



MURANG'A UNIVERSITY OF TECHNOLOGY

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND ACTUARIAL SCIENCE

UNIVERSITY ORDINARY EXAMINATION

2018/2019 ACADEMIC YEAR

**THIRD YEAR SECOND SEMESTER EXAMINATION FOR BACHELOR OF
SCIENCE IN MATHEMATICS AND COMPUTER SCIENCE AND
BACHELOR OF SCIENCE IN APPLIED STATISTICS WITH
PROGRAMMING**

AMS 320 - STATISTICAL MODELING

DURATION: 2 HOURS

DATE: 15-04-2019

TIME: 9-11 A.M.

Instructions to candidates:

1. Answer question One and Any Other Two questions.
2. Mobile phones are not allowed in the examination room.
3. You are not allowed to write on this examination question paper.

SECTION A: ANSWER ALL QUESTIONS IN THIS SECTION

QUESTION ONE (30 MARKS)

a) State and explain any three assumptions used in regression analysis. (3marks)

b) In generalized linear models the response variable distribution must be a exponential family. Given the probability mass function of binomial distribution with parameters n and p is

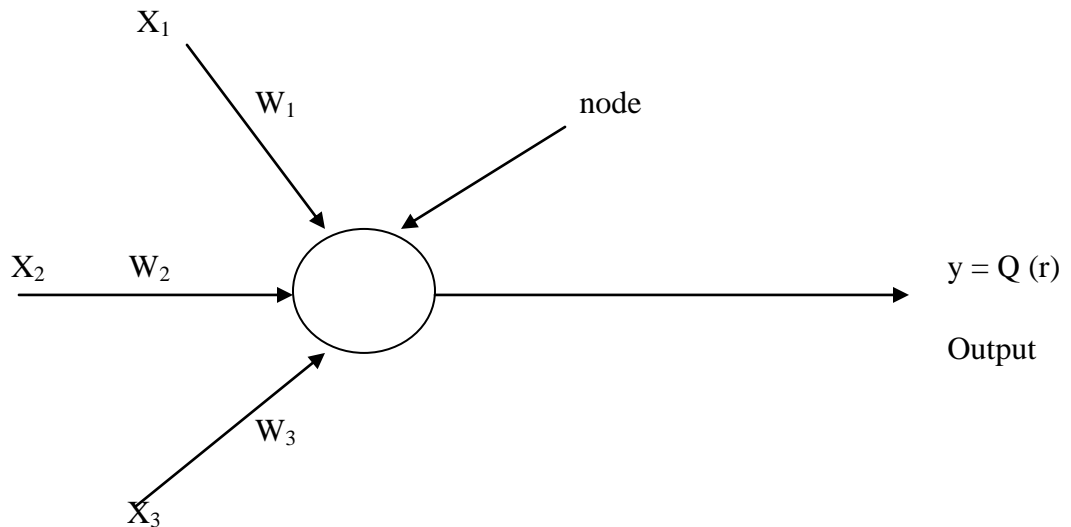
$$f(u, p) = \begin{cases} \binom{n}{u} p^u (1-p)^{n-u}, & u = 0, 1, \dots, n \\ 0, & \text{Otherwise} \end{cases}$$

Express $Bin(n, p)$ as a member of exponential family and show that the natural parameter is given by $b(u) = \ln\left(\frac{p}{1-p}\right)$ (3marks)

c) A researcher studying the characteristics of a population was given a set of data which represented the ages at the time of death for 10 people working in a given company. The ages of these people were 35, 40, 45, 45, 47, 49, 55, 57, 59, 65. Using a uniform Kernel with a bandwidth of $b = 10$, determine the Kernel density estimate of the probability of survival at the age.

(4marks)

d) The figure below shows a diagram of a single artificial neuron unit.



The model has three units $X = (x_1, x_2, x_3)$ that receives only binary signal (either 0 or 1). How many different input patterns can the node receive if it had four inputs? (2marks)

e) The data below shows the number of hours that 10 students studied for a statistics test and their score on the test.

