Active Learning

a) Multiple choice questions

1. A dependent variable is also known as a ___.

a. explanatory variable
b. control variable
c. predictor variable
d. response variable

2. In the equation y = β_0 + β_1x + u, β_0 is the __

a. dependent variable
b. independent variable
c. slope parameter
d. intercept parameter.

3. What does the equation $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x$ denote if the regression equation is y = β_0 + $\beta_1 x_1$ + u?
a. The explained sum of squares
b. The total sum of squares
c. The sample regression function
d. The population regression function

4. If the total sum of squares (SST) in a regression equation is 81, and the residual sum of squares (SSR) is 25, what is the explained sum of squares (SSE)?
a. 64
b. 56
c. 32
d. 18

5. If the residual sum of squares (SSR) in a regression analysis is 66 and the total sum of squares (SST) is equal to 90, what is the value of the R-squared?
a. 0.73
b. 0.55
c. 0.27
d. 1.2

6. The error term in a regression equation is said to exhibit homoskedasticty if _____.
a. it has zero conditional mean
b. it has the same variance for all values of the explanatory variable.
c. it has the same value for all values of the explanatory variable
d. if the error term has a value of one given any value of the explanatory variable.

7. In the equation, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, β_2 is a(n) _____.
a. independent variable
b. dependent variable
c. slope parameter
d. intercept parameter

8. Consider the following regression equation: $y = \beta_1 + \beta_2 x_1 + \beta_2 x_2 + u$. What does β_1 imply?
a.β_1 measures the ceteris paribus effect of $x_1$on x_2.
b. β_1 measures the ceteris paribus effect of y on x_1.
c. β_1 measures the ceteris paribus effect of $x_1$on y.
d. β_1 measures the ceteris paribus effect of $x_1$on u.

9. High (but not perfect) correlation between two or more independent variables is called _____.
a. heteroskedasticty
b. homoskedasticty
c. multicollinearity
d. micronumerosity

10. 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.

11. Which of the following is a statistic that can be used to test hypotheses about a single population parameter?
a. F statistic
b. t statistic
c. χ^2 statistic
d. Durbin-Watson statistic

12. Consider the equation, $Y = \beta_1 + \beta_2 X_2 + u$. A null hypothesis, $H_0: \beta_2 = 0$ states that:
a. X_2 has no effect on the expected value of β_2.
b. X_2 has no effect on the expected value of Y.
c. β_2 has no effect on the expected value of Y.
d. Y has no effect on the expected value of X_2.

13. The significance level of a test is:
a. the probability of rejecting the null hypothesis when it is false.
b. one minus the probability of rejecting the null hypothesis when it is false.
c. the probability of rejecting the null hypothesis when it is true.
d. one minus the probability of rejecting the null hypothesis when it is true.

14. Which of the following tools is used to test multiple linear restrictions?
a. t test
b. z test
c. F test
d. Unit root test

15. Which of the following is true of heteroskedasticity?
a. Heteroskedasticty causes inconsistency in the Ordinary Least Squares estimators.

b. Population R2 is affected by the presence of heteroskedasticty.
c. The Ordinary Least Square estimators are not the best linear unbiased estimators if heteroskedasticity is present.
d. It is not possible to obtain F statistics that are robust to heteroskedasticity of an unknown form.

16. Which of the following is true of the OLS t statistics?
a. The heteroskedasticity-robust t statistics are justified only if the sample size is large.
b. The heteroskedasticty-robust t statistics are justified only if the sample size is small.
c. The usual t statistics do not have exact t distributions if the sample size is large.
d. In the presence of homoskedasticity, the usual t statistics do not have exact t distributions if the sample size is small.

17. Which of the following tests helps in the detection of heteroskedasticity?
a. The Breusch-Pagan test
b. The Breusch-Godfrey test
c. The Durbin-Watson test
d. The Chow test

18. A test for heteroskedasticty 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

19. Which of the following is true of the White test?
a. The White test is used to detect the presence of multicollinearity in a linear regression model.
b. The White test cannot detect forms of heteroskedasticity that invalidate the usual Ordinary Least Squares standard errors.
c. The White test can detect the presence of heteroskedasticty in a linear regression model even if the functional form is misspecified.
d. The White test assumes that the square of the error term in a regression model is uncorrelated with all the independent variables, their squares and cross products.

20. Consider the following simple regression model: $y = \beta0 + \beta1x1 + u$. Suppose z is an instrument for x. Which of the following conditions denotes instrument exogeneity?
a. $Cov(z,u) > 0$
b. $Cov(z,x) > 0$
c. $Cov(z,u) = 0$
d. $Cov(z,x) = 0$

21. Consider the following simple regression model $y=\beta0 + \beta1x1 + u$. Suppose z is an instrument for x. Which of the following statements is true?
a. The condition $Cov(z,u) = 0$ can be tested statistically.
b. The condition $Cov(z,x) \neq 0$ cannot be tested statistically.
c. The ordinary least squares estimator is always biased if $Cov(x,u)\neq0$.
d. The ordinary least squares estimator is unbiased if $Cov(x,u)\neq0$.

