STAT 344 Project

Sameer Shankar (47555636) (GL) - Overview, Conclusion, Editor

Tina Fu (86675303) - Methodology, Appendix

Kelvin Lam (47219688) - Data Analysis, R-code

Jaskaran Singh (44025013) - Overview

November 2021

1 Overview

Rental housing for UBC is highly competitive, be it on or off campus. The rent (monthly in CAD) paid by each UBC student is a very important cost, for which they may have to plan their finances. The main purpose of this study is to obtain an estimate for the average rent paid by UBC students, on and off campus. The study aims to use a simple random sample (SRS) and a stratified sample of the same size in order to obtain data. Furthermore, a second parameter that the researchers want to calculate is the proportion of students paying more than $1000 in rent. This study is very important to carry out, as it gives an estimate of rent expenditure, which is of importance to students and families on a budget, thereby helping them plan their accommodation. The study provides a general overview of rental market prices for UBC students, which may even be useful for comparing against housing costs for students across other universities in Vancouver (or perhaps BC/other provinces).

2 Methodology

This study focuses on the amount of rent paid by UBC students living on and off campus. The target population is all UBC students who pay their own rent each month, including both undergraduate and graduate students. It is assumed that all students living on campus pay their own rent, and since the proportion of off campus students paying their own rent (that is, students off campus that do not live with their families and thus pay their own rent) is unknown, and this is estimated by using the sampled data.

2.1 SRS

The four researchers picked a random day of the week that fit everyone's schedule, and stood on the university boulevard asking random students to fill out a survey which asked students whether they lived on or off campus, and how much rent they paid.

Due to conflicting time schedules, there was only one time slot per week which worked for everyone. Therefore, the group spent 2.5 hours collecting as much data as possible, and used the total number of students surveyed in this time frame as the sample size (on average, each individual needed 2 minutes to complete the survey). After discarding responses of students who did not pay their own rent, a sample size of 55 was acquired. Therefore, $n = 55$ was used to estimate the average monthly rent paid by students $\bar{y}_p$ (continuous) and to estimate the proportion of students who pay over \$1000 in rent per month, $\hat{p}$ (binary).

2.2 Stratified Sampling

Under stratified sampling, this study tried to estimate $\bar{y}_p$ using $\bar{y}_{str}$. After completing the SRS and obtaining $n = 55$, the research group then obtained a stratified random sample, where the sample size would also be 55. Out of the 55 in the stratified random sample, 16 students were chosen for on campus (strata 1) and 39 students were chosen for off campus (strata 2). Given the aim of the study of being to find the average rent paid by UBC students, the variation between cost of living on campus vs. cost of living off campus would be very important, and hence it was logical to have the two strata be students on campus and students off campus paying their own rent. The allocation of students for the two strata is calculated by the following:

From the SRS (which was collected first), the standard deviations (SD) of rent

paid by students living on and off campus were 263 and 304, respectively (using R).

Let $S1_{guess}$ be the guessed SD for on campus students and $S2_{guess}$ be the guessed SD for off campus students. Using the SDs from SRS, $S1_{guess} = 263$ and $S2_{guess} = 304$.

The SRS data provided the following; 22 out of 36 students living off campus pay their own rent. Given the total of 56,936 students in Vancouver campus (provided by UBC Housing on their website), around 13,000 live on campus and $56,936 - 13,000 = 43,936$ off campus. Therefore, an estimate of the population living off campus, that pays their own rent, is $(22/36) \times 43,936 = 26,850$, which gives an estimated $N = 13,000 + 26,850 = 39,850$.

Let N_1 and N_2 be the proportion of the target population living on and off campus paying their own rent respectively. Given the figures above, $N_1 = 32.6\%$ and $N_2 = 67.4\%$.

