NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 1 EXAMINATION 2022-2023

MH4511 – SAMPLING AND SURVEY

November 2022 Time Allowed: 2 hours

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **FOUR (4)** questions and comprises **SIX (6)** printed pages, inclusive of an appendix on Page 6.

2. Answer **ALL** questions. The marks for each question are indicated at the beginning of each question.

3. Answer each question beginning on a **FRESH** page of the answer book.

4. This is a **RESTRICTED OPEN BOOK** exam. You are only allowed to bring in **ONE DOUBLE-SIDED A4-SIZE REFERENCE SHEET WITH TEXTS HANDWRITTEN OR TYPED ON THE A4 PAPER** (no sticky notes/post-it notes on the reference sheet).

5. Candidates may use calculators. However, they should write down systematically the steps in the workings.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Question 1. **(20 marks)**

A forester wants to estimate the total number of farm acres planted in trees for a country. Because the number of acres of trees varies considerably with the size of the farm, he decides to stratify on farm sizes.
The 240 farms in the country are placed in one of four categories according to size. A stratified random sample of 40 farms, selected by using (approximately) proportional allocation, yields the results shown in the accompanying table on number of acres planted in trees.

	H_h	n_h	$\bar{y}_h$	s_h^2
Stratum I (0-200 acres)	86	14	65	32
Stratum II (200-400 acres)	72	12	175	96
Stratum III (400-600 acres)	52	9	340	130
Stratum IV (over 600 acres)	30	5	470	270

a) Estimate the total number of acres of trees on farms in the country and determine the standard error of your estimate.
b) The study above is to be made yearly, with the bound on the error of estimation of 5000 acres. Ignoring the stratum finite population correction factor, find the approximate sample size for each stratum to achieve this bound if Neyman allocation is used.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Question 2. **(30 marks)**

A housing agent is interested to know the relative change over a two-year period in the assessed value of housing apartments in a given community. He took a simple random sample of 20 apartments from the 1000 total apartments in the community. He obtained the values (in \$100K) for this year (y_i) and the corresponding values from two years ago (x_i) for each of the 20 apartments included in the sample. Some summary statistics are given below.

$$n = 20, \quad \sum_{i=1}^{n} x_i = 154.5, \quad \sum_{i=1}^{n} y_i = 164.7 \ ,$$

$$\sum_{i=1}^{n} {x_i}^2 = 1210.55, \quad \sum_{i=1}^{n} {y_i}^2 = 1373.71 \ , \quad \sum_{i=1}^{n} x_i y_i = 1288.95.$$

a) Use an SRS estimator with only data from y_i, to estimate the average value of all the apartments in the community this year with a 95% confidence interval.

b) Use a ratio estimator to estimate the average value of all the apartments in the community this year with a 95% confidence interval. Assume that the average values of all the apartments two years ago was $\bar{x}_U = 7.9$.
[Hint: you may use the following identity.]

$$\sum_{i=1}^{n} \left(y_i - \hat{B}x_i\right)^2 = \sum_{i=1}^{n} {y_i}^2 - 2\hat{B}\sum_{i=1}^{n} x_i y_i + \hat{B}^2 \sum_{i=1}^{n} {x_i}^2$$

c) Comparing the variances of the two estimators above, how much improvement has been achieved by using the ratio estimator over the SRS estimator?

d) What would be an appropriate sample size for using the ratio estimator, if a margin of error of 0.07 is required for the estimation of the average value of all the apartments in the community this year?

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Question 3. **(30 marks)**

a) A large firm has its equipment inventories listed separately by department. From the 15 departments in the firm, 5 are randomly sampled and a simple random sample of items from each selected department are provided to an auditor, who will check to make sure that all equipment is properly identified and located. The proportion of equipment items not properly identified is of interest to the auditor. The data are given in the table below.

Department	Number of items in the Department	Number of Items Sampled	Number Items not properly Identified
1	60	15	3
2	90	20	1
3	50	10	1
4	80	30	3
5	70	25	2

Find the ratio estimate of the proportion of equipment items in the firm that are not properly identified, with its standard error.

b) Town council officials are concerned about the nuisance caused by pigeons around a hawker centre. They hire a team of investigators to estimate the number of pigeons occupying the hawker centre. With several different traps, a sample of ninety (90) pigeons is captured, tagged, and released. One week later, the process is repeated, with sixty (60) pigeons re-captured. Suppose eighteen (18) tagged pigeons are observed in the second sample.

i. Use the capture and re-capture method to estimate the number of pigeons occupying the hawker centre and its standard error.

ii. What are the assumptions needed to ensure the validity of this estimation?

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Question 4. **(20 marks)**

A forester wants to estimate the total number of trees in a certain county that are infested with a particular disease. There are ten well-defined forest areas in the county. These areas can be subdivided into plots of approximately the same size. The number of plots for each area is listed in the table below.

Area	1	2	3	4	5	6	7	8	9	10
Number of Plots	12	10	15	16	11	17	14	25	21	13

Four crews are available to conduct the survey, which must be completed in 1 day. Hence, four areas are chosen, with six plots randomly selected from each area. The data are given in the table below.

Sampled Hospital	Number of Plots Sampled	Average Number of Infected Trees per Plot	Variance of the Number of Infected Trees
1	6	13.0	13.2
3	6	7.0	5.6
7	6	11.5	5.9
9	6	4.0	5.2

a) What are the primary and secondary sampling units in this survey?
b) Suppose the four selected areas were selected using an SRS scheme. Find the unbiased estimate of the total number of infected trees in the county, along with its standard error.
c) Suppose the Lahiri's method was used to choose the 4 selected areas with probabilities proportional to the number of plots in the area and with replacement. Find the estimated total number of infected trees in the county, along with its standard error.

END OF PAPER

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Appendix

- Normal distribution

Commonly used z_α where $P(Z > z_\alpha) = \alpha$:

$z_{0.05}$	$= 1.645$	$z_{0.10}$	$= 1.283$
$z_{0.025}$	$= 1.960$	$z_{0.01}$	$= 2.328$
$z_{0.005}$	$= 2.575$	$z_{0.02}$	$= 2.054$

- T-distribution

	Area in Right Tail								
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

MH4511 SAMPLING & SURVEY

Please read the following instructions carefully:

1. **Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**

2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.

3. Please write your Matriculation Number on the front of the answer book.

4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library