22. Consider the following simple regression model $y=\beta0 + \beta1x1 + u$. 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

23. Which of the following assumptions is known as exclusion restrictions?
a. The assumption that an instrumental variable is excluded from a regression model and is correlated with the error term
b. The assumption that an instrumental variable is excluded from a regression model and correlated with an exogenous explanatory variable
c. The assumption that an exogenous explanatory variable is excluded from a regression model and is uncorrelated with the error term
d. The assumption that an endogenous explanatory variable excluded from a regression model and is uncorrelated with the error term

24. In the equation c = β_0 + β_1i + u, c denotes consumption and i denotes income. What is the residual for the 5th observation if c_5=$500 and $\widehat{c_5}$=$475?

a. $975

b. $300

c. $25

d. $50

25. In a regression model, which of the following will be described using a binary variable?

a. Whether it rained on a particular day or it did not

b. The volume of rainfall during a year

c. The percentage of humidity in air on a particular day

d. The concentration of dust particles in air

26. The binary dependent variable model is an example of a

a. regression model, which has as a regressor, among others, a binary variable.

b. model that cannot be estimated by OLS.

c. limited dependent variable model.

d. model where the left-hand variable is measured in base 2

27. In the linear probability model, the interpretation of the slope coefficient is

a. the change in odds associated with a unit change in X, holding other regressors constant.

b. not all that meaningful since the dependent variable is either 0 or 1.

c. the change in probability that Y=1 associated with a unit change in X, holding others regressors constant.

d. the response in the dependent variable to a percentage change in the regressor.

b) Conceptual questions

1) Explain what does unbiasedness means.
2) Explain what does exogenous and endogenous variable means.
3) Define and explain the concept of R-squared.
4) Explain what happens when we omit important variable from the model.
5) Explain the concept of the confidence intervals.
6) Describe the heteroskedasticity problem: explain briefly, what it is and how it affects the estimation.
7) Name and discuss the steps of one of the tests for heteroscedasticity and discuss its flows.
8) Describe Linear Probability Model and state its advantages and disadvantages.
9) Explain the concept of an instrumental variable.
10) Describe the properties that valid instrumental variable should satisfy.

c) True/False questions

1. The ideal way to analyze causation is to use experimental data.
2. The variance of the slope estimator increases as the error variance decreases.
3. If the calculated value of the *t-statistic* is greater than the critical value, the null hypothesis, H_0 is rejected in favor of the alternative hypothesis, H_1.
4. One advantage of the Linear Probability Model is that it is easy to estimate and to interpret.
5. Under the Gauss-Markov assumptions, OLS estimators are BLUE.
6. Sum of squared residuals (SSR) is greater in the unrestricted model.
7. For estimating population parameters more precisely one would prefer to have low variation in explanatory variables and high variation in error term.
8. In the Linear Probability Model (LPM) predicted values are always between 0 and 1.

9. The dummy variable coefficient for a particular group represents the estimated difference in intercepts between that group and the base group.
10. If the heteroskedasticity is present in the model, than estimates are no longer unbiased.

d) Solve the problem

Consider equation relating education, experience and race of the person to her wage in the following manner:

$$\ln(wage) = \beta_0 + \beta_1 * educ + \beta_2 * exper + \beta_3 * exper^2 + \beta_4 * black + u$$

where the variables are defined as follows:

wage	Wage of a person in USD
educ	Number of years of education
exper	Work experience in years
black	1 if the person is black, 0 otherwise

a) Explain what signs you expect for each of the parameters.
b) Suppose the equation is estimated using a random sample of 935 workers. Argue whether all of the Classical Linear Model assumptions will be satisfied.
c) The estimated output of the equation is presented below. Interpret all the coefficients. Do they have expected signs? What percentage of variation in wages do we explain using our controls?
d) According to the output, what factors are important determinants of wage?
e) Test the overall significance of the regression: explicitly state the null and alternative hypothesis and the result of the test.

f) Suppose you want to test the hypothesis that returns to education is lower for black people. How would you test it?

Model 1: OLS, using observations 1-935
Dependent variable: l_wage

	Coefficient	*Std. Error*	*t-ratio*	*p-value*
const	5.64834	0.124761	45.2733	XXXXX
educ	0.0714961	0.00660733	10.8207	XXXXX
exper	0.0139896	0.0133153	1.0506	XXXXX
sq_exper	0.000225854	0.000557475	0.4051	XXXXX
black	−0.222567	0.0384461	−5.7891	XXXXX

Mean dependent var	6.779004	S.D. dependent var	0.421144
Sum squared resid	138.9600	S.E. of regression	0.386548
R-squared	0.161155	Adjusted R-squared	0.157547
F(4, 930)	44.66678	P-value(F)	2.47e-34
Log-likelihood	−435.4839	Akaike criterion	880.9679
Schwarz criterion	905.1706	Hannan-Quinn	890.1966