NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 1 EXAMINATION 2023-2024

MH4511 – SAMPLING AND SURVEY

November 2023 Time Allowed: 2 hours

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **FIVE (5)** questions and comprises **SEVEN (7)** printed pages, inclusive of an appendix on Page 7.

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. **One double-sided A4-size reference sheet with texts handwritten or typed on the A4 paper without any attachments (e.g. sticky notes, post-it notes, gluing or stapling of additional papers) allowed in the examination hall**.

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 researcher is interested in conducting a survey to understand the amount ($) spent on delivered food per week per household in a certain estate. The estate has 10,000 households, of which 1,500 are houses (H), 2,500 are condominiums (C), and 6,000 are apartments (A). He believes that the standard deviations of the variable of interest for the tree types of households are of the relation: $S_H = 2S_C = 4S_A$. The costs for conducting an interview per household are \$50, \$30 and \$10 for houses, condominiums and apartments, respectively.

(a) What are the sampling unit and sampling frame in this survey?
(b) What are the strata in this survey?
(c) Determine the allocation proportions for the three strata, based on the optimal allocation approach.
(d) Suppose a total of \$8,000 is planned for all interviews in this survey (not including set up cost). How many households should be interviewed in each stratum?
(e) Determine the variance of the mean amount spent on delivered food per week per household for this estate when the optimal allocation approach is used and $S_A = \$100$.

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

Question 2. **(25 marks)**

A corporation is interested in estimating the total earnings (in $1,000) from the sales of certain laptop model at the end of 3-month period. The earnings figures are available for all district offices within the corporation for the corresponding 3-month period of the previous year, with an overall total of 124,800 (x $1,000). A simple random sample of 13 district offices is selected from the 120 offices within the corporation. The following table shows the data with basic summary statistics.

Office	3-mth data from previous year	3-mth data from current year
1	550	610
2	690	780
3	1400	1600
4	1020	1030
5	620	630
6	980	1050
7	880	950
8	1200	1440
9	1350	1570
10	1650	2019
11	670	980
12	720	880
13	1530	1710
Mean	**1020**	**1173**
S.D.	**372.63**	**446.28**

(a) Use an SRS estimator with only data from current year to estimate the total sales of all the offices in the corporation in the current 3-month period with a 95% confidence interval.

(b) Use a ratio estimator to estimate the total sales of all the offices in the corporation in the current 3-month period with a 95% confidence interval.

(c) Explain why the point estimate from the ratio estimator is larger than that from SRS estimator.

(d) The study above is to be made quarterly using the ratio estimator, with a 95% confidence bound on the margin of error for the total estimate equal to 2,500 (x $1,000). Find the approximate sample size required to achieve this error bound.

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

Question 3. **(25 marks)**

To improve the efficiency of the work environment, an executive of a certain company wants to estimate the number of days on medical leave taken by secretaries in the company for a one-year period. The company has 750 secretaries in 25 departments. Each department employs some secretaries, and the number of days on medical leave varies considerably from secretary to secretary. The executive decides to employ a two-stage cluster sampling scheme, using a small number of departments, and randomly selecting approximately 20% of secretaries from each. Five departments are randomly sampled. The data are summarized in the following table.

Department (i)	Number of Secretaries	Number of Secretaries Sampled	Sample Mean $\bar{y}_i$	Sample Variance s_i^2
1	45	9	10	20
2	36	7	9	16
3	20	4	7	22
4	21	4	8	26
5	28	6	13	12

(a) What are the primary and secondary sampling units in this estimation project?

(b) What is the approximate probability that a particular secretary in the company is being selected in the sample?

(c) What is the approximate probability that secretary i and secretary j from different departments both being selected in the sample?

(d) Do you think the event that secretary i is in the sample, and the event that secretary j from a different department is in the sample are independent? Justify your answer.

(e) Based on the data given in the table, use a ratio estimator to estimate the average number of days on medical leave taken by the secretaries of the company, with a standard error.

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

Question 4. **(20 marks)**

From the eight hospitals in a city, a researcher wants to sample 4 hospitals for the purpose of estimating the average number of days patients with a certain respiratory condition stayed in the hospital. The number of patients with the respiratory condition for each hospital is listed in the table below.

Hospital	1	2	3	4	5	6	7	8
Number of Patients with the Condition	12	10	35	16	11	19	14	25

Four research assistants were sent to the selected hospitals to gather information on 6 randomly selected patients within the hospital. The data are given in the table below.

Sampled Hospital	Number of Patients Sampled	Average Number of Days Stayed in Hospital	Variance of the Number of Days
1	6	13.0	13.2
3	6	7.0	5.6
6	6	11.5	5.9
8	6	6.0	5.2

(a) What are the primary and secondary sampling units in this survey?

(b) Suppose the four selected hospitals were selected using an SRS scheme. Find the unbiased estimate of the average number of days patients with the respiratory condition stayed in the hospital, along with its standard error.

(c) Suppose the Lahiri's method was used to choose the 4 selected hospitals with probabilities proportional to the number of patients with the respiratory condition in the hospital, and with replacement. Find the estimated average number of days patients with the respiratory condition stayed in the hospital, along with its standard error.

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

Question 5. **(10 marks)**

In a free concert organised for babies on an open field, the organiser wants to estimate the number of baby attendees. At the beginning of the concert, 200 event balloons were randomly given out to baby attendees. During the concert interval, 100 baby attendees are randomly sampled, and find that 25 have the concert balloon.

(a) Use the capture and re-capture method to estimate the number of baby attendees at the concert, and its standard error.
(b) What are the assumptions needed to ensure the validity of this estimation?

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						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060

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