# MACHAKOS UNIVERSITY

University Examinations 2021/2022

DEPARTMENT OF MATHEMATICS AND STATISTICS

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

THIRD YEAR SECOND SEMESTER

Date: 04/12/2022 Time: 08:30 – 10:30

---

*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

(a) Explain each of the following terms as used in sample survey, giving an example for each from a real life situation:

(i) Census;

(ii) Triangulation;

(iii) Data cleaning. (6 *marks*)

(b) Explain the importance of carrying out a *pre-test* before the main survey is rolled out, giving *four* reasons. (4 *marks*)

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

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

(d) 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 $\bar{X}$. (4 *marks*)

(e) By taking the example of a survey to be conducted in your campus, and taking the Mathematics students as your target population, and assuming a sample of size $\boldsymbol{n = 20}$ to be selected from a population of size $\boldsymbol{N = 600}$, select procedurally an actual sample, 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**

(a) (i) Differentiate between the terms *observational unit* and *sampling unit* as used in a survey, illustrating with a real life example. (4 *marks*)

(ii) Outline *five* possible sources of bias in sampling. (5 *marks*)

(b) 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*)

(c) Given a simple random sample drawn without replacement (SRS-WOR), with a sample mean $\bar{x}$ and a sample variance $s^2$, show that the variance of the sample mean $var(\bar{x})$ is given by

$$var(\bar{x}) = \left(\frac{N-n}{N}\right)\frac{S^2}{n} \qquad \text{where} \quad s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$$ (8 *marks*)

**Question Three**

(a) 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*)

(b) A survey was conducted to determine the number of mobile phone SIM cards used by households among Kenyan urban residents. A simple random sample of 26 households was drawn from a residential estate comprising 800 households. The number of phone SIM cards owned per household in the sample was as follows:

8, 6, 5, 7, 9, 4, 6, 5, 4, 5, 7, 3, 8, 9, 6, 5, 7, 9, 6, 4, 8, 3, 6, 5, 7, 4.

(i) Estimate the total number of phone SIM cards owned by the households in the residential estate; (2 *marks*)

(ii) Determine the 95% confidence limits for the population mean of the number of phone SIM cards owned by the residents; (5 *marks*)

(iii) Determine the 99% confidence limits for the population total of the number of phone SIM cards owned by the residents. (5 *marks*)

## Question Four

A random sample of size 280 is to be selected from a target population of size 5,600. This target population is divided into four strata, with corresponding attributes as shown below:

| Stratum | A | B | C | D |
|---|---|---|---|---|
| Stratum size | 1200 | 1800 | 1600 | 1000 |
| Stratum sample mean | 64 | 72 | 82 | 76 |
| Standard deviation | 10 | 14 | 8 | 6 |
| Stratum cost per unit | 36 | 25 | 64 | 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

(a) Explain the role of an *introductory letter* as an attachment to a questionnaire used in data collection in a survey. (2 *marks*)

(b) Given a two-stage sample with a sample mean $\bar{\bar{x}}$ and its corresponding population mean $\bar{\bar{X}}$, show that the sample mean $\bar{\bar{x}}$ is an unbiased estimator of the population mean $\bar{\bar{X}}$.

where $\bar{\bar{x}} = \frac{1}{n\,m}\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}$ and $\bar{\bar{X}} = \frac{1}{N\,M}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}$ (8 *marks*)

(c) The data in the table 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 residents. The population consists of 80 relatively high income earning families.

| Resident | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Income | 48 | 78 | 35 | 72 | 88 | 60 | 40 | 72 | 84 | 25 | 56 | 48 |
| Rent | 12 | 18 | 12 | 16 | 20 | 15 | 14 | 15 | 21 | 8 | 16 | 11 |

(i) Estimate the ratio estimator $R$ and the mean of the rent expenditures $Y$ from the sample.

(ii) Determine the 95% confidence interval for the ratio estimator $R$.

(10 *marks*)