$\beta_2 = \dfrac{\Sigma(X-\bar{Y})(Y_i - \bar{Y})}{\Sigma(X_2 - \bar{X})^2}$ $\quad \beta^* = 5\Sigma$

STA 2408: REGRESSION MODELLING II CAT 1

a) State the assumptions essential for the ordinary least squared estimators to be BLUE. (Be sure to explain where your assumptions are used in "BLUE").

b) In the regression $Y_i = \beta_1 + \beta_2 X_i + u_i$, suppose you multiple each X_i value by a constant, say 5. Will it change the residuals and fitted values of Y_i? Briefly explain.

c) What is the interpretation of an estimated coefficient from a simple linear regression model and that from a ~~log-log~~ loglin regression model? How could you judge a multiple regression model has a better performance in terms of statistical information?

d) i) Express the probability density function of the gamma distribution in the form of a member of the exponential family of distributions. Specify the natural and scale parameters.
ii) State the corresponding canonical link function for generalised linear modelling if the response variable has a gamma distribution

) TECHNOLOGY

e) In a two-variable regression as $Y_i = \beta_1 + \beta_2 X_i + u_i$, the standard error of $\hat{\beta}_2$ which is calculated as $Se(\hat{\beta}_2) = \dfrac{\hat{\sigma}}{\sqrt{\sum x^2}}$. Discuss two factors that could determine the standard error of $\hat{\beta}_2$ to be smaller in order to obtain the precision of estimator?

19

f)

A portfolio consists of k independent travel insurance policies. Each policy covers the policyholder's trips over one year. For policy i, the number of claims in the jth month of the covered year, Y_{ij}, is assumed to have a distribution given by

$$P(Y_{ij} = y) = \theta_{ij}(1-\theta_{ij})^y \quad \text{for } y = 0, 1, 2, \ldots$$

ild as an adult of 30

where θ_{ij} are unknown constants between 0 and 1.

(i) Write down the likelihood function and obtain the maximum likelihood estimate for the parameters θ_{ij}. [5 marks]

(ii) Show that $P(Y_{ij} = y)$ can be written in exponential family form and suggest natural parameter.

ir and the altitude h

1.60
0.95

(iii) Suppose that θ_{ij} depends on the temperature x_j recorded in the jth month. Explain why it is not appropriate to set $\theta_{ij} = \alpha + \beta x_j$. Suggest another relationship between θ_{ij} and $\alpha + \beta x_j$ that might be used.

$y\log 1 - y\log\theta_{ij}$

$= 0.1315$. assuming
f the mass density of

[8 marks]

$= m(X_i) + \epsilon_i$, where
we define $s_n(x) =$
$x)^l Y_i$. We used the

standard Gaussian kernel and $h = 0.2420$. Let $x_0 = 1$. We obtained $s_{nl}(x_0), l =, 0, 1, 2$ as $22.8038, -0.5758, 1.1189$ respectively, and $T_{nl}(x_0), l = 0, 1$ as $24.0092, 0.1631$ respectively. Compute the local linear estimators of $m(x_0)$ and $m'(x_0)$ respectively. [7 marks]

4. Given the data

X	20	30	15	25	28
Y	45.6	35.3	40.3	20.0	43.2

$y = a + bx$

where X and Y are predictor and response variables,respectively.Using a rectangular Kernel function and 3-nearest neighbour(K-NN)find the estimate of 1

$a = k \quad b = -0.1315$

$x = 4$

$y = A$

$a_1 = \dfrac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x_i^2 - (\Sigma x_i)^2}$

$a_0 = \bar{y} - a_1\bar{x}$

Σy_i

$\dfrac{\Sigma h_i}{n}$

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

CONTINUOUS ASSESSMENT TEST 2018/2019

STA 2408: REGRESSION MODELLING II

TIME: 1 HOUR

INSTRUCTIONS TO CANDIDATES:

* *Answer question al. the questions*
* *Be neat and show all your workings*

1. The height of a child at different ages Estimate the height of the child as an adult of 30

t(years)	0	5.0	8.0	12	16	18
H(inches)	20	36.2	52	60	69.2	70

[5 marks]

years of age using the growth model, $H = \frac{a}{1+be^{-ct}}$.

2. There is a functional relationship between the mass density ρ of air and the altitude h above the sea level.

Altitude above the sea level, h (km)	0.32	0.64	1.28	1.60
Mass Density, $\rho(kg/m^3)$	1.15	1.10	1.05	0.95

In the regression model $\rho = k_1 e^{-k_2 h}$, the constant k_2 is found as $k_2 = 0.1315$. assuming that the mass density of air at the top of the atmosphere is $\frac{1}{1000^{th}}$ of the mass density of air at the sea level. [8 marks]
Calculate the altitude in kilometers of the top of the atmosphere.

3. For a given sample $(X_i, Y_i), i = 1, 2, \cdots n$ from an NPR model $Y_i = m(X_i) + \epsilon_i$, where $n = 140$, $X_i \in [0.1894, 2.6855]$ and ϵ_i are i.i.d with $E\epsilon_1^2 = \sigma^2$, we define $s_n(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^l$ and $T_{nl}(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^l Y_i$. We used the standard Gaussian kernel and $h = 0.2420$. Let $x_0 = 1$. We obtained $s_{nl}(x_0), l = 0, 1, 2$ as $22.8038, -0.5758, 1.1189$ respectively, and $T_{nl}(x_0), l = 0, 1$ as $24.0092, 0.1631$ respectively. [7 marks]
Compute the local linear estimators of $m(x_0)$ and $m'(x_0)$ respectively.

4. Given the data

X	20	30	15	25	28
Y	45.6	35.3	40.3	20.0	43.2

$y = a + bx$

where X and Y are predictor and response variables,respectively.Using a rectangular Kernel function and 3-nearest neighbour(K-NN)find the estimate of

$a = k \quad b = -0.1315$
$x = h$
$y = A$

$a_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x_i^2 - (\sum x_i)^2}$

$\frac{\sum h_i}{n}$

$a_0 = \bar{y} - a_1 \bar{x}$

$\sum y_i$

i) Density function of X at point x=29.5 and at x=34

ii) regression function at point x=29.5 and at x=34 [3 marks]

[3 marks]

5. Let $X_1, X_2 \quad \epsilon\ [0,1]$ and

$$\begin{cases} Z = 1, & \text{if } X_1 = 1 \text{ or } X_2 = 1 \\ Z = 0, & \text{if } X_1 = 0 \text{ or } X_2 = 0 \end{cases}$$

be the logical XOR classification to be learned by a perception. The training set then consists of input vectors including the first co-ordinate $x_0 = 1$;

$$X^1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad X^2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad X^3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad X^4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

with correction classification$Z^{(1)} = Z^{(2)} = Z^{(3)} = 1, Z^{(4)} = 0$.
The perception with weightsw_0, w_1, w_2 classifies an object as 1 if and only if $w_0x_0 + w_1x_1 + w_2x_2 > 0$ and as 0 elsewhere. Taking $w = (0,0,0)^t$ as initial weight and $\eta = 1$ as relaxation factor, train the perception and find the weight that achieves correct classification for all input vectors in the training set.

[12marks]

(a) [Explain] the difference between the following as applied in regression analysis;
(i) Residuals and errors in a regression model (2 Marks)
(ii) Omitted and irrelevant variables in a model (2 Marks)
(iii) Linear and non linear regression models (2 Marks)
(iv) R-squared and adjusted R-squared (2 Marks)

(b) Explain four conditions that you may use to decide whether a variable should be included in the model. (6 Marks)

(c) A certain analysis using R-statistical software produced the following partial outputs (A, B and C), with Z as response variable.

OUTPUT A

Call:

lm(formula = Z ~ X1 + X2)

Coefficients:

	Estimate	Std. Error
(Intercept)	2.9858	0.1121
X1	4.5310	0.1121
X2	1.4560	0.1182

Residual standard error: 2.505 on 497 degrees of freedom

Multiple R-squared: 0.7816, Adjusted R-squared: 0.7807

F-statistic: 889.4 on 2 and 497 DF.

(Computed: $t = \frac{b_k - 0}{s.e(b_k)}$

OUTPUT B

Call:

lm(formula = Z ~ X1 + X2 + X3 + X4)

Coefficients:

	Estimate	Std. Error
(Intercept)	2.9452324	0.0440713
X1	4.5527717	0.0437007
X2	1.5315566	0.0463020
X3	2.3885817	0.0453280
X4	-0.0007734	0.0012271

tabulated: d.f

n-p degrees.

