JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF AGRICULTURAL AND FOOD SCIENCES

UNIVERSITY EXAMINATION FOR DEGREE OF BACHELOR OF SCIENCE ACTUARIAL SCIENCE

4TH YEAR 2ND SEMESTER 2021/2022 ACADEMIC YEAR

MAIN REGULAR

COURSE CODE: WAB 2312

COURSE TITLE: STATISTICAL MODELLING

EXAM VENUE: **STREAM:**

DATE: EXAM SESSION:

TIME: 2.00 HOURS

Instructions:

1. **Answer question 1 (Compulsory) and ANY other 2 questions in SECTION B**
2. **Candidates are advised not to write on the question paper.**
3. **Candidates must hand in their answer booklets to the invigilator while in the examination room.**

QUESTION ONE COMPULSORY (30 MARKS)

a) Briefly state three advantages and three disadvantages of polynomial regression (6marks)

b) Briefly state four assumptions of nonlinear regression (4marks)

c) Given the data below;

i) Write an exponential equation (4marks)

ii) Based on the data, predict f(x) when x = 10 (1mark)

X	1	2	3	4	5
f(x)	9.3	21.8	50.8	118.6	276.6

d) The decreasing value of an item that was purchased new 2008 is listed below.

Year	2008	2009	2010	2011	2012	2013	2014
Value of item in $	40	35.5	29.61	21.20	15.73	13.24	10.99

i) Write an equation relating the value of the item and the year it was purchased (4mks)

ii) Predict when the item will be worth $ 1.92 (1mark)

iii) Find a quadratic regression model for the following data: (5marks)

X	Y
3	2.5
4	3.2
5	3.8
6	6.5
7	11.5

e) Use the data below to regress the data to a second order polynomial and find the value of $\propto$ when Temperature is 70^0F (5marks)

Temperature (^{0}F)	80	40	-40	-120	-200	-280
$\propto$	6.47	6.24	5.72	5.09	4.30	3.33

QUESTION TWO (20 MARKS)

a) Suppose we have the following dataset with one response variable *y* and two predictor variables X_1 and X_2:

Y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

Required: fit a multiple linear regression model to this dataset. (10marks)

a) The random errors ε in multiple linear regression model $y = X\beta + \varepsilon$ are assumed to be identically and independently distributed following the normal distribution with zero mean and constant variance. Here y is a $n\ x 1$ vector of observations on response variable, X is a $n\ x\ K$matrix of n observations on each of the K explanatory variables, β is a $K\ X1$ vector of regression coefficients and ε is a $n\ x\ 1$ vector of random errors. The residuals $\hat{\varepsilon} = y - \hat{y}$ based on the ordinary least squares estimator of β have, in general, zero mean, non-constant variance and are not independent. Prove (10marks)

QUESTION THREE (20 MARKS)

a) It is suspected from theoretical considerations that the rate of water flow from a firehouse is proportional to some power of the nozzle pressure. Assume pressure data is more accurate. You are transforming the data

Flow rate, F(gallons/min)	96	129	135	145	168	235
Pressure, p(psi)	11	17	20	25	40	55

What is the exponent of the nozzle pressure in the regression model $F = ap^b$ (10marks)

b) When using the transformed data model to find the constants of the regression model $y = ae^{bx}$ To best fit $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, what is the sum of the square of the residuals that is minimized (5marks)

c) Find the transformed data model for the stress-strain curve $\sigma = k_1 \varepsilon e^{-k_1 \varepsilon}$ for concrete in compression, where σ is the stress and ε is the strain, (1mark)

d) Fill in the missing entries of the partially completed one-way ANOVA table. (4marks)

Source	df	SS	MS = SS/df	F-statistic
Treatments		2.124	0.708	0.75
Error	20			
Total				

QUESTION FOUR (20 MARKS)

a) Consider the multiple linear regression model y = X β + ε , E (ε) = 0, $V(\varepsilon) = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ where y is a $n\ x$ 1 vector of observations on response variable, $X\ is\ a\ n\ x\ K$ is a matrix of n observations on each of the K explanatory variables, $\beta\ is\ a\ K\ x$ 1 vector of regression coefficients and ε is a $n\ x$ 1 vector of random errors. Let h_{ii} be the $i\ th$ diagonal element of matrix $H = X(X'X)^{-1}$. Calculate The variance of the PRESS residual (10marks)

b) The sales of a company (in million dollars) for each year are shown in the table below.

x (year)	c) 2005	d) 2006	e) 2007	f) 2008	g) 2009
y (sales)	h) 12	i) 19	j) 29	k) 37	l) 45

i) Find the least square regression line y = a x + b.
ii) Use the least squares regression line as a model to estimate the sales of the company in 2012. (10marks)

QUESTION FIVE (20 MARKS)

a) Consider the linear model $y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon, E(\varepsilon) = 0, V(\varepsilon) = 1$ where the study variable y and the explanatory variables $X_1\ and\ X_2$ are scaled to length unity and the correlation coefficient between $X_1\ and\ X_2$ is 0.5. Let $b_1\ and\ b_2$ be the ordinary least squares estimators of $\beta_1\ and\ \beta_2$ respectively. Find the covariance between $b_1\ and\ b_2$ (10marks)

b) A company manufactures an electronic device to be used in a very wide temperature range. The company knows that increased temperature shortens the life time of the device, and a study is therefore performed in which the life time is determined as a function of temperature. The following data is found: (10marks)

Temperature in Celcius (t)	10	20	30	40	50	60	70	80	90

Life time in hours (y)	420	365	285	220	176	117	69	34	5

I. Calculate the 95% confidence interval for the slope in the usual linear regression model, which expresses the life time as a linear function of the temperature.
II. Can a relation between temperature and life time be documented on level 5%