

Problem Set 3

a) Multiple choice questions

1. The term ' u ' in an econometric model is usually referred to as the _____.
 - a. error term
 - b. parameter
 - c. hypothesis
 - d. dependent variable

2. Which of the following is a nonlinear regression model?
 - a. $y = \beta_0 + \beta_1 x^{1/2} + u$
 - b. $\log y = \beta_0 + \beta_1 \log x + u$
 - c. $y = 1 / (\beta_0 + \beta_1 x) + u$
 - d. $y = \beta_0 + \beta_1 x + u$

3. Which of the following is true of R^2 ?
 - a. R^2 is also called the standard error of regression.
 - b. A low R^2 indicates that the Ordinary Least Squares line fits the data well.
 - c. R^2 usually decreases with an increase in the number of independent variables in a regression.
 - d. R^2 shows what percentage of the total variation in the dependent variable, Y , is explained by the explanatory variables.

4. The normality assumption implies that:
 - a. the population error u is dependent on the explanatory variables and is normally distributed with mean equal to one and variance σ^2 .
 - b. the population error u is independent of the explanatory variables and is normally distributed with mean equal to one and variance σ .
 - c. the population error u is dependent on the explanatory variables and is normally distributed with mean zero and variance σ .
 - d. the population error u is independent of the explanatory variables and is normally distributed with mean zero and variance σ^2 .

5. The general t statistic can be written as:
 - a. $t = \frac{\text{Hypothesized value}}{\text{Standard error}}$
 - b. $t = (\text{estimate} - \text{hypothesized value})$
 - c. $t = \frac{(\text{estimate} - \text{hypothesized value})}{\text{variance}}$

d. $t = \frac{(\text{estimate} - \text{hypothesized value})}{\text{standard error}}$

6. If the error term is correlated with any of the independent variables, the OLS estimators are:

- a. biased and consistent.
- b. unbiased and inconsistent.
- c. biased and inconsistent.
- d. unbiased and consistent.

7. A _____ variable is used to incorporate qualitative information in a regression model.

- a. dependent
- b. continuous
- c. binomial
- d. dummy

8. Which of the following Gauss-Markov assumptions is violated by the linear probability model?

- a. The assumption of constant variance of the error term.
- b. The assumption of zero conditional mean of the error term.
- c. The assumption of no exact linear relationship among independent variables.
- d. The assumption that none of the independent variables are constants

9. A test for heteroskedasticity can be significant if _____.

- a. the Breusch-Pagan test results in a large p -value
- b. the White test results in a large p -value
- c. the functional form of the regression model is misspecified
- d. the regression model includes too many independent variables

10. Consider the following simple regression model $y = \beta_0 + \beta_1 x_1 + u$. The variable z is a poor instrument for x if _____.

- a. there is a high correlation between z and x
- b. there is a low correlation between z and x
- c. there is a high correlation between z and u
- d. there is a low correlation between z and u

b) Answer the following questions

- 1) Explain the meaning of “statistically significant” variable.
- 2) Describe the 2SLS method.
- 3) Explain the difference(s) between the linear probability model and probit/logit model.

c) True/False questions: indicate whether the statement is true or false.

- a. An economic model consists of mathematical equations that describe various relationships between economic variables.
- b. The variance of the slope estimator increases as the error variance decreases.
- c. A larger error variance makes it difficult to estimate the partial effect of any of the independent variables on the dependent variable.
- d. Whenever the dependent variable takes on just a few values it is close to a normal distribution.
- e. The multiple linear regression model with a binary dependent variable is called the linear probability model.
- f. The two stage least squares estimator is less efficient than the ordinary least squares estimator when the explanatory variables are exogenous.

Part 4: Solve the problem

Consider a model of the probability of being a CEO as a function of gender (variable $FEMALE=1$ if the individual is a woman, 0 otherwise), education (variable $EDUC$ =years of education), age (variable AGE) and a constant.

1. Assume that a linear probability model estimates the model. Express *Probability* ($Y=1|X$) as a function of the respective variables and unknown coefficients as implied by this model.
2. Suppose that the coefficient of AGE is estimated to be 0.045. What is its interpretation?

3. How would you adjust a model if you want to test whether education increases chance of being a CEO for women more compared to men? State the name of the test that you will use. State the null and alternative hypothesis and the rejection rule.
4. In the model from question (3) what is the effect of education on the probability of being a CEO for a woman? Write down the effect in terms of coefficients and/or variables.
5. Suppose now that the model is estimated by a probit model. Express *Probability* ($Y=1|X$) as a function of the respective variables and unknown coefficients as implied by this model.
6. How would you predict a probability of being a CEO of a 45 years old woman with 20 years of education?
7. Prediction tables from the LPM and the probit models are shown below:

LPM	$\hat{Y}_i = 0$	$\hat{Y}_i = 1$
$Y_i = 0$	600	250
$Y_i = 1$	100	50

Probit	$\hat{Y}_i = 0$	$\hat{Y}_i = 1$
$Y_i = 0$	530	300
$Y_i = 1$	100	70

Which of the two models has the highest percentage of correctly predicted outcomes? Which model is the best in predicting the outcome $Y_i=1$?