MACHAKOS UNIVERSITY

University Examinations for 2024/2025 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

SECOND YEAR FIRST SEMESTER EXAMINATION FOR DEGREE IN BACHELOR OF ARTS IN GENDER AND DEVELOPMENT BACHELOR OF ARTS IN PUBLIC ADMINISTRATION BACHELOR OF ARTS IN PSYCHOLOGY

ACU 201: INTRODUCTION TO STATISTICS IN SOCIAL SCIENCES

DATE: 20TH JANUARY, 2025 **TIME: 8.30 AM – 10.30 AM**

INSTRUCTIONS:
*Answer question **ONE (Compulsory)** and **any other TWO** questions*

QUESTION ONE (30 MARKS)

a) A sport manager uses visualization in promoting enhanced performance in college athletes. She is interested in evaluating the relative effectiveness of visualization alone versus visualization plus appropriate self-talk. An experiment is conducted with a college basketball team. Ten members of the team are selected. Five are assigned to a visualization alone group, and five are assigned to a visualization plus self-talk group. Both techniques are designed to increase foul shooting accuracy. Each group practices its technique for one month. The foul shooting accuracy of each player is measured before and one month after practice. The difference scores for each group are compared to determine the relative effectiveness of the two techniques. In the experiment described above, specify the:

i. Sample [1 mark]

ii. Statistic [1 mark]

iii. Population [1 mark]

iv. Data [1 mark]

b) State and briefly define, the four data measuring types. [4 marks]

c) Contrast between qualitative data and quantitative data. In addition, contrast between continuous data and discrete data and show how the latter two fit as subparts to either qualitative or quantitative. [6 marks]

d) Graph the empirical rule with the mean and $+/-$ one, two, and three standard deviations including the percentage of area that fit in each range. [3 marks]

e) Differentiate between random sampling and non-random sampling. [2 marks]

f) Give any three reasons for taking a Census over a sample. [3 marks]

g) Draw a scatter plot for a correlation coefficient $r = 1$. [2 marks]

h) If a company employs 3500 people and if a random sample of 175 of these employees has been taken by systematic sampling, what is the value of k? [2 marks]

i) A telephone company wants to estimate the average length of long-distance calls during weekends. A random sample of 50 calls gives a mean of $\bar{x} = 14.5$ minutes and a standard deviation as $s = 5.6$ minutes. Give a 95% confidence interval for the average length of a long-distance call during weekend. [4 marks]

QUESTION THREE (20 MARKS)

The following table shows the weight in kg of some electronics components manufactured by a company: -

28.2	26.2	25.3	27.7	27.2	28.0	29.0	25.6
27.4	26.4	24.7	26.6	26.8	28.3	27.0	28.9
25.9	28.4	26.8	27.9	25.8	27.3	29.3	23.6
25.1	28.7	26.4	25.8	25.3	24.9	26.3	26.6
24.2	25.6	27.3	26.9	27.2	26.7	27.3	26.8

a) Find the Range (R) of the data. [2 Marks]

b) Use the data to construct a grouped frequency distribution table starting with the class 23.0-24.0 [3 Marks]

c) Use the frequency table to compute: -

i) The Mean [3 Marks]

ii) The Median [3 Marks]

iii) The Mode [3 Marks]

iv) Upper quartile [3 marks]

v) Lower quartile [3 marks]

QUESTION THREE (20 MARKS)

a) Given below are the gross values, in millions of Kenyan Shilling, of milk produced and cattle raised in Kajiado county, for the ten years from 1999 through 2008.

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Milk	287	363	413	411	455	477	547	569	712	718
Cattle	177	214	212	237	238	223	223	229	252	271

i. Find the correlation coefficient for these variables. Interpret your result. [5 marks]

ii. Calculate the coefficient of determination. What does this value represent? [2 marks]

iii. Find the regression equation that predicts the value of cattle produced in a year based on the value of milk produced that year. [3 marks]

iv. Predict the value of cattle raised for a year in which the value of milk produced is 400 million shillings. [2 marks]

v. At 0.05 level of significance, test the claim that the value of milk produced can be used to predict the value of cattle raised. [4 marks]

b) A tutorial lab on campus offers free tutoring for any student on campus. Here are the GPAs for 12 randomly selected students, and the number of tutoring appointments that those students missed. Calculate r_s Spearman's rank correlation coefficient. [4 marks]

GPA	2.66	2.05	2.07	2.62	1.30	3.00	3.25	2.58	2.36	2.81	3.11	2.56
Missed Appointments	3	1	2	0	7	0	2	0	3	1	1	2

QUESTION FOUR (20 MARKS)

a) Explain any three probability sampling techniques. [6 marks]

b) Explain the three properties considered to select the best estimator. [3 marks]

c) Given the data below calculate the Variance and Standard deviation of the following frequency distribution. [6 Marks]

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80

Frequency	7	15	18	23	30	13	8	6

d) Briefly explain some common ways to present analyzed data effectively. [5 marks]

QUESTION FIVE (20 MARKS)

a) Outline the key procedures for data analysis and presentation. [4 marks]

b) An experiment was conducted to investigate the viability of equipment designed to measure the foreign of an audio source. Three independent measurements were recorded by this equipment, they were 4.1, 5.2 and 10.2. Estimate the variance with 90% confidence interval. [5 marks]

c) Suppose you plan to survey employees to find their medical expenses, you want to be 95% confident that the sample mean is within ± sh.500. A pilot study shows that the standard deviation is about sh.8000. What sample size will you use?
[3 marks]

d) A paint manufacturer wants to determine the average drying time of a new interior wall paint. If for 12 tests areas of equal size, he obtained a mean drying time of 66.3 minutes and a standard deviation of 8.4 minutes. Construct a 95% confidence interval for the true mean. [4 marks]

e) A study was being conducted to estimate the proportion of voters is a sizeable community who favored the construction of a nuclear plant. If it was found that only 140 out of 400 selected voters at random favored the project. Find a 95% confidence interval for the proportion for all voters who favored the project. [4 marks]