Residual standard error: 0.9763 on 495 degrees of freedom

Multiple R-squared: 0.967, Adjusted R-squared: 0.9667

OUTPUT C

Analysis of Variance Table

Response: Z

	Df	Sum Sq	Mean Square. SS/df	F-Ratio: $\frac{MSR}{MSE}$
Model	4	3452.925		
Residuals	495	471.8		

(i) Using output B, test whether individual independent variables are important in predicting Z (8 Marks)

(ii) Given that the predictors in output C are X1,X2,X3 and X4, test for the overall significance of the regression model. F-test. $H_0: \beta_1=\beta_2=\beta_3=\beta_4=0$ $H_1: \beta_i \neq 0$. $F = \frac{3452.9/4}{471.8/495}$ (3 Marks)

(iii) Determine whether the joint contribution of X1 and X2 in model under output B significant. (3 Marks)

(iv) With available information, find the best predictive equation. (2 Marks)

Question Two (20 Marks)

A certain data on with variables (X,Y) produced the following graphical display:

> plot(X,Y)

XLP Valid

JKUAT CONTINOUS ASSESSEMENT TESTS

NAME Mugo Grace Annette Mumbi
REG NO. SC283-1171/2011
COURSE BSC. Financial Engineering YEAR OF STUDY 4
SHEET NO.
UNIT CODE STA 2408 TITLE Regression Modelling 2. Cat 2
DATE

NOTE: This stationery will be used for Continuous Assessment work only.

It will be a breach of examination regulations to use it otherwise.

For each of the following response functions, indicate whether it is a linear response function, an intrinsically linear response function or a non linear response function. In the case of an intrinsically linear response function, state how it can be linearized by a suitable transformation.

i) $f(x,y) = exp(B_0 + B_1 X)$

Differentiate and check if partial derivatives with respect to parameters are independent of parameters

$f(x,y) = e^{B_0+B_1X}$

$\frac{\partial f(x,y)}{\partial B_0} = e^{B_0+B_1X} = exp(B_0+B_1X)$

$\frac{\partial f(x,y)}{\partial B_1} = e^{B_0+B_1X} \cdot X = [exp(B_0+B_1X)]X$

The partial derivatives are not independent of parameters. $f(x,y) = exp(B_0+B_1X)$ is non linear

Intrinsically linear response functions transformation

⇒ Take the natural log (Ln) of the response function the is non linear to linearize it ⇒

$f(x,y) = exp(B_0+B_1X)$ is non linear by test above

Linearize

$ln(f(x,y)) = ln\, e^{B_0+B_1X} = B_0 + B_1X$

$ln(f(x,y)) = B_0+B_1X$ - Test for linearity

$\frac{\partial(B_0+B_1X)}{\partial B_0} = 1$; $\frac{\partial(B_0+B_1X)}{\partial B_1} = X$

The partial derivatives after linearizing are independent of the parameters B_0 and B_1. Thus $f(x,y) = exp(B_0+B_1X)$ is an intrinsically linear response function

ii) $f(x,y) = B_0 + B_1(B_2)^{X_1} - B_2X_2$

Partial derivatives: $\frac{\partial f(x,y)}{\partial B_0} = 1$

$\frac{\partial f(x,y)}{\partial B_1} = (B_2)^{X_1}$ $\frac{\partial f(x,y)}{\partial B_2} = X_1(B_2)^{X_1-1} - X_2$

The function is non linear since partial derivatives are not independent of parameters B_0, B_1 and B_2

Linearize

$ln(f(x,y)) = ln(B_0 + B_1(B_2)^{X_1} - B_2X_2)$

$\frac{\partial(ln\, f(x,y))}{\partial B_0} = \frac{1}{B_0 + B_1(B_2)^{X_1} - B_2X_2} \cdot 1$

$\frac{\partial\, ln(f(x,y))}{\partial B_1} = \frac{1}{B_0 + B_1(B_2)^{X_1} - B_2X_2} \cdot (B_2)^{X_1}$

$\frac{\partial(ln(f(x,y)))}{\partial B_2} = \left(\frac{1}{B_0 + B_1(B_2)^{X_1} - B_2X_2}\right) \cdot X_1(B_2)^{X_1-1} - X_2$

- Partial derivatives are still not independent of parameters - function is non linear

2½/3
3

(iii) $f(x,y) = B_0 + \frac{B_1}{B_0} x$

$\frac{\partial f(x,y)}{\partial B_0} = 1 + -B_1 x B_0^{-2} = 1 - B_1 x B_0^{-2}$

$\frac{\partial f(x,y)}{\partial B_1} = x/B_0$ → Derivatives not independent of parameters thus non linear

Linearize: $\ln(f(x,y)) = \ln(B_0 + \frac{B_1}{B_0} x)$

$\frac{\partial \ln f(x,y)}{\partial B_0} = \frac{1}{(B_0 + \frac{B_1}{B_0} x)} \cdot 1 - \frac{B_1 x}{B_0^2}$

$\frac{\partial \ln f(x,y)}{\partial B_1} = \frac{1}{(B_0 + \frac{B_1}{B_0} x)} \cdot \frac{x}{B_0}$

– Partial derivatives with respect to parameters are not independent of the parameters for both the response function and transformed function.

$f(x,y) = B_0 + \frac{B_1}{B_0} x$ is Non linear

b) let $x^{(1)} = \begin{pmatrix}1\\0\\0\end{pmatrix}$ $x^{(2)} = \begin{pmatrix}1\\1\\0\end{pmatrix}$ $x^{(3)} = \begin{pmatrix}1\\0\\1\end{pmatrix}$ $x^{(4)} = \begin{pmatrix}1\\1\\1\end{pmatrix}$

with correct classification: $z^{(1)} = 0$, $z^{(2)} = z^{(3)} = z^{(4)} = 1$

Train perceptron, find weight that correctly classifies all.

Assume classification of 1 if $\Sigma w_i x_i > 0$, 0 otherwise for $w_0, w_1 \ldots w_p$

Take $w = (0\ 0\ 0)^T$ as initial weight, $\eta = 1$

$y^{(1)} = w^T x^{(1)} = (0\ 0\ 0)\begin{bmatrix}1\\0\\0\end{bmatrix} = 0 = z^{(1)}$ $\Sigma w_i x_i = 0$

Correct classification as 0 since $z^{(1)} = 0$

$y^{(2)} = w^T (x)^{(2)} = (0\ 0\ 0)\begin{pmatrix}1\\1\\0\end{pmatrix} = 0$

Classified as 0 since $\Sigma w_i x_i = 0$. But $z^{(2)} = 1$ so $x^{(2)}$ is not classified correctly

$w_{new} = w + \eta(z^{(2)} - y^{(2)})x^{(2)}$

$= \begin{pmatrix}0\\0\\0\end{pmatrix} + 1(1-0)\begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix}$

$(1\ 1\ 0)\begin{pmatrix}1\\1\\0\end{pmatrix} = 2$; w_{new} classifies $x^{(2)}$ correctly

$y^{(3)} = (1\ 1\ 0)\begin{pmatrix}1\\0\\1\end{pmatrix} = 1$; $z^{(3)}$

$\Sigma w_i x_i > 0$ so $x^{(3)}$ is correctly classified

$y^{(4)} = (1\ 1\ 0)\begin{pmatrix}1\\1\\1\end{pmatrix} = 2$; $\Sigma w_i x_i > 0$

$z^{(4)} = 1$ so $x^{(4)}$ is correctly classified.

*

Check $x^{(1)}$ using $W_{new} = (1\ 1\ 0)$

$y^{(1)} = w^T x^{(1)} = (1\ 1\ 0)\begin{pmatrix}1\\0\\0\end{pmatrix} = 1$ but $z^{(1)} = 0$

so $x^{(1)}$ is incorrectly classified

$W_{new} = w + \eta(z^{(1)} - y^{(1)})x^{(1)}$

$= \begin{pmatrix}1\\1\\0\end{pmatrix} + 1(0-1)\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\1\\0\end{pmatrix}$

$W_{new} = (0\ 1\ 0)$

$y^{(2)} = w^T x^{(2)} = (0\ 1\ 0)\begin{pmatrix}1\\1\\0\end{pmatrix} = 1$ $w x > 0$

$z^{(2)} = 1$ so $x^{(2)}$ is correctly classified.

$y^{(3)} = w^T x^{(3)} = (0\ 1\ 0)\begin{pmatrix}1\\0\\1\end{pmatrix} = 0$

But $z^{(3)} = 1$ so $x^{(3)}$ is incorrectly classified

$W_{new} = w + \eta(z^{(3)} - y^{(3)})x^{(3)}$

$= \begin{pmatrix}0\\1\\0\end{pmatrix} + 1(1-0)\begin{pmatrix}1\\0\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$

$W_{new} = (1\ 1\ 1)$

$y^4 = w^T x^4 = (1\ 1\ 1)\begin{pmatrix}1\\1\\1\end{pmatrix} = 3$ > 0

$z^4 = 1$ so x^4 is correctly classified

Check $x^{(1)}$ using $W_{new} = (1\ 1\ 1)$

$y^{(1)} = w^T x^{(1)} = (1\ 1\ 1)\begin{pmatrix}1\\0\\0\end{pmatrix} = 1$

but $z^{(1)} = 0$ so classification is incorrect.