Assume the costs for collecting samples from each stratum are the same. Given $n_h \propto \frac{N_h * S_h,guess}{\sqrt{C_h}}$,

$$n_1 = \frac{32.6\% * 263}{(32.6\% * 263 + 67.4\% * 304)} \times 55 = 16$$

$$n_2 = \frac{67.4\% * 304}{(32.6\% * 263 + 67.4\% * 304)} \times 55 = 39.$$

3 Analysis

3.1 SRS

Based on the SRS, the sample mean and standard error of the rent paid by UBC students who pay their own rent, the following 95% Confidence Interval can be constructed:

$$\bar{y}_s = \sum_{i \in S} y_i = 1079$$

$$SE[\bar{y}_s] = s/\sqrt{n} = 41.9$$

$$95\% \text{ C.I.} = (\bar{y}_s - 1.96 \times SE[\bar{y}_s], \bar{y}_s + 1.96 \times SE[\bar{y}_s]) = (997.18, 1161.42)$$

The proportion and SE of students who pay more than \$1,000 are (along with a 95% Confidence Interval):

$$\hat{p} = 0.4545$$

$$SE[\hat{p}] = \sqrt{\hat{p}(1 - \hat{p})/n} = 0.0671$$

$$95\% \text{ C.I.} = (0.3229, 0.5861)$$

Since $n/N < 0.05$, this research has ignored the finite population correction factor (FPC).

3.2 Stratified Sampling

Based on the stratified sampling, the estimates and standard errors of the rent paid by UBC students who pay their own rent for both on and off campus, the following 95% Confidence Interval can be constructed:

$$\bar{y}_{S_1} = 1074$$

$$\bar{y}_{S_2} = 1057$$

$$\bar{y}_{str} = \sum_{h=1}^{2}(N_h/N) \times \bar{y}_{S_h} = 1062$$

$$SE[\bar{y}_{str}] = \sqrt{\sum_{h=1}^{2}(N_h/N)^2 \times Var[\bar{y}_{S_h}]} = 33.7$$

$$95\% \text{ C.I.} = (996.00, 1128.11)$$

The proportion and SE of students who pay more than \$1,000 are (along with a 95% Confidence Interval):

$$\hat{p}_{str} = 0.4$$

$$SE[\hat{p}_{str}] = \sqrt{\sum_{h=1}^{2}(N_h/N)^2 \times \hat{p}_{str_h}(1 - \hat{p}_{str_h})/n_h} = 0.0663$$

$$95\% \text{ C.I.} = (0.2701, 0.5299)$$

Again, since $n_h/N_h < 0.05$ for both h, FPC is ignored.

3.3 Interpretation

Both sampling methods obtained similar estimates of the average rent paid by UBC students. However, the standard error obtained from stratified sampling method is lower than the standard error from SRS, by 19.55%.

It is known that
$$\frac{Var[\bar{y}_{str}]}{Var[\bar{y}]} \approx 1 - \frac{s_{P,B}^2}{s_P^2},$$

where $s_{P,B}^2$ is the between-strata variance.

Based on our samples, $Var[\bar{y}_{str}]/Var[\bar{y}] = (\frac{33.7}{41.9})^2 = 0.6472$, suggesting that $s^2_{P,B}/s^2_P = 0.3528$ and 35.28% of the variability of rent paid by UBC students come from the between-strata variance (i.e, the variability of students living on campus and off campus rent).

This matches our intuition: more variability in the rent costs off campus than on campus is expected. Responses from off campus students included a variety of places such as West Point Grey, Wesbrook, Dunbar, Kitsilano, and in some cases, Richmond and Burnaby. The rents across these regions are highly spread out. Comparatively, the number of residences at UBC are limited and the rents are quite similar (the rental costs are on the UBC Housing website), based on our observations. Hence, it is expected that the between-strata variance is high.

