STAT 344 Group Project:

A Statistical Study on 2019 Vancouver Property Tax

Members:	Contribution
Edward Wang, 88723788	Introduction; SRS writing & calculation; Stratified Samples writing & calculation; Conclusion & Discussion
Xinmeng Liu, 57339483	PART II
Max Mai, 15054570	Stratified Samples writing
Tianjun Wang, 47518667	Stratified Samples writing

Introduction:

For years, the property tax in the Vancouver & policies drawn around its annual increase have always been a subject of debate: whereas some claim that "Vancouver property tax rates are the lowest among major Canadian cities (*WOWA.ca, 2019*)", others complain that such expensive property tax is forcing some businesses to close down permanently (Business in Vancouver, 2020).

To address the conflicting views & better help the future investors to know their costs in terms of property tax expenses, our team dedicates this project to answering two questions:

1. What's the average property tax levied in Vancouver?
2. Is the Vancouver property tax considered expensive comparing to other major Canadian cities like Toronto?

Since the most comprehensive data available to date is published by the city Vancouver city in the year 2019, we will focus our study in the year 2019 with the following 2 parameters of interest:

1. $\overline{y_p}$: the population average of Vancouver property tax in the year 2019
2. $p_{>6326}$: the true proportion of Vancouver property tax levied in 2019 higher than 6326 CAD, the median Toronto property tax levied in the year 2019 ((*WOWA.ca, 2019*).

To give a reasonable & accurate estimate of two parameters, our team will use both a simple random sample & for each parameter, a stratified sample with optimal allocation to obtain our results. In the end, we will also evaluate our estimates & conclude our project with the more accurate set.

A Brief Overview of Data & Information Given Prior to the Sampling:

According to "Property Tax Report 2019", our group has obtained the following information of all property taxes collected the year 2019 (*Property tax report, 2021*):

- $\boldsymbol{N}$: *the total number of recorded tax collection.*
$$\boldsymbol{N} = 209466$$
- $\boldsymbol{N_{land}}$: *the total number of recorded tax collection under legal type* $\boldsymbol{land}$.
$$\boldsymbol{N_{land}} = 88851$$
- $\boldsymbol{N_{other}}$: *the total number of recorded tax collection under legal type* $\boldsymbol{other}$
$$\boldsymbol{N_{other}} = 192$$
- $\boldsymbol{N_{strata}}$: *the total number of recorded tax collection under legal type* $\boldsymbol{Strata}$
$$\boldsymbol{N_{strata}} = 120423$$

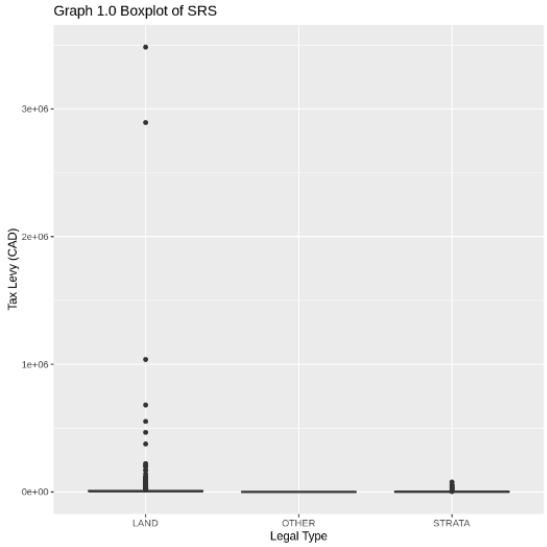
Given that these data are directly from the City of Vancouver which oversees the annual property tax levy, we treat the values listed above as population value of the listed parameters. The following relationship should also be noted:

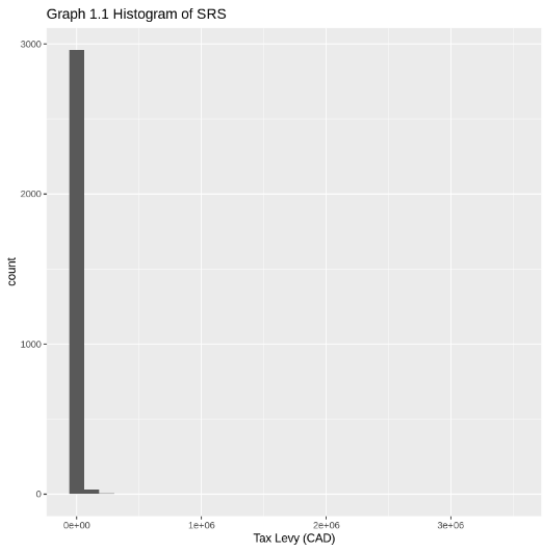
$$N = N_{land} + N_{other} + N_{strata}$$

SRS:

SRS Sampling & Overview

Since we know little about our population except for the information listed above, our team decides to take a **simple random sample** of 3000 observations (n = 3000) from all 2019 property tax collection records to see the patterns of the interested population.