The input vector $X^{(t)}_{t=1,2,\cdots,T}$ are successfully used as inputs of the perceptron and the outputs $Y^{(t)}_{t=1,2,\cdots,T}$ are compared to their targets, i.e, to the correct classification $Z^{(t)}_{t=1,2,\cdots,T}$. If one of these steps results in $Y^{(t)} = Z^{(t)}$, then the weights are not changed. However if $Y^{(t)} \neq Z^{(t)}$, then the weights vector $W = (\hat{\omega}_1, \hat{\omega}_2, \cdots, \hat{\omega}_n)^{\mathrm{T}}$ is adapted to the data in the following manner:

$$W_{\text{new}} = W + \eta(Z^{(t)} - Y^{(t)})X^{(t)},$$

η is a small relaxation factor which sometimes has to converge to zero to achieve convergence of the algorithm. The initial value W is selected arbitrarily or randomly e.g., from a uniformly distribution $(0,1)^{p+1}$.

The learning algorithm does not stop if all the input vectors have been presented to the network, but after using $X^{(T)}$ as an input, the algorithm starts all over again at the start of the sample using input $X^{(T)}$. The iteration proceeds step by step through training set several times until all objects are classified correctly by the network or some measure for the number and importance of errors becomes satisfactorily small.

XOR is an example of logical operators. XOR-function is only true if just one and only one of the input value is true, false otherwise.

Example 5.4.1. Let $p = 2$ and $X_1, X_2 \in (0,1)$. The classification to be learned is the logical XOR.

$$Z = 1, \text{ if } X_1 = 1 \text{ or } X_2 = 2$$
$$Z = 0, \text{ if } X_1 = 0 \text{ and } X_2 = 0$$

$$X^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad X^{(2)} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad X^{(3)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad X^{(4)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

with correct classification $Z^{(1)} = Z^{(2)} = Z^{(3)} = 1$, $Z^{(4)} = 0$. The preceptron with weights $\omega_0, \omega_1, \omega_2$ classifies an object as 1 if and only if (iff) $W^{\mathrm{T}}X = \omega_0 X_0 + \omega_1 X_1 + \omega_2 X_2 > 0$ and as 0 elsewhere. Taking $W = (0,0,0)^{\mathrm{T}}$ as the initial weight and we set $\eta = 1$, we find the weight that classifies correctly both inputs variables.

The steps of the delta algorithm are now;

(1). $Y^{(1)} = W^{\mathrm{T}}X^{(1)} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0 \neq Z^{(1)}$.

The $X^{(1)}$ is classified incorrectly. The weights are modified.

$$W_{\text{new}} = [0\,0\,0]^T + (1-0)[1\,1\,0]^T = [1\,1\,0]^T$$

(2). $Y^{(2)} = W^{\mathrm{T}}X^{(2)} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 1 = Z^{(2)}$.

The input $X^{(2)}$ is correctly classified.

(3). $Y^{(3)} = W^T X^{(3)} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 2 \equiv 1 = Z^{(3)}$.

The input $X^{(3)}$ is correctly classified.

(4). $Y^{(4)} = W^T X^{(4)} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1 \neq Z^{(4)} = 0$.

The input $X^{(4)}$ is classified incorrectly. The weights are modified again.

$$W_{\text{new}} = [1\,1\,0]^T + (0-1)[1\,0\,0]^T = [0\,1\,0]^T$$

(5). $X^{(1)}$ is considered as an input again and the

$$Y^{(1)} = W^T X^{(1)} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 1 = Z^{(1)}.$$

The perceptron classifies $X^{(1)}$ correctly.

(6). As $Y^{(2)} = W^T X^{(2)} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 0 \neq Z^{(1)}$.

The input $X^{(2)}$ is classified incorrectly.

$$W_{\text{new}} = [0\,1\,0]^T + (1-0)[1\,0\,1]^T = [1\,1\,1]^T$$

(7). As $Y^{(3)} = W^T X^{(3)} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 3 > 0 \equiv 1 = Z^{(3)}$.

The input $X^{(3)}$ is correctly classified.

(8). AS $Y^{(4)} = W^T X^{(4)} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1 \neq Z^{(4)} = 0$.

The input $X^{(4)}$ is classified incorrectly. The weights are modified again.

$$W_{\text{new}} = [1\,1\,1]^T + (0-1)[1\,0\,0]^T = [0\,1\,0]^T$$

Now the algorithm essentially stops as the weight vector $\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T$ achieves correct classification of all input vectors in the training set. The perceptron has learned the XOR-function of the set $(0,1)^2$.

Question 5.4.1. Let $p = 2$ and $X_1,\ X_2 \in (0,1)$ and

$$Z = 1, \text{ if } X_1 = 1 \text{ or } X_2 = 1$$
$$Z = 0, \text{ if } X_1 = 0 \text{ and } X_2 = 0$$

STA 2408: REGRESSION MODELLING II CAT 1

a) State the assumptions essential for the ordinary least squared estimators to be BLUE. (Be sure to explain where your assumptions are used in "BLUE").

b) In the regression $Y_i = \beta_1 + \beta_2 X_i + u_i$, suppose you multiple each X_i value by a constant, say 5. Will it change the residuals and fitted values of Y_i? Briefly explain.

c) What is the interpretation of an estimated coefficient from a simple linear regression model and that from a log-log regression model? How could you judge a multiple regression model has a better performance in terms of statistical information?

d) i) Express the probability density function of the gamma distribution in the form of a member of the exponential family of distributions. Specify the natural and scale parameters.

ii) State the corresponding canonical link function for generalised linear modelling if the response variable has a gamma distribution

e) In a two-variable regression as $Y_i = \beta_1 + \beta_2 X_i + u_i$, the standard error of $\hat{\beta}_2$ which is calculated as $Se(\hat{\beta}_2) = \frac{\hat{\sigma}}{\sqrt{\sum x^2}}$. Discuss two factors that could determine the standard error of $\hat{\beta}_2$ to be smaller in order to obtain the precision of estimator?

f)

A portfolio consists of k independent travel insurance policies. Each policy covers the policyholder's trips over one year. For policy i, the number of claims in the jth month of the covered year, Y_{ij}, is assumed to have a distribution given by

$$P(Y_{ij} = y) = \theta_{ij}(1-\theta_{ij})^y \quad \text{for } y = 0, 1, 2, \ldots$$

$p(q)^y$

where θ_{ij} are unknown constants between 0 and 1.

(i) Write down the likelihood function and obtain the maximum likelihood estimate for the parameters θ_{ij}.

(ii) Show that $P(Y_{ij} = y)$ can be written in exponential family form and suggest natural parameter.

$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

$\hat{a}_1 = \bar{y} - \hat{a}_2 \bar{z} =$

$\hat{a}_2 = \frac{1}{5}\hat{\beta}_2$

$y = \beta_1 + \beta_2 X_i + e$

$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

$\hat{\beta}_2 = \frac{\sum xy}{\sum x^2}$

$\bar{z} = 5\bar{x}$

$y = a_1 + a_2 z + e$

where $z = 5x$

$\hat{a}_1 = \bar{y} - a_2 \bar{z}$

$\hat{a}_2 = \frac{\sum zy}{\sum z^2} = \frac{5\sum xy}{25\sum x^2}$

$\hat{a}_2 = \frac{1}{5}\hat{\beta}_2$ $\quad a_1 = \bar{y} - \frac{1}{5}\hat{\beta}_2 \sum z^2$

d) i) Express the probability density function of the gamma distribution in the form of a member of the exponential family of distributions. Specify the natural and scale parameters.

ii) State the corresponding canonical link function for generalised linear modelling if the response variable has a gamma distribution

Solution

(i)

We need to express the distribution function of the Gamma distribution in the form:

$$f_Y(y;\theta,\varphi) = \exp\left[\frac{(y\theta - b(\theta))}{a(\varphi)} + c(y,\varphi)\right]$$

$\exp\{ Q(\theta)T(\alpha) + D(\theta) + S(\alpha)$

Suppose Y has a Gamma distribution with parameters α and λ. Then

$$f_Y(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}$$

$\Sigma\, Q(\theta)T(\alpha)$

And substituting $\lambda = \frac{\alpha}{\mu}$ we can write the density as

$$f_Y(y;\theta,\varphi) = \frac{\alpha^\alpha}{\mu^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{y\alpha}{\mu}}$$

$\frac{y}{\mu}$ α

exp

$$f_Y(y;\theta,\varphi) = \exp\left[\underbrace{\left(-\frac{y}{\mu} - \log\mu\right)\alpha}_{\Sigma\, Q(\theta)T(\alpha)} + \underbrace{(\alpha-1)\log y + \alpha\log\alpha - \log\Gamma(\alpha)}_{D(\theta)}\right]$$

$S(\alpha)$

Which is in the right form with $\theta = -\frac{1}{\mu}$; $\varphi = \alpha$; $a(\varphi) = \frac{1}{\varphi}$; $b(\theta) = -\log(-\theta)$

and $c(y,\varphi) = (\varphi - 1)\log y + \varphi\log\varphi - \log\Gamma(\varphi)$.

