BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2017-18

Course No.: ECON F241 | **Course Title: ECONOMETRIC METHODS**
Date: 06 MARCH 2018 | **Max. Marks: 30 Time: Min.90**

MID SEMESTER TEST

NOTE: Attempt all questions. Write assumptions if any. No credit without clear explanation. Answer will be graded in part on the quality and clarity of your explanation. Incorrect or irrelevant statements may be penalized, so please answer each question as clearly and concisely as possible on the answer sheet.

---

1. Carefully read and assess the following statements as TRUE or FALSE. Give a short explanation or qualification. If a statement is not true in general, but is true under some conditions, state the conditions. *No credit for just stating TRUE or FALSE*.

(4.0)

A "The Econometric model $Y_i = (B_0 + B_1X_i)\ u_i$ satisfies all the classical assumptions of the linear regression model."

B "An assumption that is crucial to the unbiasedness of least squares simple regression is that the population correlation between the disturbance and the explanatory variable is zero; so it is a good idea to test that assumption by calculating the sample correlation between the least squares residuals and the explanatory variable".

C "Since inclusion of irrelevant variables cannot bias OLS estimates, ignorance of the problem has no effect on hypothesis testing."

D "Increasing sample size changes the sampling error variance of a regression and leads to larger standard errors for the estimated coefficients."

E “The 95% confidence interval around the mean prediction is constructed by using the standard error of the estimated slope from the sample regression function.”

F “An assumption that is crucial to the unbiasedness of least squares simple regression is that the population correlation between the disturbance and the explanatory variable is zero; so it is a good idea to test that assumption by calculating the sample correlation between the least squares residuals and the explanatory variable.”

G “Omitting a relevant explanatory variable from the regression equation has no effect on the coefficient estimates on the included explanatory variables.”

H For the following sentence indicate whether it is justified and explain your reason. “because multicollinearity lowers t-statistics, all the insignificant regression coefficients should be dropped from the model because they are redundant.”

2. **For the following MULTIPLE CHOICE questions,** choose the <u>one correct best answer</u> to each question and put a tick (√) against that letter A/B/C/D. ***<u>Corrections/overwriting/illegible answers are strictly invalid.</u>***

(5.0)

1). If an estimator is unbiased, then:
A). for any given sample it will produce an estimate that is equal to the true value of the population parameter being estimated.
B). in a very large number of samples it will produce an average estimate that equals the true value of the population parameter, so on average it is correct.
C). in the class of efficient estimators it has the smallest bias.
D). it does not discriminate between different types of parameters it can be used to estimate.

2). The assumption that the error term is normally distributed: $\varepsilon_i \sim N(0, \sigma^2)$ is required in order for the least-squares estimators to be
A). Consistent.
B). Best linear unbiased estimators.
C). Best unbiased estimators.
D). Asymptotically normally-distributed (that is, approximately normally-distributed in large samples).

3). If the data observations fit exactly on a straight line, then the least-squares estimate of the variance of the error term would be
A). $\hat{\sigma}^2 = 0$.
B). $\hat{\sigma}^2 = 1$.
C). $\hat{\sigma}^2 = -1$.
D). None of the above.

4). You specify a simple classical linear regression model. The dependent variable is a measure of the health status for the t^{th} state in India. The explanatory variable is a measure of medical care spending per capita for the t^{th} state.
Consider the following argument: States with healthy populations spend less on medical care than states with unhealthy populations. If this argument is valid, then which of the following assumptions of the simple classical linear regression model will not be valid?
A). The error term is not correlated with the explanatory variable.
B). The error term has constant variance.
C). The errors are independent.
D). The error term has a normal distribution.

5). You are using a simple classical linear regression model and the OLS estimator to analyze the relationship between the hourly wage rate and years of education. Why can't you use the Z-statistic, which has a standard normal distribution, as your test statistic to test the null hypothesis that the marginal effect of education on the wage is 1.00?
A). The Z-statistic has a mean of zero.
B). The Z-statistic has a variance and standard deviation of 1.
C). The true standard error of the estimator of the marginal effect of education is unknown and unobservable.
D). Both A and B

6). You are using a simple classical linear regression model and the OLS estimator to analyze the relationship between the hourly wage rate and years of education. You have reason to believe your error term has non- constant variance. Which of the following would you expect to be the consequence of this situation?

