University of Nottingham Malaysia

BUSINESS SCHOOL

A LEVEL 2 MODULE, SPRING SEMESTER 2022-2023

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

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.	Suppose that a random sample consists of three observations, X_1, X_2 and X_3, is collected from a population with mean μ, and variance σ^2. These observations are independently and identically distributed. To estimate the population mean μ, consider the following weighted estimator τ: $$\tau = \frac{1}{5}X_1 + \frac{2}{5}X_2 + \frac{3}{5}X_3$$		
	(a)	Find the expected value of τ and explain whether τ is an unbiased estimator. [3 marks]	
		$E(\tau) = E\left(\frac{1}{5}X_1 + \frac{2}{5}X_2 + \frac{3}{5}X_3\right) = \frac{1}{5}E(X_1) + \frac{2}{5}E(X_2) + \frac{3}{5}E(X_3)$	1
		Since X_i a random variable $E(X_i) = \mu$	1
		$E(\tau) = \frac{6}{5}\mu$	0.5
		Since the expected value of the estimator does not equal to the population mean, the estimator is not an unbiased estimator.	0.5
	(b)	How could you adjust the estimator so that it becomes unbiased? Explain. [2 marks]	
		Students should demonstrate that multiplying τ with $\frac{5}{6}$ would make $E(\theta) = \mu$,	1
		and explain that now the estimator unbiased because its expected value equals to the population parameter.	1
	(c)	Find an expression for the variance of the adjusted estimator you suggested in (b). Compare it to the variance of a sample mean with the same sample size of 3. Which estimator is efficient? Explain your answer. [5 marks]	
		$var\left(\frac{5}{6}\tau\right) = var\left(\frac{1}{6}X_1 + \frac{2}{6}X_2 + \frac{3}{6}X_3\right) = \frac{1}{36}var(X_1) + \frac{4}{36}var(X_2) + \frac{9}{36}var(X_3)$	1
		Since X_i is an i.i.d $var(X_i) = \sigma^2$and their covariances are zero.	1
		$var(\theta) = \frac{14}{36}\sigma^2$	1
		$var(\bar{X}) == \frac{\sigma^2}{n} = \frac{\sigma^2}{3}$	1
		Since $var(\bar{X}) = \frac{\sigma^2}{3} < var(\theta) = \frac{14\sigma^2}{36}$, the sample mean is a efficient estimator.	1

2. A random variable *X* is normally distributed with mean 4 and variance 2. Another random variable *Y* is also normally distributed with mean 6 and variance 4. The covariance between *X* and *Y* is 3. Consider two new random variables $V=4X+3Y$ and $W=2X-4Y$.

(a) Compute $E(V)$ and $E(W)$. [2 marks]

$E(V) = E(4X + 3Y) = 4E(X) + 3E(Y) = (4 \times 4) + (3 \times 6) = 34$ 1

$E(W) = E(2X - 4Y) = 2E(X) - 4E(Y) = (2 \times 4) - (4 \times 6) = -16$ 1

(b) Compute Var(V) and Var(W) [4 marks]

$Var(V) = Var(4X + 3Y) = 4^2\, Var(X) + 3^2\, Var(Y) + (2 \times 4 \times 3)\, Cov(X,Y)$ 1

$= (16 \times 2) + (9 \times 4) + (24 \times 3) = 140$ 1

$Var(W) = Var(2X - 4Y) = 2^2\, Var(X) + 4^2\, Var(Y) - (2 \times 2 \times 4)\, Cov(X,Y)$ 1

$= (4 \times 2) + (16 \times 4) - (16 \times 3) = 24$ 1

(c) Compute correlation coefficient between *X* and *V.* Give your interpretation. [4 marks]

$\text{corr}(X,V) = \dfrac{cov(X,V)}{\sqrt{Var(X).Var(V)}}$

$cov(X,V) = Cov(X, 4X + 3Y) = 4\, Cov(X,X) + 3\, Cov(X,Y) = (4 \times 2) + (3 \times 3) = 15$ 2

$\text{corr}(X,V) = \dfrac{15}{\sqrt{2 \times 140}} = 0.896$ 1

These two variables have a strong and positive linear relationship. 1

3. Among a group of nursery students, 40% are 3 years old and 60% are 5 years old.

(a) Find the expected value and the variance of students' age. [3 marks]

X	*f*(*X*)
3	0.4
5	0.6

$E(X) = (3 \times 0.4) + (5 \times 0.6) = 4.2$ 1

$Var(X) = [(3^2 \times 0.4) + (5^2 \times 0.6)] - 4.2^2 = 0.96$ 2

(b) Random samples of two students are drawn with replacement. Construct a probability distribution table for sample means and find the expected value of sample means. [4 marks]

