Kasarani Campus
Off Thika Road
Tel. 2042692 / 3
P. O. Box 49274, 00100
NAIROBI
Westlands Campus
Pamstech House
Woodvale Grove
Tel. 4442212
Fax: 4444175

KIRIRI WOMENS' UNIVERSITY OF SCIENCE AND TECHNOLOGY

UNIVERSITY EXAMINATION, 2023/2024 ACADEMIC YEAR
THIRD YEAR, SECOND SEMESTER EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE
(MATHEMATICS)

Date: 5th December, 2023
Time: 8.30am –10.30am

KMA 319 - REGRESSION METHODS

INSTRUCTIONS TO CANDIDATES

ANSWER **QUESTION ONE (COMPULSORY)** AND **ANY OTHER TWO** QUESTIONS

QUESTION ONE (30 MARKS)

a) Determine whether the following equations are linear in parameters

i) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$ (2 marks)

ii) $Y = \beta_o e^{\beta_1 X} + \epsilon.$ (2 marks)

iii) $Y = \beta_0 X^2 + \beta_1 \sin(\beta_2 X).$ (2 marks)

b) Ordinary Least Square method of estimation parameters is based on some assumptions. Highlight four of these assumptions. (4 marks)

c) Consider no intercept model given by $y_i = \beta x_i + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$ and $\text{cov}(\epsilon_i, \epsilon_j) = 0$. Based on sample $(x_i, y_i), i = 1, 2, \ldots, n.$ Find the maximum likelihood estimator of β.

(4 marks)

d) 15 individuals were interviewed on their monthy salary (input) and expenditure (output). R output for coefficients of a simple linear equation is as shown

```
Coefficients:
            Estimate Std. Error t value
(Intercept) -8.03690    6.30940  -1.274
Salary       0.98929    0.09211  10.740
---
```

i) Write the model and interpret β_1, the coefficient of salary. (3 marks)

ii) Test for significance of the parameters at 1% level of significance. (3 marks)

e) Consider the data on the price and the number items purchased for a particular product

Price (KES) (X)	10	15	20	25	30	35
Number of items purchased (Y)	200	170	130	100	100	90

Fit the non-linear model $Y = \beta_0 e^{\beta X}$ to the data. Hence predict the number of items purchased when the price is 40.

(5 marks)

f) In a regression analysis problem with two parameters β_0 and β_1, the following was obtained

$$(X'X)^{-1} = \frac{1}{336}\begin{pmatrix} 8 & 36 \\ 36 & 204 \end{pmatrix}$$

and the residual variance $\hat{\sigma}^2 = 140.4225$. Find

i) $\widehat{Var}\,(\hat{\beta}_0)$ (1 mark)

ii) $\widehat{Var}(\hat{\beta}_1)$ (1 mark)

iii) $\widehat{Cov}\,(\hat{\beta}_0, \hat{\beta}_1)$ (1 mark)

iv) $\widehat{Var}\,\left(\hat{\beta}_0 + \hat{\beta}_1\right)$ (2 marks)

QUESTION TWO (20 MARKS)

Let the linear regression model be given by $y_i = \alpha + \beta x_i + \epsilon_i$ where α and β are the parameters of the model and ϵ is a random variable with mean 0, variance σ^2 and $Cov\left(\epsilon_i\, \epsilon_j\right) = 0, i \neq j\ i, j = 0, 1, 2, \ldots, n$. Based on a random sample $(x_i, y_i), i = 1, 2, 3, \ldots$, show that;

a) Ordinary least square estimates of;

i) α is $\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$. (4 marks)

ii) β is $\hat{\beta} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{S_{xy}}{S_{xx}}$. (4 marks)

b) $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators of α and β respectively. (6 marks)

c) $var(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$. (3 marks)

d) $Var\,(\hat{\alpha}) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$. (3 marks)

QUESTION THREE (20 MARKS)

Research was conducted to determine whether there the Quantity Sold depend on Price and Advertising. The data on these variables are given in the following table/

Quantity Sold	8500	4700	5800	7400	6200	7300	5600
Price	300	750	450	300	750	450	600
Advertising	420000	30000	60000	75000	480000	270000	135000

A multiple regression analysis was adopted and R output is as shown

```
> summary(Model)

Call:
lm(formula = Quantity.Sold ~ Price + Advertising, data = data)

Residuals:
      1        2       3       4       5       6       7
 -23.01   223.95 -465.94  239.12  -52.73  204.94 -126.33

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.536e+03  3.869e+02  22.062  2.5e-05 ***
Price       -5.571e+00  6.644e-01  -8.386  0.00111 **
Advertising  3.948e-03  6.956e-04   5.676  0.00476 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 310.5 on 4 degrees of freedom
Multiple R-squared:  0.9617,    Adjusted R-squared:  0.9426
F-statistic: 50.27 on 2 and 4 DF,  p-value: 0.001464
```

Use the output given to answer the following questions.

a) Write simple linear regression model.

(2 marks)

b) Predict the Quality sold when the price is KES 400 and advertising cost is KES 80,000.

(2 marks)

c) Based on the p-value, test the hypotheses on significance of individual parameters at 5% level of significance.

(5 marks)

d) Test on the adequecy of the overall model at 5% level of significance.

(3 marks)

e) Determine 95% confidence intervals of the parameters.

(6 marks)

f) What percentage of variation in the quality sold is determined by price and adverising cost?

(2 marks)

QUESTION FOUR (20 MARKS)

Consider the data on the price and the number items purchased for a particular product;

Price (KES) (X)	10	15	20	25	30	35
Number of items purchased (Y)	200	170	130	100	100	90

a) Determine Simple linear regression equation of Y on X. (5 marks)

b) Estimate the residual variance, σ^2. (3 marks)

c) Estimate the standard error of the parameters of the model. (3 marks)

d) Test for significance of the individual parameters. (5 marks)

e) Obtain the ANOVA table, hence comment on the significance of the overall model. (4 marks)

QUESTION FIVE (20 MARKS)

a) An automobile insurance company wants to use gender (x1=0 if female, 1 if male) and traffic penalty points (x2) to predict the number of claims (y). The observed values of these variables for a sample of six motorists are given by;

x1	0	0	0	1	1	1
x2	0	1	2	0	1	2
y	1	0	2	1	3	5

You are to use the following model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$. Determine

i) $X'X$. (3 marks)

ii) $(X'X)^{-1}$ (4 marks)

iii) $X'y$ (2 marks)

iv) $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$ (3 marks)

b) Discuss the concept of multicollinearity in multiple regression analysis; meaning, causes and effects and ways of reducing these effects. (8 marks)