A). Your study will not be internally valid because your estimate of the effect of education will be biased.

B). Your study will not be internally valid because your standard error of the estimate of the effect of education will be incorrect, and therefore any inferences you make or hypotheses you test may be incorrect.

C). Your study will be internally valid but not externally valid, because it will not generalize to the population in which you are interested.

D). Your study will be both internally and externally valid, because non-constant error variance does not affect the validity of your conclusions.

7). The Gauss-Markov Theorem tells us that, under appropriate assumptions, the least squares estimator of β in the usual linear regression model:

A Is a linear estimator, and therefore is "best".

B Has the smallest bias among all possible linear estimators for this parameter vector.

C Is most efficient among all possible linear and unbiased estimators of this parameter.

D Is most efficient among all possible unbiased estimators that have a Normal sampling distribution.

8). All of the following are possible effects of multicollinearity EXCEPT

A. the variances of regression coefficients estimators may be larger than expected

B. the signs of the regression coefficients may be opposite of what is expected

C. a significant F ratio may result even though the t ratios are not significant

D. the VIF is zero

9). Which of the following options is not a source of Multicollinearity :

A. Data collection method employed

B. Model specification

C. Over determined model

D. Sampling technique

10). Which of the following best describes the idea of multicollinearity?

A. At least one X variable is closely related to the Y variable.

B. The equation relating the Y variable to the X variables is not useful.

C. There is no correlation at all between the X variables and the Y variable.

D. Some of the X variables are very closely related to one another.

3. Suppose that X and Z are independently distributed standard normal random variables. Let $Y = X(X - 2Z)$.

What is $E(Y \mid X)$? Compute mean of Y (μ_Y) and What is E ($Z\ Y$)?

(2.0)

4. Consider the case that the regression function does not have an intercept. If we know that the population regression function is: $Y_i = \beta_1 X_i + u_i$. Where $\beta_0 = 0$

a What is $E(Y_i \mid X_i)$? What is $E(Y_i \mid X_i = 0)$?

b Derive the ordinary Least Squares estimator for the β_1?

c Obtain the estimator for the β_1 by using the moment condition $E(u_i X_i) = 0$ and compare with the estimator obtained at Part b.

(2.0)

5.

A For each of the following scatter plots involving random samples for variables X and Y, Answer the following questions:

a Do X and Y appear to be independent?

b Would the OLS intercept estimate $\widehat{\beta 0}$ be positive, negative, or close to 0?

c The relationship between X and Y appears to be positive and significant, negative and significant, or insignificant?

d Would the R^2 coefficient of the OLS estimated model be big (> 0.9), small (< 0.1), or intermediate?

e Is there any classical assumption other than correct specification which appears to be violated? (Write at most one.)

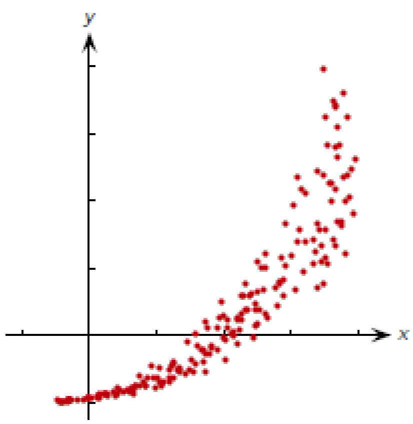

Figure.1

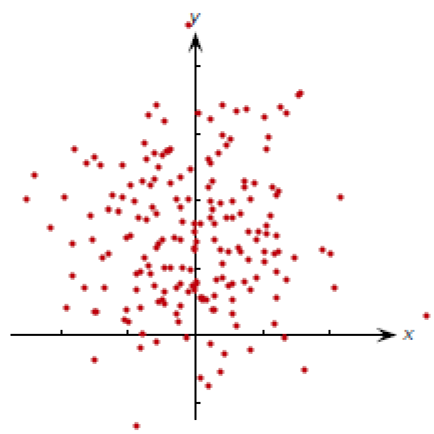

Figure.2

1	Independence:			Independence:	
2	β_0			β_0	
3	β_1			β_1	
4	R^2			R^2	
5	Assumptions			Assumptions	

B Define R^2 and $\overline{R}^2$, and what are their important property? Show how they are related?

