



ALUPE UNIVERSITY
COLLEGE

...Bastion of Knowledge...

P. O.Box 845-50400 Busia(K)
principal@auc.ac.ke
Tel: +254 741 217 185
+254 736 044 469
off Busia-Malaba road

**OFFICE OF THE DEPUTY PRINCIPAL
ACADEMICS, STUDENT AFFAIRS AND RESEARCH**

UNIVERSITY EXAMINATIONS

2020 /2021 ACADEMIC YEAR

THIRD YEAR FIRST SEMESTER REGULAR EXAMINATION

**FOR THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS WITH
COMPUTING)**

COURSE CODE: STA 316

COURSE TITLE: APPLIED REGRESSION ANALYSIS

DATE: 15/03/2021

TIME: 1400 – 1700 HRS

INSTRUCTION TO CANDIDATES

- SEE INSIDE

THIS PAPER CONSISTS OF 4 PRINTED PAGES

PLEASE TURN OVER

REGULAR-MAIN EXAM

STA 316: APPLIED REGRESSION ANALYSIS

STREAM: ASC

DURATION: 3 Hours

INSTRUCTION TO CANDIDATES

Answer **ALL** questions from section A and any **THREE** from section B.

SECTION A [31 Marks]. Answer ALL questions.

QUESTION ONE [15 Marks]

- a) When is linear regression used? [2 Marks]
- b) Minimization of errors to get a best fit line have explored several possibilities. Give any three of these possibilities. [3 marks]
- c) A linear regression model is given by $Y_i = \beta_0 + \beta_1 x + \varepsilon_i$. Give the interpretation β_0 and β_1 [2 Marks]
- d) State any two assumptions for simple linear regression. [2 Marks]
- e) Write the linear hypothesis and test statistic that can be tested using the F -test. [3 Marks]
- f) What are the key aspects that make ANOVA models differ from ordinary regression models? [3 Marks]

QUESTION TWO [16 Marks]

- a) What are residuals and their interpretation of residuals plots in Method of Least Squares. [4 Marks]
- b) Given n points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ and the equation of straight line $y = a + bx$, derive the normal equations. [6 Marks]
- c) Describe the following giving the expression and interpretation for each
 - i) Coefficient of determination, R^2 [3 Marks]
 - ii) Coefficient of variation, CV [3 Marks]

SECTION B [39 Marks] Answer any THREE questions]

QUESTION THREE [13 Marks]

Suppose crickets make their chirping sound by sliding one wing cover very rapidly back and forth over the other. The following table lists fifteen frequency - temperature observations recorded for the striped ground cricket.

Observation no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Chirps per second (x)	22	17	19.8	18.4	17.1	15.5	14.7	17.1	15.4	16.2	13.0	17.2	16.0	17.0	13.4
Temperature y ($^{\circ}F$)	84.6	70.6	93.3	84.3	80.6	75.2	69.7	82	69.4	83.3	83.6	82.6	80.6	83.5	77.3

- a) Determine the best straight line for approximating the xy – relationship and sketch a graph. [7 Marks]
- b) Calculate the residuals and sketch a residuals plot. [6 Marks]

QUESTION FOUR [13 Marks]

A study was conducted on estimating heating oil used for a single family home in the month of January based on average temperature and amount of insulation in inches.

Oil(Gal)	10	20	19	12	11
Temperature	6	12	10	8	12
Insulation	34	40	32	36	34

- a) Fit a multiple regression model to this data [11 Marks]
- b) Interpret your model [2 Marks]

QUESTION FIVE [13 Marks]

- a) Write down the multiple regression model and describe it. [4 Marks]
- b) Give the general outline of ANOVA table for the two way classification fixed effect model with two factors A and B. [4 Marks]
- c) A multiple regression model developed for estimating amount of heating oil used by each household based on average temperature and amount of insulation, the following STATA output was obtained. Write a regression equation and give an interpretation of each of the three coefficient values. [5 Marks]

```
. mvreg heatingoil = temperature insulation
```

Equation	Obs	Parms	RMSE	"R-sq"	F	P
heatingoil	15	3	26.01378	0.9656	168.4712	0.0000

heatingoil	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
temperature	-5.436581	.3362162	-16.17	0.000	-6.169133	-4.704029
insulation	-20.01232	2.342505	-8.54	0.000	-25.1162	-14.90844
_cons	562.151	21.0931	26.65	0.000	516.1931	608.1089

QUESTION SIX [13 Marks]

Infant mortality is often used as a general measure of the quality of healthcare for children and mothers. It is reported as the rate of deaths of new borns per 1000 live births. Data recorded for fifteen health centres allow us to build regression models to help understand or predict infant mortality. The variables available for our model are child death rate (deaths per 100,000 children aged 1–14), percent of teens who are high school dropouts (ages 16–19), percent of low–birth weight babies (lbw), teen birth rate (births per 100,000 females ages 15–17).

Variables (i)	Deaths (j)			
	1	2	3	4
Child death rate	8	11	15	14
High school dropouts	12	16	17	11
Low–birth weight	12	22	16	10
Teen deaths	6	14	17	15

- a) Obtain the least squares estimates of the model parameters [3 Marks]

- b) Is there any significance between the package designs and construct ANOVA table.(Use $F_{(4,16)} = 5.80$) [10 Marks]

QUESTION SEVEN [13 Marks]

- a) An assessment on kids was conducted using a two simple factor study in which the effect of gender and age on solving basic problems was considered. Gender and age are in two and three levels respectively as shown in the table below.

Factor A: Gender	Factor B: Age		
	Infant	Toddler	Pre-school
$i = 1$ (Male)	9	12	17
$i = 2$ (Female)	7	10	13

- Compute $(\alpha\beta)_{13}$ and $(\alpha\beta)_{23}$ [4 Marks]
- b) Suppose a two way classification for fixed effects is given by;

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

Where $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$ and $k = 1, 2, \dots, n$

Show that i) $\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$ and [5 Marks]

ii) $\hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$ [4 Marks]