BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2015-16

Course No.: ECON F241 **Course Title: ECONOMETRIC METHODS**
Date: 18 March 2016 **Max. Marks: 30; Time: 90min**

MID SEMESTER TEST (REGULAR)

INSTRUCTIONS: **Answer all questions**. *This exam is closed-book and closed-notes. You may use a calculator on this exam. Start* answering each question on a separate page and *Attempt all parts of the same question at one place.* **Write the answers clearly, legibly and highlight your final answer**. **BE BRIEF**. **All the Best!!**

1) A) Are the following statements TRUE or FALSE? Provide a short explanation. Just by writing TRUE or FALSE, you will not receive any credit. Provide a correct explanation. Weak justifications will receive minimal credit. **(5.0)**

a) Correlation is a measure of the degree of linear relatedness of two variables. If Y and X are uncorrelated, then they are statistically independent

b) The ratio of two variables that are distributed according to a chi-square distribution is distributed according to a t-distribution.

c) Sum of squares, such as total sum of squares, error sum of squares and regression sum of squares, follow a F-distribution

d) The bus schedule states that the bus stops at 8:10am at Pilani bus stop. An econometrician has collected 100 observations on actual bus arrival times; the sample average is 8:14am and the sample variance is 144. Statistically speaking, the bus schedule is accurate.

e) In OLS estimation, the predicted value of Y_i depends on the value of X_i.

f) The Econometric model $Y_i = (\beta_0 + \beta_1 X_i)u_i$ satisfies all the classical assumptions of the linear regression model.

g) The width of the confidence interval for the OLS estimator of the slope coefficient in a simple linear regression model doesn't depend on the sample size n.

h) 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.

i) Since inclusion of irrelevant variables cannot bias Ordinary Least Squares estimates, ignorance of the problem has no effect on hypothesis testing.

j) The fact that residuals in the linear model estimated by least-squares have zero mean is a consequence of assuming that the expected value of the error term is zero.

B) For the following questions, choose the correct best answer and put a tick (√) against the corresponding letter A/B/C/D/E. Corrections/Overwriting/Illegible answers are strictly invalid. Each question carries 0.50 marks. (5.0)

1. Estimates of the same population parameters can differ from one sample to another. This phenomenon is called
A. measurement error.
B. rounding error.
C. sampling variation.
D. population variance.
E. error in judgment

2. If an estimator is unbiased, then its mean square error (MSE)
A. must be greater than its variance.
B. must be equal to its variance.
C. must be less than its variance.
D. can be greater or less than its variance.
E. must be identically zero.

3. Which assumption is required for the sum of the least-squares residuals to equal zero?
A. $E(\varepsilon_i|X_i) = 0$.
B. Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant.
C. No autocorrelation: $Cov(\varepsilon_i,\varepsilon_j) = 0$ for $i\neq j$.
D. All of the above.
E. None of the above.

4. Suppose a production function is estimated of the form $y_i = \beta_1 + \beta_2 X_i$, where Y_i denotes output in kilograms and X_i denotes labor input. Now suppose the output data are converted to pounds (there are about 2.2 pounds to a kilogram) and the equation is re-estimated. Which of the following are true?
A. $\hat{\beta}_1$ will increase by a factor of 2.2 .
B. $\hat{\beta}_2$ will increase by a factor of 2.2 .
C. The sum of squared residuals will increase by a factor of $(2.2)^2$.
D. The r^2 value will be unaffected.
E. All of the above.

5. If we add more regressors to an equation and re-estimate it, what will surely rise?
A. the sum of squared residuals.
B. the sum of the residuals.
C. R-square.
D. Theil's adjusted R-square.
E. all of the above.

6. According to the following model: $Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 X_{i2}^2$, a one-unit increase in X_{i2} will cause Y_i to increase by about
A. β_2 units.
B. $(\beta_1 + \beta_2)$ units.
C. $(\beta_2 + \beta_3)$ units.
D. $(\beta_2 + \beta_3 x_{i2})$ units.
E. $(\beta_2 + 2\beta_3 x_{i2})$ units.

7. Taking into account the formula for the variance of the slope estimator in a simple regression model, how might you choose to design your data collection to ensure that you have more-accurate slope estimates?
A. Make sure all of your X observations are close to the mean of X.
B. Make sure your X observations are widely dispersed around the mean of X.
C. Make sure that you have a lot of large and small values of Y in your sample.
D. Make sure that the values of Y in your sample are all very close to the mean of Y, so you have good resolution due to dense data.
E. None of the above

8. Which of the four maintained hypotheses of ordinary least squares regression makes it easy to calculate the variance of the slope estimator in a simple regression (across all possible samples that might be drawn)?
A. Expected value of the regression error term is zero.
B. Covariances among different error terms are zero.
C. Variance of the error term is constant.
D. B. and C
E. None of the above

9. If you reject a joint null hypothesis using the F-test in a multiple hypothesis setting, then
A. a series of *t*-tests may or may not give you the same conclusion.
B. the *F*-statistic must be negative.
C. all of the hypotheses are always simultaneously rejected.
D. the regression is always significant
E. None of the above

