BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

SECOND SEMESTER 2020-2021

ECON F241 ECONOMETRIC METHODS

PROBLEM SHEET 1 15 FEB 2021

Multiple Choice Questions

1. If E denotes the expectation the variance of a random variable X is denoted as?
   a) $(E(X))^2$
   b) $E(X^2)-(E(X))^2$
   c) $E(X^2)$
   d) $2E(X)$

2. The random variables X and Y have variances 0.4 and 0.5 respectively. Let Z= 5X-2Y. The variance of Z is?
   a) 3
   b) 8
   c) 12
   d) 1

3. Which of the following statements best describes the relationship between a parameter and a statistic?
   a) A parameter has a sampling distribution with the statistic as its mean.
   b) A parameter has a sampling distribution that can be used to determine what values the statistic is likely to have in repeated samples.
   c) A parameter is used to estimate a statistic.
   d) A statistic is used to estimate a parameter.

4. Which of the following statements best explains the expected value of a random variable
   a) value that has the highest probability of occurring.
   b) mean value of the number of observations made available to us
   c) mean value over an infinite number of observations of the variable.
   d) most common value over an infinite number of observations of the variable.

5. Which of the following is equal to Cov(X + Y, X − Y), where X and Y are random variables on a sample space S?
   a) Var(X) − Var(Y)
   b) $Var(X^2) - Var(Y^2)$
   c) $Var(X^2) + 2Cov(X,Y) + Var(Y^2)$
   d) $Var(X^2) - 2Cov(X,Y) + Var(Y^2)$

6. Which of the following is equal to Var(2X −3Y), where X and Y are random variables on S?
   a) 4Var(X) + 12Cov(X, Y) + 9Var(Y)
   b) 2Var(X) − 3Var(Y)
   c) 2Var(X) + 6Cov(X,Y) + 3Var(Y)
   d) 4Var(X) − 12Cov(X,Y) + 9Var(Y)
   e) 2Var(X) − 6Cov(X, Y) + 3Var(Y)

7. The regression model includes a random error or disturbance term for a variety of reasons. Which of the following is NOT one of them?

a) measurement errors in the observed variables
b) omitted influences on Y (other than X)
c) linear functional form is only an approximation
d) the observable variables do not exactly correspond with their theoretical counterparts
e) there may be approximation errors in the calculation of the least squares estimates

8. Which of the following assumptions about the error term is not part of the so called "classical assumptions"?

a) it has a mean of zero
b) it has a constant variance
c) its value for any observation is independent of its value for any other observation
d) it is independent of the value of X
e) it has a normal distribution

9. In a linear regression problem, we are using “R-squared” to measure goodness-of-fit. We add a feature in linear regression model and retrain the same model. Which of the following option is true?

a) If R Squared increases, this variable is significant
b) If R Squared decreases, this variable is not significant
c) Individually R squared cannot tell about variable importance. We can’t say anything about it right now
d) None of these

10. A fitted regression equation is given by Yhat = 20 + 0.75X. What is the value of the residual at the point X=100, Y=90?

a) 5
b) -5
c) 10
d) -10

11. The least squares estimator of the slope coefficient is unbiased means:

a) the estimated slope coefficient will always be equal to the true parameter value
b) the estimated slope coefficient will get closer to the true parameter value as the size of the sample increases
c) the estimated slope coefficient will be equal to the true parameter if the sample is large
d) the mean of the sampling distribution of the slope parameter is zero
e) if repeated samples of the same size are taken, on average their value will be equal to the true parameter

12. Which one of the statement is true regarding residuals in regression analysis?

a) Mean of residuals is always zero
b) Mean of residuals is always less than zero
c) Mean of residuals is always greater than zero
d) There is no such rule for residuals.

Conceptual questions

1. Why does the sample variance underestimate the population variance? Tell qualitatively.

2. Correlations vs Covariance. Explain the difference.

3. Explain qualitatively and quantitative why R squared never decreases on adding more variables.

4. An estimator is a random variable that is calculated from a random sample that either provides a point estimate or an interval estimate for some population parameter. T/F

5. If a random sample of *n* measurements is selected from a population with mean 'mu' the expected value of the sample mean is always equal to 'mu' regardless of the sample size n. T/F

Numericals/proofs/short answer type

1. Let E[X] = 1, E[X2] = 3, E[XY] = −4 and E[Y] = 2. Find Cov(X, 2X + Y).

2. Let $Z_1, Z_2, \ldots, Z_n$ be independent normal random variables with mean 0 and variance 1. Let $\chi = Z_1^2 + \cdots + Z_n^2$. Using that χ is the sum of independent random variables, compute both the mean and variance of χ.

3. Suppose a variable Y is an exact linear function of X:

$$Y = \lambda + \mu X$$

where λ and μ are constants, and suppose that Z is a third variable. Show that $\rho_{XZ} = \rho_{YZ}$

4. Show that, when you have n observations, the condition that the generalised estimator (λ1X1 + · · · + λnXn) should be an unbiased estimator of μX is λ1 + · · · + λn = 1.

A researcher has international cross-section data on aggregate wages, W, aggregate profits, P, and aggregate income, Y, for a sample of n countries. By definition,

$$Y_i = W_i + P_i$$

The regressions

$$\hat{W}_i = a_1 + a_2 Y_i$$
$$\hat{P}_i = b_1 + b_2 Y_i$$

are fitted using OLS regression analysis. Show that the regression coefficients will automatically satisfy the following equations:

$$a_2 + b_2 = 1$$
$$a_1 + b_1 = 0$$

5. Explain intuitively why this should be so.

6. We change the units of measurement of X such that the new measure, A, is related to the original one by $A_i = \mu_2 X_i$. Show that the new estimate of the slope coefficient is β_2/μ_2, where β_2 is the slope coefficient in the original regression.

7. BITS Pilani's Placement Unit is trying to figure out how much coding practice students need so as to get a good job in IT. They conduct a survey and their analysis gives them the following equation:

$$CTC(hat) = 1{,}181.33 + 0.967 * CCHrs \qquad R^2 = 0.973$$

Where CTC is in '000 Rs. and CCHrs is in number of hours.

a) Interpret the coeefficients and comment on their sign and magnitude.
b) As a student you are trying to decide whether it is worth putting in effort in coding practice. From this analysis, what decision should you make. How much variation in the CTC can be attributed to the number of hours spent in competitive coding (CC)?
c) Ashok spent 3 hrs daily for 2 weeks doing competitive coding (CC). What CTC can he expect?

Interesting read / Case Study

In the introductory lectures, we discussed the importance and applications of econometrics. Here is a very interesting article which talks about how critical econometric modelling has become to the financial industry.

Recipe for Disaster: The Formula That Killed Wall Street – Felix Salmon

Link: https://www.wired.com/2009/02/wp-quant/