X_i	$\bar{X}$	$f(\bar{X})$

2

		3,3 | 3 | 0.16 3,5 | 4 | 0.24 5,3 | 4 | 0.24 5,5 | 5 | 0.36 $E(\bar{X}) = (3 \times 0.16) + (4 \times 0.48) + (5 \times 0.36) = 4.2$	2
	(c)	Find the variance the sample means. What is your observation? [3 marks]	
		$Var(\bar{X}) = [(3^2 \times 0.16) + (4^2 \times 0.48) + (5^2 \times 0.36)] - (4.2)^2 = 0.48$ Observation: $Var(\bar{X}) = 0.48 = \frac{0.96}{2} = \frac{\sigma^2}{n}$	2 1

SECTION B

Answer any **two** questions from this section

4. The following table shows the results of a regression that estimate the percentage of household budget spent on tobacco in Belgium using the number of kids and adults in the household as independent variables.

```
Model 1: OLS, using observations 1-300
Dependent variable: share

                 Coefficient   P-values (2-tailed test)
      const      3.7899        0.0001
      nkids      -0.2184       0.0606
      nadults    -0.5067       0.0090

 Sum squared residual  2177.816    Sum squared total   3107.614
 R-squared             0.2992      Adjusted R-squared  0.2968
```

The variables are defined as follows:

'share' – the percentage of household budget spent on tobacco,

'nkids' – number of children in the household,

'nadult' – number of adults in the household.

(a) Interpret the estimated coefficient of 'nadults'. [6 marks]

Students should provide an accurate interpretation of the estimated coefficient correctly stating that change in the dependent variable is percentage point.

(b) Test, at the 5% significance level, whether the number of children in a household is negatively related to the percentage of household budget spent on tobacco. State the null and alternative hypotheses clearly. [6 marks]

Students should transform the given p-value for a one-tailed test and conduct the hypothesis testing. They should reject the null hypothesis and provide a correct conclusion since p-value for one-tailed test 0.0303 is less than the level of significance 5%. Students should show all steps of the hypothesis testing.

(c) Conduct an F-test for overall significance of the above regression model at the 5% significance level. State the null and alternative hypotheses clearly. [6 marks]

Students should calculate the required test-statistic using the given value of R-squared or SSE and SST. Students must set hypotheses correctly.

F-test stat=63.4 > $F_{0.05,\,2,\,297}$ =3.03; Reject the null.

They should reject the null hypothesis and provide a correct conclusion. Students should show all steps of the hypothesis testing.

(d) Would you expect the variance of error terms in the above regression model to be

		homoscedastic (constant)? Explain. [5 marks]
		Students should explain that variation in the dependent variable is likely to be higher when the value of the explanatory variables. Since the data set employed is a cross-sectional data involving different household with different number of kids and adults, the regression model is expected to be heteroscedastic.
	(e)	Using the squared residuals (uhat^2) obtained from the estimation of the original regression as the dependent variable, an auxiliary regression was estimated to conduct a White's general test. The R-squared of the auxiliary model is 0.0442. Conduct the test for heteroscedasticity at the 1% significance level. State the null and alternative hypotheses clearly. [6 marks]
		Students should set appropriate null and alternative hypotheses for White's test and calculate the test statistic. Then, identify the Chi-square critical value using the given statistics and provide their decision and conclusion. Test statistics = 13.26 $\chi^2_{5,0.01} = 15.086$ Decision: Do not reject the null; Conclusion: No heteroscedasticity Students should show all steps of the hypothesis testing.
	(f)	Based on statistical relationships presented in the above regression results, can you conclude that the household spending on tobacco is influenced by the included independent variables? Explain. [6 marks]
		Students should explain that although regression analysis deals with the dependence of one variable on other variables, a statistical relationship in itself cannot logically imply causation. Students should explain that the model specification must base on a priori condition or theoretical considerations to ascribe causality.

5. A researcher estimated two regression models to examine the determinants of 'wage' (hourly wage in Malaysia Ringgit) using a sample of 55 individuals. The independent variable 'educ' represents the years of schooling; 'gender' is a dummy variable equals to 1 for female employee and 0 for male employee; 'exper' and 'expersq' represent the years of experience and the square of the years of experience, respectively. The results are presented in the table below.

Variables	**Model A**	**Model B**
const	−11.6243	−20.6651
	(0.014)	(0.009)
educ	2.1425	1.7844
	(0.001)	(0.002)
gender	−2.6243	−3.6034
		(0.092)
exper	-	3.4488
		(0.047)
exper-sq	-	−0.0479
		(0.021)
Residual Sum of Squares (SSE)	8514.23	7501.47
R^2	0.3544	0.4098

Note: Estimated coefficients of each model are given in the table. Numbers in parentheses are corresponding P-values (2-tailed test). *SSE* represents the Residual (Error) Sum of Squares.

