



MACHAKOS UNIVERSITY

University Examinations 2022/2023 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR SECOND SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE IN MATHEMATICS AND COMPUTER SCIENCE

BACHELOR OF SCIENCE IN STATISTICS AND PROGRAMMING

BACHELOR OF SCIENCE IN ACTUARIAL SCIENCE

BACHELOR OF SCIENCE IN MATHEMATICS

BACHELOR OF EDUCATION

BACHELOR OF ARTS

SMA 365: DESIGN AND ANALYSIS OF SAMPLE SURVEYS

DATE: 21/04/2023

TIME: 02:00 – 04:00 PM

Instructions to the Candidate:

1. Answer **Question 1** and any other **two** questions.
2. You need a Scientific Calculator and Statistical Tables for this paper

QUESTION ONE (COMPULSORY) (30 MARKS)

- (a) Explain each of the following terms as used in sample survey, giving an example from a real life situation:
- (i) Triangulation;
 - (ii) Data cleaning. (4 marks)
- (b) (i) Explain the importance of carrying out a *pre-test* before the main survey is rolled out, giving *four* reasons. (4 marks)
- (ii) Outline any **three** ways in which a *questionnaire* can be administered to the respondents during the collection of data in a survey. (3 marks)

- (c) Construct a short questionnaire for a survey designed for the collection of primary data from the respondents, covering the following variables: (8 marks)

- Sex
- Age in years
- Level of education (highest)
- Monthly salary
- Marital status
- Water source for domestic use: tap, well, river, lake, dam, etc.
- Employment status: permanent, contract, casual.

- (d) Given a stratified random sample with a stratified sample mean $\bar{X}_{st}$ and a population mean $\bar{X}$, show that the stratified mean $\bar{X}_{st}$ is an unbiased estimator of the population mean $\bar{X}$. (3 marks)

$$\text{where } \bar{X}_{st} = \frac{1}{N} \sum_{i=1}^m N_i \bar{x}_i \quad \text{and} \quad \bar{X} = \frac{1}{N} \sum_{i=1}^m X_i$$

- (e) A survey is to be conducted among *boda boda* operators in Machakos Town. Taking the registered operators as the target population, with a population of size $N = 500$ from which a sample of size $n = 25$ is to be selected. Select an actual sample procedurally, stating clearly the range of your sampling frame and the selected sampling units, using each of the following sampling techniques:
- (i) Simple random sampling; (5 marks)
- (ii) Systematic sampling. (3 marks)

QUESTION TWO (20 MARKS)

- (a) (i) Explain the term *questionnaire* as used in a survey. (2 marks)
- (ii) Outline *five* possible sources of sampling bias in a survey. (5 marks)
- (b) Outline *three* features of random sampling techniques and *three* features of non-random sampling techniques in a survey. (3 marks)

- (c) Given a simple random sample drawn without replacement with a sample variance given by

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{and its population variance given by } S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2.$$

Show that the sample variance s^2 is an unbiased estimator of the population variance S^2 .

(10 marks)

QUESTION THREE (20 MARKS)

- (a) Given a population of size $N = 5$ comprising the data values 24, 21, 18, 27, 15, from which a random sample of size $n = 3$ is to be selected, determine all the possible samples of size 3 which can be drawn from the population.

Hence, using the selected samples, prove that the sample mean $\bar{x}$ is an unbiased estimator of the population mean $\bar{X}$. (8 marks)

- (b) A survey was conducted on the number of residents in a certain village with several homesteads. A simple random sample of 20 homesteads was selected from the village which has a total of 2,500 homesteads. The number of people per homestead in the sample was as follows:

13, 10, 18, 11, 20, 15, 10, 14, 12, 16, 20, 12, 15, 14, 18, 11, 12, 15, 14, 10

- (i) Estimate the *mean* number and the *total* number of the people in the village; (2 marks)
- (ii) Determine the 95% confidence interval for the population mean of the number of people in the village; (5 marks)
- (iii) Determine the 99% confidence interval for the population total of the number of people in the village. (5 marks)

QUESTION FOUR (20 MARKS)

A survey is to be conducted among employees of a certain county government, with four sub-counties A, B, C and D. A random sample of 140 employees is to be selected from the target population of 2,800 employees. This target population is divided into four strata based on the four sub-counties, with their corresponding attributes as shown in the table below:

Stratum	A	B	C	D
Stratum size	800	900	600	500
Stratum sample mean	64	75	80	74
Standard deviation	4	5	8	3
Stratum cost per unit	25	49	36	16

- (a) Determine the sample size for each stratum, if the sample is to be drawn from the population using stratified random sampling with:
 - (i) proportional allocation; (2 marks)
 - (ii) optimum allocation with fixed cost; (4 marks)
 - (iii) optimum allocation with variable cost. (6 marks)
- (b) By using the sample sizes based on *proportional allocation* obtained in (a) (i) above:
 - (i) Determine the 99% confidence interval for the population mean; (6 marks)
 - (ii) Determine the variance of the estimate of the population total. (2 marks)

QUESTION FIVE (20 MARKS)

- (a) Explain each of the following sampling errors which may occur in a sample survey, illustrating each with an example from a real life situation.
 - (i) frame error;
 - (ii) response error. (6 marks)

- (b) The data in the table below shows observations on income (X) and rent (Y) for the total monthly income and monthly house rent expenditure for a random sample of 12 urban households in an estate. The population consists of 250 households in the estate.

Household	A	B	C	D	E	F	G	H	J	K	L	M
Income	56	80	36	72	84	60	48	75	56	36	54	48
Rent	14	20	12	18	24	15	16	25	18	08	16	12

- (i) Determine the ratio estimator R and the mean of the rent expenditures Y from the sample.
- (ii) Determine the 95% confidence interval for the population ratio estimator R .

(14 marks)