



MACHAKOS UNIVERSITY

University Examinations 2021/2022 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 MATHEMATICS

BACHELOR OF EDUCATION

BACHELOR OF ARTS

SMA 365: DESIGN AND ANALYSIS OF SAMPLE SURVEYS

DATE: 17/12/2021

TIME: 8:30 – 10:30 AM

INSTRUCTION:

Answer Question One and Any Other Two Questions

You need a Scientific Calculator and Statistical Tables for this paper

QUESTION ONE (COMPULSORY) (20 MARKS)

- a) Explain *any two* of the following terms as used in sample survey:
- Sampling frame;
 - Statistical induction;
 - Data cleaning. (4 marks)
- b) i Explain *three* sampling errors which may occur in a sample survey, illustrating each with an example from a real life situation. (6 marks)
- ii Explain the term *triangulation* as used in the collection of primary statistical data in a survey, illustrating with two examples from a real life situation. (3 marks)
- c) It is generally assumed that multi-stage sampling eliminates bias in a sample. However, there is no guarantee that this will be achieved. Account for this statement, illustrating with a real life example. (3 marks)
- d) i Given a simple random sample drawn without replacement (SRS-WOR) with a sample mean $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and a population mean $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$, show that the sample mean $\bar{x}$ is an unbiased estimator of the population mean $\bar{X}$. (3 marks)

- ii Given a stratified random sample with a stratified sample mean given by

$$\bar{X}_{st} = \frac{1}{N} \sum_{i=1}^m N_i \bar{x}_i \text{ and a population mean given by } \bar{X} = \frac{1}{N} \sum_{i=1}^m X_i, \text{ show that the}$$

stratified mean $\bar{X}_{st}$ is an unbiased estimator of the population mean $\bar{X}$. (3 marks)

- e) Given a population of size $N = 5$ comprising the data values 27, 24, 18, 30, 21, 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)

QUESTION TWO (20 MARKS)

- a) Differentiate between each of the following terms as used in a statistical survey, illustrating with a real life example in each case:
- Observational unit and sampling unit;
 - Under-coverage and over-coverage. (8 marks)
- b) Explain the importance of carrying out a *pre-test* before the main survey is rolled out. Justify this statement giving **two** reasons. (4 marks)
- c) Given a simple random sample drawn without replacement with a sample variance given by $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ and its corresponding population variance $S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$. Show that the sample variance s^2 is an unbiased estimator of the population variance S^2 . (8 marks)

QUESTION THREE (20 MARKS)

- a) Describe multi-stage sampling technique as used in the collection of statistical data in a survey study, illustrating with an example from a real life situation. (4 marks)
- b) Construct a questionnaire designed for the collection of primary data from respondents covering each of the following variables: (8 marks)
- Age
 - Sex
 - Level of education (highest level)
 - Monthly Salary
 - Marital status
 - Water source for domestic use
 - Duration of employment
 - Employment status: permanent, contract, casual.

- c) By taking the example of a survey to be conducted in your campus, and taking the Bachelor of Science students from the Department of Mathematics as your target population, and assuming a sample of size $n = 25$ to be selected from a population of size $N = 750$, select procedurally an actual sample stating clearly the range of your sampling frame and the selected sampling units using:
- Simple random sampling; (4 marks)
 - Systematic random sampling. (4 marks)

QUESTION FOUR (20 MARKS)

A random sample of size 320 is to be selected from a target population of size 4,800. This target population is divided into four strata as shown below:

Stratum	A	B	C	D
Stratum size	1200	1350	1500	750
Stratum sample mean	64	72	82	76
Standard deviation	11	14	9	6
Stratum cost per unit	16	25	64	16

- Determine the sample size for each stratum if the sample is to be drawn from the population using stratified random sampling with:
 - proportional allocation; (2 marks)
 - optimum allocation with fixed cost; (4 marks)
 - optimum allocation with variable cost. (6 marks)
- Determine the 99% confidence interval for the population mean (*use the sample sizes based on proportional allocation obtained in (a) (i) above*). (8 marks)

QUESTION FIVE (20 MARKS)

- Explain the role of an *introductory letter* as an attachment to a questionnaire. (2 marks)
- Given a two-stage sample in which the sample mean is given by $\bar{\bar{x}} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m x_{ij}$ and its corresponding population mean is given by $\bar{\bar{X}} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M x_{ij}$. Show that the sample mean $\bar{\bar{x}}$ is an unbiased estimator of the population mean $\bar{\bar{X}}$. (8 marks)

c) A simple random sample of 20 homesteads was selected from a village comprising 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 total number of people in the village; (2 marks)
- ii. Determine the 95% confidence limits for the mean number of people in the village; (4 marks)
- iii. Determine the 99% confidence limits for the total number of people in the village. (4 marks)