MACHAKOS UNIVERSITY

University Examinations 2019/2020 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

………..YEAR …… SEMESTER SPECIAL /SUPPLEMENTARY EXAMINATION FOR

BACHELOR OF ENVIRONMENT STUDIES

ESU 302: STATISTICS

DATE: **TIME:**

INSTRUCTION: *Answer Question ONE which is compulsory and any other TWO Questions*

QUESTION ONE (30 MARKS)

a)   Differentiate the following terms as they apply in scientific research

   i.   Sample and a population

   ii.   Response and independent variables

   iii.   Descriptive and inferential statistics

   iv.   Null hypothesis and Alternative hypothesis

   v.   Type I and Type II error   (10 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

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

   ii   Highlight three assumptions for regression modeling   (3 marks)

QUESTION TWO (20 MARKS)

a) The data below is a summary of the slim possible participants weights in kilograms before and after the season.

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

Test the hypothesis that on average the exercise did not result to any significant weight loss at 10% significance level   (8 marks)

b) Below was the daily harvest of a tropical plant during a particular week in the year 2017.

Harvested fruits	11-20	21-30	31-40	41-50	51-60
Frequency	3	6	11	3	2

Determine;

i.   Average harvest per day

ii.   Median

iii.   Mode and

iv.   Standard deviation.   (12 marks)

QUESTION THREE (20 MARKS)

a) The information relates health records from County XYZ randomly selected

Gender	Males	Females
Average lifespan	50	58
Standard deviation	12	9
Sample size	100	150

Determine;

i.   The gender with the higher dispersion lifespan

ii.   The combined standard deviation

iii.   State the hypotheses for testing life expectancy for the two genders

iv.   Do the life span differ significantly for the two gender (use α-level = 1%).

                                                                   (12 marks)

b) 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) Differentiate between a discrete random variable and a continuous random variable

(4 marks)

b) Suppose a fair coin is tossed twice and by letting X represent the number of Heads that show up. Determine $E(X)$ and $Var(X)$ (8 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	$F = \dfrac{180.23}{7.15} = 25.207$
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 (1 mark)

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 α=5% (4 marks)