LAIKIPIA UNIVERSITY

UNIVERSITY EXAMINATIONS

FIRST SEMESTER 2025/2026 ACADEMIC YEAR

THIRD YEAR EXAMINATION FOR THE DEGREES OF BACHELOR OF SCIENCE (STATISTICS)

STAT 314: ECONOMETRIC MODELS

STREAM: R ***TIME: 2 HRS***

DAY: FRIDAY [8.30 – 10.30 A.M] ***DATE: 30/01/2026***

THIS QUESTION PAPER CONSISTS OF FIVE (5) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO.

Instructions: Answer Question ONE (compulsory) and any Other TWO questions

- **Show your working clearly**
- **Write your registration number before you open the answer booklet**

QUESTION ONE (30 MARKS)

a) Data for 51 U.S. states was used to examine the relationship between violent crime rate (violent crimes per 100,000 persons per year) and the predictor variables of urbanization (percentage of the population living in urban areas) and poverty rate. A predictor variable indicating whether or not a state is classified as a Southern state (1 = Southern, 0 = not) was also included. Some Minitab output for the analysis of this data is shown below (with some information intentionally left blank).

Predictor	Coefficient	SE Coef	T- statistics	P-value
Constant	-321.9	148.2	-2.17	0.035
Urban	4.689	1.654	2.83	0.007
Poverty	39.34	13.52	2.91	0.006
South(S=1)	-649.3	266.96	-2.43	0.019
Urban*South	12.05	2.871	4.2	0.0000
Poverty*South	-5.838	16.671	–0.35	0.728
	Analysis of Variance			
Source	DF	SS	MS	F
Regression	5	2060459	412091	----
Residual Error	45	882169	19604	
Total	50	2942628		

i) Obtain an equation that represents the fitted relationship between crime, urbanization, and poverty for Southern states? **(3 Marks)**

ii) Predict the violent crime rate for a Southern state with an urbanization of 55.4 and a poverty rate of 13.7. **(2 Marks)**

iii) Predict the violent crime rate for a non-Southern state with an urbanization of 65.6 and a poverty rate of 8.0. **(2 Marks)**

iv) Calculate the ANOVA F test statistic value **(2 Marks)**

v) Find the critical F value for the ANOVA F test **(2 Marks)**

vi) Which predictors should probably be removed from the model to improve it? **(2 Marks)**

b) Explain the role of the following areas in econometric analysis

i) Economic theory **(3 Marks)**

ii) Mathematical economics **(3 Marks)**

iii) Economic Statistics **(3 Marks)**

iv) Mathematical Statistics **(3 Marks)**

c) Determine whether the following models are linear in the parameters, or the variables, or both. Which of these models are linear regression models?

i) $Y_i = \beta_1 + \beta_2^3 X_i + \mu_i$ **(1 Mark)**

ii) $Y_i = \beta_1 + e^{\beta_1 + \beta_2 x_i + \mu_i}$ **(1 Mark)**

iii) $Y_i = \beta_1 + \beta(\frac{1}{X_i}) + \mu_i$ **(1 Mark)**

iv) $Y_i = \beta_1 + \beta_2 InX_i + \mu_i$ **(1 Mark)**

v) $Y_i = \beta_1 - \beta_2 \ln(\frac{1}{X_i}) + \mu_i$ **(1 Mark)**

QUESTION TWO (20 MARKS)

a) Explain the four key functions and benefits of adjusted R squared **(8 Marks)**

b) Explain the assumptions made in classical linear regression models **(12 Marks)**

QUESTION THREE (20 MARKS)

i) Explain the two ways in which a GLM generalizes a linear model **(6 Marks)**

ii) Discuss the properties of different link functions, explaining why the log link might be problematic when the response variable has zero observations. **(6 Marks)**

c) A scientific foundation wanted to evaluate the relation between y= salary of researcher (in '000 of dollars), x1= number of years of experience, x2= an index of publication quality, x3=sex (Male =1, Female=0) and x4= an index of success in obtaining grant support. A sample of 35 randomly selected researchers was used to fit the multiple regression model. Parts of the computer output appear below.

Predictor	Coeffient	SE Coef	T
Constant	17.846931	2.001876	8.915
Years	1.10313	0.359573	3.068
Papers	0.32152	0.037109	
Sex	1.5934	0.687724	2.317
Grants	1.288941	0.298479	4.318
	s = 1.75276	R-sq = 92.3%	adj R-sq = 91.4%

i) Describe the value of the expected p-value of the ANOVA F test in terms of size (large, very large, small, very small) giving reasons. **(2 Marks)**

iii) Determine the variables that significantly explains variations in salary **(4 Marks)**

v) Determine the salary for male, and for female. **(2 Marks)**

QUESTION FOUR (20) MARKS

a) Critically appraise each of the following statements relating to heteroskedasticity in a regression model:

(i) "In the presence of heteroskedasticity the OLS estimator of the coefficient vector is inefficient and this causes the usual t tests and F tests to be invalid." **(3 Marks)**

(ii) "The OLS estimator is unbiased but it is inconsistent if the errors are heteroskedastic." **(3 Marks)**

iii) "One way of eliminating the problems associated with heteroskedasticity is to transform the model so that the errors are normally distributed and uncorrelated with the regressors. **(3 Marks)**

iv) "Sometimes apparent heteroskedasticity can be caused by a mathematical misspecification of the regression model. This can happen, for example, if the dependent variable ought to be logarithmic, but a linear regression is estimated." **(3 Marks)**

QUESTION FIVE (20 MARKS)

a) An econometric model contains an error term ut . Explain six reasons why the error term must be included in a regression model **(12 Marks)**

b) Explain **five** features of generalised linear models **(8 Marks)**