



# MACHAKOS UNIVERSITY

University Examinations 2019/2020 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

**SUPPLEMENTARY EXAMINATION FOR THE DEGREE OF BACHELOR  
OF SCIENCE ENVIROMENT  
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) Descriptive and inferential statistics
  - iii) Sampling unit and Sample frame (6 marks)

- b) (i) Determine the regression equation of Y on X given that

$$\sum xy = 205, \sum x^2 = 2150, \bar{x} = 35 \text{ and } \bar{y} = 7. \quad (5 \text{ marks})$$

- (ii) Highlight three assumptions for regression analysis (3 marks)

- c) 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   | 10 | 8  | 12 | 15 | 6 | - | -  | - |
| Females | 8  | 12 | 13 | 9  | 3 | 8 | 11 | 3 |

- i. Determine whether gender influence the class performance significantly
  - ii. Test the hypothesis that the mean score for females is 14.9 (9 marks)
- d) Highlight four limitations of statistics in research (3 marks)
- e) Compute the standard error of the mean given the data 5,7,9,7,2. (4 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;

- Average harvest per day
  - Median
  - Mode and
  - 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;

- The gender with the higher dispersion lifespan
- 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  $\alpha$ -level = 1%).
- (12 marks)

- a) 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)
- a) 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) Discuss four functions of statistics in research
- (8 marks)
- b) The lecturer presumed that the average class attendance was 65 students. He did call register for 10 days and observed following: 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68. The t-test output was as given

| The sample t-test output, for the test value=65 |       |    |                 |                 |                          |        |
|-------------------------------------------------|-------|----|-----------------|-----------------|--------------------------|--------|
|                                                 | t     | df | Sig. (2-tailed) | Mean difference | 95% CI of the difference |        |
|                                                 |       |    |                 |                 | Lower                    | Upper  |
| Attendance                                      | 2.825 | 9  | 0.020           | 2.500           | 0.4979                   | 4.5021 |

- i. State the hypothesis for the above scenario
- ii. Based on the t-value and the sig.(2-tailed) value make statistical conclusion
- iii. Interpret the 95% CI lower and upper difference values (6 marks)

(c) The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 and 9.6 liters. Find a 95% confidence interval for the mean of all such containers, assuming an approximate normal distribution. (6 marks)