Imtahan Suallari

1. Discuss the structure of Economic data: Cross Sectional Data, Time Series Data and Panel or Longitudinal Data
2. Provide one example for each structure of Data that is used in Econometric Analysis. Discuss the problems that are associated with Time Series Data.
3. We have simple linear regression Model, which estimate the effect of training on Wage level as following:

Explain the factors that can be included in u. Why do we assume that expectation of error term is 0? Is it strong assumption? Explain the coefficients of above Model. Do you expect correlation between training and u? Why and why not?

1. Simple Linear Regression Model which estimate the effect of fertilizers on production level of crop is given as following:

Explain the factors that can be included in u. Why do we assume that expectation of error term is 0? Is it strong assumption? Explain the coefficients of above Model. Do you expect correlation between fertilizer and u? Why and why not?

Construct fitted value and residual graph for the following cases.

1. U=2.3 , training=5
2. U=-0.8 , training=4
3. U=1, training=3
4. U=-2, training=7

Discuss the meaning of fitted value and error term.

Construct fitted value and residual graph for the following cases.

1. U=2 , fertilizer=4
2. U=-1.5 , fertilizer=5
3. U=-2, fertilizer=3
4. U=3, fertilizer=7

Discuss the meaning of fitted value and error term.

1. Discuss total sum of squares (SST), Explained sum of squares (SSE) and Residual sum of squares (SSR).
2. Discuss the 3 Assumptions that we made for Simple Linear Regression. Discuss the potential problems related to these assumptions.
3. Drive coefficients by applying OLS method.
4. What is predicted birth wight when *cigs*=0? What about when cigs=10? Comment on the difference
5. To predict a birth weight of 85 ounces, what would be cigs have to be? Comment
6. Does this simple regression necessarily captures a causal relationship between the child´s birth weight and smoking habit?
7. Interpret the coefficient of the following functional forms:
8. List and explain the Gauss-Markov Assumptions for Simple Regression. Explain the importance of these Assumptions. What is unbiased estimators?
9. Let`s kids denote the number of children ever born to a woman and let *educ* denote years of education for the woman. A simple model relating fertility to years of education is

1. What kinds of factors are contained in u? Are these likely to be correlated with level of education?
2. Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain

|  |  |
| --- | --- |
| GPA | ACT |
| 3 | 4 |
| 2 | 3 |
| 4 | 3 |
| 4 | 6 |

1. Estimate the relationship between GPA and ACT using OLS. Obtain the intercept and slope estimates in the equation

Comment on the direction of relationship and interpret the result.

|  |  |
| --- | --- |
| Crop yields | Fertility |
| 10 | 2 |
| 3 | 1 |
| 3 | 4 |
| 4 | 5 |

Estimate the relationship between Crop Yields and Fertility using OLS. Obtain the intercept and slope estimates in the equation

 Comment on the direction of relationship and interpret the result.

1. Complete the following for the (*x*, *y*) pairs of data points (1, 5), (3, 7), (4, 6), (5, 8), and (7, 9).
a. Prepare a scatter plot of these data points.
b. Compute *b*1.
c. Compute *b*0.
d. What is the equation of the regression line?
2. Sample Dataset is given as following:

 4, 1, 1, 3, 3, 5, 15, 20, 30

1. Find median and variance
2. Find 1st, 2nd and 3rd Quartiles
3. What is shape of this distribution?
4. The profit for a production process is equal to $1,000 minus two times the number of units produced. The mean and variance for the number of units produced are 50 and 90, respectively. Find the mean and variance of the profit

19. Complete the following for the (*x*, *y*) pairs of data points (2, 5), (4, 7), (1, 6) and (5, 6)
 a. Prepare a scatter plot of these data points.
 b. Compute *b*1.
 c. Compute *b*0.
 d. What is the equation of the regression line?

e. What is expected change in y variable if x changes by 2 units?

20. A manufacturer of detergent claims that the contents of boxes sold weigh on average at least 16 ounces. The distribution of weight is known to be normal, with a standard deviation of 0.4 ounce. A random sample of 16 boxes yielded a sample mean weight of 15.84 ounces. Test at the 10% significance level the null hypothesis that the population mean weight is at least 16 ounces.

21. A company that receives shipments of batteries tests a random sample of nine of them before agreeing to take a shipment. The company is concerned that the true mean lifetime for all batteries in the shipment should be at least 50 hours. From past experience it is safe to conclude that the population distribution of lifetimes is normal with a standard deviation of 3 hours. For one particular shipment the mean lifetime for a sample of nine batteries was 48.2 hours. Test at the 10% level the null hypothesis that the population mean lifetime is at least 50 hours.