Thus the natural parameter is $\frac{1}{\mu}$ ignoring the minus sign, and the scale parameter is α.

(ii) The corresponding link function is $\frac{1}{\mu}$

$$\log \exp\left\{ \frac{\alpha^\alpha}{\mu^\alpha \Gamma\alpha}\, y^{\alpha-1}\, e^{-\frac{y\alpha}{\mu}} \right\}$$

$$\alpha\log\alpha - \alpha\log\mu - \log\Gamma\alpha + (\alpha-1)\log y - \frac{y\alpha}{\mu}$$

$$\alpha\left(-\log\mu - \frac{y}{\mu}\right)$$

e) In a two-variable regression as $Y_i = \beta_1 + \beta_2 X_i + u_i$, the standard error of $\hat{\beta}_2$ which is calculated as $Se(\hat{\beta}_2) = \frac{\hat{\sigma}}{\sqrt{\sum x^2}}$. Discuss two factors that could determine the standard error of $\hat{\beta}_2$ to be smaller in order to obtain the precision of estimator?

Solution

(i) As X spread out from the mean, $(\sum x^2)\uparrow \Rightarrow Se(\hat{\beta}_2)\downarrow$

(ii) As $\hat{\sigma}^2 = \frac{\sum \hat{u}^2}{n-1}$, when $n \to \infty \Rightarrow \hat{\sigma}^2 \downarrow \Rightarrow Se(\hat{\beta}_2)\downarrow$

f) A portfolio consists of k independent travel insurance policies. Each policy covers the policyholder's trips over one year. For policy i, the number of claims in the jth month of the covered year, Y_{ij}, is assumed to have a distribution given by

$$P(Y_{ij} = y) = \theta_{ij}(1-\theta_{ij})^y \quad \text{for } y = 0,1,2,\ldots$$

where θ_{ij} are unknown constants between 0 and 1.

(i) Write down the likelihood function and obtain the maximum likelihood estimate for the parameters θ_{ij}.

Solution

$\frac{1}{n-1} \quad \Sigma e^2.$

(i) The likelihood is

$$L = \prod_{i=1}^{k} \prod_{j=1}^{12} \theta_{ij}(1-\theta_{ij})^{y_{ij}}$$

Where y_{ij} is the number of claims on the ith policy in the jth month.

Taking the logarithm of L we have

$$\log L = \sum_{i=1}^{k}\sum_{j=1}^{12}\left[\log\theta_{ij} + y_{ij}\log(1-\theta_{ij})\right]$$

and so

$$\frac{\partial \log L}{\partial \theta_{ij}} = \frac{1}{\theta_{ij}} - \frac{y_{ij}}{1-\theta_{ij}}$$

And setting the derivative to zero we find $1-\hat{\theta}_{ij} = y_{ij}\hat{\theta}_{ij}$ so that

$$\hat{\theta}_{ij} = \frac{1}{1+y_{ij}}$$

(ii) Show that $P(Y_{ij} = y)$ can be written in exponential family form and suggest its natural parameter.

Solution

$$P(Y_{ij} = y) = \theta_{ij}(1-\theta_{ij})^{y} = e^{\log\theta_{ij} - y\log(1-\theta_{ij})}$$

The natural parameter is $\log(1-\theta_{ij})$.

(iii) Suppose that θ_{ij} depends on the temperature x_j recorded in the jth month. Explain why it is not appropriate to set $\theta_{ij} = \alpha + \beta x_j$. Suggest another relationship between θ_{ij} and $\alpha + \beta x_j$ that might be used.

Solution

The range of $\alpha + \beta x_j$ is $(-\infty, +\infty)$ which means it is not suitable for modelling parameters $\theta_{ij} \in [0,1]$.

A possible relationship to consider is $\log \dfrac{\theta_{ij}}{1-\theta_{ij}} = \alpha + \beta x_j$.

W-1-2-60-1-6

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

UNIVERSITY EXAMINATIONS 2018/2019

FOURTH YEAR FIRST SEMESTER EXAMINATION FOR

THE DEGREE OF BACHELOR OF SCIENCE IN STATISTICS/BACHELOR OF SCIENCE IN FINANCIAL ENGINEERING/ BACHELOR OF SCIENCE IN BIOSTATISTICS/ BACHELOR OF SCIENCE IN ACTUARIAL SCIENCE.

STA 2408: REGRESSION MODELLING II

DATE: DECEMBER 2018 TIME: 2 HOURS

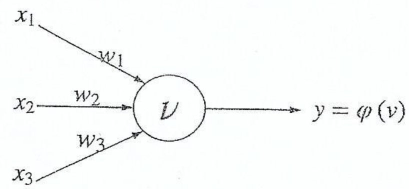
INSTRUCTIONS TO CANDIDATES:

1. *Answer question ONE and any other two questions*
2. *Be neat and show all your workings*
3. *All questions except question one carry equal marks*

This paper consists of 5 printed pages
STACS Examination board 2018/2019.

QUESTION ONE

(a) (i) Below is a diagram of a single artificial neuron (unit)



Flowchart (1).png

The node has three inputs x=(x_1, x_2, x_3) that receive only binary signals (either 0 or 1). How many different input patterns can this code receive? What if the node had four inputs? Five? Can you give a formula that computes the number of binary input patterns for a given number of inputs. [3 marks]

(ii) Suppose that a credit card company decided to deploy a new system for assessing credit worthiness of its customers. The new system is using a feed-forward neural network with a supervised learning algorithm. Suggest in a form of essay what should the bank have before the system can be used? Discuss the problem associated with this requirement. [3 marks]

(b) The height of a child at different ages Estimate the height of the child as an adult of 30

t(years)	0	5.0	8.0	12	16	18
H(inches)	20	36.2	52	60	69.2	70

years of age using the growth model, $H = \frac{a}{1+be^{-ct}}$. [5 marks]

(c) There is a functional relationship between the mass density p of air and the altitude h above the sea level.

Altitude above the sea level, h (km)	0.32	0.64	1.28	1.60
Mass Density, $\rho(kg/m^3)$	1.15	1.10	1.05	0.95

In the regression model $\rho = k_1 e^{-k_2 h}$, the constant k_2 is found as $k_2 = 0.1315$. assuming that the mass density of air at the top of the atmosphere is $\frac{1}{1000^{th}}$ of the mass density of air at the sea level.
Calculate the altitude in kilometers of the top of the atmosphere. [8 marks]

(d) For a given sample $(X_i, Y_i), i = 1, 2, \cdots n$ from an NPR model $Y_i = m(X_i) + \epsilon_i$, where $n = 140$, $X_i \in [0.1894, 2.6855]$ and ϵ_i are i.i.d with $E\epsilon_1^2 = \sigma^2$, we define $s_n(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^l$ and $T_{nl}(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^l Y_i$. We used the standard Gaussian kernel and $h = 0.2420$. Let $x_0 = 1$. We obtained $s_{nl}(x_0), l =, 0, 1, 2$ as $22.8038, -0.5758, 1.1189$ respectively, and $T_{nl}(x_0), l = 0, 1$ as $24.0092, 0.1631$ respectively. Compute the local linear estimators of $m(x_0)$ and $m'(x_0)$ respectively. [7 marks]

(e) The following result are from a perspective study that were considered in building a logistic regression model for predicting capsule=1 that included psa, age, and gleason in the model(model 1). Part of the resulting SAS output follows:

Figure 1: Model 1

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	514.289	411.208
SC	518.229	426.969
-2 Log L	512.289	403.208

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-6.3896	1.4976	18.2045	<.0001
psa	1	0.0266	0.00894	8.8442	0.0029
age	1	-0.0208	0.0188	1.2351	0.2664
gleason	1	1.0790	0.1611	44.8373	<.0001

(i) Write the resulting logistic regression equation for model 1 [1 mark]

(ii) What is the predicted probability of having a capsule=1 for a 69-year old man with psa level of 10mg/ml and a gleason score of 5, according to model 1? [1 mark]

(iii) What does the intercept from the model tell you? [1 mark]

(iv) Calculate the odds ratio and 95% confidence interval for psa from the model. Interpret. [1 mark]

QUESTION TWO (20 MARKS)

(a) In Kernel regression estimation,one may choose a deterministic or stochastic design depending on the type of the problem at hand.

