KIRIRI WOMENS' UNIVERSITY OF SCIENCE AND TECHNOLOGY
UNIVERSITY EXAMINATION, 2023/2024 ACADEMIC YEAR

FOURTH YEAR, FIRST SEMESTER EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE
(MATHEMATICS)

Date: 7th December, 2023
Time: 11.30am –1.30pm

KMA 400 - RESEARCH METHODOLOGY

INSTRUCTIONS TO CANDIDATES

ANSWER **QUESTION ONE (COMPULSORY)** AND **ANY OTHER TWO** QUESTIONS

QUESTION ONE (30 MARKS)

Read the following essay then use it to answer the questions that follow.

Voter turnout in the Kenya has been decreasing steadily over the past ten years, in contrast to other areas of the country. According to surveys conducted by local non-profits, turnout is lowest among those under 25 years of age. Low voter turnout has been shown to have negative associations with overall civic engagement. It is becoming an area of increasing concern in many democracies. When specific groups of citizens lack political representation, they are likely to become more excluded over time, leading to an erosion of trust in democratic institutions. Addressing this problem will have practical benefits, and will contribute to understanding of this widespread phenomenon. There have been some effective attempts at engaging these groups. In the last two elections, major parties increased their campaigning efforts. In addition, voter education with the aim of informing citizens on the need for participation in choosing the leadership has been increased. However, these interventions have yet to have any significant effect on turnout. An attempt to determine the main factors that lead to low voter turnout as well as the possible ways of solving this problem is necessary.

Required:

a)   In reference to the passage above, what could be the research topic?

(2 marks)

b)   What is the main problem of concern that motivates this research?

(2 marks)

c)   In line with the characteristics of good research objectives in (c), outline two possible specific objectives to the problem in the passage.

(2 marks)

d)   Discuss two possible research designs that can be adopted in this research.

e)   Briefly explain any five data collection methods you would use.

(5 marks)

f)   Outline the possible steps that you would follow when conducting this research.

(6 marks)

g)   Highlight **FOUR** characteristics of a good research that your research on the above topic must have.

(4 marks)

h)   To determine whether the voter turnout is really magnified problem for under 25 years of age, comparative research was conducted to compare willingness of under 25 of age and those between 25 and 40 years to turnout for voting. In 10 selected counties, a sample was taken and each sampled individual was asked to rate their possibilities of turning out to vote in a scale of 0-10 (0 not likely and 10 for highly likely). The following average rates were obtained as follows.

Under 25	5	3	8	2	4	6	7	5	4	4
25-40	8	7	6	4	4	7	7	6	8	5

Use the ordinary sign test to determine whether the average rate of turnout for under 25 is significantly less than that of between 25-40 years of age. Use 5% level of significance.

(5 marks)

QUESTION TWO (20 MARKS)

a)   Errors in sampling can be either systematic or sampling. Differentiate between the two and give examples of the causes for each of the two and how the two errors can be reduced.

(6 marks)

b)   Sampling methods are broadly classified as probability or non-probability sampling methods. Distinguish between the two classifications, hence elaborate on any **FOUR** probability sampling designs you may use in data collection.

(8 marks)

c)   To ensure that the questioner captures the required information, it must be designed to be as good as possible. Briefly explain any **SIX** characteristics of a good questionnaire.

(6 marks)

QUESTION THREE (20 MARKS)

a)   You are a public health researcher interested in social factors that influence heart disease. You survey 500 towns and gather data on the percentage of people in each town who smoke, the percentage of people in each town who bike to work, and the percentage of people in each town who have heart disease. Because you have two independent variables and one dependent variable, and all your variables are quantitative, you decide to use multiple linear regression to

analyze the relationship between them. You decide to analyze the data using excel and obtain the following summary output.

Regression Statistics	
Multiple R	0.9898
R Square	0.9796
Adjusted R Square	0.9795
Standard Error	0.6540
Observations	498

ANOVA

	df	SS	MS	F	Significance F
Regression	2	10176.57	5088.28	11895.24	0
Residual	495	211.7403	0.427758		
Total	497	10388.31			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	14.9846	0.0801	186.9882	0	14.8272	15.1421
biking	-0.2001	0.0013	-146.525	0	-0.2028	-0.1975
smoking	0.17833	0.0035	50.3866	5.2E-197	0.1713	0.1853

Use the output given to answer the following questions.
i)   Write simple linear regression model.

(2 marks)
ii)   Interprate the parameters of biking and smoking.

(2 marks)
iii)   Predict the percentage of those with heart decease for a town whose 16% of the population bike to work and 4% are smokers.   (2 marks)
iv)   Based on the p-value, test the hypotheses on significance of individual parameters at 5% level of significance.

(4 marks)
v)   Test on the adequecy of the overall model at 5% level of significance.

(3 marks)
vi)   What percentage in variation in percentages of heart desease is determined by biking and smoking?   (2 marks)
b)   There are several factors to consider before choosing a method of data analysis. Highlight any five of these factors.   (5 marks)

QUESTION FOUR (20 MARKS)

a) Reference the following materials using American Psychological Association style.
   i) Journal
      Authors: Kemei Anderson Kimutai, Christopher Ouma Onyango and Mike Wafula.
      Journal: Science Journal of Applied Mathematics and Statistics, **Volume**: 9 **Issue:** 5
      Pages: 126–132. **Title:** Estimation of finite population total in presence of missing
      values in two-phase sampling. **Year:** 2021   (4 marks)

   ii) Text Book
      Year: 2002. *Title: Statistical Analysis with Missing Data.* **Authors: Second:** Little, R.
      J. A. **First:** Rubin, D. B. **Publisher:** Wiley.

                                                                                      (3 marks)
   iii) Conference proceedings
      Authors: Ma, Y. and Fu, Y. **Year:** 2004. **Title:** A study on distance metrics for k-
      nearest neighbor imputation. *Conference: Proceedings of the 4th International
      Conference on Machine Learning and Data Mining.*
                                                                                      (3 marks)

b) Illustrate the four levels of measurements using an example. Hence explain how these levels
   of measurements determine the method of data analysis to be used.   (10 marks)

QUESTION FIVE (20 MARKS)

a) Choosing a research problem requires a careful consideration since it has a great impact on the
   research process. Discuss the steps to be followed when formulating a research problem.
                                                                                      (10 marks)
b) In hypothesis testing one can either adopt parametric or non-parametric tests. Highlight the
   advantages of adopting nonparametric tests.

                                                                                      (4 marks)
c) You have collected the following data on ownership of Small Micro enterprises (SME).

Sex	SME Owners	Not SME Owner
Men	100	400
Women	50	450

Does the act of owning a SME dependent on sex? Test the hypothesis at 5 % significance level.

(6 marks)