10. The slope estimator, β_2, has a smaller standard error, other things equal, if
A. there is more variation in the explanatory variable, *X*.
B. there is a large variance of the error term, *u*.
C. the intercept, β_1 , is small.
D. the sample size is smaller.
E. All the above

2) Consider the linear regression: $Y_i = \beta_0 + \beta_1 X_i + u_i$, i = 1, ..., n.
Suppose that the following assumptions hold:
A1. (X_i, u_i) are i.i.d. for i = 1, 2, ..., n
A2. Both X_i and u_i are continuous random variables
A3. $E[u_i|X_i] = 0$ with probability one for i = 1, ..., n
A4. $E[X^4_i] < \infty$ and $E[u^4_i] < \infty$, i.e., large outliers are rare.
A5. $Var(u_i|X_i) = \sigma^2_u$ for every i, i.e., u_i is homoskedastic
Suppose further that you have decided to estimate the slope coefficient β_1 using the estimator

$$\widetilde{\beta}_1 = \frac{1}{3}\frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}$$

a) Derive a formula for the conditional variance $Var(\tilde{\beta}_1 \mid X1, ..., Xn)$ of $(\tilde{\beta}_1)$
b) Compare the conditional variance formula you have derived in part (a) with the conditional variance of the OLS estimator, i.e., $Var(\hat{\beta}_1 \mid X1, ..., Xn)$ of $\hat{\beta}_1$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}.$$

Which estimator has the bigger conditional variance $\tilde{\beta}_1$ or $\hat{\beta}_1$? Is this a violation of the Gauss-Markov Theorem? Why or why not?

(4.0)

3) Suppose a researcher was interested in how well ones height predicts ones weight. Hence, he will collect data on 110 people in order to run the simple regression. The Model $Y_i = \beta_1 + \beta_2 X_i + u_i$, i = 1, ..., 110. where Yi denotes the weight of the ith individual and Xi denotes the height of the ith individual. From this data, you calculated the following quantities:

$$\sum_{i=1}^{110} Y_i = 17375 \quad \text{and} \quad \sum_{i=1}^{110} \left(Y_i - \overline{Y}\right)^2 = 94228.8\ .\ \sum_{i=1}^{110} X_i = 7665.5 \quad \sum_{i=1}^{110} \left(X_i - \overline{X}\right)^2 = 1248.9$$

$\sum_{i=1}^{110} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = 7625.9$ ……………where Y Bar and X Bar denote the respective sample means.

a) From the information given above, calculate the OLS estimates for β_1 and β_2. Show your work.

b) Calculate the (unadjusted) R^2 measure for this regression and explain its meaning. Show your work.

(4.0)

4) An econometrician estimated the following demand equation: $\ln q_t = \beta_0 + \beta_1 \ln p_t + \varepsilon_t$ (1)

where q_t is the per capital quantity consumed in period t (measured in grams) and p_t is the price in period t (measured in rupees/gram). Suppose a sample of data for 42 time periods is used to obtain the following estimated demand equation (with standard errors in parentheses): $\ln \hat{q}_t = -5.0 - 0.4 \ln p_t$ (2)

(1.6) (0.2375)

The other values are: $\hat{\sigma}^2 = 0.18, R^2 = 0.49$

Answer the following questions.

a) Interpret β_1 in (1) using q_t and p_t. Interpret $R^2 = 0.49$ in (2).

b) Test the following hypothesis using the significance test approach at the 1% significance level. The critical value of the test statistic is 2.423

$H_0 : \beta_1 \geq 0$

$H_1 : \beta_1 < 0$

What happens if one uses the interval approach at the 10% significance level? Find the p-value of the test statistic.

d) If q_t is now measured in kilograms, how do the change affect the two estimates in Model (2). Give your reasoning.

(6.0)

5) Suppose you regress Y on X_2, X_3 X_4, and X_5 as following:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_1$$

Y = the number of Oil wells drilled (Thousands); X_2 = price at the wellhead in the previous period (in constant Rupees)
X_3 = domestic output (Rs. millions of barrels per day); X_4 = GNP constant rupees (Rs. billions)
X_5 = trend variable
The regression result obtained from software is

Dependent Variable: Y
Method: Least Squares
Date
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.854596	8.895196	-1.107856	0.2781
X2	2.701012	0.695769		0.0006
X3		0.937314	3.264226	0.0031
X4	-0.016060	0.008179		0.0604
X5	-0.022701	0.272306	-0.083368	0.9342

R-squared	0.580377	Mean dependent var	10.64613
Adjusted R-squared	0.515819	S.D. dependent var	2.351515
S.E. of regression		Akaike info criterion	3.969390
Sum squared resid	69.61077	Schwarz criterion	4.200678
Log likelihood	-56.52554	F-statistic	
Durbin-Watson stat	0.933888	Prob(F-statistic)	0.000107

(a) Fill in the missing numbers due to the malfunction of printer.

(b) How would you interpret this result is good or not? How would you interpret the coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$?

(c) Would you reject the hypothesis that the domestic output (X_3) has the effect of 3.00 on wildcat drilled (Y)? And why you are using the t-test but not using the normal distribution test? The critical value of the test statistic is 2.056 at appropriate degrees of freedom.

(6.0)