



# MACHAKOS UNIVERSITY

University Examinations 2018/2019

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE

THIRD YEAR EXAMINATION FOR THE DEGREE OF  
BACHELOR SCIENCE ENVIROMENT

ESU 302: STATISTICS

DATE: 27/9/2019

TIME: 2:00 – 4:00 PM

**INSTRUCTION:** Answer Question *ONE* which is compulsory and any other *TWO* Questions  
**QUESTION ONE (30 MARKS)**

- a) Discuss three advantages of sampling over census (6 marks)
- b) Differentiate the following terms as they apply in scientific research
- i) Response and independent variables
  - ii) Descriptive and inferential statistics
  - iii) Null hypothesis and Alternative hypothesis (6 marks)
- b) In a class sample of 12 students, the instructor wanted to establish whether gender influenced their performance. Given below are their test scores per gender, at  $\alpha = 5\%$

Males	8	12	13	9	3	-	-
Females	10	8	12	15	6	8	11

- i. Determine whether gender influence the class performance significantly
  - ii. Test the hypothesis that the mean score for females is 11.9 (10 marks)
- c) (i) Determine the regression equation of Y on X given that

$$\sum xy = 130, \sum x^2 = 2400, \bar{x} = 60 \text{ and } \bar{y} = 4. \quad (5 \text{ marks})$$

(ii) Highlight three assumptions for regression modeling

(3 marks)

**QUESTION TWO (20 MARKS)**

- a) The data below is a summary of copy typing speed per minute before and after in service training.

Participant	A	B	C	D	E	F	G	H
Rate Before	85	124	172	123	111	139	99	77
Rate after	105	160	175	143	156	127	95	100

Test the hypothesis that on average the training did not result to any improvement on the typing speed. (8 marks)

- b) Below are the ABC co. daily sales during a particular week in the year 2014.

<i>Sales</i> <i>(KSH.000)</i>	11-20	21-30	31-40	41-50	51-60
Frequency	3	6	11	3	2

Determine;

- i. Average sales per day
- ii. Median
- iii. Mode and
- iv. Standard deviation. (12 marks)

**QUESTION THREE (20 MARKS)**

- a) The lecturer presumed that the average class attendance was 65 students on average. He did call register for 10 days and observed the following: 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68. Test the hypothesis that  $\mu=65$  (4 marks)
- b) Discuss any four types of test-statistics and highlight when appropriate to use each (8 marks)
- c) Using the data below determine if there was any correlation between the sodium concentration and Opuntia area coverage. (8 marks)

Sodium conc.	1.75	5.83	5.33	4.67	7.17	5.50	9.33	6.83	7.50	10.80	11.30	11.40
Opuntia area	14.2	30.1	71.2	77.5	75.9	121.8	132.1	159.0	181.9	184.3	194.6	219.1

**QUESTION FOUR (20 MARKS)**

- a) The function  $\hat{y} = 3.5 + 1.23x_1 + 0.78x_2$  is model of predicting the expected yield (kg '000') per hectare of a certain crop as influenced by fertilizer  $k63(x_1)$  and the humidity level ( $x_2$ ). Interpret the values 3.5, 1.23 and 0.78 (6 marks)
- b) Differentiate the following terms as apply in social research.
- i. Type I and Type II error
  - ii. Sample and population
  - iii. Two tailed and one tailed test (6 marks)
- c) The table below shows the scores of eight interviewees on aptitude and subject matter knowledge

Interviewee	A	B	C	D	E	F	G	H
Aptitude test score	15	20	28	12	40	60	20	80
Subject matter score	40	30	50	30	20	10	30	60

Compute the rank coefficient of correlation (8 marks)

**QUESTION FIVE (20 MARKS)**

- a) Highlight three principles of experimental design (3 marks)
- b) By citing examples distinguish between discrete and numeric variables (6 marks)
- c) A manufacturer wished to compare the performance of different machine brands, their performance difference is summarized in ANOVA table below.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean sum of Squares	F – value
Between machines	$4 - 1 = 3$	540.69	180.23	$F = \frac{180.23}{7.15} = 25.207$
Error	$16 - 4 = 12$	85.75	7.15	
Total	$16 - 1 = 15$	626.44		

- i) Based on the table above state the number of machine brands compared (1marks)
- ii) Explain the two sources of variation in column one of the table (2 marks)
- iii) When is it appropriate to use ANOVA and F-test for data analysis (4 marks)
- iv) Was there a significant difference in their performance at  $\alpha=5\%$  (4 marks)