(i) Describe a deterministic design model and give its kernel estimate of the mean function. [3 marks]

(ii) Describe a stochastic design model and give its kernel estimate of the mean function [3 marks]

(b) A steel cylinder at 80^0F of length 12" is placed in a commercially available liquid nitrogen bath (-315^0F). If the thermal expansion coefficient of steel behaves as a second order polynomial function of temperature and the polynomial is found by regressing the data below.

Temperature, $T(^0F)$	Thermal expansion Coefficient, α ($\mu\ in/in/\ ^0F$)
-320	2.76
-240	3.83
-160	4.72
-80	5.43
0	6.00
80	6.47

(i) Fit the data to $\alpha = a_0 + a_1T + a_2T^2$. [10 marks]

(ii) Calculate the reduction in the length of the cylinder in inches. [4 marks]

QUESTION THREE (20 MARKS)

(a) Given the data

X	20	30	15	25	28
Y	45.6	35.3	40.3	20.0	43.2

where X and Y are predictor and response variables,respectively.Using a rectangular Kernel function and 3-nearest neighbour(K-NN)find the estimate of

i) Density function of X at point x=29 [3 marks]

ii) regression function at point x=29 [3 marks]

(b) We have a sample $(X_i, Y_i), i = 1, 2, \cdots, n$ generated from the NPR model $Y_i = m(X_i) + \epsilon_i$, $i = 1, 2, \cdots, n$ where $m(x)$ is an unknown smooth function and $X_1 < X_2 <, \cdots, < X_n$. Let $\boldsymbol{Y} = [Y_1, \cdots, Y_n]^T$ be the response vector and $\hat{Y}=[\hat{Y}_1, \cdots, \hat{Y}_n]^T$ be the estimated response vector, where $\hat{Y}_i = \hat{m}(X_i)$, $i = 1, 2, \cdots, n$ for some non-parametric estimator $\hat{m}(x)$. When $\hat{m}(x)$ is a linear smoother, we have $\hat{\boldsymbol{Y}} = \boldsymbol{AY}$ where $\boldsymbol{A}$ is known as the associated smoother matrix and df=trace(A) is known as the associated degrees of freedom, measuring how complex the fitting model is.

(i) First assume that $\hat{m}(x)$ is the usual regression spline smoother constructed based on the p-th order truncated power basis $\Phi(x)$ using K distinct knots $\tau_1 < \tau_2 < \cdots \tau_K$. When $n > K + p$ and p fixed, show that df will increase as K increasing. [2 marks]

(ii) Assume now that $\hat{m}(x)$ is the N-W estimator using a bandwidth $h > 0$ and a symmetric kernel $K(.)$ which is a pdf. Show that when K is fixed and n is sufficiently large, df will decrease as h increasing. [3 marks]

(iii) Assume now that $\hat{m}(x)$ is the cubic smoothing spline smoother with a smoothing parameter λ. Show that df will decrease as λ increasing. [3 marks]

(iv) Assume now that $\hat{m}(x)$ is the P-spline smoother with p-th order truncated power basis $\Phi(x)$ using K distinct knots $\tau_1 < \tau_2 < \cdots < \tau_K$ and a smoothing parameter λ. Show that when $n > p + K$, df will decrease as λ increases. [6 marks]

QUESTION FOUR (20 MARKS)

(a) Let $X_1, X_2 \epsilon [0,1]$ and

$$\begin{cases} Z = 1, & \text{if } X_1 = 1 \text{ or } X_2 = 2 \\ Z = 0, & \text{if } X_1 = 0 \text{ or } X_2 = 0 \end{cases}$$

be the logical XOR classification to be learned by a perception. The training set then consists of input vectors including the first co-ordinate $x_0 = 1$;

$$X^1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad X^2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad X^3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad X^4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

with correction classification$Z^{(1)} = Z^{(2)} = Z^{(3)} = 1, Z^{(4)} = 0$.
The perception with weightsw_0, w_1, w_2 classifies an object as 1 if and only if $w_0x_0 + w_1x_1 + w_2x_2 > 0$ and as 0 elsewhere.Taking $w = (0,0,0)^t$ as initial weight and $\eta = 1$ as relaxation factor,train the perception and find the weight that achieves correct classification for all input vectors in the training set. [12marks]

(b) We have a sample (X_i, Y_i), $i = 1, 2, \cdots, n$ from an NPR model $Y_i = m(X_i) + \epsilon_i$ where $E(\epsilon_i|X_1, \cdots, X_n) = 0$ and $E(\epsilon_i^2|X_1, \cdots, X_n) = \sigma^2(X_i)$ and $X_1, X_2, \cdots, X_n$ has a pdf $f(x)$. Moreover, $E(\epsilon_i\epsilon_j|X_1, \cdots, X_n) = 0$ for $i \neq j$.Define

$$C_1(h) = -\frac{2}{n}\sum_{j=1}^{n} \epsilon_j[\hat{m}_h(X_j) - m(X_j)], \quad C_2(h) = -\frac{2}{n}\sum_{j=1}^{n} \epsilon_j[\hat{m}_h^{(-j)}(X_j) - m(X_j)]$$

Where $[\hat{m}_h(X_j)$ is the N-W estimator of $m(x)$ at X_j and $\hat{m}_h^{(-j)}$ is the N-W of $m(x)$ at X_j obtained using all the data except (X_j, Y_j).
Find the asymptotic expression of $E(C_1(h)|X_1, \cdots, X_n)$ [8mks]

W1-2-60-1-6

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

UNIVERSITY EXAMINATIONS 2017/2018

FOURTH YEAR FIRST SEMESTER EXAMINATIONS FOR THE DEGREE OF BACHELOR OF SCIENCE IN FINANCIAL ENGINEERING

STA 2408: REGRESSION MODELLING II

DATE: JANUARY 2018 TIME: 2 HOURS

INSTRUCTIONS TO CANDIDATES:

1. *Answer question ONE (section A) and any other two questions in section B.*
2. *Be neat and show all your workings*
3. *All questions except question one carry equal marks*

This paper consists of 4 printed pages.

SECTION A (30 MARKS)

1. (a) For the regression model

$$Y = X\beta + \epsilon$$

with $\epsilon \sim N(0, \sigma^2 W)$ where W is not an identity matrix, a student argues that the ordinary multiple regression cannot be applied and one should regress $W^{-1}Y$ against $W^{-1}X$ by ordinary multiple regression to estimate β. Do you agree with the student? If so, show how this can be done [6 marks]

(b) An Actuarial Science student is trying to analyze data from a particular stock using an equation of the form $y = f(x) = \frac{a_0}{x} + \frac{a_1}{x^2}$. The measured values of (x, y) are listed in the table below

x	1	2	3	4	5	6	7	8	9	10
y	3	7	0.9	6.8	0.7	0.96	19	37	0.8	0.1

Use nonlinear regression method to determine a_0 and a_1 [7 marks]

(c) Define the term Artificial Neural Network (ANN) [2 marks]

(d) Stock prices (Y, in dollars) are assumed to be affected by the annual rate of dividends of stock (X). A simple linear regression analysis ($Y_i = \beta_o + \beta_1 X_i + \epsilon_i$) was performed on 21 observations where ϵ_i were assumed to follow $\sim iidN(0, \sigma^2)$ and summary statistics were as listed below

$$\bar{X} = 0.4, \bar{Y} = 4, S_X^2 = 0.1, S_Y^2 = 20, S_{XY} = 1.25$$

Perform an ANOVA analysis for the level of significance 5% and give your conclusions[7 marks]

(e) Let $Y = (9, 4, 1, 3)^\top$, $X_1 = (9, 3, 9, 4)^\top$, $X_2 = (5, 3, 5, 4)^\top$

(i) Find the matrix $(X^\top X)^{-1}$ [2 marks]

(ii) Fit a regression model $Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \epsilon$. Give the fitted coefficients and estimate of error variance [4 marks]

(iii) Construct the 95% confidence interval for $E(Y|X_1 = 6, X_2 = 4)$ [3 marks]

2. In a study of the effect of hormones on the productivity of a certain variety of tomato plant, ten plants were treated at different hormone strengths and their yields, in Kgs, noted. A simple linear regression model is fitted in R, with the depended variable (yield in kgs) stored as the vector y and the independent variable (hormone strength) stored as the vector x. Some of the R commands and edited output are shown below

```
> lm1 = lm(y~x)
> summary(lm1)

Coefficients:
              Estimate    Std. Error
(Intercept) -3.72335     0.28072
```

```
x          0.42999    0.01899

> qt(0.975,8)
[1] 2.306004
```

