Mid Term Exam

MSDA3055

Linear Regression and Time Series

Oct 25, 2021

**Question 1**

Consider the regression model

Suppose you know that . Derive a formula for the least squares estimator of .

**Answer:**

As given in the question β1 = 4

Yi = 4+β2Xi + Ui

Taking the above as the estimated model

Applying sum of squared errors:

……………… consider it as **equation x**

Now differentiating x with respect to Beta1:

= 0

-2xi

β1 =

Thus, when beta1 is 4:

β1 =

**Question 2**

A researcher has data for 50 countries on (the average number of newspapers purchased per adult in one year) and (GDP per capita) measured in US $, and fits the following regression:

Where and . The researcher realizes that GDP has been underestimated by $100 in every country and that should have been regressed on , where

Explain, with mathematical proofs, how the following components of the output would have differed:

1. The coefficient of GDP
2. The intercept

**Answer A:**

**The coefficient of the slope estimate of GDP remains unchanged.**

Text, letter

Description automatically generated

**Answer B**

**The new intercept decreases by 0.2 and is equal to 23.8.**

Text, letter

Description automatically generated

**Question 3**

Suppose that

where is the number of crimes committed in town and is the number of police officers working in town.

You, the researcher, wish to interpret this regression as a causal model of the effect of on .

1. How would you interpret ? (Hint: You may need to use the derivative chain rule.)
2. How would your answer to part (a) change if the model had been defined with instead of ?

Answer:

**Answer A: β2 is interpreted as linear multiplier**

**Answer B: Then we can say that it is a logarithmic multiplier along X - axis**

**Question 4**

Consider the California Test Score Data Set (given in excel file in Question 4 sheet paper.)

1. How many observations do you have in the data set?

**420**

1. Consider the variable avginc (the average district income) in column O, measured in 1000s of dollars:
2. Define a new variable, income, which is the variable avginc multiplied by 1000.

**In excel it is defined**

1. What is the mean and standard deviation of avginc?

|  |  |
| --- | --- |
| mean | 15.31659 |
| standard deviation | 7.22589 |

1. What is the mean and standard deviation of income?

|  |  |
| --- | --- |
| mean | 15316.59 |
| standard deviation | 7225.89 |

1. What is the correlation between avginc and math score in column R? What is correlation between income and math score in column R? Are the two correlations the same or different? Explain.

**The correlation between avginc and math score and correlation between income and math score will be same.**

**Derived in the attached excel.**

1. Run a simple OLS linear regression with income as dependent variable and math score as independent variable and interpret its coefficient estimates.

|  |  |
| --- | --- |
| Beta1 | -160742 |
| Beta2 | 269.4742 |

1. Interpret goodness-of- fit for OLS linear regression.

**Answer: 0.489191 (Calculated in Excel)**



**Question 5**

Consider the regression model

Where the total number of observations is 25, standard deviation of the intercept is 1.1 and the standard deviation of the slope is 3.2.

Test the claim that the slope must be less than 30 in a significant level of 1%.

Answer:

**Question 6**

Which of the following statements is TRUE concerning OLS estimation?

1. OLS minimizes the sum of the vertical distances from the points to the line
2. OLS minimizes the sum of the square of the vertical distances from the points to the line.
3. OLS minimizes the sum of the horizontal distances from the points to the line
4. OLS minimizes the sum of the square of the horizontal distances from the points to the line.

**Answer: B**

**Question 7**

The residual from a standard regression model is defined as

1. The difference between the actual value, , and the mean, .
2. The difference between the fitted value, , and the mean, .
3. The difference between the actual value, , and the fitted value, .
4. The square of the difference between the fitted value, , and the mean, .

**Answer: C**

**Question 8**

Which of the following statements best describes the algebraic representation of the fitted regression line?

1. +

**Answer is C**

**Question 9**

In a multiple regression model, the adjusted

1. Cannot be negative.
2. Will never be greater than the regression .
3. Equals the square of correlation coefficient
4. Cannot decrease when an additional explanatory variable is added.

**Answer: B**

**Question 10**

If we multiply both and by 1000 and re-estimate the regression, the intercept coefficient and its standard error will

1. Increase by 1000 times
2. Decrease by 1000 times
3. Remain same
4. Increase by () times

**Answer: C**

**Question 11**

In a simple linear regression model, the elasticity of Y with respect to X is given by:

1. (X/Y)
2. X
3. (1/Y)

**Answer: C**

**Question 12**

Given , state which of the following statements is true:

1. measures the change in per unit change in , holding the value of constant.
2. measures the change in per unit change in , and not of any effect that may have on .
3. Both (a) and (b) are true.
4. Neither (a) nor (b) is true.

**Answer: C**

**Question 13**

When comparing of two regression models, the models should have the same

1. X variables.
2. Y variables.
3. Error term.
4. Beta coefficients.

**Answer: C**

**Question 14**

If in , both and are standardized variables, the intercept term will be

1. Positive
2. Negative
3. Between -1 and +1
4. Equal to zero

**Answer: A**

**Question 15**

Which one of the following is NOT an assumption of the classical linear regression model?

1. The explanatory variables are uncorrelated with the error terms.
2. The disturbance terms have zero mean.
3. The dependent variable is not correlated with the disturbance terms.
4. The disturbance terms are independent of one another.

**Answer: C**