\right)^2 n\sigma^2 = \frac{n}{(n-2)^2}\sigma^2$ Since $\frac{n}{(n-2)^2} < \frac{1}{n}$, $Var(\tau) < Var(\bar{x})$. But the estimator τ is not an unbiased estimator. It cannot be compared to the variance of sample mean for efficiency.	0.5 mark 0.5 mark 1 mark 1 mark
	(c)	Is τ a consistent estimator for μ? Explain your answer. [4 marks]	
		An estimator is said to be consistent if the estimator is unbiased (or asymptotically unbiased), and its variance of reduces to zero when the sample size approaches infinity. When $n \to \infty$, $E(\tau) = \frac{n}{n-2}\mu = \frac{\infty}{\infty}\mu = \mu$ $(\because \infty - 2 \approx \infty)$ Therefore τ is an asymptotically unbiased estimator. When $n \to \infty$, $Var(\tau) = \frac{n}{(n-2)^2}\sigma^2 = \frac{\infty}{(\infty-2)^2}\sigma^2 \approx \frac{\sigma^2}{\infty} = 0$ Therefore τ is a consistent estimator.	1 mark 1 mark 1 mark 1 mark

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 Var(A) = 0.5 and 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]	
		$E(P) = E[\omega A + (1-\omega)B] = 0.45$ $\omega E(A) + (1-\omega)E(B) = 0.45$	1 mark
		$\omega 0.3 + (1-\omega)0.5 = 0.45$ $0.4\omega + 0.5 - 0.5\omega = 0.45$	1 mark
		$\omega = 0.25$ $1-\omega = 0.75$	1 mark
	(b)	Find the variance of the return on the portfolio suggested in part (a). [2 marks]	
		$Var(P) = Var[0.25A + 0.75B]$	1 mark
		$= (0.25^2 \times 0.5) + 0.75^2 \times 0.8 = 0.48125$	1 mark
	(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]	
		$Var(P) = Var[\omega A + (1-\omega)B]$ $= \omega^2 Var(A) + (1-\omega)^2 Var(B)$ $= (\omega^2 \times 0.5) + [(1 - 2\omega + \omega^2) \times 0.8]$ $= 0.5\omega^2 + 0.8 - 1.6\omega + 0.8\omega^2$ $= 1.3\omega^2 - 1.6\omega + 0.8$	2 marks
		To find ω that minimise $Var(P)$ $var(P) = 1.3\omega^2 - 1.6\omega + 0.8$ $\frac{d\,var(P)}{d\omega} = 2.6\omega - 1.6 = 0$ $\omega = 0.6154$	2 marks
		$\frac{d^2\,var(P)}{d\omega^2} = 2.6 > 0; \text{Minimum}$	1 mark
		To minimise the variance of the return on portfolio, 61.54% of wealth must be invested in Apple (A).	

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]

X	*P(X)*
4	0.5
3	0.3
2	0.2

1 mark

$\mu = E(X) = (4 \times 0.5) + (3 \times 0.3) + (2 \times 0.2) = 3.3$ 1 mark

$\sigma^2 = Var(X) = (4^2 \times 0.5) + (3^2 \times 0.3) + (2^2 \times 0.2) - (3.3^2) = 11.5 - 10.89 = 0.61$ 2 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]

sample outcome	mean	sample outcome	mean	sample outcome	mean
4,4	4	3,4	3.5	2,4	3
4,3	3.5	3,3	3	2,3	2.5
4,2	3	3,2	2.5	2,2	2

1 mark

Probability distribution of sample mean

$(\bar{X})$	4	3.5	3	2.5	2
P($\bar{X}$)	=0.5×0.5 =0.25	(0.5×0.3) + (0.3×0.5) =0.3	(0.5×0.2) + (0.3×0.3) + (0.2×0.5) =0.29	(0.3×0.2) + (0.2×0.3) =0.12	=0.2×0.2 =0.04

2 marks

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

[3 marks]

$E(\bar{X}) = (4 \times 0.25) + (3.5 \times 0.3) + (3 \times 0.29) + (2.5 \times 0.12) + (2 \times 0.04) = 3.3$ 2 marks

Observation: $E(\bar{X}) = \mu$ 1 mark

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.

$$\widehat{AHE} = 16.806 - 2.248 FEMALE + 1.624 EDU + 0.282(FEMALE \times EDU)$$

S.E (5.6462) (0.1456) (0.5461) (0.1524)

$R^2 = 0.54$

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]

Students should provide an accurate interpretation of the estimated coefficient -2.248

Example: Holding the years of education constant, average hourly earnings of a female employee is US$2.248 less than a male employee.

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

[5 marks]

Students should explain that the relationship between education and averaging hourly earnings is moderated by gender. To capture the moderating effect of the gender, a slope-dummy is included.

