1. \*Explain the essences of the random variable, experiment and outcome. Give an example describing all of them.
2. \*What is the difference between discrete and continuous random variables?
3. \*What is expected value of the random variable? Give an example about that as well.
4. \*\*What are the main properties of the expected value?
5. \*Explain the difference between population and sample. Why there is a need for sampling?
6. \*What is econometrics and why there is a need for that?
7. \*\*What are the main goals of using econometric methods?
8. \*\*How does one go about *structuring* **an empirical economic analysis**?
9. \*What are the main differences between the cross-sectional and time-series data?
10. \*What is the ceteris paribus effect? What is its role in the econometric models?
11. \*Form a simple econometric model and give explanations of every elements of the model.
12. \*\*Why is it assumed that the changes in error term (u) and in other variables have to be zero, while analyzing the effect of the independent variable?
13. \*\*What are the restrictive assumptions made for having fair ceteris paribus effect, while forming a regression.
14. \*\*\*Derive the OLS estimates of B0 and B1.
15. \*\*\*Prove that SST = SSE + SSR.
16. \*\*How the nonlinearities are incorporated in simple regression? Give the examples for that.
17. \*\*How many assumptions there are for unbiased simple regression function? Why there is a need for those assumptions?
18. \*\*\*Prove the unbiasedness of OLS by considering the first four assumptions for SLR.
19. \*\*Explain the assumption of the homoscedasticity.
20. \*\*Give examples for both cases of homoscedasticity and heteroskedasticity assumptions and show the difference on the graph.
21. \*What are the multiple regression model and its differences from the simple regression model?
22. \*\*Derive the OLS estimates for multiple regression function. What are the first order conditions of OLS?
23. \*How does the ceteris paribus work in the multiple regression models?
24. \*\*How the changes in more than one independent variables are calculated in the multiple regression models?
25. \*\*There are 2 models of:

y=b0+b1\*X1+u

y=b0+b1\*X1+b2\*X2+u

Are b1`s the same in both models? Why/why not?

1. \*How the multiple regression and the simple regression is differed? Why do we need multiple regression analysis, while we have the simple one? Explain your answer clearly.
2. \*\*\*What is the “partialling out” effect? How the B1 is calculated (formula) in below case:
3. \*\*In which cases, B1`s in the simple regression and multiple regression analysis are the same? Explain your opinion broadly.
4. \*How the R2 is affected (increase or decreasing) when the new independent variable is added to the model? And why? What is the better alternative to use for the analysis?
5. \*What are the Gauss Markov assumptions? Why these assumptions are needed? Explain one of them in your answer.
6. \*\*In a study relating college grade point average to time spent in various activities, you distribute a survey to several students. The students are asked how many hours they spend each week in four activities: studying, sleeping, working and leisure. Any activity is put into one of the four categories, so that for each student, the sum of hours in the four activities must be 168. The model is:

What assumption is violated in this model and how would you solve the violation problem?

1. \*\*\*The following equation represents the effects of tax revenue mix on subsequent employment growth for the population of regions in a country:

,

where growth is the percentage change in employment over a ten-year period, is the share of property taxes in total tax revenue, is the share of income tax revenues, and is the share of sales tax revenues. All of these variables are measured in 1980. The omitted share, , includes fees and miscellaneous taxes. By definition, the four shares add up to one. Other factors would include expenditures on education, infrastructure, and so on (all measured in 1980). Why must we omit one of the tax share variables from the equation? And Give a careful interpretation of .

1. \*\*Suppose that you are interested in estimating the ceteris paribus relationship between y and x1. For this purpose, you can collect data on two control variables, x2 and x3. (For concreteness, you might think of y as final exam score, x1 as lecture attendance, x2 as GPA up through the previous semester, and x3 as SAT or ACT score.) Let be the simple regression estimate from y on x1 and let be the multiple regression estimate from y on x1, x2, x3.

If x1 is highly correlated with x2 and x3 in the sample, and x2 and x3 have large partial effects on y, would you expect and to be similar or very different? Explain.

1. \*\*What is the normality assumption about? Explain the essence, mainly.
2. \*How the t-test is conducted? Explain with the definitions of the terms.
3. \*What is the p-value of the coefficient and how it defines the statistical significance?
4. \*\*\*Suppose you have the following model and you are interested in joint effect of trade turnover of Georgia and Russia with Azerbaijan on Azerbaijan GDP:

GDP – is Azerbaijan GDP;

tradeGEO – is trade turnover of Georgia with Azerbaijan;

tradeRUS – is trade turnover of Russia with Azerbaijan;

exchangerate – AZN per USD;

How would you formulate your hypothesis and your model in order to find the joint effects of two independent variables in the initial regression model?

