PREDICTIVE ANALYTICS ISM-E1003

Homework 1, due date 11.03. at 10:00

What is the correct answer? Correct answer 4 points, incorrect answer –1 point, no answer 0 point.

1. What does a positive linear relationship between *x* and *y* in a simple regression imply?

(a) Increases in the independent variable are usually accompanied by increases in the regressor

(b) The relationship between *x* and *y* cannot be explained by a straight line

(c) Decreases in the independent variable is usually accompanied by increases in the regressors

(d) Increases in the regressor are usually accompanied by increases in the dependent variable.

2. Suppose you have 5-year annual data on the excess returns on a fund manager’s portfolio (‘fund ABC’) and the excess returns on a market index (where  is the return on fund ABC,  is the risk-free rate and  is the return on the market index):

|  |  |  |
| --- | --- | --- |
| Year *t* | Excess return on fund ABC | Excess return on market index |
| 1 | 14.0 | 16.0 |
| 2 | 32.0 | 21.7 |
| 3 | 11.6 | 6.0 |
| 4 | 21.2 | 16.2 |
| 5 | 17.4 | 11.0 |

What is the estimated alpha () for Fund ABC?

(a) 2.3

(b) 3.3

(c) 4.3

(d) 5.3.

3. Given the data in Question 2, what is the estimated beta () of Fund ABC?

(a) 3.1

(b) 2.1

(c) 1.1

(d) None of the above.

4. If our regression equation is ***y = Xb + u***, where we have *T* observations and *k* regressors, what will be the dimension of  using the standard matrix notation

(a) *T* x *k*

(b) *T* x 1

(c) *k* x 1

(d) *k* x *k*.

Question 5 refers to the following regression estimated on 64 observations:

*yt = b*1 *+ b*2*X*2*t + b*3*X*3*t + b*4*X*4*t + ut*

5. Which of the following null hypotheses could we test using an *F*-test?

(i) *b*2 = 0

(ii) *b*2 = 1 and *b*3 + *b*4 = 1

(iii) *b*3*b*4 = 1

(iv) *b*2 -*b*3 -*b*4 = 1.

(a) (i) and (ii) only

(b) (ii) and (iv) only

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

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

For Question 6, you are given the following data



The regression equation is

*yt = b*1 *+ b*2*X*2*t + b*3*X*3*t + ut*

6. Which of the following is the correct value for ?

(a) 2.89

(b) 1.30

(c) 0.84

(d) We cannot determine the value of  from the information given in the question.

7. 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.

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 assumption of homoscedasticity can be written mathematically as

(a) 

(b) 

(c) 

(d) .