(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]

Incorrect statement. Students should explain that if the data fall naturally into s subgroups, then s-1 dummy variables can be created, and the value of the intercept represents the reference group (male workers in this specification).

(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.

$$\widehat{AHE} = 18.026 + 1.846 EDU$$

S.E (5.6628) (0.3612)

$R^2 = 0.32$

Residual Sum of Squares (RSS) = 2891.88

[7 marks]

Conduct a F-test for joint-significance. Calculated F-test statistic = 5.25 and the critical value for DF1=2, DF2=50 at 1% significance level is 5.06. The null hypothesis should be rejected. Students should set the null and alternative hypotheses accurately, state their conclusion clearly.

	(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]
		No. Students should calculate the estimated coefficients of 'EDU" for both male (1.624) and female (1.906) and conclude that the return to education is higher for female workers.
	(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]
		Student should identify the issue as omitting variables bias and describe its consequences in the given context.

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]
		Students should provide an accurate interpretation of the estimated coefficient -2.248 *Example: Holding other variables constant, 1 percentage **point** increase in inflation rate would lead to an average decrease of 4150 units in the sales of new residential properties in Malaysia.*
	(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]
		This is a one-tailed hypothesis test. The observed p-value for the two-tailed test is 0.1153. Since p-value for the one-tailed test 0.0577 is smaller than the level of significance 10% (0.1), the null hypothesis is rejected. Students should set the null and alternative hypotheses accurately, state their conclusion clearly.
	(c)	Conduct the test of overall significance of the model at the 5% significance level. State your hypotheses and conclusion clearly. [7 marks]
		Conduct an F-test for overall significance. Calculated F-test statistic = 41.8 and the critical value for DF1=2, DF2=26 at 5% significance level is 3.37. The null hypothesis should be rejected. Students should set the null and alternative hypotheses accurately, state their conclusion clearly.
	(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]
		Due to a high correlation between 'interest' and 'inflation', there might be a problem of multicollinearity. *Students should discuss the consequences of multicollinearity.* • *The regression estimates are unbiased and consistent, and their standard errors are correctly estimated (BLUE), although these tend to be large.* • *It is likely that the usual t-tests are not reliable* • *Estimators may be very sensitive to the addition or deletion of a few observations* • *There are clearly problems of interpretation of the contribution of individual explanatory variable.* *And the following remedies for Multicollinearity. A good answer should include how to transform the variable.* • *Do nothing (OLS estimators are still BLUE)* • *Drop a redundant variable(s) (but be mindful of model specification error if a relevant variable is dropped)* • *Increase the sample size* • *Transform the variables*
	(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]
		DL=1.143; DU=1.652. Since DW of 1.0952 lies between 0 and DL, the model suffers positive autocorrelation. Students should set null and alternative hypotheses clearly and provide an accurate conclusion.
	(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]
		Students should explain how including dummy variables for the pandemic years can control the impact of COVID-19 pandemic using an example of correct model specification.

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]
		Students should provide an accurate interpretation of the coefficient in a log-linear model. *Example: Holding other variables constant, an additional kid in a family would lead to an average of 9.42% decrease in travelling distance of a family in Malaysia when they take vacation.*
	(b)	Interpret the estimated coefficient of the variable 'ln *income*' (logarithm of family's annual income). [6 marks]
		Students should provide an accurate interpretation of the coefficient in a log-log model. *Example: Holding other variables constant, 1% increase in family's annual income would lead to an average of 0.9595% increase in travelling distance of a family in Malaysia when they take vacation.*
	(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]

		This is a one-tailed hypothesis test. The t-statistic is –2.9874. Since the absolute value of the test statistic is greater than the critical value of 2.345 (approximate based on n=200), the null hypothesis is rejected. Students should set the null and alternative hypotheses accurately, state their conclusion clearly.
	(d)	Define heteroscedasticity and explain why the OLS assumption of homoscedasticity might be violated in the model estimated above. [5 marks]
		Students are expected to define heteroscedasticity correctly. They should explain that the assumption of Homoscedasticity is often violated when using cross-sectional data as the size of the economic unit, such as income, becomes larger, there is more uncertainty associated with the outcomes y (in this case the travelling distance).
	(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]
		Students should describe the steps of White's general test for heteroscedasticity with appropriate hypotheses, correct specification for auxiliary regression model, associated chi-square critical value, decision rule and the example of conclusion.
	(f)	Explain the rationales for transforming distance, income, and age variables into logarithmic format. [5 marks]
		Students should explain that the variables have the characteristic that they are positive and often have distributions that are positively skewed, with a long tail to the right. The relationship between distance and income could also be non-linear in coefficients.