

Regression Analysis, MA 272**Answer Three (3) Questions Only****Time: 3 Hours**

1. a) Define the following:
 - i. A scatter plot.
 - ii. Sample correlation coefficient.
 - iii. Coefficient of determination.
 - b) What is an outlier? How can a point be classified as an outlier without a graphical plot.
 - c) Distinguish between Type I and Type II errors.
2. The table below presents the mean heights of 161 children in a village aging 18 to 29 months.

Age (month)	18	19	20	21	22	23	24	25	26	27	28	29
Height (cm)	76.1	77.0	78.1	78.2	78.8	79.7	79.9	81.1	81.2	81.8	82.8	83.5

- a) Draw a scatter plot of the data and interpret it.
- b) Calculate the best-fit line.
- c) Compute the sample correlation coefficient and interpret it.

3. The following are advertised sale prices of colour televisions at Anderson's shop

Size (inches)	9	20	27	31	35	40	60
Sale Price (\$)	76.1	77.0	78.1	78.2	78.8	79.7	79.9

- a) Use the Ordinary Least Squares method to deduce the model parameters, $\hat{\beta}_0$ and $\hat{\beta}_1$ and interpret them.
- b) Find the estimated sale price for a 32 inch television.
- c) Assess whether the linear relationship is statistically significant based on the null hypothesis $H_0 : \hat{\beta}_1 = 0$.

4. Consider the sample data:

Y	X_1	X_2
78.5	7	2.6
74.3	1	2.9
104.3	11	5.6
87.6	11	3.1
95.9	7	5.2
109.2	11	5.5
102.7	3	7.1

a) Use the Ordinary Least Squares minimisation principle to show that the multiple regression analysis model coefficients are given as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$$

b) Perform Analysis of Variance (ANOVA) test to show whether the model coefficients are significant or not.

c) Hence or otherwise, find the average value of the mean at the point $(x_1 = 2.31, x_2 = 3.42)$.

Examiner
(Paul Boye)