



NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF SCIENCE AND TECHNOLOGY EDUCATION

DEPARTMENT OF SCIENCE, MATHEMATICS AND TECHNOLOGY EDUCATION

BACHELOR OF EDUCATION

SURVEYING AND SAMPLING TECHNIQUES

PST4134

Main Examination Paper

NOVEMBER 2024

This examination paper consists of 6 pages

Time Allowed: 3 hours

Total Marks: 100

Special Requirements: Statistical Tables

Examiner's Name: Sesilani Nkomo

External Examiner: Dr Sunzuma G

INSTRUCTIONS

1. This paper consists of **NINE** questions. Answer **ALL** questions in section A and at **most three** questions in section B.
2. All answers should be presented in good style.
3. Begin each full question on a new page.

MARK ALLOCATION

QUESTION	MARKS
1-4	40
5.	20
6.	20
7.	20
8.	20
9.	20

SECTION A [40 marks]

Candidates may attempt **ALL** questions in this section

- A1.** Contrast the following pairs of terms:
- (a) Pre-testing and Pilot survey. [3]
 - (b) Non-sampling errors and sampling errors. [3]
 - (c) Data cleaning and data entry. [3]
 - (d) Closed questions and open-ended questions. [3]
 - (e) Probability sampling and Non-probability sampling. [3]
- A2.** Briefly describe any two (2) non-probability sampling techniques. [5]
- A3.** (a) State three methods for checking data entry and their relative merits. [6]
- (b) What is meant by the term village mapping in socio-economic research. [2]
- © Criticise the following questions proposed for inclusion in a questionnaire?
- (i) Don't you agree that democracy is an unachievable proposition in Africa?
 - 1 Yes
 - 2 No [3]
 - (ii) What is the highest level of education attained by your parents?
 - 1 Primary level
 - 2 Secondary level
 - 3 Tertiary level Veterinarians. [3]
- A4.** The Ministry of Agriculture through the Grain Marketing Board (GMB) desires to estimate the total maize yield for the farming season 2016 to 2017 from Zimbabwe's ten provinces.
- (a) What is the advantage of using stratified random sampling over simple random sampling technique in collecting this information? [2]
 - (b) Discuss the relative merits of using personal interviews as a method of data collection for the study. [2]
 - (c) What is the target population for study? [1]
 - (d) What could be used as the sampling frame for the survey? [1]

SECTION B [60 marks]

Candidates may attempt at most **THREE** questions in this section

B5. (a) Distinguish between cluster sampling and stratified sampling. [3]

(b) A mathematician developed a test designed to assess the attitudes of students towards mathematics at a certain school with 100 classes. She sampled 15 classes and tested every member of the sampled class. The results are shown below:

Class i	No. of students m_i	Total Score y_i	No. interest in mathematics a_i
1	30	1400	15
2	25	1000	10
3	35	1500	10
4	20	1400	12
5	28	1400	10
6	30	1200	13
7	32	1800	12
8	35	1400	14
9	36	1600	10
10	32	1200	11
11	35	1700	10
12	40	1800	8
13	20	1000	10
14	25	1000	12
15	38	2100	10

(i) Estimate the average score at the school [3]

(ii) Estimate the variance of the average score. [6]

(iii) Find the sample size required to estimate the proportion of students interested in mathematics with a bound of 0.04 on the error of estimation and 5% level of significance. [8]

B6. A farmer has citrus orchard of $N = 2000$ nartjies trees. The farmer wishes to estimate the average yield (in bushels) from the orchard, from a 1-in-10 systematic sample of trees. The data from this survey are listed in the table below.

Tree Sample	Yield (bushels) y	y^2
1	3.5	11.55
2	4.0	16.00
3	2.0	4.00
.	.	.
.	.	.
.	.	.
199	3.9	15.21
200	2.8	7.84
	700	2500

- (a) Estimate, μ , the average yield of nartjies trees. [2]
- (b) Construct a 95% confidence interval for the mean yield. [6]
- (c) Suppose the farmer is now interested in estimating the total yield from the orchard. Estimate, τ , and place a bound on the error of estimation. Use 5% level of significance. [8]
- (d) Determine the sample size required to estimate, μ , with a bound on the error of estimation of $d = 0.01$ using 5% level of significance. [4]

B7. (a). The results of stratified random samples for a characteristic X are summarised in a table as follows:

Stratum (i)	N_i	\bar{y}_i	a_i	\hat{p}_i	n_i	s_i^2
1	1000	30.125	15	0.6	25	26.3
2	1200	35.725	25	0.71	35	14.7
3	900	25.125	10	0.5	20	19.1
4	1400	30.725	30	0.75	40	15.6

- (i) Estimate the population total (τ) and its associated variance. [6]
- (ii) Find a 98% confidence interval for the population proportion that possess characteristic X . [9]
- (b) A researcher wishes to estimate the average income for employees in a large organisation. Records have employees listed by seniority, and generally speaking, salary increases with seniority. Discuss the advantages of using stratified random sampling over simple random sampling in this case. [5]