1. \*\*How is the joint significance of more than one independent variables checked in the models? What kind of test is used? Explain it clearly.
2. \*\*What is the F-statistic about? Why is it always positive? Explain it clearly.
3. \*\*\*Suppose you are given the following model and elements:

H0: B1=0; B2=0

H1: H0 is not true

n= 56; (Number of observations)

F critical value for 5% significance level is 4.0012.

Show the restricted model and calculate the F-stat in order to know whether to reject the null hypothesis. Make a conclusion on the result.

1. \*\*\*Following table with three models is given where standard errors are in parentheses below the coefficients. The variable mktval is market value of the firm, profmarg is profit as a percentage of sales, ceoten is years as CEO with the current company and comten is total years with the company. Please, read the model and explain it briefly. Then give an explanation whether longer tenure with the company, holding the other factors fixed, is associated with a lower salary.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dependent Variable: log(*salary*)** | | | |
| **Independent variables** | (1) | (2) | (3) |
| **Log(sales)** | 0.224  (0.027) | 0.158  (0.040) | 0.188  (0.040) |
| **Log(mkvtal)** | - | 0.112  (0.050) | 0.100  (0.049) |
| **Profmarg** | - | -0.023  (0.0022) | -0.0022  (0.0021) |
| **Ceoten** | - | - | Put a value to calculate |
| **Comten** | - | - | Put a value to calculate |
| **Intercept** | 4.94  (0.20) | 4.62  (0.25) | 4.57  (0.25) |
| **Observation number** | 177 | 177 | 177 |
| **R-squared** | 0.281 | 0.304 | 0.353 |

1. \*\*Consider the given table and comment on R-squares of different model. May the R-squared be the reliable indicator of important variable? Can it be used to compare different models and point out the importance of added variable?

|  |  |  |  |
| --- | --- | --- | --- |
| **Dependent Variable: log(*salary*)** | | | |
| **Independent variables** | (1) | (2) | (3) |
| **Log(sales)** | 0.224  (0.027) | 0.158  (0.040) | 0.188  (0.040) |
| **Log(mkvtal)** | - | 0.112  (0.050) | 0.100  (0.049) |
| **Profmarg** | - | -0.023  (0.0022) | -0.0022  (0.0021) |
| **Ceoten** | - | - | 0.0171  (0.0055) |
| **Comten** | - | - | -0.0092  (0.0033) |
| **Intercept** | 4.94  (0.20) | 4.62  (0.25) | 4.57  (0.25) |
| **Observation number** | 177 | 177 | 177 |
| **R-squared** | Put a value to calculate | Put a value to calculate | Put a value to calculate |

1. \*\*Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (roe, in percentage form), and return on the firm`s stock (ros, in percentage form):

In terms of model parameters, state the null hypothesis that, after controlling for sales and roe, ros has no effect on CEO salary. State the alternative that better stock market performance increases a CEO`s salary.

1. \*The variable rdintens is expenditures on research and development (R&D) as a percentage of sales. Sales are measured in millions of euros. The variable profmarg is profits as a percentage of sales. The following equation is estimated:

N=32; R-squared=0.099

Interpret the coefficient on log(sales). In particular, if sales increase by 10%, what is the estimated percentage point change in rdintens? Is this an economically large effect?

1. \*\*Consider the multiple regression model with three independent variables, under the classical linear model assumptions MLR.1 through MLR.6 (6 assumptions):

You would like to test the null hypothesis:

Define and . Write regression equation involving , , and that allows you to directly obtain .

1. \*\*\*In the simple regression model under Gauss Markov assumptions, we argued that the slope estimator, , is consistent for . Using , show that plim=.

Hint\*: You need to use the consistency of , and the law of large numbers, along with the fact that .

1. \*The following model allows the return to education to depend upon the total amount of the parents’ education, called pareduc:

Show that the proportionate effect on wage of another year of education is .

1. \*The estimated equation is

N=1 230; R-squared=0.195;

If someone’s parents have 32 years of total education, by what percentage is that person’s return to education estimated to exceed that of someone whose parents have 24 years of education?

1. \*\*When pareduc is added as a separate regressor, we get

N=1 230; R-sqaured=0.195;

Now how does the return to education depend on parent education?

1. \*Suppose we want to estimate the effects of alcohol consumption (alcohol) on university grade point average (uniGPA). In addition to collecting information on grade point averages and alcohol usage, we also obtain attendance information (say, percentage of lectures at‑ tended, called attend). A standardized test score (say, SAT) and post-sixteen GPA (psGPA) are also available.

Should SAT and psGPA be included as explanatory variables? Explain why and how would you check whether it would better to add them.

