Estimating the Total Land Value of Port Moody

Yutao Shen (301251469)

Zhengyang Sun (301338540)

Linwan Xu (301243591)

STAT 410

Sampling Project

Simon Fraser University

Table of contents

Introduction

In 1989, actress Kim Basinger and her partners paid $20 million for the town of Braselton Georgia, and turned it into a tourist attraction (Wikipedia). However, after five years, for reasons of bankruptcy, Kim Basinger and her partners transferred this town for a million dollars (Wikipedia). In just five years, the land value of a small town has shrunk 20 times. Nowadays, how to understand the real land value of an area? Therefore, our group assumes that a wealthy businessman wants to buy all the land of Port Moody, and predicts how much money the businessman will spend to accomplish this magnificent feat. Our group firstly determines the data to be selected and carries out a data collection. Then, we use Two-Stage Cluster Sampling to analyze the data and find out the variance and confidence interval (CI) of the data. Moreover, we summarize the difficulties and problems we have encountered and find out various situations that may cause errors. The purpose of our group is to find the total land value of Port Moody, excluding the building costs.

Data selection

Before selecting the specific data, our team firstly determines a series of definitions, such as sampling frame and target population. As shown below:

Target population: lands in Port Moody

Observation unit: lands within a zip code area

Sampling frame: list of all zip codes in Port Moody

Sampling unit: zip codes are on the Sampling frame

Sampled population: lands within the zip code area from the sampling frame

After these definitions, we start to collect data formally. We find the map of Port Moody and divide it into 12 different areas according to the street trend, zone 1 to zone 12 (Figure 1). Then we find the daily garbage collection map of Port Moody, and discover that zone 12 is not in the garbage collection scope (Figure 2) (2020 Waste Collection Schedule). Moreover, according to the BC assessment website, we realize that the land value of zone 12 is visible, so there is no need to estimate at all. Therefore we exclude zone 12 from our population and only take data from zone 1 to 11. After completing the zone selection, we find the postal code (zip code) contained in each street from Dataset Canada. Furthermore, according to the location of each road, 705 postal codes are divided into their areas (Table 1). Then, according to the random selection software, five of these 11 areas are randomly selected (zone 1, 2, 5, 7, 9). In these five zones, we randomly select 25% of the zip code in each zone, and then we find all the land values contained in each zip code according to the BC assessment website and Google map. Moreover, since we only seek the value of the land itself, the apartments are not included in the data we have taken. Whenever the zip code we take contains apartments, we will automatically delete the zip code. Secondly, because the park belongs to public facilities, the land value of the park itself is not high and cannot be sold; therefore, the land value of the apartment is not included in the data we have taken.
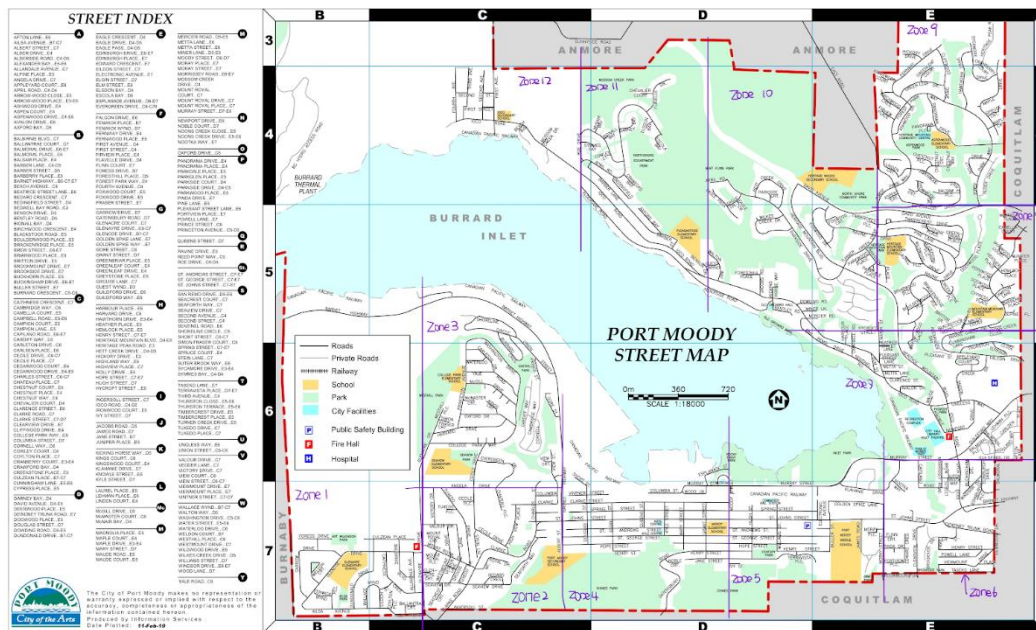
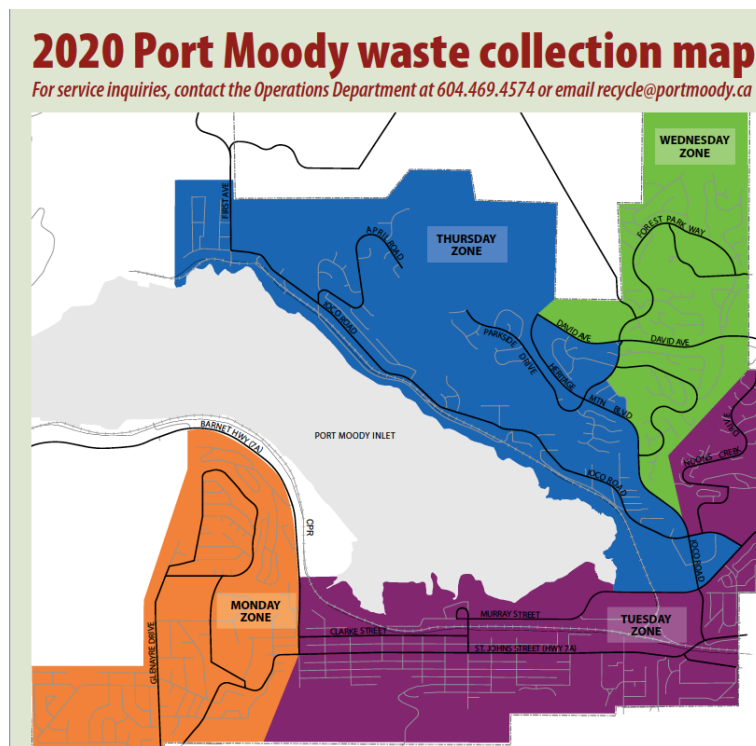**Figure 1:** 12 zones of Port Moody map