- B8.** (a) The NUST bursar's department is interested in estimating the amount of money paid by students per semester. The University has an enrolment of 8000 students. A simple random sample of $n = 300$ students is selected. The sample average payment was 280 dollars with a sample variance of 100.
- (i) Estimate the total amount (τ) of money paid by the students per semester. [3]
 - (ii) Place a bound on the error of estimation at 5% significance level. [3]
 - (iii) Determine the sample size required to estimate the total amount paid per semester, with a bound on the error of estimation $d = 4000$ at 5% significance level [4]
- (b) Suppose y_1, y_2, \dots, y_n denotes a simple random sample from a population with mean μ and variance σ^2 .
- (i) show that \bar{y} is unbiased for μ [2]
 - (ii) $Var(\bar{y}) = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$, [4]
 - (iii) Hence or otherwise prove that the sample size required to estimate the mean is given by:

$$n = \frac{N\sigma^2}{\frac{d^2}{z^2}(N-1) + \sigma^2} \quad [4]$$

END OF QUESTION PAPER

SURVEY TECHNIQUES

Sampling procedure	Estimator	Variance	Sample Size [bound d]
Simple			
Mean	$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$	$Var(\bar{y}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$	$n = \frac{N\sigma^2}{\frac{d^2}{z^2}(N-1) + \sigma^2}$
Total	$\hat{t} = N\bar{y}$	$Var(\hat{t}) = N^2 Var(\bar{y})$	$n = \frac{N\sigma^2}{\frac{d^2}{z^2 N^2}(N-1) + \sigma^2}$
Proportion	$\hat{p} = \frac{1}{n} \sum_{i=1}^n y_i$	$Var(\hat{p}) = \frac{\hat{p}\hat{q}}{n-1} \left(\frac{N-n}{N} \right)$	$n = \frac{Npq}{\frac{d^2}{z^2}(N-1) + pq}$
Stratified			
Mean	$\bar{y}_{st} = \frac{1}{N} \sum N_i \bar{y}_i$	$Var(\bar{y}_{st}) = \frac{1}{N^2} \sum N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \frac{s_i^2}{n_i}$	$n = \frac{\sum N_i^2 \sigma_i^2 / w_i}{\frac{d^2}{z^2} + \sum N_i \sigma_i^2}$
Total	$\hat{t}_{st} = \sum N_i \bar{y}_i$	$Var(\hat{t}_{st}) = \sum N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \frac{s_i^2}{n_i}$	$n = \frac{\sum N_i^2 \sigma_i^2 / w_i}{\frac{d^2}{z^2} + \sum_{i=1}^L N_i \sigma_i^2}$
Proportion	$\hat{p}_{st} = \frac{1}{N} \sum N_i \hat{p}_i$	$Var(\hat{p}_{st}) = \frac{1}{N^2} \sum N_i (N_i - n_i) \frac{\hat{p}_i \hat{q}_i}{n_i - 1}$	$n = \frac{\sum N_i^2 p_i q_i / w_i}{\frac{d^2}{z^2} + \sum N_i p_i q_i}$
Systematic			
Mean	$\bar{y}_{sy} = \frac{1}{n} \sum_{i=1}^n y_i$	$Var(\bar{y}_{sy}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$	$n = \frac{N\sigma^2}{\frac{d^2}{z^2}(N-1) + \sigma^2}$
Total	$(\hat{t}_{sy}) = N\bar{y}$	$Var(\hat{t}_{sy}) = N^2 Var(\bar{y})$	$n = \frac{N\sigma^2}{\frac{d^2}{z^2 N^2}(N-1) + \sigma^2}$
Proportion	$\hat{p}_{sy} = \frac{1}{n} \sum_{i=1}^n y_i$	$Var(\hat{p}_{sy}) = \frac{\hat{p}_{sy} \hat{q}_{sy}}{n-1} \left(\frac{N-n}{N} \right)$	$n = \frac{Npq}{\frac{d^2}{z^2}(N-1) + pq}$
Cluster			
Mean	$\bar{y} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i}$	$Var(\bar{y}) = \left(\frac{N-n}{NnM^2} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}m_i)^2}{n-1}$	$n = \frac{N\sigma_c^2}{\frac{d^2}{z^2} N\bar{M}^2 + \sigma_c^2}$
Total	$\hat{t} = M\bar{y}$	$Var(\hat{t}) = M^2 Var(\bar{y})$	$n = \frac{N\sigma_c^2}{\frac{d^2}{z^2 N} + \sigma_c^2}$
Proportion	$\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i}$	$Var(\hat{p}) = \left(\frac{N-n}{NnM^2} \right) \frac{\sum_{i=1}^n (a_i - \hat{p}m_i)^2}{n-1}$	$n = \frac{N\sigma_c^2}{\frac{d^2}{z^2} N\bar{M}^2 + \sigma_c^2}$
Stratified	Allocation		
	Proportional	Fixed cost c_i or variance	Neyman
Mean	$n_i = n \frac{n_i}{N}$	$n_i = n \frac{N_i \sigma_i / c_i}{\sum_{j=1}^L N_j \sigma_j / \sqrt{c_j}}$	$n_i = n \frac{N_i \sigma_i}{\sum_{j=1}^L N_j \sigma_j}$
proportion	$n_i = n \frac{n_i}{N}$	$n_i = n \frac{N_i \sqrt{p_i q_i} / \sqrt{c_i}}{\sum_{j=1}^L N_j \sqrt{p_j q_j} / \sqrt{c_j}}$	$n_i = n \frac{N_i \sqrt{p_i q_i}}{\sum_{j=1}^L N_j \sqrt{p_j q_j}}$