Unique Paper Code	:	32377908
Name of the Paper	:	Econometrics (DSE – 2(B))
Name of the Course	:	B.Sc. (Hons.) Statistics
Semester	:	V
Duration	:	3 hours
Max Marks	:	75

Instructions for Candidates

1. Answer any **FOUR** questions.
2. All Questions carry **equal marks**.
3. The question paper has total 4 pages including this page

1. Suppose in the regression model:

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t,$$

γ_{12} which depicts coefficient of correlation between X_1 and X_2 is zero.

   (a) Would you suggest to run the following regressions:

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + U_{1t}$$

$$Y_t = \gamma_0 + \gamma_2 X_{2t} + U_{2t}$$

   Justify your answer.

   (b) Also examine whether:

   (i) $\hat{\alpha}_1 = \hat{\beta}_1$ and $\hat{\gamma}_2 = \hat{\beta}_2$

   (ii) $\hat{\alpha} = \hat{\alpha}_0$ and $\hat{\alpha} = \hat{\gamma}_0$

   Support your answer with appropriate expressions and their mathematical proof.

2. Suppose that a researcher , using data on class size (CS) and average test scores from 50 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 640.3 - 4.93 \text{ x } CS$$

$$R^2 = 0.11, \text{ and estimate of error variance} = 75.69$$

Last year a classroom had 21 students, and this year it has 24 students. What is the regression's prediction for the change in the classroom average test score?

Further, the sample average class size across the 50 classrooms is 22.8. What is the sample average of the test scores across the 50 classrooms?

Also find the sample standard deviation of test scores across the 50 classrooms.

3. Statement: "Absence of high pair-wise correlation(s) among regressors confirms absence of multicollinearity"

Is this statement true for a linear regression model with two independent variables? Give detailed reasons to support your answer.

Discuss the validity of the statement when we have more than two regressors in the model? In this case, if you think that the information provided by pair-wise correlations is insufficient, what kind of additional information will be required to conclude about the presence or absence of multicollinearity?

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

4. A study was performed to assess the impact of employment size (X) on the average compensation (Y) and following regression was obtained

$$\hat{Y}_i = 3417.833 + 148.767\, X_i \qquad (*)$$
$$(81.136) \qquad (14.418)$$
$$\text{p-value} = \quad (<0.001) \qquad (<0.001) \qquad R^2 = 0.9383$$

The analyst suspected violation in one of the assumptions and hence fitted the model

$$\frac{\hat{Y}_i}{\sigma_i} = 3406.639 \frac{1}{\sigma_i} + 154.153 \frac{X_i}{\sigma_i} \qquad (**)$$
$$(80.983) \qquad (16.959)$$
$$\text{p-value} = \quad (<0.001) \qquad (<0.001) \quad R^2 = 0.9993$$

What violation is the analyst suspecting? Identify the specific form of this violation for which regression (**) provides a solution. Explain any one test to detect this violation. What are the repercussions of such a violation?

5. For the General Linear model $\underset{nx1}{Y} = \underset{nx3}{X}\,\underset{3x1}{\beta} + \underset{nx1}{U}$ with usual assumptions, obtain the restricted least square estimator of $\underset{\sim}{\beta}$ along with its variance, under the set of linear restrictions given by $\begin{bmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix} \underset{\sim}{\beta} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

6. In an Election research, data were obtained for 72 districts of California State legislature and included the total number of registered voters by district, their party affiliation, the number of votes received by each candidate and the identity of the incumbent. The following are the results of three regression runs (numbers in parenthesis are *t* - values)

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt

All districts:

$$\text{WSV} = 0.240 + \underset{(4.82)}{0.174}\,\text{WSTE} + \underset{(4.60)}{0.141}\,\text{WSRV} + \underset{(7.01)}{0.075}\,\text{I} \quad ; \quad R^2 = 0.535,\ n = 72$$

Incumbent districts:

$$\text{WSV} = 0.329 + \underset{(3.67)}{0.157}\,\text{WSTE} + \underset{(6.07)}{0.409}\,\text{WSRV} \quad ; \quad R^2 = 0.440,\ n = 55$$

Nonincumbent districts:

$$\text{WSV} = 0.212 + \underset{(3.39)}{0.234}\,\text{WSTE} + \underset{(3.21)}{0.399}\,\text{WSRV} \quad ; \quad R^2 = 0.515,\ n = 17$$

where WSV = Winners share of total votes cast, WSTE = Winners share of total advertising expenditures, WSRV = Proportion of registered voters that are registered to the winner's political party and I $= \begin{cases} 1 & \textit{for incumbent district} \\ 0 & \textit{for Nonincumbent district} \end{cases}$

(a) Interpret the regression coefficients for *all districts*. Why is the coefficient of WSTE different in the three equations?

(b) ~~Explain exactly what the t-value means~~What do you infer from the t-value~~.~~? Interpret R^2 and enunciate on why it is different for each equation~~?~~.

(c) Why does the Incumbency variable appear only in the first equation?

(d) Could this model be used productively to predict? What insights could a candidate get from the model?

Formatted: Font: (Default) Times New Roman, 12 pt, Complex Script Font: Times New Roman, 12 pt, Italic