22. A random sample of n = 25 is obtained from a population with variance s2, and the sample mean is computed. Test the null hypothesis H0: m = 100 versus the alternative hypothesis H1: m >100 with a = 0.05. Compute the critical value xc and state your decision rule for the following options.

a. The population variance is s2= 225.

b. The population variance is s2 = 900.

c. The population variance is s2 = 400.

d. The population variance is s2 = 600.

23. A random sample of 172 marketing students was asked to rate, on a scale from 1 (not important) to 5 (extremely important), health benefits as a job characteristic. The sample mean rating was 3.31, and the sample standard deviation was 0.70. Test at the 1% significance level the null hypothesis that the population mean rating is at most 3.0 against the alternative that it is larger than 3.0.

24. A random sample of 1,562 undergraduates enrolled in management ethics courses was asked to respond on a scale from 1 (strongly disagree) to 7 (strongly agree) to this proposition: Senior corporate executives are interested in social justice. The sample mean response was 4.27, and the sample standard deviation was 1.32. Test at the 1% level, against a two-sided alternative, the null hypothesis that the population mean is 4.

25. The accounts of a corporation show that, on average, accounts payable are $125.32. An auditor checked a random sample of 16 of these accounts. The sample mean was $131.78 and the sample standard deviation was $25.41. Assume that the population distribution is normal. Test at the 5% significance level against a two sided alternative the null hypothesis that the population mean is $125.32.

26. Dicsuss the ``Partialling-Out“ in Multiple Regression context and the reasons for its usage.

27. Discuss the main Assumptions for unbiased OLS estimators.

28. We have multiple linear regression Model, which estimate the effect of training and education on Wage level as following:

Explain the factors that can be included in u. Why do we assume that expectation of error term is 0? Is it strong assumption? Explain the coefficients of above Model. Do you expect correlation between education and u? Why and why not? If there is a correlation between training and education, how would it affect the estimators and what solutions do you suggest for this problem?

29. Simple Linear Regression Model which estimate the effect of fertilizers on production level of crop is given as following:

 Ln(

Explain the factors that can be included in u. Why do we assume that expectation of error term is 0? Is it strong assumption? Explain the coefficients of above Model. Do you expect correlation between fertilizer and land quality? Why and why not? If there is a correlation between training and education, how would it affect the estimators and what solutions do you suggest for this problem?

30. Consider the following equation.

a) Interpret the coefficient of each variable and error term. What are the necessary assumptions that we made about error term in order to get unbiased estimators.

b) Do you expect the correlation between IQ level and education? Why does the existence of correlation torture the estimation? How would you like to solve this problem?

c) Let`s assume that regression of Education on IQ level is given as following:

Education1=11 IQ level1=20

Education2=12 IQ level2=15

Education3=10 IQ level3=50

Based on above information find the estimated wage level for 1, 2 and 3rd observations.

31. Consider the following equation.

a) Interpret the coefficient of each variable and error term. What are the necessary assumptions that we made about error term in order to get unbiased estimators.

b) Do you expect the correlation between Income and Expenditure? Why does the existence of correlation torture the estimation? How would you like to solve this problem?

c) Let`s assume that regression of Expenditure on Income level is given as following:

Expenditure1=13 Income1=21

Expenditure2=10 Income2=10

Expenditure3=25 Income3=50

Based on above information find the estimated level of Economic Growth for 1, 2 and 3rd observations.

32. The following model is used to estimate the effect of some variables on price level of houses in a certain street:

a) What is expected the sign of and?

b) Do you expect the multicollinearity problem in this estimation model?

c) If and =0.89, interpret these coefficients?

d) If and =0.67. Number of bedrooms and square feet is 4 and 90 respectively. Old number of bedrooms and square feet is 3 and 45. What is the change in the value of price?

e) Discuss perfect collinearity and Homoscedasticity assumptions.

33. What is multicollinearity problem? What problems do they create for unbiasedness of OLS estimators? Discuss the ways to solve multicollinearity problem in regression analysis.

34. What is misspecification problem? What problems does it create for OLS estimators? Discuss the ways to overcome this problem.

35. Explain the test that we use to check the significance of each estimators in multiple linear regression. Furthermore, discuss the test that we use for checking joint significance. Explain your answer.

36. Discuss the homoscedasticity and normally distributed residuals assumptions in multiple regression context. Why do they important? Discuss the importance of perfect collinearity assumption.