Hours studied (X)	3	10	15	5	11	7	12	20	1	17
Test scores (Y)	32	60	75	40	65	45	60	90	20	80

i) Find the regression equation of least squares that approximates the regression of the test scores on the number of hours studied. (6marks)

ii) Predict the test score of a student who studied for 16 hours. (2marks)

f) Consider the data below

X	4	5	7	10
Y	8	10	15	20

Calculate $f(6)$ using linear Spline interpolation. (5marks)

g) State any five properties of a fitted regression model. (5marks)

SECTION B – ANSWER ANY TWO QUESTIONS IN THIS SECTION

QUESTION TWO (20 MARKS)

a) Let Y_1, Y_2, \dots, Y_n be independent normal random variables with mean $\beta_0 + \beta_1 x_1, \beta_0 + \beta_2 x_2, \dots, \beta_0 + \beta_n x_n$, and σ^2 . Find the point estimators of β_0, β_1 and σ^2 . (9marks)

b) Given that $\hat{\beta}_1 = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$. Show that $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}^2$ are unbiased estimators of β_0, β_1 and σ^2 respectively, where $i = 1, \dots, n$ (11marks)

QUESTION THREE (20 MARKS)

a) Show that the least square estimates of the multiple regression coefficients in matrix form are given by

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

Where X' is the transpose of X and $(X'X)^{-1}$ is the inverse of $(X'X)$ (6marks)

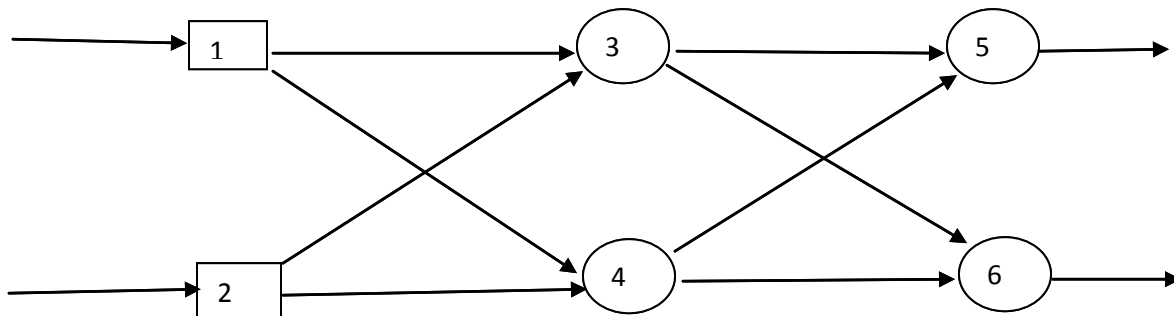
b) The following data show the number of bedrooms, the number of washrooms and the price at which a random sample of eight family houses was sold recently. Use matrix notations in (a) above to determine the least square estimates of multiple regression coefficients for the following data. (14marks)

Number of Bedrooms (X_1)	Number of washrooms (X_2)	Price in (\$)
5	3	90,000
4	2	88,000
3	1	75,000
6	2	85,000

QUESTION FOUR (20 MARKS)

a) The data below show the weights of 10 mothers at the time they delivered their first borns in a given hospital. Weight 55, 60, 65, 65, 67, 69, 75, 77, 79, 85. Using triangular kernel density with bandwidth $h = 10$, find the Kernel smoothed density estimate $\hat{f}(70)$. (4marks)

b) The following diagram represents a feed forward neural network with hidden layers.



The table below lists all the weights in the network

$$W_{13} = -2 \qquad W_{35} = 1$$

$$W_{23} = 3 \qquad W_{45} = -1$$

$$W_{14} = 4 \qquad W_{36} = -1$$

$$W_{24} = -1 \qquad W_{46} = 1$$

Each of the nodes 3, 4, 5 and 6 uses the following activation function

$$\psi(r) = \begin{cases} 1 & \text{if } r \geq 0 \\ 0 & \text{Otherwise} \end{cases}$$

Where r denotes the weighted sum of a node. Each input nodes (1 and 2) can only receive binary values (0 or 1). Calculate the output of the network for each of the input patterns. (6marks)

Pattern	P_1	P_2	P_3
Node 1	0	1	0
Node 2	0	0	1

c) State any three properties of a Kernel density function (3marks)

d) The data below shows the weight in kg of 10 bulls sold by a farmer to a given slaughter company.

700, 715, 718, 620, 735, 737, 737, 745, 775, 785, 790

Use the smoothed Kernel with uniform Kernel of bandwidth 15 to find the estimate

i) $\hat{f}(735)$ (3marks)

ii) $\hat{F}(635)$ (4marks)