Name: Enrolment No:	

University of Petroleum & Energy Studies (UPES)
School of Business (SoB)
End-Semester Examination – May 2023

Program: BA (Hons.) Economics | **Semester: IV**
Subject / Course: Introductory Econometrics | **Maximum Marks: 100**
Course Code: ECON 2017 | **Duration: 03 Hours**

INSTRUCTIONS:

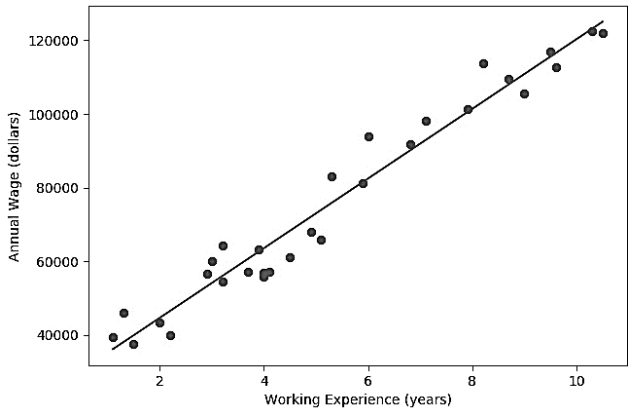
- **This is a CLOSED-BOOK EXAM. Only Non-scientific calculator is allowed.**
- **Cellphones / Tablets / Laptops / Books / Notes etc. are NOT allowed.**
- **All questions are compulsory. If Choice is there, it is *indicated within the question as* OR.**
- **Your answers must be "brief & to the point."**

Q. No.	Questions	Marks	COs
	SECTION A **10Q x 2M = 20 Marks**		
Q 1.	Two events, A and B, are said to be mutually exclusive if: **A)** $P(A \mid B) = 1$ **B)** $P(B \mid A) = 1$ **C)** $P(A \,\&\, B) = 1$ **D)** $P(A \,\&\, B) = 0$	2	CO1
Q 2.	Type I error occurs when we: **A)** reject a false null hypothesis. **B)** reject a true null hypothesis. **C)** do not reject a false null hypothesis. **D)** do not reject a true null hypothesis.	2	CO1

Q 3.	The violation of the assumption of constant variance of the residual is known as: **A)** The variance of the errors is not constant. **B)** The variance of the dependent variable is not constant. **C)** The errors are not linearly independent of one another. **D)** The errors have non- zero mean.	2	CO1
Q 4.	Autocorrelation is generally occurred in: **A)** Cross-section data. **B)** Time series data. **C)** Pooled data. **D)** None of the above.	2	CO1
Q 5.	In the regression function $\boxed{Y = \alpha + \beta X + \varepsilon}$: **A)** X is the regressor. **B)** Y is the regressor. **C)** α is the regressor. **D)** ε is the regressor.	2	CO1
Q 6.	BLUE is referred as the **A)** Best Linear Unbiased Estimator. **B)** Best Linear Unconditional Estimator. **C)** Basic Linear Unconditional Estimator. **D)** Both **B)** and **C)**.	2	CO1
Q 7.	Data on one/more variables collected at a given point of time is known as: **A)** Panel data. **B)** Time series data. **C)** Pooled data. **D)** Cross-section data.	2	CO1
Please Turn Over			

Q 8.	Probability of occurrence of an event lies between **A)** –1 and 0. **B)** –1 and 1. **C)** 1 and 0. **D)** –100 and 100.	2	CO1
Q 9.	A sure way of removing multicollinearity from the model is to: **A)** Work with panel data. **B)** Drop variables that cause multicollinearity in the first place. **C)** Transform the variables by first order of differencing them. **D)** Obtaining additional sample data.	2	CO1
Q 10.	The coefficient of determination, R^2 shows: **A)** The proportion of the variation in the dependent variable Y is explained by the independent variable X. **B)** The proportion of the variation in the dependent variable X is explained by the independent variable Y. **C)** The proportion of the variation in ε is explained by the independent variable X. **D)** Both **A)** and **C)**.	2	CO1
SECTION B **4Q x 5M = 20 Marks**			
Q 11.	True or False? Briefly justify the reasoning. *"If a fair coin is tossed many times for independent trials, and the last eight tosses are all tails, then the chance that the next toss will be tails is somewhat less than 50%."*	5	CO2
Q 12.	What is a *Null Hypothesis* (H_0) and an *Alternative Hypothesis* (H_1)? Using a relevant example, briefly explain these two concepts.	5	CO2
Please Turn Over			