Figure 2: Garbage collection map of Port Moody area

Zone	# zip code	# selected zip code	real number of selected zip code (without apartments and parks)
1	57	14	14
2	97	24	18
5	54	13	10
7	89	22	14
9	48	11	8

Table 1: Date of zone and number of zip codes

Methodology

Method: Two-Stage Cluster Sampling

We use the Two-Stage Cluster Sampling to design the survey. A two-stage cluster sample is obtained by first selecting a sample of clusters, and selecting a sample of elements from each sampled cluster (Lohr, 2019). Firstly, we collect the population elements into disjoint subpopulations, called primary sampling units (Lohr, 2019). In the second-stage sampling units (SSU), there may be clusters of elements, for each PSU in the first-stage sample (Lohr, 2019). A sample of SSU is drawn from each PSU, every element in the selected SSU is surveyed.

It can be possible to observe each ssu in the sampled psus, we should to estimate the individuals' psu totals by Formula (1), and then use the Formula (2) to do an unbiased estimator of the population total. Also for the estimate the population total, the variance of $\hat{t}_{unb}$ is combined by two parts: the variance from one-stage cluster sampling (Formula 4) and the variance inside psus (Formula 5).

$$\hat{t}_i = \sum_{j \in \mathcal{S}_i} \frac{M_i}{m_i} y_{ij} = M_i \bar{y}_i \qquad \text{-----(1) (Lohr, 2019)}$$

$$\hat{t}_{unb} = \frac{N}{n} \sum_{i \in \mathcal{S}} \hat{t}_i = \frac{N}{n} \sum_{i \in \mathcal{S}} M_i \bar{y}_i = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}_i} \frac{N}{n} \frac{M_i}{m_i} y_{ij}. \qquad \text{-----(2) (Lohr, 2019)}$$

$$\hat{V}(\hat{t}_{unb}) = N^2 \left(1 - \frac{n}{N}\right) \frac{s_t^2}{n} + \frac{N}{n} \sum_{i \in \mathcal{S}} \left(1 - \frac{m_i}{M_i}\right) M_i^2 \frac{s_i^2}{m_i}.$$

-----(3) (Lohr, 2019)

$$s_t^2 = \frac{1}{n-1} \sum_{i \in \mathcal{S}} \left(\hat{t}_i - \frac{\hat{t}_{unb}}{N}\right)^2$$

-----(4) (Lohr, 2019)

$$s_i^2 = \frac{1}{m_i - 1} \sum_{j \in \mathcal{S}_i} (y_{ij} - \bar{y}_i)^2$$

-----(5) (Lohr, 2019)

