Register Number:
Date:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE – 27
M.SC(BIG DATA ANALYTICS) – I SEMESTER
SEMESTER EXAMINATION – OCTOBER 2021
(Examination conducted in January-March 2022)

BDADE 3521*:* INTRODUCTION TO ECONOMETRICS & FINANCE

TIME: 2.5 Hrs **MAX MARKS: 70**

This question paper contains SIX printed pages and THREE parts

PART A

Answer ALL questions (MCQ) **20 X 1 = 20**

1. Which is an example of cross-sectional data:
    a. Data on India for the period 1995 to 2005
    b. Data on China for the period 1995 to 2005
    c. Data on India and China only for the period 1995
    d. None of these
2. The OLS residuals
    a. can be calculated using the errors from the regression function.
    b. can be calculated by subtracting the fitted values from the actual values.
    c. are unknown since we do not know the population regression function.
    d. should not be used in practice since they indicate that your regression does not run through all your observations
3. For a regression model, the Explanatory Sum of Squares is 12.97 while the Residual Sum of Squares is 5.13. The R^2 is approximately:
    a. 1
    b. 2.12
    c. 0.71
    d. 0.12
4. Consider the following regression line: $\widehat{Grades}$= 698.9 – 2.28 *STR where STR is the Student-Teacher Ratio. The standard error of the slope coefficient is 0.52. The t-statistic of the slope coefficient is:
    a. -1.96
    b. 1.96
    c. -4.38

d. 2.28

5. If the explanatory variable (X) is correlated with the disturbance term (u), then the OLS estimator is:
    a. Biased
    b. Unbiased
    c. Very efficient
    d. Very efficient and Unbiased
6. The R^2 for a regression model is 0.77. The R^2 number implies that:
    a. The slope coefficient is 0.77
    b. Dependent variable (Y) changes by 77 percent for 100% change in explanatory variables (Xs)
    c. 77% of variation in dependent variable (Y) is explained by the explanatory variables (Xs) in the model
    d. Model is useless as the value is less than 1
7. If our model has perfect multicollinearity, one option is to:
    a. Use instrumental variables
    b. Drop one of the two variables that have exact linear relationship.
    c. Abandon the model
    d. Use intuition to guess the estimates
8. Variance Inflation Factor (VIF) is used to detect:
    a. Degree of polynomial of the dependent variable
    b. Degree of polynomial of the explanatory variable
    c. Multicollinearity
    d. Stationarity
9. Identification problem in econometrics refers to
    a. identity crisis of the explanatory variables
    b. identity crisis of the dependent variables
    c. inability to recover structural parameters from reduced form equations
    d. inability to recover reduced form parameters from structural equations
10. In estimating SEM (Simultaneous Equation models), we find Proxy variable(s) for the explanatory variable that may be uncorrelated with the error term. These Proxy variables are known as:
    a. Coefficients of reduced form
    b. Parameters of structural form
    c. Instrumental variables
    d. Explanatory variable

11. A regression model that includes the lagged values of the dependent variable in its explanatory variables is called:
    a. Distributed lag model
    b. Auto-regressive model
    c. Fixed effects model
    d. Probit model
12. The test statistic used to study the direction of causality between two time-series variables is:
    a. Bilateral cause-effect test
    b. Granger causality test
    c. Almond test
    d. Nonsensical test
13. Which of the following procedures is the most appropriate for estimation of over-identified SEM (Simultaneous Equation models):
    a. OLS
    b. MLE
    c. 2SLS or IV
    d. GLS
14. In simultaneous equation model, the number of equation to be estimated is
    a. One more than the number of endogenous variables
    b. Equal to the number of endogenous variables
    c. Depends on the underlying economic theory
    d. Equal to the number of endogenous and exogenous variables
15. In testing causality of two time-series variables (X & Y), it is assumed that:
    a. At least one variable is stationary
    b. Both variables are stationary
    c. Neither are stationary
    d. Some unobserved variable has constant mean
16. The Ho that we test in Hausman test is:
    a. Fixed Effects and Random Effects estimator are no different
    b. Fixed Effects and Random Effects estimator are both zero
    c. Fixed Effects and Random Effects estimator are not equal to zero
    d. Fixed Effects and Random Effects estimator are different
17. If in panel data, individual-specific error term is correlated with the Xs, then estimates will likely be unbiased if we employ
    a. Random Effects model