Q 13.	Using a relevant example, briefly explain the difference between *Two-Tailed* & *One-Tailed* Tests.	5	CO2
Q 14.	A recent research survey done by Dr. Chakraborty asked 15,292 randomly sampled registered Indian voters about their political affiliation (Rightist, Leftist, or Independent) and whether or not they identify as 'Swing Voters.' • 15% of respondents identified as Independent, • 33% identified as Swing Voters, and • 21% identified as both. What percent of voters are Independent OR Swing Voters? Show your calculation.	5	CO2
	SECTION C **3Q x 10M = 30 Marks**		
Q 15.	Below are the final exam scores of 20 *Introductory Econometrics* students. 79 83 57 82 94 83 72 74 73 71 66 89 78 81 78 81 88 69 77 79 **(a)** What is the mean score? [2 points] **(b)** What is the median score? [2 points] **(c)** What is the mode? [2 points] **(d)** What is the Standard Deviation (S.D.)? [2 points] **(e)** Draw a free-hand histogram for the score distribution. [2 points] **OR** Using the regression function $\boxed{Y_i = \alpha + \beta X_i + \varepsilon_i}$ write down the **key five assumptions of the Ordinary Least Squares (OLS)** and briefly explain each one of them.	10	CO3
	Please Turn Over		

Q 16.	Prove that $\boxed{\sum(Y_i - \bar{Y})(X_i - \bar{X}) = \sum(Y_i - \bar{Y})X_i = \sum Y_i X_i - n\bar{X}\bar{Y}}$. *Given*: 1. $\boxed{\bar{X} = \sum X_i / n}$ 2. $\boxed{\bar{Y} = \sum Y_i / n}$ 3. $\boxed{Y_i = \alpha + \beta X_i + \varepsilon_i}$	10	CO3
Q 17.	To study the relationship between the wage (dependent variable) and working experience (independent variable), we use a linear regression model: $\boxed{Wage_i = \alpha + \beta * Experience_i + \varepsilon_i}$. In this study, we use 30 data points [$i = 1 \rightarrow 30$], where the annual salary (in USD) ranges from \$39,343 to \$1,21,872 and the years of experience range from 1.1 years to 10.5 years. Looking at this scatterplot below, we can imagine that the relationship in this sample is pretty close to linear.  A quick MS Excel regression exercise spits out the following equation: $\boxed{\widehat{\boldsymbol{Wage}}_{\boldsymbol{i}} = 25792.20 + 9449.96 * Experience_i}$. **Calculate and interpret** the values of $\boxed{\widehat{\boldsymbol{Wage}}_{\boldsymbol{i}}}$ for a typical worker i from the sample: **1.** When $Experience_i$ $\boxed{\uparrow}$ by 1 extra year, i.e., $\frac{\partial Y_i}{\partial X_i}$. [1 point] **2.** When $Experience_i = 0$ year. [3 points] **3.** When $Experience_i = 5$ years. [3 points] **4.** When $Experience_i = 10$ years. [3 points]	10	CO3
Please Turn Over			

