MH4511 Sampling & Survey

Tutorial 8 AY2025/26 Semester 1

Problem 8.1

An economic survey is designed to estimate the total amount spent on utilities for households in a small housing estate. Because no list of households is available, cluster sampling is used, with divisions (blocks) forming the clusters. The survey decides to sample blocks within the estate, and then to sample households within the block. A simple random sample without-replacement of 5 blocks is selected from the 20 blocks of the estate. Households within the sampled blocks are sampled using SRS without-replacement. The data are summarized in the following table.

Block i	Number of Households	Number of Households Sampled	Sample Mean of Cost for Block i	Estimated Total Cost for Block i	Sample Variance of Cost for Block i
1	60	12	101	6,060	25
2	60	12	72	4,320	18
3	52	10	121	6,292	26
4	78	15	99	7,722	20
5	55	11	170	9,350	29
Total	**305**	**60**		**33,744**	

We also have $s_r^2 = 213{,}000$ and $s_t^2 = 2{,}000{,}000$.

a) What are the primary and secondary units in this survey?
b) State the sampling weight for each of the secondary sampling unit of the primary sampling units for this sample.
c) Based on the information given, estimate the total amount spent on utilities for households in this estate and calculate the standard error for your estimate.
d) Based on the information given, estimate the mean amount spent on utilities per households in this estate and calculate the standard error for your estimate.

Problem 8.2

A food manufacturer has 90 plants located throughout the country and wants to estimate the average number of hours that the manufacturing machines were down for repairs in the past months. He knows that he has a combined total of 4500 machines in all plant. Because the plants are widely scattered, and each plant contains many machines, and checking the repair record for each machine would be time consuming. So, with the available time and money, he decides to randomly sample 6 plants and approximately 20% of the machines in each plant. The summary data are shown in the table below.

Plant Index	Number of Machines	Sampled Number of Machines	Average Downtime	Sample Variance
1	50	10	5.40	11.40
2	65	13	4.00	10.70
3	45	9	5.70	16.80
4	48	10	4.80	13.30
5	52	10	1.30	11.20
6	58	12	3.80	15.00

a) What are the primary and secondary sampling units in this estimation project?
b) Using an unbiased estimator, estimate the average downtime per machine and calculate the standard error of your estimate. (We may assume that the estimated between cluster variance of the cluster total is $s_t^2 = 768$.)

Problem 8.3

An auditor wanted to sample sick leave records of a large firm in order to estimate the average number of days of sick leave per employee over the past quarter. The firm has eight divisions, with varying numbers of employees per division. Because number of days of sick leave used within each division should be highly correlated with the number of employees, the auditor decided to sample $n = 3$ divisions with probabilities proportional to number of employees.

Division	Number of Employees	Total Number of Sick Leave for Sampled Divisions
1	1200	--
2	450	--
3	2100	4320
4	860	--
5	2840	--
6	1910	4160
7	390	--
8	3200	5790
Total	12,950	

Suppose the three sampled divisions were $3, 6 \text{ and } 8$, and the total number of sick days are given in the above table.

a) Use the SRS estimator to estimate the average number of sick days used per person for the entire firm and find its standard error.
b) Use the ratio estimator to estimate the average number of sick days used per person for the entire firm and find its standard error.
c) Use the sampling with unequal probability estimator to estimate the average number of sick days used per person for the entire firm and find its standard error.