Costs were defined to be on the expensive side if students paid more than $1,000 rent per monthly. Unfortunately, the study was unable to conclude that the mean and median rent paid by UBC students is higher/lower than $1,000, based on both sampling methods. For the mean, $1,000 was included in the 95% confidence intervals from both vanilla estimates $\bar{y}_s$ and stratified estimates $\bar{y}_{str}$. For the median, 50% was included in the 95% confidence intervals from both vanilla estimates $\hat{p}$ and stratified estimates $\hat{p}_{str}$. The main reason behind the large standard error was that the sample sizes were small. The researchers noticed that the lower bound of the CI for $\bar{y}_s$ and $\bar{y}_{str}$ are close to $1,000 ($997.18 and $996.00 respectively). $1,000 may be excluded from the C.I. if more surveys were conducted, which would allow the researchers to be 95% that the average rent paid by UBC students is higher than $1,000.

3.4 Advantages and Disadvantages of each method

There are multiple advantages and disadvantages for each method used. The study group compared the cost and difficulty of collecting samples, as well as

the accuracy of the estimate in both sampling methods.

The SRS incurred a lower sampling time cost than stratified sampling. Even though both sampling methods have the same sample size and each students should use roughly the same amount of time to complete the survey, researchers had to classify on campus and off campus students in stratified sampling. For the SRS, the researchers could simply ask any students at the main street and discard those who live with their family. However, for stratified sampling, the researchers have to know the stratum the students belong to (either living on campus or off campus) before giving them the correct version of the survey (on campus/off campus). Samples would have to be discarded if they do not belong to the right stratum. Thus, stratified sampling had a greater time cost than SRS.

It was also easier to collect samples in the SRS than the stratified sample. For the SRS, the researchers collected surveys on the main mall street (near fountain). Many students were willing to fill out the survey. For stratified sampling, the researchers collected samples of on campus students and off campus students at main mall street and the bus loop respectively. Students at the bus loop were usually rushing to their bus or class, and understandably, more unwilling complete the survey.

The stratified sample generated a more accurate result. As mentioned in the previous subsection, stratified sampling reduces the standard error of the estimate by 19.55%, thanks to the large proportion of the between-strata variance in the population variance. The standard error of the proportion is also reduced by 1.29%.

4 Conclusion and Discussion

Overall, the research suggests that estimates for the average rent cost per month for UBC students is $1070 and $1062 from the SRS and stratified sample respectively. The researchers are 95% confident that the true rent cost per month for UBC students lies between ($997.18, $1161.42) and ($996, $1128.11), which are the 95% CIs for the SRS and stratified sample respectively.

The research also produces estimates for the proportion of students at UBC paying more than $1000 per month on rent being 0.4545 and 0.4 from the SRS and stratified sample respectively. The researchers are 95% confident that the true proportion of UBC students paying more than $1000 per month per month on rent lies between (0.3229,0.5861) and (0.2701,0.5299), which are the 95% CIs for the SRS and stratified sample respectively.

While the researchers made attempts to ensure that the samples were collected were large and done so randomly, however there are limitations to the study. Firstly, due to the school/work schedules of the four researchers, and taking into account exams/assignments, there was a significant time constraint. With the time constraint in place, the sample sizes for both SRS and stratified sampling were limited. For example, as the sample sizes were rather small, the within-strata variance for students on campus in the estimation of $\bar{y}_{str}$ is likely to be unrepresentative of the true within-strata variance of students on campus, as there were only 16 students surveyed for this strata. With better planning and more time, the researchers could have scheduled two sampling sessions for each type of sampling, and a sample size of 100 could have been obtained. The small sample size is also the reason behind why the confidence intervals for the proportion of students paying more than $1000 for rent per month are very wide. With a larger sample size, there would be a smaller margin of error in the intervals.

Furthermore, another limitation is that some students that live off campus may not come to campus to attend classes (perhaps they have online classes or they want to skip classes), and so that may distort our sample (it may not be representative of the population).

The results of this experiment may not be generalized to larger populations easily. For example, the average rent paid by UBC would be significantly different from the average rent paid by residents in Vancouver, given that this would include adults, and since students do not work full-time, therefore not earning as much as adults, they will not be able to afford to live in affluent areas. Moreover, UBC students are likely to rent places that are fairly close to UBC in order to reduce their commute, thereby skewing their housing prices to be representative of areas near UBC (West Point Grey, Kitsilano and Dunbar), whereas Vancouver residents would also represent prices in Burnaby, Downtown, etc. which would make the estimates different from the ones obtained in the study.