From the above visualisations of the SRS collected, we have noticed that there are extreme values present in our dataset, most notably in the legal type "land" & "strata." The table of the average property tax of each stratum below also supports this view: (Evaluate the mean in the discussion part)

$\overline{y_{SRS,land}}$	19087.022
$\overline{y_{SRS,other}}$	733.000
$\overline{y_{SRS,strata}}$	2831.623

SRS Estimates & Results:

After sampling, we have obtained the following estimates from our SRS:

Estimating $\overline{y_p}$ with $\overline{y_{SRS}}$:

$\overline{y_{SRS}}$	9782.116
$SE(\overline{y_{SRS}})$	1593.182
95% $CI\ of\ \overline{y_{SRS}}$	$(6659.48, 12904.8)$

(Note: all calculations & formulas will be in the **Appendix 1**)

- $\overline{y_{SRS}}$: $the\ SRS\ estimate\ of\ the\ average\ property\ tax\ in\ Vancouver$ 2019

- $SE(\overline{y_{SRS}})$: $the\ standard\ error\ of\ \overline{y_{SRS}}$

In summary, from SRS gathered, we estimate that the average property tax in Vancouver 2019 ($\overline{y_p}$) with our sample average $\overline{y_{SRS}} = 9782.116$. Given that our sample size is large (n > 30), by the Central Limit Theorem, we are 95% confident that the true $\overline{y_p}$ falls in the interval $(6659.48\ , 12904.8)$.

Estimating $p_{>6326}$ with $p_{SRS,>6326}$:

$p_{SRS,>6326}$	0.242
$SE(P_{SRS,>6326})$	0.00777
95% $CI\ of\ p_{SRS,>6326}$	(0.2267 ,0.2572)

- p_{SRS}: $the\ SRS\ proportion\ of\ Vancouver\ property\ tax$ 2019 $larger\ than$ 6,326 CAD
- $SE(p_{SRS})$: $the\ standard\ errror\ of\ p_{SRS}$

Similarly, we estimate the true proportion of Vancouver property tax 2019 larger than 6,326 CAD ($p_{>6326}$) with our sample proportion $p_{SRS,>6326} = 0.242$. Also, given that our sample size is large ($np_{SRS} > 10$ & $n(1 - p_{SRS}) > 10$), by Central Limit Theorem, we are 95% confident that the true population proportion ($p_{>6326}$) will fall in the interval $(0.2267\ ,0.2572)$.

For Optimal Allocation of Stratified Sampling:

For Stratified Sampling with the parameter of interest $\overline{y_p}$

We also get the following values and use them as estimates of $S_{land}, S_{strata}, S_{other}$, which are going to help us to determine the number of observations drawn from each stratum when applying the optimal allocation in the stratified sampling:

$S_{SRS,land}$	133785.561
$S_{SRS,strata}$	4056.390
$S_{SRS,other}$	1036.619

- S_{land}: the population standard deviation of 2019 Vancouver property tax under legal type land.
- S_{other}: the population standard deviation of 2019 Vancouver property tax under legal type other.
- S_{strata}: the population standard deviation of 2019 Vancouver property tax under legal type strata.
- $S_{SRS,land}$: the SRS standard deviation of 2019 Vancouver property tax under legal type land.

- $S_{SRS,other}$: the SRS standard deviation of 2019 Vancouver property tax under legal type other.
- $S_{SRS,strata}$: the SRS standard deviation of 2019 Vancouver property tax under legal type strata.

