KIRIRI WOMEN'S UNIVERSITY OF SCIENCE AND TECHNOLOGY

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 IN MATHEMATICS
KMA 319: REGRESSION METHODS

Date: 14th August, 2023
Time: 2.30pm – 4.30pm

INSTRUCTIONS TO CANDIDATES
ANSWER QUESTION ONE (COMPULSORY) AND ANY OTHER TWO QUESTIONS
QUESTION ONE (30 MARKS)

a) Define the following terms. Giving appropriate examples
   i) Liner model (2 Marks)
   ii) Non-linear model (2 Marks)

b) State assumptions of the simple linear model. (5 Marks)

c) Distinguish between maximum likelihood estimation and ordinary least squares estimation of the parameter of a simple linear model (3 Marks)

d) Consider the following simple linear model:
$Y_i = \beta_o + \beta_1 X_i + e_i \quad ,i = 1,2,...,n$ Where β_0, β_1 are constants, $E(e) = 0, \text{var}(e) = \sigma^2$
Suppose that $Y \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$, derive the maximum likelihood estimators of β_0 and β_1 (6 Marks)

e) For the model in (d) show that $\hat{\beta}_1$ is unbiased estimate of the parameter β (5 Marks)

f) Let $Y=(7,4,1,3)^T$, $Z_1=(4,3,9,4)^T$ and $Z_2=(6,3,7,4)^T$ Fit a regression model $Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$. Give the fitted coefficients and estimate on error variance (7 Marks)

QUESTION TWO (20 MARKS)

The following data in Table 2 refers to the number of claims (X) received by a motor insurance company in a week and the number of settlements (Y) of these claims in the following week during 10 randomly selected weeks in a year.

X	100	110	120	130	140	150	160	170	180	190
Y	45	51	54	61	66	70	74	78	85	89

Table 2

A regression model $y_i = \beta_0 + \beta_1 x_i + e_i$ where e_i's are $N(0, \sigma^2)$ is to be fitted on the above data.

a) Obtain the estimate of β_0 and β_1 (9 Marks)

b) Obtain the estimate of σ^2 (3 Marks)

c) Test the hypothesis $H_0 : \beta_0 = 0$ and $H_1 : \beta_0 \neq 0$ (5 Marks)

d) Obtain the 99% confidence interval of β_0 (3 Marks)

QUESTION THREE (20 MARKS)

In a study involving two covariates and a response variable, the following data were obtained as shown in Table 3;

X_1	6	7	7	8	10	10	8
X_2	4	20	20	10	10	2	1
Y	49	55	50	42	17	26	16

Table 3

a) Write down the predictor matrix X and the response vector y

(2 Marks)

b) Compute $X^T X$ and $X^T y$

(4 Marks)

c) Obtain the estimated vector of slopes and hence write down the regression equation. Given that the adjoint matrix of $X^T X$ is:

(5 Marks)

$$Adj(X^T X) = \begin{bmatrix} 209558 & -22872 & -2282 \\ -22872 & 2658 & 168 \\ -2282 & 168 & 98 \end{bmatrix}$$

d) Estimate the error variance of the model

(4 Marks)

e) Construct the 95% confidence interval for $E\left[Y \middle| X_1 = 11, X_2 = 5\right]$

(5 Marks)

QUESTION FOUR (20 MARKS)

a) Describe how to perform regression analysis using R software

(3 Marks)

b) In order to study the amount of body fat (Y) a statistician took measurements of the triceps (X1), thigh (X2) and midarm (X3) of 20 women; and performed analysis by running R code $lm(Y \sim X1 + X2 + X3)$ and obtained the output presented in Output 1.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -32.32719    0.71288 -45.348  <2e-16 ***
X1            0.83303    0.01779  46.833  <2e-16 ***
X2            0.52401    0.01234  42.459  <2e-16 ***
X3            0.02638    0.01836   1.437    0.17
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1935 on 16 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9985
F-statistic:  4263 on 3 and 16 DF,  p-value: < 2.2e-16
```

Output 1

Use the output to answer the questions below:

i) Write down the fitted model (2 Marks)

ii) what are the values of $\beta's$ and interpret (4 Marks)

iii) What is the value of R-Squared and interpret (3 Marks)

iv) Which of the independent variables are statistically significant and why?

(2 Marks)

v) Is the model statistically significant?

(3 Marks)

vi) Determine the amount of body fat of the women in Table 4;

(3 Marks)

Triceps Skinfold	**Thigh Circumference**	**Midarm Circumference**
43.1	29.1	11.9
49.8	28.2	22.8
51.9	37.0	18.7

Table 4

QUESTION FIVE (20 MARKS)

a) What is logit? How can you transform logit to probability?

(4 Marks)

b) A researcher is interested in how variables, such as GRE (Graduate Record Exam) scores, GPA (Grade Point Average) and prestige undergraduate institution, effect admission into graduate school. The response variable, admit/don't admit, is a binary variable. The researcher performs a logistic model using R function $glm(\ )$ and obtains the Output 2

i) Write down the R code that could have generates the Output 2. Assume the data is contained in a file binary.csv saved in a folder KMA320 in drive C

(4 Marks)

ii) Write the resulting logistics regression equation

(2 Marks)

iii) What does the intercept; coefficients of GRE and GPA from the model tell you?

(6 Marks)

iv) What is the predicted probability of having being admitted with a GRE of 500, GPA of 3.54 and a rank of 2 on prestige of undergraduate institution?

(4 Marks)

```
Call:
glm(formula = admit ~., family = binomial, data = adm)

Deviance Residuals:
    Min      1Q  Median      3Q     Max
-1.6268  -0.8662  -0.6388   1.1490   2.0790

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.989979   1.139951  -3.500 0.000465 ***
gre          0.002264   0.001094   2.070 0.038465 *
gpa          0.804038   0.331819   2.423 0.015388 *
rank2       -0.675443   0.316490  -2.134 0.032829 *
rank3       -1.340204   0.345306  -3.881 0.000104 ***
rank4       -1.551464   0.417832  -3.713 0.000205 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 499.98  on 399  degrees of freedom
Residual deviance: 458.52  on 394  degrees of freedom
AIC: 470.52

Number of Fisher Scoring iterations: 4
```

Output 2

a)