37. What are Interaction terms? What is interpretation of them? Why do we use it in economic and business applications?

38. Consider the following regression:

a) Interpret the each coefficient.

b) Calculate the change in the wage level if education increases by 3 years and experience increases 1.5 years.

c) Calculate the change in wage level if experience decreases 1.5 years and education increases by 1.5 years.

d) Do you expect multicollinearity problem in the above equation? Why and Why not?

39. Discuss R2 the coefficient of determination. Discuss the adjusted R2.

40.

1. What is predicted birth rate when *cigs*=0 and no vitamins? What about when cigs=15 and vitamins=7? Comment on the difference
2. To predict a birth weight of 125 ounces, what would be vitamins have to be? Comment
3. Explain the error term
4. Do you expect multicollinearity problem in the above equation? Why and Why not?

41. Consider the following Regression results.



a) Interpret the coefficients of this regression. Which variables are dummy?

b) Which variables is significant at 5% significance level. Check significance at 60% significance level. Formulate Null and Alternative hypothesis for each variable.

c) Write the estimated equation based on the information given in the above. Interpret confidence interval of Education and Male. Construct 98% confidence Interval for Education and Male.

d) Interpret the R2. Can this model be used to predict wage level? Why adjusted R square and R square is different in above mentioned regression? Discuss.

42. Consider the following Regression results.



a) Interpret the coefficients of this regression. Which variables are dummy?

b) Which variables is significant at 2% significance level. Check significance at 44% significance level. Formulate Null and Alternative hypothesis for each variable.

c) Write the estimated equation based on the information given in the above. Interpret confidence interval of Age and Married. Construct 90% Confidence Interval for Children and Married variable.

d) Interpret the R2. Can this model be used to predict wage level? Why adjusted R square and R square is different in above mentioned regression? Discuss.

43. Consider the following Regression results.



a) Interpret the coefficients of this regression. Which variables are dummy?

b) Which variables is significant at 8% significance level. Check significance at 60% significance level. Formulate Null and Alternative hypothesis for each variable.

c) Write the estimated equation based on the information given in the above. Interpret confidence interval of Wage and Married. Construct 95% Confidence Interval based on South and Age variable.

d) Interpret the R2. Can this model be used to predict wage level? Why adjusted R square and R square is different in above mentioned regression? Discuss.

44. Consider the following Regression results.



a) Interpret the coefficients of this regression. Which variables are dummy?

b) Which variables is significant at 16% significance level. Check significance at 25% significance level. Formulate Null and Alternative hypothesis for each variable.

c) Write the estimated equation based on the information given in the above. Interpret confidence interval of Saving and Investment. Construct 80% Confidence Interval based for Saving and Investment.

d) Interpret the R2. Can this model be used to predict wage level? Why adjusted R square and R square is different in above mentioned regression? Discuss.

45. Consider the following Regression results.



a) Interpret the coefficients of this regression. Which variables are dummy?

b) Which variables is significant at 4% significance level. Check significance at 54% significance level. Formulate Null and Alternative hypothesis for each variable.

c) Write the estimated equation based on the information given in the above. Interpret confidence interval of Location and Investment. Construct 94% Confidence Interval for Location and Investment.

d) Interpret the R2. Can this model be used to predict wage level? Why adjusted R square and R square is different in above mentioned regression? Discuss.

46. Consider the following regression results.



a) Interpret the coefficients of this regression. Which variables are dummy?

b) Which variables is significant at 8% significance level. Check significance at 22% significance level. Formulate Null and Alternative hypothesis for each variable.

c) Write the estimated equation based on the information given in the above. Interpret confidence interval of Age and TQDKscore. Construct 95% Confidence Interval for studyinghours and TQDK score.

d) Interpret the R2. Can this model be used to predict wage level? Why adjusted R square and R square is different in above mentioned regression? Discuss.

47. Consider the following regression results.



a) Interpret the coefficients of this regression. Which variables are dummy?

b) Which variables is significant at 10% significance level. Check significance at 12% significance level. Formulate Null and Alternative hypothesis for each variable.

c) Write the estimated equation based on the information given in the above. Interpret confidence interval of TQDK score and Studying Hours. Construct 90% Confidence Interval for TQDK score and Studying hours.

d) Interpret the R2. Can this model be used to predict wage level? Why adjusted R square and R square is different in above mentioned regression? Discuss.

48. Discuss the dummy variable, categorical variable and interaction terms in multiple regression context. Explain the reasons that they are used for?

49.

Calculate the predicted wage level for the following cases.