b. Pooled model
c. Fixed effects model
d. Any model will suffice

18. The random walk model $Y_t = Y_{t-1} + u_t$:
    a. Is difference stationary
    b. Is not difference stationary
    c. is incorrectly specified above
    d. is stationary
19. Writing each simultaneous system equation in terms of only exogenous or predetermined variables gives us the Reduced Form equation. Running OLS on the Reduced form equation will give us _____ estimates:
    a. Biased
    b. Unbiased
    c. Inflated
    d. Deflated
20. The First Stage in 2SLS checks if the instrument is strongly correlated with the
    a. Error term
    b. Dependent variable
    c. Endogenous explanatory variable
    d. White noise

PART B

Answer ANY SIX questions **6 X 5 = 30**

21. Show that the random walk is non stationary. Describe the unit root test.
22. A popular test for auto-correlation is the Durbin Watson (DW) test. The DW statistic is given by $d = \frac{(\sum \widehat{e_t} - \widehat{e_{t-1}})^2}{\sum \widehat{e_t}^2}$ where $\widehat{e_t}$ is the residual.
    a. What is auto-correlation?
    b. What is intuition behind why d=2 implies no auto-correlation?
23. What conditions are required for a variable to be a valid instrument? Which condition can be easily tested?
24. What do we mean by co-integration in time series context? Describe the Engle-Granger co-integration test briefly.
25. Consider the following models (X is a continuous variable and D a dummy variable):
    a. $Y = \beta_0 + \beta_1 X + u$
    b. $Y = \beta_0 + \beta_1 X + \theta D + \gamma(X * D) + u$

Graphically show the above and discuss when would use these models?

26. Use simple regression framework to explain the residual term? How is the residual term used to derive the OLS estimator? [Note: you do **not** need to derive.]
27. Your data has weight and height from 29 female and 81 male students at your university. You also know the number of siblings they have. You consider a new theory that children who have more siblings come from poorer families and will have to share the food on the table. You decide to hypothesize that peers with many siblings will weigh less, on average, for a given height. In addition, you believe that the muscle/fat tissue composition of male bodies suggests that females will weigh less, on average, for a given height. To test these theories, you perform the following regression:

    $\widehat{Studentw}$ = - 229.92 – 6.52 *Female + 0.51 *Sibs+ 5.58 *Height, R^2=0.50

    where Studentw is in pounds, Height is in inches, Female takes a value of 1 for females and is 0 otherwise, Sibs is the number of siblings.

    a. Does the intercept make sense?
    b. Interpret the coefficients and the R^2.
    c. Set up the null and mention the test to test:
        i. $H_0: \beta_{Female} = 0$
        ii. $H_0: \beta_{Female} = \beta_{Height} = 0$

28.

    a. State the order condition.
    b. Consider the following equations:

    $$Q^d{}_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_t$$

    $$Q^s{}_t = \beta_0 + \beta_1 P_t + \beta_2 R_t + v_t$$

    where Q and P are endogenous variables, and, I and R are exogenous variables. Use order condition to examine which equation is identified-exactly, over or not?

PART C

Answer ANY TWO questions **2 X 10 = 20**

29.

    a. Write the expectation moment condition between the error term (u) and explanatory variable (X) for the OLS estimator to be unbiased?

b. Consider the following mode: $Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + u_t$ where u_t, the error term is auto-correlated. Explain why the slope coefficient using OLS is biased.

30. Discuss how the Error Correction Model incorporates long-run and short-run characteristics.

31.

a. Show and discuss how if we have panel data, we can deal with any bias due to time-invariant unit-specific characteristics to deal with it.
b. State the different assumptions under which we use (i) Fixed Effects and (ii) Random Effects.