6. A researcher in the Econometrics Class wants to know about the expenditure on food in a month for students at BITS Pilani, Pilani Campus. In particular he wants to estimate the mean of their food expenditure. Now we randomly sample (n) 225 students. The sample mean food expenditure obtained is $\overline{Y}$ = 962:5 with standard deviation of the sample s = 88:8: He want to test the hypothesis that H_0: $\mu = 1000$ against H_1 $\mu \neq 1000$.

a What test statistic do we use? What is its distribution if the null is true?

b Calculate the statistic and carry out the test at 5% significance level. (Critical value is 1.96)

c Calculate the 95% two-sided confidence interval for the population mean food expenditure

(3.0)

7. Given sample values of (X, Y) equal to (0, 0), (2, 4) and (4, 2), a simple ordinary least squares (OLS) regression of Y on X and intercept gives $\widehat{\beta 1}$ = 1.0 and $\widehat{\beta 2}$ = 0.5. There is no need to derive these estimates.

Answer the following giving the appropriate formulas.

a Compute the standard error for this regression?

b Calculate se $(\widehat{\beta 2})$, the standard error of $(\widehat{\beta 2})$.

c Calculate a 95 percent confidence interval for β3

[For t with 1, 2 and 3 degrees of freedom, the critical values are 12.706, 4.303 and 3.182.]

d Calculate R^2 for this regression.

(4.0)

8. Using data on a random sample of 30 houses in a particular neighborhood the following relation between the sales price in 1000$ as dependent variable Y and the interior area of the house in 100 square foot and the lot size in 100 square foot as the independent variables X_1 and X_2 respectively, is estimated with OLS (standard errors in parentheses)

$$Y = 20.1 + 13.1X_1 + 9.12X_2 \qquad s = 2.22$$
$$(5.111) \quad (2.221) \quad (3.419)$$

a What is the interpretation of the coefficient of X_1?

b Can X_2 be omitted from the relation? Why (not)?

c A homeowner in the neighborhood wants to build an addition that adds 100 square foot to the interior area of the house. The cost of this addition is 7000$. Test whether the net gain (increase in sales price minus cost) is significantly different from 0 at the 5% level, i.e. with level of significance of 5%. Hint: As in a. pay attention to the units of measurement of the variables.

d If $\sum_{i=1}^{30}(Y_i - \overline{Y})^2 = 300$

test with level of significance 5% whether the coefficients of interior area and lot size are both equal to 0. The critical value of test statistic is 3.35.

(4.0)

9. An Econometrician run the multiple regressions and got the following output:

Dependent Variable: S
Method: Least Squares
Date: 02/28/02 Time: 10:04
Sample: 1 38
Included observations: 38

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.223923	1.935467	4.765736	0.0000
T	-0.786568	0.590256	-1.332588	0.1918
E	7.906074	3.660143	2.160045	0.0381
P	0.000408	0.000495	0.823516	0.4161
H	-0.018971	0.002674	-7.094425	0.0000

R-squared	0.709460	Mean dependent var	8.438421
Adjusted R-squared	0.674243	S.D. dependent var	2.621777
S.E. of regression	1.496382	Akaike info criterion	3.766057
Sum squared resid	73.89229	Schwarz criterion	3.981529
Log likelihood	-66.55509	F-statistic	20.14541
Durbin-Watson stat	2.146911	Prob(F-statistic)	0.000000

a) Your friend expects the coefficient of P to be positive. Carry out the formal hypothesis statement, test the hypothesis by using the 5% level of significance, and make your conclusion. The 5% significant level of critical test statistic value is **1.697**

b) Your friend also argues that for every car at least should have 8.5 units of drag for safety in the road. Construct the confidence interval for the coefficient of E, carry out the hypothesis statement and test by using the 1% level of significance and make your conclusion. The critical value of the test statistic at 0.01, 33df is **-2.457.**

c) Explain why you need to use the F test. What is your conclusion from the F hypothesis test? The critical value of $\mathbf{F^{c}_{(0.05,\ 4,\ 33)}} = \mathbf{2.69}$.

(3.0)