PREDICTIVE ANALYTICS ISM-E1003

EXAM 14.04. 2022, 13:00-16:00

Instructions: A calculator is allowed in the exam. Mark your answer with X.

Correct answer 4 points, incorrect answer -1 point, no answer 0 point.

Student name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Student number:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Consider the following linear model y = a + bx, where a = 0.0043 and b = 1.2050. The covariance matrix of the estimates is:

|  |  |  |
| --- | --- | --- |
|  | a | b |
| a | 3.6539e-05 | -7.2296e-05 |
| b | -7.2296e-05 | 1.9954e-02 |
|  |  |  |

The corresponding *t*-statistics are:

1. **0.7114 and 8.5307**
2. 0.8434 and 2.907
3. 10.8480 and 7.7711
4. 117.68 and 60.390

2. Which of these is NOT a reason for adding a disturbance term to a regression model ?

(a) Some determinants of the effect variable may be omitted from the model

(b) Some determinants of the effect variable may be unobservable

(c) **Some determinants of the independent variable may be omitted from the model**

(d) There may be errors in the way that the dependent variable is measured which cannot be modelled.

3. The estimators  and  determined by OLS will be the Best Linear Unbiased Estimators (BLUE) if which of the following assumptions hold?

(I) The errors have zero mean

(II) The variance of the errors is constant and finite over all values of the independent variable(s)

(III) The errors are linearly independent of one another

(IV)There is no relationship between the error and corresponding independent variables

(a) I and II only

(b) I, II and III only

(c) II, III and IV only

(d) **I, II, III, and IV**

4. Standard errors

(a) Give us an idea of the deviation of the errors from their mean

(b) Measure the reliability of the independent variables

(c) **Give us an idea of the precision of estimates of and**

(d) Measure the reliability of the dependent variables.

5. Which one of the following is the most appropriate as a definition of *R*2 in the context that the term is usually used?

1. It is the proportion of the total variability of *y* that is explained by the model
2. **It is the proportion of the total variability of *y* about its mean value that is explained by the model**
3. It is the correlation between the fitted values and the residuals
4. It is the correlation between the fitted values and the mean.

6. Suppose that the value of *R*2 for an estimated regression model is exactly one. Which of the following are true?

1. All of the data points must lie exactly on the line
2. All of the residuals must be zero
3. All of the variability of *y* about its mean has been explained by the model
4. The fitted line will be horizontal with respect to all of the explanatory variables.
5. (ii) and (iv) only
6. (i) and (iii) only
7. **(i), (ii), and (iii) only**
8. (i), (ii), (iii), and (iv).

7. Which of the following would NOT be a potential remedy for the problem of multicollinearity between regressors?

(a) Removing one of the explanatory variables

(b) **Transforming the data into logarithms**

(c) Transforming two of the explanatory variables into ratios

(d) Collecting higher frequency data on all of the variables.

8. If the residuals of a regression on a large sample are found to be heteroscedastic which of the following might be a likely consequence?

(i) The coefficient estimates are biased

(ii) The standard error estimates for the slope coefficients may be too small

(iii) Statistical inferences may be wrong.

(a) (i) only

(b) **(ii) and (iii) only**

(c) (i), (ii), and (iii)

(d) (i) and (ii) only.

9. A normal distribution has coefficients of skewness and excess kurtosis which are, respectively,

(a) **0 and 0**

(b) 0 and 3

(c) 3 and 0

(d) Will vary from one normal distribution to another.

 

10. The graphs above are time series plots of residuals from two separate regressions. Which of these combinations is true?

(a) **A shows negative autocorrelation and B shows positive autocorrelation**

(b) A shows positive autocorrelation and B shows negative autocorrelation

(c) A shows heteroscedasticity and B shows homoscedasticity

(d) A shows homoscedasticity and B shows heteroscedasticity.

11. Assuming a researcher runs the following regression  where is residual from a regression. If the researcher conducts a hypothesis test with null hypothesis of  against an alternative hypothesis of , what type of test is he or she conducting?

(a) Test for heteroscedasticity

(b) **Test for autocorrelation**

(c) Test for non-normality

(d) Test for homoscedasticity.

12. Which of the following models can be estimated using ordinary least squares?

(i) An AR(1)

(ii) An ARMA(2,0)

(iii) An MA(1)

(iv) An ARMA(1,1).

1. (i) only
2. **(i) and (ii) only**
3. (i), (ii), and (iii) only
4. (i), (ii), (iii), and (iv).
5. A variable *x* is defined as ­­\_\_\_\_\_\_\_\_ if its value is determined outside of the equation or system of equations. What is the blank?

(a) Endogenous

(b) **Exogenous**

(c) Homogeneous

(d) Heterogeneous.

1. Which of these is an approach used to determine the appropriate lag lengths of VAR models?

(a) Graphically plotting the time series of the data

(b) Selecting the number of lags that maximises the information criteria

(c) **Selecting the number of lags that minimises the information criteria**

(d) None of the above.

1. Which of the following are consequences of using non-stationary data in regressions?

(I) Shocks will be persistent

(II) It can lead to spurious regressions

(III) *t*-ratios will not follow a *t*-distribution

(IV) The *F*-statistic will not follow an *F*-distribution.

(a) I only

(b) I and II only

(c) I, II, and III only

(d**) I, II, III, and IV**.