5 Appendix

5.1 Graphs

Histogram (Under SRS)

Histogram (Under Stratified Sampling)
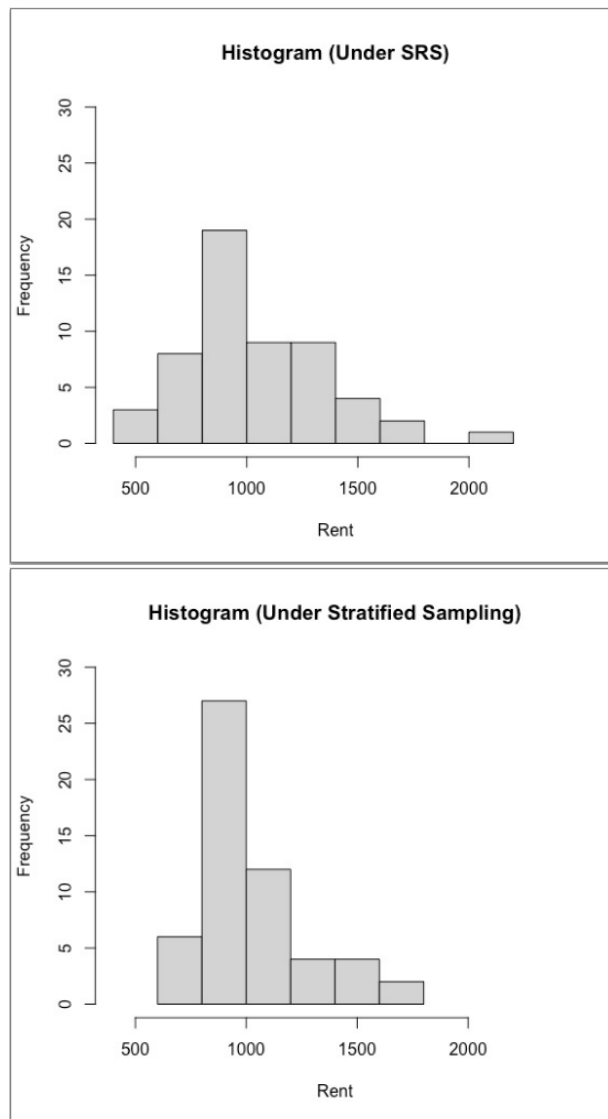
5.2 SRS data