For Stratified Sampling with the parameter of interest $p_{>6326}$:

Similarly, we use the following values as the estimates of $S_{land,>6326}, S_{strata,>6326}, S_{other,>6326}$

$S_{SRS,land,>6326}$	0.499987697962188
$S_{SRS,strata,>6326}$	0.222989191556536
$S_{SRS,other,>6326}$	0

- $S_{land,>6326}$: the population standard deviation of proportion of 2019 Vancouver property tax greater than 6326 CAD under legal type land.
- $S_{strata,>6326}$: the population standard deviation of proportion of 2019 Vancouver property tax greater than 6326 CAD under legal type strata.
- $S_{other,>6326}$: the population standard deviation of proportion of 2019 Vancouver property tax greater than 6326 CAD under legal type other.
- $S_{SRS,land,>6326}$: the SRS standard deviation of proportion of 2019 Vancouver property tax greater than 6326 CAD under legal type land.
- $S_{SRS,strata,>6326}$: the SRS standard deviation of proportion of 2019 Vancouver property tax greater than 6326 CAD under legal type strata.
- $S_{SRS,other,>6326}$: the SRS standard deviation of proportion of 2019 Vancouver property tax greater than 6326 CAD under legal type other.

Stratified Sampling & Estimates for the Parameter $\overline{y_p}$:

Stratified Sampling & Overview:

Given that each property tax record is assigned with a specific legal type & each type has very different averages, we decide to use legal type as our strata. To obtain the number of samples needed for each stratum (in total the sample size would still be 3000, equal to that of the SRS), we use the optimal allocation method with the values of $S_{SRS,land}, S_{SRS,strata}, S_{SRS,other}$:

n_{land}	2882
n_{strata}	118
n_{other}	0

Stratified Sample Results:

With the sample size in each stratum determined, we obtain the estimates ($\overline{y_{str,land}}, \overline{y_{str,strata}}$) & their associated standard error from the subsamples & combine them to form our stratified estimates.

$\overline{y_{str}}$	8780.30634781997
$SE(\overline{y_{str}})$	790.554977380863
95% $CI\ for\ \overline{y_{str}}$	$(7320.82, 10329.8)$

- $\overline{y_{str}}$: $the\ stratified\ estimate\ of\ the\ average\ property\ tax\ in\ Vancouver$ 2019
- $SE(\overline{y_{str}})$: $standard\ error\ of\ \overline{y_{str}}$

In summary, from the stratified sample, we estimate that the average property tax in Vancouver 2019 ($\bar{y}_p$) with the sample average $\overline{y_{str}} = 8780.31$. Given that our sample size is large (n>30), by the Central Limit Theorem, we are 95% confident that the true $\bar{y}_p$ falls in the interval $(7320.82, 10329.8)$.

Stratified Sampling & Estimates for the Parameter $p_{>6326}$:

Same as Stratified Sampling & Estimates for the Parameter $\overline{y_p}$, here we will also use legal type as our strata, and by applying the optimal allocation method with $S_{SRS,land,>6326}, S_{SRS,strata}, S_{SRS,other,>6326}$, we have calculated the number of samples needed for each stratum:

$n_{land,>6326}$	1870
$n_{other,>6326}$	0
$n_{other,>6326}$	1130

Stratified Sample Results:

With the sample size in each stratum determined, we obtain the estimates ($p_{str,land,>6326}, p_{str,strata>6326}$) & their associated standard error from the subsamples & combine them to form our stratified estimates.

$p_{str,>6326}$	0.237760850250982
$SE(p_{str,>6326})$	0.00590426811187634
95% $CI\ for\ p_{str,>6326},$	$(0.226188, 0.249333)$

- $p_{str,>6326}$: the stratified sample proportion of Vancouver property tax 2019 larger than 6,326 CAD
- $SE(p_{str,>6326})$: the standard error of $p_{str,>6326}$

In conclusion, we estimate the true proportion of Vancouver property tax 2019 larger than 6,326 CAD ($p_{>6326}$) with our sample proportion $p_{str,>6326} =$ 0.2378. Also, given that our sample size is large ($np_{SRS} > 10$ & $n(1 - p_{SRS}) > 10$), by Central Limit Theorem, we are 95% confident that the true population proportion ($p_{>6326}$) will fall in the interval $(0.226188, 0.249333)$.

Conclusion & Discussion:

Parameters \ Sampling Type	STR with optimal Allocation	SRS
$SE\ of\ estimated\ \overline{y_p}$	790.554977380863	733.000
$SE\ of\ estimated\ p_{>6326}$	0.00590426811187634	0.00777

Given that the stratified samples with optimal allocation yield more accurate estimates (with smaller standard error), we decide to use their values as the results of our project. Thus, for parameter $p_{>6326}$, since 0.5 (namely, more than 50% of people paid more than 6326 CAD for their property tax in 2019 Vancouver) is not inside our confidence interval $(0.226188, 0.249333)$, we can safely claim that it is highly unlikely that the Vancouver property tax is more expensive than that in Toronto in 2019. As for the estimated $\overline{y_p}$, using the results from the stratified sample, we claim that we are 95% confident that the true $\overline{y_p}$ should fall side the interval $(7320.82, 10329.8)$. However, at the same time, we worry that in fact average or mean may not be a good estimator to measure individual property tax paid given that properties with different legal types pay very different rates on average.