SECTION D
2Q x 15M = 30 Marks

Q 18.	In order to **<u>minimize the errors</u>** in the *Ordinary Least Squares (OLS)*, algebraically solve the values of $\boxed{\hat{\alpha}}$ and $\boxed{\hat{\beta}}$ as follows: 1. $\boxed{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}}$ *and* 2. $\boxed{\hat{\beta} = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}}$ <u>*Given*</u>: 1. $\boxed{Y_i = \alpha + \beta X_i + \varepsilon_i}$ 2. $\boxed{\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i}$ 3. $\boxed{\sum \hat{\varepsilon}_i = 0}$ 4. $\boxed{\sum \hat{\varepsilon}_i X_i = 0}$ 5. $\boxed{\bar{X} = \sum X_i / n}$ 6. $\boxed{\bar{Y} = \sum Y_i / n}$ <u>*Note*</u>*: You must show all the necessary steps, e.g., setting up the Minimization problem, and then deriving the First Order Conditions (FOCs), solving the system of equations, and finding the values* $\hat{\alpha}$ *and* $\hat{\beta}$*, etc.*	15	CO4

Please Turn Over

Q 19.	**<u>CRITICAL THINKING</u>:** After the famous *FIFA World Cup Qatar 2022* crown to Lionel Messi's Argentina, a controversy of systematic bias in sports is being much discussed. In football, referees may have significant power and influence over the final outcome of a match through goals validated, penalties, red and yellow cards, off-side and foul calls, extra minutes to injury time, etc. among other actions. The referee's actions on the field, if biased, may be contrary to the best interest of the sport. As these biases increase, the sport may lose popularity as fans question the fairness of the competition, with the direct consequence of revenue loss for *FIFA*.  Dr. Chakraborty is interested in investigating whether individual referees have a systematic effect on the final score of a football match, using the Hero Indian Super League (ISL) data for 2014-2022. Suppose you are working as Chakraborty's Research Assistant (RA). You have the following information: **1.** There are 11 football clubs in India playing the Hero ISL *[see below]*.  	15	CO4

2. Each club plays 20 *League matches*.
3. Each club faces the same opponent twice, 1 home match, another away match.

   e.g., Bengaluru FC plays against FC Goa twice, 1 match in the *Sree Kanteerava Stadium, Bengaluru*, another match in the *Jawaharlal Nehru Stadium, Goa*.

4. After *League matches*, 2 *Knockouts*, 2 *Semi-finals*, and 1 *Final* are played.
5. There are 11 referees, 1 each from the home city/state of a club.

Chakraborty asks you to use your knowledge of hypothesis testing and regression estimation to **propose an econometric model** for the research claim: "*the home referee is a significant factor affecting the match score in favour of the home club*" in the format of a typical **OLS**: $\boxed{Y_i = \alpha + \beta R_{ji} + \delta_s X_{si} + \varepsilon_i}$ where Y_i is the value of 'dependent' variable $\forall\ i$, α is the constant/intercept, β is the coefficient for R_{ji} = Referee form city/state j in the match i: the 'independent' variable, δ_s are the coefficients for all the relevant control variables: $X_s = X_1, X_2, X_3 \ldots \forall\ s\ \&\ i$; *where*

- i represents the *league matches* $[i = 1 \rightarrow 20]$,
- s represents the *control variables,* a maximum of 3 to 4 controls, and
- $\boxed{j = \begin{cases} 0: R_{0i} = away\ referee\ in\ a\ home\ match\ i \\ 1: R_{1i} = home\ referee\ in\ a\ home\ match\ i \end{cases}}$

Note: *Please mention what exactly your proposed variables are, including the control variables? What type of variables are they, i.e., binary, categorical, continuous, etc.? And, what values they could be assigned with? What should be the expected signs of β, i.e., +ve or, -ve? What should be the expected signs of δs, i.e., +ve or, -ve?*

Hint:

- *Y_i = score differences in the $Match_i$*
- *$\Longrightarrow$ usually, a continuous variable*
- *$\Longrightarrow$ values varying from – 5 to +5, i.e., the score difference of a typical match in between Bengaluru FC (or, any other team) and Goa FC (or, any other team) ranges between (– 5, +5). e.g.,*
  Final Score: Bengaluru FC – Goa FC = – 3 when Bengaluru FC loses the match by 3 goals.
  Final Score: Bengaluru FC – Goa FC = +2 when Bengaluru FC wins the match by 2 goals.