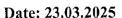The Open University of Sri Lanka
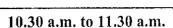
B.Sc/B.Ed. DEGREE, CONTINUING EDUCATION PROGRAMME

No Book Test (NBT) 2024/2025

Level 05 - Applied Mathematics

ADU5301 - Regression Analysis I

Date: 23.03.2025   **10.30 a.m. to 11.30 a.m.**

Instructions

- This examination is of **one-hour** duration.

- Answer **all** questions.

- The examination paper has two parts. Part A consists of 10 multiple choice questions. Pat A is allocated 40 marks and Part B, consisting of a structured essay type question is allocated 60 marks, distributed as indicated.

- **Detach Part A, underline the most appropriate answer to each question in this section, and return along with the answers to Part B.**

- Non programmable calculators are permitted.

- Where appropriate, you may assume that the model is fitted using the method of

  least squares, and $\widehat{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{\sum(x_iy_i)-(\sum x_i)(\sum y_i)/n}{\sum x_i^2-(\sum x_i)^2/n}$ and $\widehat{\beta}_0 = \frac{\sum y_i}{n} - \widehat{\beta}_1\frac{\sum x_i}{n}$.,

  in the usual notation.

Part A

1) Which of the following statements is most appropriate about fitting the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, using the method of least squares to a response variable that is negatively correlated with the predictor variable?
   The estimate for the slope parameter, $\widehat{\beta}_1$ is
   a) negative.
   b) positive.
   c) non-negative
   d) non-positive

2) When the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ is fitted to a response variable, using the predictor variable x, least squares estimate for the slope parameter is 0.5. Which of the following statements is most appropriate?
   Associated with one unit of increase in the predictor variable,
   a) the response variable reduces by 0.5 units.
   b) the response variable increases by 0.5 units.
   c) the response variable may increase or decrease by 0.5 units, depending on the level of the predictor variable
   d) the change in the response variable is not correctly described by any of the statements in (a), (b), (c) above.

3) Which of the following statements is most appropriate for the least squares fitted line for the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$?
    a) When the predictor variable takes the value x_i, the vertical distance between the fitted line and the population regression line is ϵ_i.
    b) The fitted line passes through the point $(\bar{x}, \bar{y})$, where $\bar{x}$ and $\bar{y}$ denote the average of the predictor and response variable values respectively.
    c) The fitted line **always** passes through the point $(0, \beta_0)$.
    d) All of the above statements (a), (b) and (c) are correct.

4) Which of the following statements is most appropriate about the ordinary residuals $(y_i - \hat{y}_i)$ from a least squares model fit of the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$?
    a) The magnitude of the residuals of extreme observations is **always** large.
    b) The residuals **always** randomly scatter about the zero line.
    c) Plots of residuals against the predictor variable and plots of residuals against the predictor variable will have the same pattern.
    d) All of the above statements in (a), (b) and (c) are correct.

5) Which of the following statements is correct about the estimates for the intercept and slope parameters computed by fitting the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ to a set of data for which the mean of the predictor variable $\bar{x}$ is positive.
    a) The estimates for the intercept and slope parameters will be negatively correlated.
    b) The estimates for the intercept and slope parameters will be positively correlated.
    c) The estimates for the intercept and slope parameters will have no correlation.
    d) The estimates for the intercept and slope parameters will be positive or negative, depending on whether the slope of the fitted line is positive or negative.

6) Which of the following statements is most appropriate about a least squares model fit of the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, when the model assumptions are valid?
    a)   The sum of the fitted values equals the sum of the observed values.
    b)   The residuals are **always** uncorrelated with the fitted values.
    c)   The residuals **always** have the same variance, regardless of the level of the predictor variable.
    d)   The statements (a), (b) and (c) are all valid.

7) Which of the following statements is most appropriate about a least squares model fit of the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$?
    a) If the residual of an observation is zero, the model exactly predicts the observed value.
    b) If the residual of an observation is positive, the data value lies above the population regression line.
    c) If the residual of an observation is negative, the data value lies below the population regression line.
    d) All of the statements (a), (b) and (c) are correct.

8) Suppose a simple linear regression model fitted the data well and the model assumptions are valid. Identify the **incorrect** statement.

    a) Points in a scatter plot of the response variable against the predictor variable should have a linear trend.

    b) The partial derivatives of the fitted regression function, with respect to each model parameter do **not** involve the model parameters.

    c) The least squares residuals against the predictor variable should **not** have any clear pattern.

    d) The least squares residuals against the fitted values should **not** have any clear pattern.

9) Which of the following plots is most appropriate for detecting extreme data values in a data set with two variables? Underline the correct plot.

    a) Scatter plot of response against the predictor variable

    b) Scatter plot of residuals against the predictor variable

    c) Scatter plot of residuals against the response variable

    d) Scatter plot of residuals against the fitted values

10) Which of the following plots is most appropriate for examining whether the selected regression function fits the data well? Underline the correct plot

    a) Scatter plot of response against the predictor variable

    b) Scatter plot of residuals against the predictor variable

    c) Scatter plot of residuals against the response variable

    d) Scatter plot of residuals against the fitted values

Part *B*

1) A researcher is investigating whether adding a food preservative increases the shelflife of dried coconut milk powder and aims to quantify the relationship between the amount of preservative and the resulting shelf life. Varying amounts of the preservative, ranging from 1 mg to 10 mg, were added to 100-gram samples of coconut milk powder, and the shelf life (in weeks) of each sample was measured. As a control, the shelf life of samples without the preservative was also measured.

The following summary statistics were computed from the data.:

$$n = 40, \quad \sum x = 196.5, \quad \sum y = 498 ,$$

$$\sum xy = 3007.55, \quad \sum x^2 = 12149.75, \quad \sum y^2 = 7474.28$$

The researcher seeks your assistance in fitting the simple linear regression model,

$y_i = \beta_0 + \beta_1 x_i + \epsilon_{i,}$ where x_i denote the amount of preservative added (mg per 100 gram sample) and y_i denote the shelflife of the sample (in weeks).

i) State the assumptions required to obtain the least squares estimates for the slope and intercept parameters in the regression function.

(10 marks)

ii) Given that the least squares estimate for the slope parameter is 2.205, determine the equation of the fitted line, based on the method of least squares.

(15 marks)

iii) Describe two advantages of using replicates in the study.

(10 marks)

iv) Suppose the researcher had only three control replicates, and the observed shelflives of these three replicates were 2 weeks, 2 weeks, and 3 weeks.
   a) Estimate the mean shelflife of a sample that received no preservative, **without using the fitted model** and
   b) Compare the value obtained in part (a), with the value you get from the fitted model.

(10 marks)

v) Out of the two estimates obtained in part (iv), which estimate would you recommend? Justify your answer, clearly stating any conditions under which your recommendation is more appropriate.

(15 marks)