(a) Write down the equation of the model that has been fitted to the data, defining your notations carefully. State the distribution of any error terms in your model [5 marks]

(b) If X is the design matrix for the model fitted in R, then

$$(X^{\top}X) = \begin{pmatrix} 10.60278 & 154.047 \\ 154.047 & 2320.782 \end{pmatrix}$$

Also,

$$\sum_{i=1}^{10} y_i = 26.76 \qquad \sum_{i=1}^{10} x_i y_i = 424.33$$

where y_i is the ith yield and x_i is the ith hormone strength. Give suitable calculations that show how the parameter estimates have been obtained in the R output [5 marks]

(c) If $x_3 = 11$ and $y_3 = 1.66$, calculate the corresponding fitted value and residual [4 marks]

(d) Test the hypothesis that there is no relationship between hormone strength and yield, stating your conclusion clearly. State the size of your hypothesis test [4 marks]

(e) Calculate the 95% confidence interval for the intercept and gradient in the regression model [2 marks]

3. (a) Consider data from the simple linear model $y_i = \beta_o + \beta_1 x_i + \epsilon_i$, $i = 1, 2, \cdots, n$ where x_i's are fixed constants, β_o, β_1 are the unknown coefficients and ϵ_i's are unobserved i.i.d random variables from $N(0, \sigma^2)$. Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ be the least squares estimators of β_0 and β_1 respectively. Find the distribution of the vector $(\widehat{\beta}_0, \widehat{\beta}_1)$ [6 marks]

(b) Suppose the model is modified by including more independent variables so that the model is now written as $Y = X\beta + \epsilon$ and assuming same assumptions for the error model distribution.

(i) Show that the vector covariance of residuals $e = y - X\widehat{\beta}$ and y is given by $cov(e, y) = \sigma^2 M$ where M is the idempotent and symmetric matrix [4 marks]

(ii) Suppose the variance of the error term (σ^2) is unknown, demonstrate how such can be estimated from the model given in b(i) above [7 marks]

(c) Suppose we have a data set (x_i, y_i) for $i = 1, 2, \cdots, n$. Consider two different models $y_i = \alpha + \beta x_i^2 + \epsilon_i$ and $y_i = \alpha + \beta x_i^2 + \gamma \exp^{x_i} + \epsilon_i$. Compare the residual sum of squares of the two models. Explain [3 marks]

4. The data given in the following table are the numbers of deaths from AIDS in Kenya for 12 consecutive quarters starting from the second quarter of 1998

Quarter (i)	1	2	3	4	5	6	7	8	9	10	11	12
Number of deaths (n_i)	1	2	3	1	4	9	18	23	31	20	25	37

(a) (i) Draw a scatterplot of the data

(ii) Comment on the nature of the relationship between the number of deaths and the quarter in this early phase of the epidemic [4 marks]

(b) A statistician has suggested that a model of the form

$$E\left[N_i\right] = \gamma i^2$$

might be appropriate for these data, where γ is a parameter to be estimated form the above data. She has proposed two methods for estimating γ given in (i) and (ii) below

(i) Show that the least squares estimate of γ, is obtained by minimizing $q = \sum_{i=1}^{12}\left(n_i - \gamma i^2\right)^2$ is given by

$$\widehat{\gamma} = \frac{\sum_{i=1}^{12} i^2 n_i}{\sum_{i=1}^{12} i^4}$$

(ii) Show that an alternative (weighted) least squares estimate of γ obtained by minimizing $q^* = \sum_{i=1}^{12} \frac{\left(n_i - \gamma i^2\right)^2}{i^2}$ is given by

$$\tilde{\gamma} = \frac{\sum_{i=1}^{12} n_i}{\sum_{i=1}^{12} i^2}$$

~~(c) Noting that $\sum_{i=1}^{12} i^4 = 60,710$ and $\sum_{i=1}^{12} i^2 = 650$, calculate $\widehat{\gamma}$ and $\tilde{\gamma}$ for the above data~~ [8 marks]

(d) To asses whether the single parameter model which was used in part (b) is appropriate for the data, a two parameter model is now considered. The model is of the form $E\left(N_i\right) = \gamma i^{\theta}$ for $i = 1, 2, \cdots, 12$

(i) To estimate the parameters γ and θ, a simple linear regression of the form $E\left(Y_i\right) = \alpha + \beta x_i$ is used , where $x_i = \log(i)$ and $Y_i = \log(N_i)$ for $i = 1, 2, \cdots, 12$. Relate the parameters γ and θ to the regression parameters α and β

(ii) The least squares estimates of α and β are -0.6112 and 1.6008 with standard errors 0.4586 and 0.2525 respectively. Using the value for the estimate of β, conduct a formal statistical test to asses whether the form of the model suggested in (b) is adequate [8 marks]

Unbiasedness $\Rightarrow E(\hat{\beta}) = \beta$

Consistency $\Rightarrow \lim_{n \to \infty} (\hat{\beta} - \beta) = 0$

Asymptotic normality $\Rightarrow \hat{\beta} = \beta + (X^T X)^{-1} X^T U.$

SECOND SEMESTER, 2017/2018
Campus: MAIN

2014
2015

STATISTICS AND ACTUARIAL SCIENCE

STA 2403: REGRESSION MODELS II

NOTE: Answer **All** questions

they are known not to assume the regression function

1. Consider the non-parametric regression model

$$Y_i = m(x_i) + \epsilon_i$$

where $x_i = \frac{i}{n}$ for $i = 1, 2, \cdots, n$ and $\epsilon_1, \cdots, \epsilon_n$ are independent with mean 0 and variance σ^2. We are interested in estimating the regression function $m(x)$. State the weighted least squares problem that is solved by the local polynomial kernel estimator $\widehat{m}_h(x;p)$ of degree p and bandwidth h. Give an expression, in terms of matrices $\mathbf{W}$ and $\mathbf{X}$ and a vector $\mathbf{Y}$, all of which you should define, for a vector whose first component is $\widehat{m}_h(x;p)$. Assume that m is twice continuously differentiable on $[0,1]$, that the kernel $\mathbf{K}$ is non-negative, symmetric, continuously differentiable on $[-1,1]$ and zero outside $[-1,1]$, and that $h \leftarrow 0$ as $n \leftarrow \infty$ but $nh^2 \leftarrow \infty$ as $n \leftarrow \infty$. Show that, for $x \in (0,1)$,

$$E\{\widehat{m}_h(x;0)\} - m(x) = \frac{1}{2}h^2\mu_2(\mathbf{K})m''(x) + o(h^2)$$

as $n \leftarrow \infty$, where $\mu_r(\mathbf{K}) = \int_{-1}^{1} y^r \mathbf{K}(y)dy$.
You may assume that under the stated conditions

$$\frac{1}{nh}\sum_{i=1}^{n}(x_i - x)^r \mathbf{K}\left(\frac{x_i - x}{h}\right) = \begin{cases} h^r\mu_r(\mathbf{K}) + o(h^r) & \text{if r is even} \\ O(h^{r-1}/n) & \text{if r is odd} \end{cases}$$

Now suppose that $z_n = \alpha h$ for some $\alpha \in [0,1)$. Derive the corresponding expansion of the bias of $\widehat{m}_h(z_n;0)$ in this case assuming that $\mu_{r,\alpha}(\mathbf{K}) = \int_{-\alpha}^{1} y^r \mathbf{K}(y)dy$ then

$$\frac{1}{nh}\sum_{i=1}^{n}(x_i - \alpha h)^r \mathbf{K}\left(\frac{x_i - \alpha h}{h}\right) = h^r\mu_{r,\alpha}(\mathbf{K}) + o(h^r)$$

$$\frac{1}{nh}\sum_{i=1}^{n}|x_i - \alpha h|^r \mathbf{K}\left(\frac{x_i - \alpha h}{h}\right) = O(h^r)$$

State the order of the bias of the local linear kernel estimator $\widehat{m}_h(\cdot;1)$ both at an interior point $x \in (0,1)$ and at a sequence of boundary points $z_n = \alpha h$

2. (a) A regression model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ is to be fitted to a set of data consisting of the n-tripples (y_i, x_{1i}, x_{2i}) for $i = 1, 2, \cdots, n$. Derive the normal equations for obtaining the Least Squares estimates of α, β_1 and β_2

(b) The plasma lipid level of total cholesterol (y) was recorded for a sample of 8 patients before they received drug therapy. The investigator wanted to see how, if at all, these levels depend on the weight (x_1) and the age (x_2) of the patient. From analyses of the data, the following regression sums of squares were obtained

Regression on weight alone	0.1486
Regression on age alone	4.0439
Regression on weight and age	4.1139

y - cholestrol
x_1 - weig

The total sum of squares about the mean was 4.9150. Using backwards elimination or otherwise, obtain the best model for describing the relationship between Y and the x-variables