PART II:

As it is possible to construct competing size alpha tests, the LRT can be uniformly dominated. Thus, some articles have argued that it is the LR criterion that produces the “inferior” statistical procedure, but this is wrong. Instead, this paper argues that it is actually the allegedly superior test flawed. In some cases, the "superior" test provides unnecessary and inappropriate inferences. This manifests itself in the fact that the more powerful or most powerful scale tests required by Neyman-Pearson theory may not be scientifically appropriate. Even if it is true that the LR criterion is not infallible and not universally satisfactory for hypothesis-testing problems, it still serves as a generally reasonable preferred option for non-Bayesian parametric hypothesis testing problems. Therefore, in scientific inquiry, we should pay equal attention to “intuition”, which is sometimes as important as size, power, and unbiasedness. Combined with the paper and daily statistical learning experience, I believe that the implement of statistics is constantly looking for better statistic procedure and method ,instead of just focusing on test and its properties such as unbiasedness and power.

Appendix 1: Calculation Steps & Formulas Used

SRS Estimates:

Estimate of $\overline{y_p}$ with $\overline{y_{SRS}}$:

$$\overline{y_{SRS}} = \frac{\sum_{i=1}^{3000} y_i}{3000} = 9782.116, y_i: property\ tax\ of\ a\ certain\ observation\ of\ SRS$$

$$SE(\overline{y_{SRS}}) = \sqrt{(1 - \frac{n}{N}) * \frac{S_{SRS}^2}{n}} = 1593.182, S_{SRS}: sample\ standard\ deviation\ of\ property\ tax$$

$$95\%\ CI\ for\ \overline{y_{SRS}} = \overline{y_{SRS}} \pm 1.96 * SE(\overline{y_{SRS}}) = (6659.48\ , 12904.8)$$

Estimate of $p_{>6326}$ with $p_{SRS,>6326}$:

$$p_{SRS,>6326} = \frac{\#\ of\ observations\ with\ property\ tax\ larger\ than\ 6{,}326\ CAD}{total\ number\ of\ observations} = 0.242$$

$$SE(P_{SRS}, > 6326) = \sqrt{\frac{p_{SRS}(1-p_{SRS})}{n} * (1 - \frac{n}{N})} = 0.00777$$

$$95\%\ CI\ of\ p_{SRS} = p_{SRS} \pm 1.96 * SE(p_{SRS}) = (0.2267\ ,0.2572)$$

Stratified Sampling & Estimates for the Parameter $\overline{y_p}$:

$$n_{land}: n_{strata}: n_{other} = N_{land} * S_{SRS,LAND}: N_{strata} * S_{SRS,strata}: N_{other} * S_{SRS,other}$$

$$n_{land} = 3000 * \frac{N_{land} * S_{SRS,LAND}}{N_{land} * S_{SRS,LAND} + N_{strata} * S_{SRS,strata} + N_{other} * S_{SRS,other}} = 2881.54 \approx 2882$$

$$n_{strata} = 3000 * \frac{N_{strata} * S_{SRS,strata}}{N_{land} * S_{SRS,LAND} + N_{strata} * S_{SRS,strata} + N_{other} * S_{SRS,other}} = 118.414 \approx 118$$

$$n_{other} = 3000 * \frac{N_{other} * S_{SRS,other}}{N_{land} * S_{SRS,LAND} + N_{strata} * S_{SRS,strata} + N_{other} * S_{SRS,other}} = 0.048247 \approx 0$$

Stratified Sample Results:

$\overline{y_{str,land}}$	17342.3168459403
$SE(\overline{y_{str,land}})$	1838.7310716822
$\overline{y_{str,strata}}$	2477.04720338983
$SE(\overline{y_{str,strata}})$	224.476100277419