(a) Interpret the estimated coefficient of 'gender' in Model B. [6 marks]

Students should provide an accurate interpretation of the estimated coefficient of dummy variable 'gender'.

(b) Interpret the estimated coefficient of 'educ' in Model B. [5 marks]

Students should provide an accurate interpretation of the estimated coefficient the variable 'educ'.

(c) Explain why the researcher included the square of experience in Model B. [4 marks]

Students should explain that the relationship between wage and experience variables could be non-linear. To capture the effect of non-linearity in variable, exper-sq is included.

(d) Using the results of Model B, test the hypothesis that on average female employees earn less than their male counterparts at the 5% significance level. State the null and alternative hypotheses clearly. [6 marks]

Since one-tailed p-value for 'age' is 0.046 and smaller than 0.05 (5%), there is enough evidence to reject the null hypothesis. It can be concluded that on average female employees earn less than their male counterparts. Students should show all steps of the hypothesis testing. Students should show all steps of the hypothesis testing.

(e) Conduct a F-test for the joint significance of 'exper' and 'exper-sq' at 5% significance level. State the null and alternative hypotheses clearly. [7 marks]

		Students should calculate the required test-statistic using the given value of R-squared. Students must set hypotheses correctly. $F_{test} = 3.375 > F_{0.05,df1=2,df2=50} = 3.18$, reject the null hypothesis. They should reject the null hypothesis and provide a correct conclusion. Students should show all steps of the hypothesis testing.
	(f)	Based on the regression results of Model B and the test conducted in part (e), explain the nature of the relationship between wage and experience. Find the years of experience at which the hourly wage of an individual is the highest, holding other variables constant. [7 marks]
		A square term of experience is included because of the potential non-linear relationship between wage and experience. Since 'exper' and 'expersq' are individually and jointly significant, the relationship between wage and experience appears to be non-linear. Hourly wage rate increase with years of experiences but at a decreasing rate and the relationship has an inverted U-shape. Holding other variables constant, the effect of experience on the hourly wage rate can be written as: $wage = 3.4488 exper_i - 0.0479 exper_i^2$ $\frac{d(wage)}{d(exper)} = 3.4488 - 0.0958 exper = 0; exper = 36;$ $\frac{d^2(wage)}{d(exper)^2} = -0.0958 < 0;$ $wage$ is maximised at $exper = 36$.

6. The following table presents the results of an OLS estimation that investigates the determinants of foreign direct investment in Malaysia for the period 1976-2015. The dependent variable is the natural logarithm of the inflows of foreign direct investment to Malaysia in 1987 prices (ln-FDI_t).

Model 1: OLS, using observations 1981-2020 (T=40)

Variables	**Coefficients**	**P-values**
ln-$PGDP_t$	0.5782	0.0042
$Inflation_t$	−0.0298	0.0027
$Interest_t$	−0.0329	0.0267
Constant	1.7534	0.0864

Residual Sum of Squares (SSE)	17.2804
D-W	2.7743

The independent variables are defined as follows:
ln-$PGDP_t$: Natural logarithm of the per capita GDP of Malaysia
$Inflation_t$: Annual inflation rate of Malaysia in percentage
$Interest_t$: Interest rate in Malaysia in percentage

(a) Interpret the estimated coefficient of '$\ln\text{-}PGDP_t$'.

[6 marks]

Students should provide an accurate interpretation of the estimated coefficient based on log-log specification.

(b) Interpret the estimated coefficient of '$Inflation_t$'.

[6 marks]

Students should provide an accurate interpretation of the estimated coefficient based on log-lin specification and change in inflation rate is percentage point.

(c) Test the hypothesis that an increase in interest rate reduces the inflows of FDI to Malaysia at 5% significance level. State your null and alternative hypothesis clearly.

[6 marks]

Students should conduct a one-tailed test and provide all steps in hypothesis testing process. They should reject the null hypothesis and provide a correct conclusion.

(d) Another researcher suggests that the impacts of inflation rate and interest rate on the inflow of FDI could be the same. Explain how you could conduct a hypothesis test to confirm this suggestion.

[6 marks]

Students should explain the steps in a test for linear restriction including the correct hypotheses and decision rule.

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

[6 marks]

		Since Durbin-Watson statistic of 2.7743 is between $4 - d_L$ and 4, we reject the null hypothesis. There is evidence of a negative correlation of residuals. Student should set the null and alternative hypotheses accurately, provide all steps in hypothesis testing process, and state their conclusion clearly.
	(f)	Based on the result of the test in part (e) of this question, comment on the reliability and validity of the regression estimates. [5 marks]
		Students should identify the consequences of violating the assumption of no autocorrelation.