STAT 521: Homework Assignment 4

Due on April 4, 2024

Problem 1: (20 pt)

A city block is divided into 100 blocks from which 5 blocks are selected with replacement and with probability proportional to the number of households enumerated in a previous census. Within each sampled block, the average household income and the average household size (=number of people in the household) are obtained from the sampled blocks. The following table presents a summary of information obtained from the sample blocks.

Table 1: Summary of information obtained from the sampled households

Block	Block Size	Average Household income $(\times 10^{-3}\$)$	Average Household size
1	50	30	2
2	60	70	4
3	47	80	5
4	50	50	4
5	70	60	4

1. What is the estimated average household income and its estimated variance?

2. What is the estimated per capita income (= income per person) and its estimated variance? (You may need to use a Taylor linearization.)

Problem 2:

Suppose that we have a population of clusters with equal size M. Suppose that the population has the following ANOVA structure as summarized in the folloiwng table.

Table 2: ANOVA table

Source	d.f.	Mean Sum of Square
Between Clusters	49	6,218
Within Clusters	450	2,918

1. Find the cluster size M.

2. Compute the intracluster correlation coefficient.

3. What is the variance of the mean estimator under this cluster sampling?

4. Compute the design effect of this sampling design and give an interpretation.

Hint for (3): You can ignore n_I/N_I term. That is, assume that $n_I/N_I = 0$.

Problem 3: (20 pt) A statistician wishes to carry out a survey on the quality of health care in the cardiology service of hospitals. For that, he selects by simple random sampling of $n = 100$ hospitals among the $N = 1,000$ hospitals listed and then, in each of the selected hospitals, he collects the opinions of all the cardiology patients.

1. We consider that each cardiology unit is comprised of exactly $M = 50$ beds and that the 95% confidence interval on the true proportion P of dissatisfied patient is:

$$P \in [0.10 \pm 0.018],$$

   (that signifies in particular that, in the sample, 10% of patients are dissatisfied with the quality of care). How do you estimate the intracluster correlation coefficient ?

2. How would the accuracy of the statistician's survey on satisfaction evolve if, all at once, there are $M = 25$ beds and $n = 200$ hospitals are selected in the sample using the same sampling design ?