- $\overline{y_{str,land}}$: the stratified estimate of the average property tax in Vancouver 2019 under legal type land.
- $SE(\overline{y_{str,land}})$: standard error of $\overline{y_{str,land}}$
- $\overline{y_{str,strata}}$: the stratified estimate of the average property tax in Vancouver 2019 under legal type strata.
- $SE(\overline{y_{str,strata}})$: standard error of $\overline{y_{str,strata}}$

$$\overline{y_{str}} = \frac{N_{land}}{N} * \overline{y_{str,land}} + \frac{N_{strata}}{N} * \overline{y_{str,strata}} = 8780.30634781997$$

$$SE(\overline{y_{str}}) = \sqrt{\left(\frac{N_{land}}{N}\right)^2 * SE\left(\overline{y_{str,land}}\right)^2 + \left(\frac{N_{strata}}{N}\right)^2 * SE\left(\overline{y_{str,strata}}\right)^2} = 790.55497738086$$

Stratified Sample Results:

$p_{str,land,>6326}$	0.505347593582888
$SE(p_{str,land,>6326})$	0.0114394567047594
$p_{str,strata>6326}$	0.0407079646017699
$SE(p_{str,strata>6326})$	0.00585098050147294

- $p_{str,land,>6326}$: the stratified sample proportion of Vancouver property tax 2019 larger than 6,326 CAD under legal type land.
- $p_{str,strata>6326}$: the stratified sample proportion of Vancouver property tax 2019 larger than 6,326 CAD under legal type strata.

$$p_{str,>6326} = \frac{N_{land}}{N} * p_{str,land,>6326} + \frac{N_{strata}}{N} * p_{str,strata>6326} = 0.237760850250982$$

$$SE\left(p_{str,>6326}\right) = \sqrt{\left(\frac{N_{land}}{N}\right)^2 * SE\left(p_{str,land,>6326}\right)^2 + \left(\frac{N_{strata}}{N}\right)^2 * SE\left(p_{str,strata>6326}\right)^2}$$

