

SMA2205

**NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY**  
**DEPARTMENT OF APPLIED MATHEMATICS**  
**B.Sc. PART II HONOURS EXAMINATIONS 2002**

**SMA 2205 –STATISTICS III**

**APRIL/ MAY 2002**

3 HOURS (100 Marks)

This paper has 4 pages

Answer **FOUR** questions: Question 1 in **SECTION A** (28 Marks) and **THREE** other questions from **SECTION B** (24 Marks each). Where a question contains subdivisions, the mark value of each subdivision is indicated in brackets.

Candidates are advised not to spend more than one hour on Question 1. Cordless calculators may be used. Candidates are allowed to bring into the examination a list of formulae approved by the examiner. **GOOD LUCK!**

**SECTION A (COMPULSORY)**

1. (a) Define the following as they relate to sample surveys and give an example in each case:
- i) Target population (2 marks)
  - (ii) Study population (2 marks)
  - (iii) Sampling units (2 marks)
  - (iv) Sampling frame (2 marks)
- (b) In simple random sampling we normally draw one unit at a time without replacement from a finite population of size  $N$  using random digits until our required sample of size  $n$  is achieved.
- (i) Show that the probability of obtaining such a sample is the same as the probability of selecting any one of  $\binom{N}{n}$  possible samples of size  $n$  from a population size of  $N$ . (4 marks)

- (c) From a list of 25 000 names and addresses of registered voters in a certain constituency a simple random sample of 300 names showed on investigation 42 ghost (deceased) registered voters. Estimate the total number of ghost voters in the constituency. Compute the standard error of this estimate. (6 marks)
- (d) In a private library the books are kept on 130 shelves of similar size. The number of books on 15 shelves picked at random are found to be 28, 23, 25, 33, 31, 18, 22, 29, 30, 22, 26, 20, 21, 28, 25.
- (i) Estimate the total number,  $Y_T$ , of books in the library. Compute the standard error of this estimate. (4 marks)
- (ii) Suppose the estimate in (i) is not accurate enough; we want to be 95% certain that a simple random sample estimate of  $Y_T$  is within 100 of the true value of  $Y_T$ . How many shelves should be included in the sample? (4 marks)

**SECTION B (Answer any THREE questions in this section)**

2. All the farms in a certain province in the country are stratified by farm size and the mean number of hectares of wheat per farm in each stratum with the following results:

Farm Size (ha)	No. of Farms	Mean Wheat (ha)	Std Deviation
0 – 20	368	2.7	2.1
21 – 40	425	8.1	3.6
41 – 60	389	12.1	3.9
61 – 80	316	16.9	5.1
81 – 100	174	20.8	6.1
101 – 120	98	25.2	6.5
$\geq 121$	138	31.8	9.1

- (a) Compute the overall population mean hectareage of wheat  $\bar{Y}$  and population variance  $S^2$ . (6 marks)
- (b) For a sample of 100 farms compute the sample size in each stratum under stratified simple random sampling with:
- (i) proportional allocation, and (5 marks)
- (ii) Neyman allocation. (5 marks)

SMA2205

3. The following finite population contains three strata with  $N_h = 4$  and  $n_h = 2$  in each stratum. The variable of interest is ( $Y$ ), and ( $X$ ) is an auxiliary variable.

	STRATUM					
	I		II		III	
	Y	X	Y	X	Y	X
	2	2	2	1	3	1
	3	4	5	4	7	3
	4	6	9	8	9	4
	11	20	24	23	25	12
Totals	20	32	40	36	44	20
$R_h$	0.625		1.111		2.200	

- (a) Compute the variance of the estimate of the population total under each of the following schemes:
- (i) Simple random sampling; (2 marks)
- (ii) Stratified simple random sampling; (4 marks)
- (iii) Combined ratio: stratified random sampling; (6 marks)
- (iv) Separate ratio: stratified random sampling. (6 marks)
- (b) Compare the efficiencies of (ii), (iii) and (iv) with that of (i). Give reasons why the most efficient has turned out to be so in this particular case. (6 marks)
4. In studying lung function in a group of 560 workers in a coal mine, an estimate was required of the mean value of some relevant measure  $Y$ . A simple random sample of 10 workers was chosen and their  $Y$  values,  $y_i$ , determined by an appropriate test. A note was also made of the heights,  $x_i$ , of the 10 workers. The results were as follows:

$y_i$	3.0	3.5	3.3	3.1	4.1	3.2	3.7	2.9	3.9	3.4
$x_i$ (cm)	173	183	170	175	160	157	168	180	178	163

From routine medical records the average height for the group of 560 mine workers is known to be 173.2 cm.

- (a) Estimate the population mean,  $\bar{Y}$ , from the data using the following:

- (i) the ratio estimator; (9 marks)
- (ii) the linear regression estimator. (9 marks)

5. (a) Discuss fully, giving examples where possible, the fundamental principles of designing a postal questionnaire. (12 marks)
- (b) Guided by the principles you have discussed in (a) design a postal questionnaire whose main objective is to obtain information on the prevalence and causes of HOMOSEXUALISM in Zimbabwe. (12 marks)
6. Write short notes on each of the following:
- (a) Reasons for stratification in sample surveys. (8 marks)
- (b) Systematic sampling. (8 marks)
- (c) Sources of errors in sample surveys. (8 marks)

-----\*\*\* END OF QUESTION PAPER \*\*\*-----