MURANG'A UNIVERSITY OF TECHNOLOGY

SCHOOL OF PURE, APPLIED AND HEALTH SCIENCES

DEPARTMENT OF MATHEMATICS AND ACTUARIAL SCIENCE

UNIVERSITY ORDINARY EXAMINATION

2020/2021 ACADEMIC YEAR

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

AMS 320 – STATISTICAL MODELING

DURATION: 2 HOURS

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) Explain the following terms giving a relevant example in each case:

i. Semi studentized residual (2 marks)

ii. Generalized linear model (2 marks)

iii. Random effects model (2 marks)

iv. Polynomial regression (2 marks)

v. Analysis of covariance (2 marks)

b) Show that for multiple linear regression, the method of least squares and the maximum likelihood estimation under normal assumption results in equivalent estimators. (5 marks)

c) A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selects 10 women. The results follow; X is age and Y is a measure of muscle mass.

X	63	43	46	41	78	54	48	68	42	57
Y	8.4	10.6	11.3	11.9	5.4	10.0	9.2	7.0	10.1	9.1

i. Fit a simple linear regression model to the data. (4 marks)

ii. Test for the significance of regression. (4 marks)

iii. Compute R^2 for this model and comment on it. (2 marks)

d) The data below relates to the management fees that a leading mutual fund pays to its investment advisors to manage its assets. The fees paid depends on the net asset value of the fund.

Fee (%)	0.52	0.51	0.48	0.46	0.44	0.42	0.41	0.40	0.39	0.39	0.38	0.37
Asset (millions)	0.5	5	10	15	20	25	30	35	40	45	55	60

Fit an exponential regression model to that data. (5 marks)

SECTION B – ANSWER ANY TWO QUESTIONS IN THIS SECTION

QUESTION TWO (20 MARKS)

a) Explain the following violations of the classical linear regression model, describing how they can be identified, their effects and remedial actions that can be taken.

i. Multicollinearity (3 marks)

ii. Heteroscedasticity (4 marks)

iii. Autocorrelation (4 marks)

b) A criminologist studying the relationship between level of education and crime rate collected the following data from a random sample of 10 counties. X is the percentage of individuals in the county having at least high school certificate and Y is the crime rate last year.

X	74	81	81	87	66	68	81	82	75	82
Y	8.5	8.4	8.2	6.3	9.1	6.6	5.9	7.9	6.5	6.8

Perform the Brown-Foray test on the data and make relevant conclusion. (5 marks)

c) A staff analyst for a manufacturer of microcomputer components has compiled monthly data for the past 8 months on the value of industry production of processing units that use these components (x_t in million KES) and the value of the firm's components used (y_t in thousands KESS). The analyst believes that a simple linear regression model is appropriate but anticipates positive autocorrelation. The data follow:

x_t	10.3	10.2	10.1	9.8	9.7	9.4	9.8	10.2
y_t	2.05	2.03	2.01	1.95	1.94	1.89	1.99	2.05

Conduct the Durbin-Watson test for autocorrelation using $\alpha = 0.05$ (4 marks)

QUESTION THREE (20 MARKS)

a) Describe the six departures from the linear regression model that can be studied by residuals suggesting an appropriate remedial measure for each case. (12 marks)

b) A study was done on the effects of dietary supplement on the growth rates of rats. Here $X = dose\ of\ dietary\ supplement$ and $Y = growth\ rate$. The resulting data are shown below.

X	10	10	15	15	20	20	25	25	25	30	35	35
Y	73	78	85	88	90	91	87	86	91	75	65	63

i) Fit a simple linear regression model to the data. (3 marks)

ii) Conduct an F test for lack of fit for the model in (i) above. (5 marks)

QUESTION FOUR (20 MARKS)

Consider the petrol consumption data in the table below:

Automobile	y	x_1	x_2
Apollo	8.1	5.7	1.8
Omega	7.2	5.7	1.3
Nova	8.5	4.1	1.6
Monarch	7.8	5.8	1.8
Duster	8.5	3.7	1.5
Skyhawk	9.4	3.8	1.4
Monza	9.1	4.3	1.4
Sirocco	14.8	1.5	0.9
Pacer	8.4	4.2	1.5
Rabrat	8.6	2.3	1.2
Granada	7.6	4.9	1.8
Eldorado	6.1	8.2	2.4

a) Fit a multiple linear regression model to petrol consumption (y, kilometers per liter) on engine displacement (x_1, liters) and vehicle weight (x_2, tons). (8 marks)

b) Construct the analysis of variance table and test for the significance of regression. (4 marks)

c) Calculate R^2 and R^2_{adj} for this model. (2 marks)

d) Find a 95% confidence interval for β_1 and β_2 (2 marks)

e) Compute the t statistics for testing $H_0: \beta_1 = 0$ and $H_0: \beta_2 = 0$. What conclusions can you draw? (2 marks)

f) Find a 95% confidence interval for a new observation on petrol consumption when $x_1 = 3.5$ and $x_2 = 2.0$ (2 marks)