Appendix 2: R Code:

```
library(tidyverse)
library(infer)
property_tax_data <- read.csv("property-tax-report-2016-2019.csv", sep=";")
property_tax_data_tidy <- property_tax_data %>%
                          select(c(LEGAL_TYPE, ZONING_CLASSIFICATION, TAX_ASSESSMENT_YEAR, CURRENT_LAND_VALUE, TAX_LEVY)) %>%
                          drop_na()
head(property_tax_data_tidy)
N <- nrow(property_tax_data_tidy)
N
## Length for LEGAL_TYPE:
attach(property_tax_data_tidy)
N_legal <- tapply(TAX_LEVY, LEGAL_TYPE, length)
N_legal
N_ZONING <- tapply(TAX_LEVY, ZONING_CLASSIFICATION, length)
N_ZONING
Nland <- 88851
NOther <- 192
NSTRATA <- 120423
###SRS sample
set.seed(3)
n = 3000
SRS_sample <- property_tax_data_tidy %>%
              rep_sample_n(size = n, replace = FALSE)


prop_srs <- nrow(SRS_sample %>% filter(TAX_LEVY > 6326)) / nrow(SRS_sample)
prop_srs
SE_prop_srs <- sqrt((1-n/N) * prop_srs*(1- prop_srs) / n)
SE_prop_srs
```

```
### Visualisations
SRS_boxplot <-   SRS_sample %>%
                 ggplot(aes(x = LEGAL_TYPE, y = TAX_LEVY)) +
                 geom_boxplot() +
                 ggtitle("Graph 1.0 Boxplot of SRS") +
                 labs(x = "Legal Type", y = "Tax Levy (CAD)")
SRS_histogram <- SRS_sample %>%
                 ggplot(aes(x = TAX_LEVY)) +
                 geom_histogram() +
                 ggtitle("Graph 1.1 Histogram of SRS") +
                 xlab("Tax Levy (CAD)")

SRS_boxplot
SRS_histogram
y_s_hat <- SRS_sample %>%
           summarize(y_s_bar = mean(TAX_LEVY), se_y_s_hat = sqrt((1 - n/N) * var(TAX_LEVY)/n))
y_s_hat
```

```
##S_guess_legal:
S_guess_legal <- SRS_sample %>%
                 group_by(LEGAL_TYPE) %>%
                 summarize(S_h_guess = sd(TAX_LEVY))
S_guess_legal
mean_legal <- SRS_sample %>%
                 group_by(LEGAL_TYPE) %>%
                 summarize(mean = mean(TAX_LEVY))
mean_legal
### Proportion varainces
prop_LAND <- nrow(SRS_sample %>% filter(LEGAL_TYPE == "LAND" & TAX_LEVY > 6326))/nrow(SRS_sample %>% filter(LEGAL_TYPE == "LAND"))
prop_OTHER <- nrow(SRS_sample %>% filter(LEGAL_TYPE == "OTHER" & TAX_LEVY > 6326))/nrow(SRS_sample %>% filter(LEGAL_TYPE == "OTHER"))
prop_STRATA <- nrow(SRS_sample %>% filter(LEGAL_TYPE == "STRATA" & TAX_LEVY > 6326))/nrow(SRS_sample %>% filter(LEGAL_TYPE == "STRATA"))
sd_LAND <- sqrt(prop_LAND * (1 - prop_LAND))
sd_LAND
sd_OTHER <- sqrt(prop_OTHER*(1 - prop_OTHER))
sd_OTHER
sd_STRATA <- sqrt(prop_STRATA *(1 - prop_STRATA))
sd_STRATA

### LEGAL_TYPE variance:
y_s_bar <- y_s_hat$y_s_bar
y_ph_bar_legal_srs <- SRS_sample %>%
                      group_by(LEGAL_TYPE)%>%
                      summarize(mean = mean(TAX_LEVY))
```

```r
### Stratified Estimate:
set.seed(10)
str_land <- property_tax_data_tidy %>%
            filter(LEGAL_TYPE == "LAND") %>%
            rep_sample_n(size = 2882, replace = FALSE)
head(str_land)
str_STRATA <- property_tax_data_tidy %>%
              filter(LEGAL_TYPE == "STRATA") %>%
              rep_sample_n(size = 118, replace = FALSE)
head(str_STRATA)
str_land_mean <- mean(str_land$TAX_LEVY)
se_land_mean <- sqrt((1- 2882/Nland)*var(str_land$TAX_LEVY) / 2882)
str_land_mean
str_STRATA_mean <- mean(str_STRATA$TAX_LEVY)
se_STRATA_mean <- sqrt((1- 118/NSTRATA)*var(str_STRATA$TAX_LEVY) / 118)
str_STRATA_mean
str_estimate <- (N_legal[1] / N)*str_land_mean + (N_legal[3]/ N)*str_STRATA_mean
as.numeric(str_estimate)
str_estimate_se <- sqrt((N_legal[1] / N)^2*se_land_mean^2 + (N_legal[3]/ N)^2 * se_STRATA_mean^2)
as.numeric(str_estimate_se)
set.seed(15)
str_prop_land <- property_tax_data_tidy %>%
            filter(LEGAL_TYPE == "LAND") %>%
            rep_sample_n(size = 1870, replace = FALSE)
str_prop_strata <- property_tax_data_tidy %>%
              filter(LEGAL_TYPE == "STRATA") %>%
              rep_sample_n(size = 1130, replace = FALSE)
prop_land_str <- nrow(str_prop_land %>% filter(LEGAL_TYPE == "LAND" & TAX_LEVY > 6326))/nrow(str_prop_land %>% filter(LEGAL_TYPE == "LAND"))
se_LAND_prop <- sqrt((1 - 1870/Nland)* prop_land_str * (1 - prop_land_str)/1870)
prop_strata_str <- nrow(str_prop_strata %>% filter(LEGAL_TYPE == "STRATA" & TAX_LEVY > 6326))/nrow(str_prop_strata %>% filter(LEGAL_TYPE == "STRATA"))
se_STRATA_prop <- sqrt((1- 1130/NSTRATA)*prop_strata_str * (1 - prop_strata_str)/ 1130)
str_prop <- (N_legal[1] / N)*prop_land_str + (N_legal[3]/ N)*prop_strata_str
as.numeric(str_prop)
str_prop_se <- sqrt((N_legal[1] / N)^2*se_LAND_prop^2 + (N_legal[3]/ N)^2 * se_STRATA_prop^2)
as.numeric(str_prop_se)
```

Reference:

1. *Vancouver Property Tax 2019 | Calculator & Rates - WOWA.ca*. (n.d.). Wowa Leads Inc. https://wowa.ca/taxes/vancouver-property-tax

2. *Property Tax Calculator & Estimator 2019 | WOWA.ca*. (n.d.). Wowa Leads Inc. https://wowa.ca/calculators/property-tax

3. Business in Vancouver. (2020, February 11). *Vancouver businesses still shoulder unfair property tax burden*. https://biv.com/article/2020/02/vancouver-businesses-still-shoulder-unfair-property-tax-burden

4. *Property tax report 2016-2019*. (2021, May 8). https://opendata.vancouver.ca/explore/dataset/property-tax-report-2016-2019/information/?refine.report_year=2019