NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 1 EXAMINATION 2024-2025

MH4511 – SAMPLING AND SURVEY

November 2024 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 pilot study was conducted to understand whether the citizens agree to get a Mpox vaccine in a certain region with a large number of citizens. Suppose that 75% of the residents can be contacted by phone, 15% by email and 10% by face-to-face interview. A researcher took a survey that included 250 responders. The following table shows the summary statistics.

Interview Method	Sample Size	Number of citizens agree to get a vaccine
By Phone	100	45
By Email	100	30
Face-to-face	50	10
Total	250	85

(a) Assume that all the 250 responders were selected using simple random sampling from the entire population in the region. What would be a 95% confidence interval for the proportion of the citizens who would agree to get a vaccine shot?

(b) Suppose the responders were selected using a stratified sampling by interview method, estimate the proportion of the citizens who would agree to get a vaccine shot with a standard error.

(c) Suppose the cost of getting an observation from face-to-face interview is 4 times that of using phone or email (in the ratio of 1:1:4 for phone, email and face-to-face, respectively), what would be the appropriate allocation ratios of the sample size for the three strata if a new study is to be conducted in the same region?

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

Question 2. **(20 marks)**

The number of diseased trees in a certain forest is of interest to the management. Due to the heavy resources required to identify those trees, the forest area (a total of 16800 units) is divided into 120 unequal sized strip plots. A random sample of 5 plots was selected. The plot area and the number of diseased trees in the plot are listed in the table below.

Plot	Area of Plot (unit)	Number of Diseased Trees
1	110	8
2	120	22
3	130	36
4	140	45
5	150	84
Mean	**130**	**39**
S.D.	**15.81**	**28.81**

(a) Use an SRS estimator with only data for the number of diseased trees to estimate the total number of diseased trees in the forest with a 95% confidence interval.

(b) Use a ratio estimator to estimate the total number of diseased trees in the forest with a 95% confidence interval.

(c) Explain why the point estimate of the ratio estimator is different from that of SRS estimator.

(d) With a 95% confidence bound on the margin of error for the total estimate equal to 900, find the approximate sample size required for the ratio estimator 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)**

As part of the effort to understand the impact of social media on Junior College (JC) students, a sociologist wants to estimate the average number of hours a JC student spent on social media per day. There are 20 Junior Colleges with a total of 25,000 students.

The number of hours spent on social media varies considerably from student to student within JC, but the mean difference between JCs is believed to be small. Hence, the sociologist decides to employ a two-stage cluster sampling scheme, using a small number of JCs, and randomly selecting approximately 10% of students from each JC. Five JCs are randomly sampled. The data are summarized in the following table.

Junior College (i)	Number of Students	Number of Students Sampled	Sample Mean $\bar{y}_i$	Sample Variance s_i^2
1	1050	100	4.7	20
2	1250	120	5.2	16
3	800	80	3.9	22
4	1090	110	6.1	26
5	1100	110	4.3	12

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

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

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

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

(e) Based on the data given in the table, use an unbiased estimator to estimate the average number of hours spent on social media per day by JC students, with a standard error.

(f) Based on the data given in the table, use a ratio estimator to estimate the average number of hours spent on social media per day by JC students, with a standard error.

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

Question 4. **(20 marks)**

A dog lover likes to know the proportion of households that has a pet dog in her neighbourhood. There are 10 blocks of households in her neighbourhood, the number of households are listed below.

Block #	1	2	3	4	5	6	7	8	9	10
Number of Households	320	175	255	130	200	240	125	135	250	170

A sample of the 4 blocks was selected, and a simple random sample (SRS) of households from each selected block were checked to see whether there is a pet dog in the house.

Sampled Block	Number of Households Sampled	Number of Households with a Pet Dog
1	60	27
4	28	7
6	45	18
9	50	23

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

(b) Suppose the 4 selected blocks above were selected using an SRS scheme. Find the unbiased estimate of the proportion of households that has a pet dog, along with its standard error.

(c) Suppose the Lahiri's method was used to choose the 4 selected blocks with probabilities proportional to the number of households in the block with replacement. Find the estimated proportion of households that has a pet dog, along with its standard error.

(d) Explain why the unbiased estimate for the proportion in Part (b) is larger than that in Part (c).

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

Question 5. **(15 marks)**

In a survey to understand the usage of mobile devices, the number of housing units, the number of residents, and the total number of mobile devices within the block for a random sample of 5 blocks was selected from a small city with 25 blocks are reported in the table below.

Block	Number of Housing Units	Number of Residents	Number of Mobile Devices
1	12	40	76
2	14	38	73
3	8	24	40
4	24	56	120
5	12	35	77
Total	70	193	386

Estimate the average number of mobile devices per resident in the city. What is the standard error of your estimate?

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

CONFIDENTIAL

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