Result

After collecting the data, we construct an Excel file, and put all information in it which includes zone ID, total zone (cluster) number N, selected zone(cluster) number n, total number of zip code in each zone M, selected number of zip code in each zone m, as well as the corresponding land value y, the selected zip codes with their corresponding land values are shown in the Appendix. We first estimate population total t_i for each of the cluster by using Formula (1); we have $\hat{t}_1 = 421592764.3$, $\hat{t}_2 = 1086394611$, $\hat{t}_3 = 1426099500$, $\hat{t}_4 = 605927257.1$, $\hat{t}_5 = 1512768000$; then from the population total we get an unbiased estimator of population total tunb is equal to 11116120692 by the Formula (2). We would like to calculate the variance of $\hat{t}_{unb}$ by using Formula (3); noticed that there are S_t^2 and S_i^2 in this formula we need to calculate first; to obtain S_t^2, we use Formula (4), then we get result of S_t^2 is 1.72258E+17; Using Formula (5) we get S_i^2 for each of the cluster; these are $S_1^2 = 23097802765550$, $S_2^2 = 201862499937908$, $S_3^2 = 635397039958333$, $S_4^2 = 3012601488352$, and $S_5^2 = 117395457428571$. Plugging these numbers to Formula (3), we get the estimated variance of the unbiased population total is

2.89748E+18; thus square root the variance, we have the standard error of the unbiased

total population is 1702199236. Finally, we have a 95% confidence interval(CI)

(9413921455, 12818319928) by using $\hat{t}_{unb} \pm 1.96 se(\hat{t}_{unb})$

$\hat{t}_i$	Si^2	St^2			
421592764.3	2.30978E+13	1.72E+17		$\hat{t}_{unb}$	11116120692
1086394611	2.01862E+14			$\hat{var}(\hat{t}_{unb})$	2.90E+18
1426099500	6.35397E+14			$SE(\hat{t}_{unb})$	1702199236
605927257.1	3.0126E+12				
1512768000	1.17395E+14			95% CI	(9413921455, 12818319928)

Table 2: Result of Sampling project

Discussion & Conclusion

For our data, oil land (zone 12) value is constant, it does not affect estimated value and

estimated CI. We directly add the oil industry land value (121279000); therefore, our

final total land value in 12 zones is 11237399692, and the CI changes to (9535200455,

12939598928). In this investigation, our group encounters some problems that may affect

the error. Firstly, some zip code areas span two zones; therefore, we can only estimate the

proportion of each zone based on the map area. Moreover, there are some errors between

Google map and the BC assessment map. It is possible that the housing conditions of

some zip code areas are slightly different due to different data collection times. Although

the land value difference between the two is small, these errors may also affect the final

result. Furthermore, some streets contain the same zip code; therefore, we can only delete

some of the same zip codes, resulting in some small errors in the land value of some

postal code areas where having the same postal code area. These three errors affect our

final results. If we can find more detailed maps in the future, our group can further

improve the data to make the final results more perfect. Moreover, if you are a billionaire, you can buy port moody.

Appendix

The Selected Zones and Zip codes

Zone	M_i	m_i	y
1	57	14	4813000, 8933000, 13319000, 13256000, 3863000, 7272000, 9215000, 1100100, 3878000, 4953000, 8953000, 17589000, 5010000, 1395000
2	97	18	5651000, 5939000, 6667000, 6314000, 13675000, 7939000, 4904000, 2589000, 4891000, 8903000, 65043000, 5944000, 13433000, 22738000, 8037000, 7226000, 4216000, 7490000
5	54	10	6773000, 13281000, 5553000, 37792000, 38977500, 41703000, 19658000, 85152000, 8090000, 7113000, 15520000
7	89	14	7028000, 11544000, 8847000, 17055000, 6851000, 575400, 969000, 796000, 7090000, 11121000, 4218000, 1754000, 7076000, 1898000
9	48	8	21203000, 36565000, 21707000, 23253000, 48228000, 45057000, 32465000, 23650000

Reference

2020 Waste Collection Schedule. (n.d.). Retrieved from https://www.portmoody.ca/en/resourcesGeneral/2020-calendar/2020-City-of-Port-Moody-Waste-Collection-Schedule.pdf

BC Assessment. (n.d.). Retrieved from https://www.bcassessment.ca/

Dataset Canada. (n.d.). Retrieved from https://www.geopostcodes.com/Greater_Vancouver?loc=Port Moody

Lohr, S. L. (2019). Sampling: design and analysis (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Wikipedia (2020, April 4). Braselton, Georgia. Retrieved from https://en.wikipedia.org/wiki/Braselton,_Georgia