1. \*What is the dummy variable about? Give an example using dummy in the regression and explain the model.
2. \*There is a give model below. Explain the case when d=1 and d=0 in the graph. What is the additional value in the model when d=1?
3. \*\*We have the following regression model:

(235.11) (0.018) (5.86) (11.21) (0.134) (Put a value to calculate)

n=706; R-squared=0.123;

t-critical value of 5% confidence level for this model is 1.984 and the values under each coefficient are their standard errors.

The variable sleep is total minutes per week spent sleeping at night, totwrk is total weekly minutes spent working, educ and age are measured in years, and male is a gender dummy. All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence?

1. \*\*Having the following model:

(0.22) (0.009) (0.0059) (0.006) (Put a value to calculate) (Put a value to calculate)

n= 1 388; R-squared=0.0472

t-critical value of 5% confidence level for this model is 1.984 and the values under each coefficient are their standard errors.

How much more is a white child predicted to weigh than a nonwhite child, holding the other factors in the first equation fixed? Is the difference statistically significant?

1. \*\*Given model is:

(6.29) (3.83) (0.53) (Put a value to calculate) (Put a value to calculate) (Put a value to calculate)

n=4137; R-squared=0.0858

The variable sat is the combined SAT score, hsize is size of the student’s high school graduating class, in hundreds, female is a gender dummy variable and black is a race dummy variable equal to one for blacks and zero otherwise.

Holding hsize fixed, what is the estimated difference in SAT score between nonblack females and nonblack males?

1. \*\* Consider the model

n=6763; R-squared=0.202

Using this equation, find the value of totcoll such that the predicted values of log(wage) are the same for men and women.

1. What is heteroskedasticity? Why is it important to have it, although there is unbiased and consistent OLS estimates without this assumption?
2. \*\*What are the consequences of heteroskedasticity? Explain it clearly.
3. \*\*\*Prove that , when there is homoscedasticity.
4. \*\*\*Derive the equation of  when there is heteroskedasticy in the regression.
5. \*\*Explain how to test the heteroskedasticity? What are the well-known tests for that to use and their differences?
6. \*\*\*Consider a linear model to explain monthly beer consumption:

Write the transformed equation that has a homoskedastic error term and explain how you came up with that equation.

1. \*\*The variable smokes is a binary variable equal to one if a person smokes, and zero otherwise. We estimate a linear probability model for smokes:

At what point does another year of age reduce the probability of smoking? Explain the way you find it.

1. \*\*\*There are different ways to combine features of the Breusch-Pagan and White tests for heteroskedasticity. One possibility not covered in the text is to run the regression

where the are the OLS residuals and the are the OLS fitted values. Then, we would test joint significance of Xi1, Xi2, …, Xik and . (Of course, we always include an intercept in this regression.)

Explain why the R-squared from the regression above will always be at least as large as the R-squareds for the BP regression and the special case of the White test.

1. \*What tests are used in order to detect general functional form misspecification? Explain the conduction of the test on the example.
2. \*What is the difference between nested and nonnested models? What tests are used to detect the right functional forms of these models? Explain it clearly.
3. \*How to test against nonnested models? What approaches there are? Explain them.
4. \*Consider that the model is misspecified due to the unavailable data of an independent variable. What method is needed to be used to solve the misspecification? Explain the method and the way solved the problem clearly.
5. Let math10 denote the percentage of students at a Michigan high school receiving a passing \*score on a standardized math test. We are interested in estimating the effect of per student spending on math performance. A simple model is

where poverty is the percentage of students living in poverty. The variable lnchprg is the percentage of students eligible for the federally funded school lunch program. Why is this a sensible proxy variable for poverty?

1. \*What is the difference between time-series and cross sectional analysis? Give a data samples for both and show the difference on their data as well.
2. \*\*State the Gauss-Markov assumptions for time series analysis and explain each of them briefly. Show the difference of those with that of Gauss-Markov assumptions of cross sectional analysis as well.
3. \*Assume that the dependent and independent variables are trending together over time. Why the relationship is not always causal? Identify the reasons and solutions for that.
4. \*\*When the three event indicators befile, affile, and afdec are dropped from the below equation, we obtain R-squared = Put a value to calculate and R(adjusted)-squared= Put a value to calculate. Are the event indicators jointly significant at the 10% level, considering F-critical value is about 2.13?

n=132; R-squared = Put a value to calculate; R(adjusted)-squared= Put a value to calculate.

1. \*What are stationary and non-stationary processes about? Explain the difference and each element in it.
2. \*\*Explain the weakly dependent time series analysis in terms of covariance stationary process. Why it is important to have weak dependence in time series analysis?