3. The data given in the following table are the numbers of deaths from AIDS in Kenya for 12 consecutive quarters starting from the second quarter of 1998

Quarter (i)	1	2	3	4	5	6	7	8	9	10	11	12
Number of deaths (n_i)	1	2	3	1	4	9	18	23	31	20	25	37

(a) (i) Draw a scatterplot of the data

(ii) Comment on the nature of the relationship between the number of deaths and the quarter in this early phase of the epidemic

(b) A statistician has suggested that a model of the form

$$E[N_i] = \gamma i^2$$

might be appropriate for these data, where γ is a parameter to be estimated form the above data. She has proposed two methods for estimating γ given in (i) and (ii) below

(i) Show that the least squares estimate of γ, is obtained by minimizing $q = \sum_{i=1}^{12} (n_i - \gamma i^2)^2$ is given by

$$\hat{\gamma} = \frac{\sum_{i=1}^{12} i^2 n_i}{\sum_{i=1}^{12} i^4}$$

10,002 $= 0.2585$

(ii) Show that an alternative (weighted) least squares estimate of γ obtained by minimizing $q^* = \sum_{i=1}^{12} \frac{(n_i - \gamma i^2)^2}{i^2}$ is given by

$$\tilde{\gamma} = \frac{\sum_{i=1}^{12} n_i}{\sum_{i=1}^{12} i^2}$$

$= 174/650 = 0.2677$

(c) Noting that $\sum_{i=1}^{12} i^4 = 60,710$ and $\sum_{i=1}^{12} i^2 = 650$, calculate $\hat{\gamma}$ and $\tilde{\gamma}$ for the above data

$\sum_{i=1}^{12} i^2 n_i = \sum_{i=1}^{12} i^2 \gamma_i^2$ $\hat{\gamma} = 0.$

CONTINUED

(d) To asses whether the single parameter model which was used in part (b) is appropriate for the data, a two parameter model is now considered. The model is of the form $E(N_i) = \gamma i^{\theta}$ for $i = 1, 2, \cdots, 12$

(i) To estimate the parameters γ and θ, a simple linear regression of the form $E(Y_i) = \alpha + \beta x_i$ is used , where $x_i = \log(i)$ and $Y_i = \log(N_i)$ for $i = 1, 2, \cdots, 12$. Relate the parameters γ and θ to the regression parameters α and β

(ii) The least squares estimates of α and β are -0.6112 and 1.6008 with standard errors 0.4586 and 0.2525 respectively. Using the value for the estimate of β, conduct a formal statistical test to asses whether the form of the model suggested in (b) is adequate

Non-Linear   Linear

$\hat{\beta}_0 = \bar{y} - \frac{\hat{\beta}_1}{n}\sum_{i=1}^{n} 1/x_i$ $\beta_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} \frac{y_i}{x_i} - \bar{y}}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} - \sum_{i=1}^{n}\frac{1}{x_i^2}}$ $\beta_1 = \dfrac{\sum xy - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$

$\hat{\beta}_1 =$

1 2 3 4 5 6 7 8 9 10

3 7 0.9 0.8 0.7 0.06 19 37 0.8 0.1

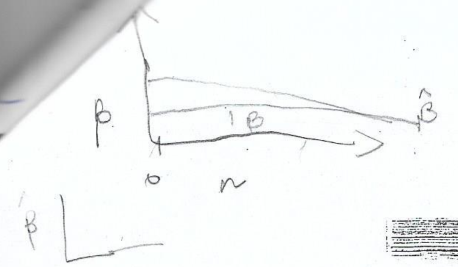
3 7/2 0.3 0.2 0.14 0.01 19/7 37/8 0.0889 0.01

$= \dfrac{14.5882 - 6.936}{0.292897 - 1.54976} = \dfrac{7.6522}{-1.256863}$

$= -6.088$

$6.936 + \frac{6.088}{10}(2.92897)$

$= 8.719$



W1-2-60-1-6

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

UNIVERSITY EXAMINATIONS 2016/2017
FIRST SEMESTER THIRD YEAR DEGREE IN BACHELOR OF SCIENCE IN STATISTICS AND FINANCIAL ENGINEERING

STA 2408: REGRESSION MODELING II

Date: DECEMBER 2016 Time: 2hrs

Instructions: Attempt Question One and Any Other Two Questions

QUESTION ONE (30 MARKS)

a) What is the interpretation of an estimated coefficient from a simple linear regression model and that from a log-log regression model? How could you judge a multiple regression model has a better performance in terms of statistical information? (5 marks)

b) Among all the econometric problems, heteroscedasticity seems to be the most abstract and difficult to understand. Use your words to explain briefly what is heteroscedasicity and what are the consequences of heteroscedasticity (5 marks)

c) What is a biased estimator? (2 marks)

d) Show that $\hat{\beta}_1 = \beta_1 + \dfrac{\sum (X_i - x_i)e_i}{\sum (X_i - x_i)^2}$ and explain under what conditions $\hat{\beta}_i$ would be biased. (4 marks)

e) What is a consistent estimator? What conditions must be fulfilled for an estimator to be consistent? Illustrate your answer with appropriate diagram. (4 marks)

f) Consider the model : $Y_t = \beta_0 + \beta_1 X_t + u_t \qquad (t = 1,2,\ldots,n)$
Where u satisfies the basic stochastic assumption.

u is random

$\mathrm{E}[u] = 0$

$\mathrm{E}[u^2] = \sigma_u^2$ Constant (homoscedasticity)

$$\frac{\hat{\delta}}{\sqrt{Y_n + \bar{X}}}$$

$E[u_i u_j] = 0 \quad (i \neq j)$ (Serial independence)

How the properties of would $\hat{\beta}_1$ be affected in the following cases:

i) The number of observations (n) is increased. (2 marks)

ii) The variance of X (dispersion of values of X) increases. (2 marks)

iii) All the $X's$ have the same value. (2 marks)

g) Suppose autocorrelation is existed in the data and it was ignored in the estimation. So what are the consequences of the OLS estimated results? (4 marks)

multicollinearity

QUESTION TWO (20 MARKS)

Autocorrel

a) Let $X^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $X^{(2)} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ $X^{(3)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $X^{(4)} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ with correct classification $Z^{(1)} = Z^{(2)} = Z^{(3)} = 1 \quad , Z^{(4)} = 0$

Train the perceptron, and find the weight that achieves correct classification for all inputs vectors in the training set. Assume that the perceptron with weights

$w_0, w_1, \ldots w_p$ classifiers as objects as 1 iff $\sum w_i x_i > 0$ and 0 elsewhere

Take $w = (0 \quad 0 \quad 0)^T$ as the initial weight and $\eta = 1$ (10 marks)

b) What are the practical consequences of imperfect multicollinearity? (3 marks)

- large R^2

c) What is meant by heteroscedasticity? What are its effects on OLS estimators and their variances? (4 marks)

d) In the regression $Y_i = \beta_1 + \beta_2 X_i + u_i$, suppose you multiple each X_i value by a constant, say 5. Will it change the residuals and fitted values of Y_i? Briefly explain. (3 marks)

QUESTION THREE (20 MARKS)

a) Suppose the true model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e_i$$

But you estimate

$$Y_i = \alpha_0 + \alpha_1 X_1 + u_i$$

If you use observations of Y at $X = -3, -2, -1, 0, 1, 2, 3$, and estimate the "incorrect" model, what bias will result in these estimates? (6 marks)

$a_0 = \bar{y} - \hat{a}_1$

b) What is meant by intrinsically linear and intrinsically nonlinear regression Models? Give some examples. (4 marks)

c) What is the difference between OLS and nonlinear least-squares (NLLS) estimation? (3 marks)

d) Derive the BLUP properties of the mixed models (7 marks)

QUESTION FOUR (20 MARKS) ✓

a) In the context of Generalised Linear Models, consider the exponential distribution with

$$f(x)=\frac{1}{\mu}e^{-x/\mu}, \; x>0$$

i) Show that $f(x)$ can be written in the form of an exponential family of distributions. ✓ (2 marks)

ii) Determine the canonical parameter θ, the variance function and the dispersion parameter. (4 marks)

The following regression equation was calculated for a class of 24 MSC finance students.

$$\hat{Y} = \overset{\beta_0}{3.1} + \overset{\beta_1}{0.021}X_1 + \overset{\beta_1}{0.075}X_2 + 0.043X_3$$

Standard error		(0.019)	(0.034)	(0.018)

Where: Y = Students score on a theory examination

X_1 = Students rank (from the bottom) in high school

X_2 = Students verbal aptitude score

X_3 = a measure of the students character.