SRS data

Timestamp	Oncampus?	residence oncampus	residence offcampus	Rent	YearofStudy	Family
10/19/2021 10:22:42	No		West pont grey	1100	Year 2	No
10/19/2021 10:23:57	Yes	Orchard Commons		1200	Year 1	No
10/19/2021 10:28:59	No		Burnaby	0	Year 1	Yes
10/19/2021 10:29:52	No		Tsawwassen	0	Year 2	Yes
10/19/2021 10:31:22	Yes	Tptem park		456	Year 1	No
10/19/2021 10:33:18	Yes	totempark		456	Year 1	No
10/19/2021 10:33:33	Yes	Orchard Commons	I dont live in Canada	917	Year 1	No
10/19/2021 10:34:13	Yes	Marine		975	Year 3	No
10/19/2021 10:35:04	No		s cha nue e s g y	850	Year 4	No
10/19/2021 10:36:40	No		Richmond	0	Year 2	Yes
10/19/2021 10:36:54	Yes	marine drive		1000	Year 3	No
10/19/2021 10:37:24	Yes	Ponderosa		600	Year 3	No
10/19/2021 10:38:38	Yes	Walter Gage		900	Year 2	No
10/19/2021 10:46:14	Yes	On campus, Marine drive residence		932	Year 3	No
10/19/2021 10:47:29	No		East	1000	Year 2	No
10/19/2021 10:48:04	Yes	Totem park	N/a	800	Year 1	No
10/19/2021 10:49:42	Yes	Orchard Commons		1700	Year 2	No
10/19/2021 10:51:27	Yes	Marine		932.5	Year 3	No
10/19/2021 10:52:45	No		West Point grey	960	Year 4	No
10/19/2021 10:53:56	Yes	Wesbrook		1300	Year 4	No
10/19/2021 10:56:23	Yes	Orchard commons		1400	Year 1	No
10/19/2021 10:57:17	Yes	Marine drive		1000	Year 4	No
10/19/2021 11:00:19	No		Richmond	0	Year 1	Yes
10/19/2021 11:00:47	No		Richmond	0	Year 1	Yes
10/19/2021 11:02:07	Yes	Vanier		800	Year 1	No
10/19/2021 11:03:37	No		Kitsalano	1500	Year 1	No
10/19/2021 11:05:01	Yes	Ponderosa		1160	Year 3	No
10/19/2021 11:05:07	Yes	UBC but not residence		0	Year 2	Yes
10/19/2021 11:07:28	No		University	1240	Year 2	No
10/19/2021 11:08:42	No	NA	Vancouver general hosp	500	Year 3	Yes
10/19/2021 11:09:16	No		Surrey	0	Year 2	Yes
10/19/2021 11:09:53	No		West Vancouver	0	Year 1	Yes
10/19/2021 11:10:28	Yes	Totem park		900	Year 1	No

5.3 Stratified sampling data

On-Campus data

Timestamp	residence_oncampus	Rent	Year of Study	Family
10/30/2021 15:21:33	Exchange	1350	Year 5+	No
11/2/2021 10:37:56	Fairview Crescent	1500	Year 3	No
11/2/2021 10:48:03	5516 dalhousie road	1200	Year 2	No
11/2/2021 10:48:52	totem	1300	Year 1	No
11/2/2021 11:07:23	Gage	1350	Year 2	No
11/2/2021 11:07:35	marine drive residence	930	Year 2	No
11/2/2021 11:08:02	Marine drive	930	Year 2	No
11/2/2021 11:10:26	Spirit park osoyoos cresc	1951	Year 3	Yes
11/2/2021 11:11:49	Marine. Drive	1200	Year 4	No
11/2/2021 11:11:57	Place vanier	800	Year 1	No
11/2/2021 11:13:12	Ponderosa	940	Year 4	No
11/2/2021 11:14:28	Marine Drive	950	Year 3	No
11/2/2021 11:14:36	OC	900	Year 1	No
11/2/2021 11:15:24	Marine drive	962	Year 3	No
11/2/2021 11:16:15	Ponderosa Commons	1140	Year 4	No
11/2/2021 11:17:56	Ritsumeikan house	995	Year 1	No
11/2/2021 11:18:39	Walter Gage	920	Year 3	No
11/2/2021 11:19:45	Thunderbird	850	Year 2	No
11/2/2021 11:22:38	Marine Drive	932.51	Year 2	No
11/2/2021 11:23:46	Ponderosa	1200	Year 3	No
11/2/2021 11:25:26	Ponderosa Commons	1100	Year 3	No
11/2/2021 11:35:53	Vanier	800	Year 4	No
11/2/2021 11:50:25	Totem park	850	Year 3	No
11/2/2021 11:50:35	Ponderosa	1200	Year 1	No
11/2/2021 12:02:07	Spruce House, Ponderos	960	Year 3	No
11/2/2021 12:18:04	Exchange	1000	Year 4	No