1. training=5 and education=10, person is living in South and Male
2. training=4 and education=12, person lives in north and she is female
3. training=3 and education=8, person lives in South and she is female
4. training=7 and education=9, person lives in West and he is Male

Interpret the coefficients of each variable.

50.

Calculate the predicted average grade for the following cases. (Male and Capital is dummy variables)

1. Number of Hours=9, person studies in small city and he is Male
2. Number of Hours=7, student studies in capital and she is female
3. Number of Hours=8, student studies in capital and he is male
4. Number of Hours=15, student studies in small city and she is female

Interpret the coefficients of each variable.

**51**. Discuss the proxy variable and the reason that we use it. Moreover, explain conditions that are required for good proxy variable. Explain your answer by using one example.

**52**. Using the data in HPRICE1.RAW estimate two alternative models for housing prices. The First one has all variables in level form:

 Price = + \*lotsize + \*sqrft +\*bdrms + u (1)

And the second one uses the logarithms of all variables except bdrms :

 lnPrice = + \*lnlotsize + \*lnsqrft +\*bdrms + u (2)

a) Perform the RESET test for each model and explain the testing methodology.

b) Employ the testing methodology, for nonnested alternatives, offered by Davidson MacKinnon.

**53**. The Öle WAGE2.RAW contains information on monthly earnings, education, several demographic variables, and IQ scores for 456 men in 1980. As a method to account for omitted ability bias, we add IQ to a standard log(wage) equation:

log(wage) = + \*educ +\*exper + \*tenure + \*married +\*south +\*urban + \*black + u

log(wage) = + \*educ +\*exper + \*tenure + \*married +\*south +\*urban + \*black + \*IQ + u

a) Which variables are dummy variables? Interpret the coefficient of dummy variables.

b) Why do we use IQ variable in the second equation? Does it increase the precision of the model? Why and Why not?

**54.** What is functional form of misspecification? Does it cause biased estimator? Why and why not? Explain your answer by using example.

**55**. Explain the each of the following test separately.

a) Regression specification error test (RESET)

b) Davidson-MacKinnon Test.

**56.** Explain the ways to find the effect of interaction terms and quadratic terms in the regression. Moreover, explain the importance of using interaction and quadratic term.

**57**. Consider the following model.

Find the effect of education on wage level. If education=5 years, predict wage level. Furthermore, explain the error term. In which value of education, we have maximum wage level? Draw the graph.

**58**. Consider the following model.

Find the effect of investment on GDP level by fixing the effect of Savings. If investment=25 years, predict GDP level. Furthermore, explain the error term. In which value of Investment, we have maximum GDP level? Draw the graph.

**59.** Consider the following model.

a)

b)

c)

d)

Find the effect of education on wage level. If education=12 years, predict wage level. Furthermore, explain the error term. In which value of education, we have maximum or minimum wage level? Draw the graph

**60.** Consider the following model.

a)

b)

Find the effect of education on wage level. If education=15 years and tenure=5 years, predict wage level. Furthermore, explain the error term. In which value of education, we have maximum or minimum wage level if tenure=3 years? Draw the graph.

**61.** In the Neal-Johnson paper the log earnings of young American men is regressed on age and a Black and a Hispanic dummy. Here are the results (appropriately estimated standard errors in parentheses):

log(wage) = c + 0.12\*age – 1.23\*black – 0.67\*Hispanic

 (0.014) (0.026) (0.030)

a) What is the meaning of the black coefficient?

b) Does the coefficient on the black variable measure the extent of labor market discrimination? Why or why not?

Neal and Johnson re-estimated their regression with comprehensive test scores (AFQT ). The
test scores were measured at around age 18. They contain many skill elements, including IQ.
AFQT is normalized (mean=0, std=1) and is entered into the regression in a quadratic form.
c) Do you think AFQT can be a proxy for unobservable that cause trouble in the original
regression? If not, why? If yes, state exactly what it would proxy.

**62**. Suppose that you would like to estimate whether elite high schools add more value to their students than other high schools. Your data contains a test score on student competence measured at the end of high school, gender, parental education, and whether the student graduated from an elite high school.

(a) Write down a regression model that may enable you to estimate the effect of elite schools on student´s achievement.

(b) State the assumptions under which OLS consistently estimates the effect in this regression. Do you think they are satisfied in this case?

(c) Could you use elementary school grade point average as a proxy variable for unobservable? Would its inclusion lead to consistent estimation of the effect?