$\beta_0 \pm t_{(1-\alpha)} * se(\beta_i)$

i) Calculate the t ratio and the 95% confidence interval for each regression coefficient. (5 marks)

ii) What assumptions did you make in (a) above? How reasonable are they? (3 marks)

iii) Which regressor gives the strongest evidence of being statistically discernible? (3 marks)

iv) In writing up a final report, should one keep the first regressor in the equation, or drop it? Why? (3 marks)

$\pm \beta_0 * t$ t calc 0.05

2

W1-2-60-1-6

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
University Examinations 2014/2015
FIRST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF STATISTICS AND ACTUARIAL SCIENCE

STA 2408: REGRESSION AND MODELLING II

DATE: AUGUST, 2014 **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question ONE (Compulsory) and Any Other TWO.

1. (a) Define the term Artificial Neural Network (2 marks)

 (b) (i) What is meant by training an artificial neural network (2 marks)

 (ii) Prove that if there exists a set of connection weights w^* which are able to perform the transformation $y = d(x)$, the perceptron learning rule will converge to some solution in a finite number of steps for any initial choice of weights (6 marks)

 (c) Consider the model

$$Y = X\theta + \epsilon \quad (1)$$

 where X is an $n \times p$ matrix of rank p, θ is a p vector and ϵ is an n-vector of $i.i.d$ variables with zero mean and finite variance. Assume the regression model (1) has uncorrelated errors with mean 0 and variance σ^2. Let $\widehat{Y} = X\widehat{\beta}$ denote the estimated value. Show that

$$\sum var(\widehat{Y}_i) = \sigma^2 p$$

 (3 marks)

 (d) Using appropriate examples

 (i) Define a random effects model (3 marks)

 (ii) Hierarchical models (3 marks)

 (iii) Suppose data are scores in a mathematics exams given to four ninth grade classes from each of the fifteen high schools in Thika. Aside from the differences between boys and girls (which could be modeled by fixed effects) there would undoubtedly be three sources of variability (i) among schools (ii) among classes within each school and (iii) among pupils within each class. Let the exam score of pupil k (of gender t) in class j of school i be Y_{tijk}. With a brief discussion, give a model that could be used to model the exams scores (4 marks)

 (e) An often used method for regularization is `ridge regression`. That is, when there is no well-defined solution of the normal equations in linear regression $(X^\top X)\theta = X^\top y$. In this scenario, it is sought to minimize, w.r.t. θ, the equation

$$||Y - X\theta||^2 + \lambda||\theta||^2$$

 for some $\lambda > 0$.

 i.) Show that, for an appropriate range of λ, the ridge regression estimator is:

$$\widehat{\theta} = (X^\top X + \lambda I_p)^{-1} X^\top Y$$

 Interpret, relative to least squares, the new `optimization function`. (3 marks)

CONTINUED

ii.) Find the ridge regression solution for the data below for a general value of λ and for the straight line model $Y = \beta_o + \beta_1 X + \epsilon$ (only apply the ridge-penalty to the slope parameter, not ot the intercept). Show that when λ is chosen as 4, the ridge solution fit is $\widehat{Y} = 40 + 1.75X$

Data: $X^\top = (X_1, X_2, \cdots, X_8)^\top = (-2, -1, -1, -1, 0, 1, 2, 2)$ and $Y^\top = (Y_1, Y_2, \cdots, Y_8)^\top = (35, 40, 36, 38, 40, 43, 45, 43)$

(4 marks)

2. (a) A gamma distribution is known to be a possible model that can be used to model rainfall storms in a long-rain season with a pdf of the form

$$f(x) = \frac{1}{\Gamma(\alpha)\lambda^\alpha x^{\alpha-1}\exp^{-\lambda x}} \qquad x > 0$$

Starting with the likelihood function, show and or design a Newton-Raphson algorithm that can be used to estimate the parameters of the function fitted to rainfall data and provide an R-pseudocode for the estimation process

(8 marks)

(b) Venables and Ripleys (1999) contains a data set describing weights (y_i) of obese patients after different number of days (x_i) since the start of a weight reduction program. Venables and Ripleys (1999) suggests the following model to the dataset

$$y_i = \beta_0 + \beta_1 \exp^{-\beta_2 x_i} + e_i$$

Juma, a financial engineering student at JKUAT would like to implement Newton-Raphson algorithm. Give an R-pseudocode that can be used to estimate the parameters of the non-linear model given by Venables and Ripleys (1999)

(12 marks)

3. (a) Provide a proof that given a model of the form

$$y = f(x; \beta) + \epsilon(\beta)$$

where f is known to be a non-linear continuous function and X as a matrix of covariates and $\epsilon(\beta)$ is the disturbance. Suppose there exists a hull for an objective function for minimizing the squared errors; though without guaranteed uniqueness of the solution; there exists a solution that minimizes the non-linear least squares criterion function given by

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i; \beta))^2$$

(8 marks)

(b) Describe the general procedure followed while finding solutions to a non-linear regression problem in the case we adopt the use of Newton-Raphson algorithm

(6 marks)

(c) Discuss the consistency of a Non-linear Least Square Estimator

(6 marks)

4. Cellular phone plans provide a clear case of missing data in estimation problems. Safaricom has to offer "plans" of options, bundling together four or five options (example internet, SMS etc) for one price or alternatively selling them separately. Safaricom has offered a four-option plan in some areas and five-option plan (including the four, plus one more...) in another area. In each of the areas, customers were asked to choose their favorite plans and results were tabulated as shown below

Plan 1					Plan 2				
1	2	3	4	5	1	2	3	4	5
26	36	24	5	-	56	18	29	5	-
41	28	34	10	-	4	24	32	24	10
53	39	12	11	-	11	30	22	23	7
27	59	17	0	-	27	71	10	0	31
32	39	12	0	-	47	4	29	11	-

Describe an EM algorithm with relevant probability distribution to help the company come up with a viable pricing policy. (20 marks)

Kmaths

SECOND SEMESTER, 2014
Campus: MAIN

STATISTICS AND ACTUARIAL SCIENCE

STA 2408: REGRESSION MODELING II

NOTE: Answer **All** questions on the answer booklet provided. Clear workings and illustrations will earn maximum marks.

1. (a) Define the term Artificial Neural Network

   (b) (i) What is meant by training an artificial neural network

   (ii) Prove that if there exists a set of connection weights w^* which are able to perform the transformation $y = d(x)$, the perceptron learning rule will converge to some solution in a finite number of steps for any initial choice of weights

   (c) Suppose data are scores in a mathematics exams given to four ninth grade classes from each of the fifteen high schools in Kacheliba. Aside from the differences between boys and girls (which could be modeled by fixed effects) there would undoubtedly be three sources of variability (i) among schools (ii) among classes within each school and (iii) among pupils within each class. Let the exam score of the pupil k (of gender t) in class j of school i be Y_{tijk}. With a brief discussion, give a model that could be used to model the exams scores

   (d) Find the ridge regression solution for the data below for a general value of λ and for the straight line model $Y = \beta_o + \beta_1 X + \epsilon$ (only apply the `ridge`-penalty to the slope parameter, not ot the intercept). Show that when λ is chosen as 4, the ridge solution fit is $\widehat{Y} = 40 + 1.75X$

   Data: $X^\top = (X_1, X_2, \cdots, X_8)^\top = (-2, -1, -1, -1, 0, 1, 2, 2)$ and $Y^\top = (Y_1, Y_2, \cdots, Y_8)^\top = (35, 40, 36, 38, 40, 43, 45, 43)$

2. (a) Provide a proof that given a model of the form

$$y = f(x; \beta) + \epsilon(\beta)$$

   where f is known to be a non-linear continuous function and X as a matrix of covariates and $\epsilon(\beta)$ is the disturbance. Suppose there exists a hull for an objective function for minimizing the squared errors; though without guaranteed uniqueness of the solution; there exists a solution that minimizes the non-linear least squares criterion function given by

$$Q_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i; \beta))^2$$

   From the book

   (b) Describe the general procedure followed while finding solutions to a non-linear regression problem in the case we adopt the use of Newton-Raphson algorithm

   (c) Discuss the consistency of a Non-linear Least Square Estimator

3. Suppose that $Q_T(\beta)$ is quadratic in β:

$$Q_T(\beta) = a + b^\top \beta + \beta^\top C \beta$$

where a is scalar, b is a vector and C a symmetric, positive definite matrix. Find the parameter estimators from the given objective function. Suppose that $E(Q_T(\beta))$ is a continuous function on the compact parameter space Θ_1 such that β_o is its unique, global minimum. Suppose also that $E(Q_T(\widehat{\beta})) = \inf_{\Theta_1} E(Q_T(\beta))$. Show that $\widehat{\beta}$ converges in probability to β_o. Discuss the consistency and asymptotic normality of the estimators.