**63**. The question is returns to education. Education is endogenous because of unobserved ability. Give an example for a proxy variable. (State the definition of a proxy and argue that your choice satisfies it). Does including it into the model make the OLS estimator of the parameter on education consistent for the returns to education?

**64.** Consider the following regression table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Wage | coefficients | Standard Error | T-test | P-value | 95% Confidence Level |
| Constant | 350.54 | 45 |  \*\* | 0.2 | (2.34; 5.67) |
| Tenure | 43.56 | \*\* |  6.7 | 0.011 | (1.23; 9.43) |
| Experience | \*\* | 38.23 |  -2.1 | 0.34 | (-1.43; 5.46) |
| Education | 45.9 | 8.87 |  \*\* | 0.53 | (2.34: 7.82) |

a) Fill the \*\* in the above table.

b) Formulate Null hypothesis and check the significance of variables if significance level is 10%.

c) Write the estimated equation based on the information given above.

**65.** Consider the following regression result.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Wage | coefficients | Standard Error | T-test | P-value | 95% Confidence Level |
| Constant | 23.4 | 12.5 |  \*\* | 0.039 | (2.33; 5.67) |
| Tenure | 198.56 | \*\* |  3.5 | 0.21 | (1.23; 9.43) |
| Experience | \*\* | 38.23 |  7.6 | 0.09 | (-1.43; 5.46) |
| Education | 25.90 | \*\* |  4.85 | 0.15 | (2.34: 7.82) |

a) Fill the \*\* in the above table. Find standard deviation of tenure, experience and education if n=16.

b) Formulate Null hypothesis and check the significance of variables if significance level is 20 %.

c) Write the estimated equation based on the information given above.

**66.** Consider the following regression result.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Wage | coefficients | Standard Error | T-test | P-value | 95% Confidence Level |
| Constant | 8.56 | 6.7 |  \*\* | 0.04 | (5.67; 14.78) |
| lnTenure | 123.52 | \*\* |  6.5 | 0.32 | (1.23; 9.43) |
| lnExperience | \*\* | 4.65 |  8.6 | 0.12 | (-1.43; 5.46) |
| Education | 56.90 | \*\* |  4.85 | 0.005 | (2.34: 7.82) |

a) Fill the \*\* in the above table. Find standard deviation of tenure, experience and education if n=100.

b) Formulate Null hypothesis and check the significance of variables if significance level is 1 %.

c) Interpret Confidence Interval for each variable.

**67.** Simple Linear Regression Model which estimate the effect of fertilizers on production level of crop is given as following:

 Log(

a) Interpret the coefficient of fertilizer and land quality.

b) Find the exact percentage change in the level of crop yield if fertilizer and land quality increases by one unit.

c) Why do we use logarithmic transformation of dependent variable?

**68**. Consider the following regression result.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Ln(Wage) | coefficients | Standard Error | T-test | P-value | 95% Confidence Level |
| Constant | 0.04 | 0.2 |  \*\* | 0.04 | (5.67; 14.78) |
| Ln(Tenure) | 0.92 | \*\* |  6.5 | 0.002 | (1.23; 9.43) |
| Ln(Experience) | \*\* | 0.16 |  8.6 | 0.12 | (-1.43; 5.46) |
| Ln(Education) | 2.45 | \*\* |  4.85 | 0.005 | (2.34: 7.82) |

a) Fill the \*\* in the above table. Find standard deviation of tenure, experience and education if n=30.

b) Formulate Null hypothesis and check the significance of variables if significance level is 0.5%.

c) Write the estimated equation based on the information given above and interpret confidence Interval

**69**. What is the purpose for the use of beta coefficients? How can we derive these coefficients? What is the main advantage of it compared to simple regression coefficients?

**70**. Discuss the main advantages and disadvantages of using logarithmic functional form.

**71**. Simple Linear Regression Model which estimate the effect of fertilizers on production level of crop is given as following:

 Log(

a) Interpret the coefficient of education and tenure.

b) Find the exact percentage change in the level of wage if education and tenure increases by one unit. What is the difference?

c) Why do we use logarithmic transformation of dependent variable?

**72.** Consider the following regression result.

 zprice= - 0.456\*znox – 0.143\*zcrime + 1.43\*zrooms – 0.235\*zdist – 2.43\*zstratio

a) Why do we use standardized beta coefficients?

b) Interpret the each coefficients.

**73.** Why do we use Do-File in STATA software? What is difference between Data Editor and Data Browser option?

**74**. Give a brief information about codebook, tab, sum and describe command. Explain each of them.

**75.** Explain endogeneity problem that we observe in multiple regression analysis. How can we solve this problem?