Off-Campus Data

Timestamp	residence_offcampus	Rent	Year of Study	Family
2/11/2021 10:40	Marine Drive	1200	Year 4	No
2/11/2021 10:41	West Point Grey	920	Year 3	No
2/11/2021 10:46	Point Grey	0	Year 3	Yes
2/11/2021 10:55	Mainland	875	Year 3	No
2/11/2021 11:13	North shore	0	Year 3	Yes
2/11/2021 11:21	Oakridge	1100	Year 3	No
2/11/2021 11:22		0	Year 5+	No
2/11/2021 11:33	Surrey	600	Year 2	Yes
2/11/2021 11:33	New Westminster	0	Year 2	Yes
2/11/2021 11:34	South Vancouver	800	Year 1	No
2/11/2021 11:34	Mt. Pleasant	1800	Graduate	No
2/11/2021 11:35	Coquitlam	0	Year 4	Yes
2/11/2021 11:36	Point grey	650	Year 3	No
2/11/2021 11:37	Kits	1200	Year 4	No
2/11/2021 11:38	Surrey	0	Year 4	Yes
2/11/2021 11:38	Kerrisdale	0	Year 2	Yes
2/11/2021 11:41	Dunbar	1100	Year 3	No
2/11/2021 11:42	Kits	1000	Year 2	No
2/11/2021 11:42	North van	0	Year 3	Yes
2/11/2021 11:42	Downtown	3000	Year 3	Yes
2/11/2021 11:43	West point grey	860	Year 4	No
2/11/2021 11:43		0	Year 2	Yes
2/11/2021 11:44	Oakridge	980	Year 4	Yes
2/11/2021 11:47	Fairview	0	Graduate	Yes

13

5.4   R script

```
### SRS
# Read data from the csv file
data <- read.csv("SRS.csv")
# Discard the samples not in the population
sample <- data[data$Family == "No", ]
# Indicator of sample rent larger than $1,000
sample$exp <- ifelse(sample$Rent > 1000, 1,0)
n <- as.numeric(nrow(sample))


# sample mean of the monthly rent
avg_srs <- mean(sample$Rent)
avg_srs
# Standard Error (SE) of the estimate
se_srs <- sd(sample$Rent)/sqrt(n)
round(se_srs,2)
# 95% C.I. of the sample mean
CI_srs <- c(avg_srs - 1.96*se_srs, avg_srs + 1.96*se_srs)
round(CI_srs,2)


# p_hat
prop_srs <- mean(sample$exp)
round(prop_srs,4)
# SE of p_hat
se_prop_srs <- sqrt(prop_srs*(1-prop_srs)/n)
round(se_prop_srs,4)
# 95% C.I. of p_hat
CI_prop_srs <- c(prop_srs - 1.96*se_prop_srs, prop_srs + 1.96*se_prop_srs)
round(CI_prop_srs,4)
```

```
### Code for estimating the population size and within-strata variances
for stratified sampling
# Number of off-campus students paying their own rent in the sample
offcampus_rent <- sum(ifelse(data$Family=="No" & data$Oncampus. == "No",1,0))
# Number of off-campus students in the sample
offcampus <- sum(ifelse(data$Oncampus. == "No",1,0))
# Estimated proportion of off-campus students paying their own rent
prop_offrent <- offcampus_rent/offcampus
prop_offrent


# Total number of UBC Vancouver students
ttl <- 56936   #Source: https://www.ubc.ca/about/facts.html
# Total number of on-campus students paying their own rent
N_on <- 13000   #Source: https://vancouver.housing.ubc.ca
# Estimated total number of off-campus students paying their own rent
N_off <- round((ttl-N_on)*prop_offrent,0)
# Population size
N <- N_on + N_off
N


# Guessed sample SD of rent paid by on-campus students
oncampus_sd <- sd(sample[sample$Oncampus. == "Yes","Rent"])
oncampus_sd
# Guessed sample SD of rent paid by off-campus students
offcampus_sd <- sd(sample[sample$Oncampus. == "No","Rent"])
offcampus_sd


# Optimal allocation
n_on <- round(n*(oncampus_sd*N_on)/(oncampus_sd*N_on + offcampus_sd*N_off),0)
n_off <- n - n_on
```

```r
satrified_n <- c(n_on, n_off)
satrified_n

### Stratified Sampling
set.seed(1214)
data_on <- read.csv("oncampus.csv")
data_on <- data_on[data_on$Family == "No", ]
sample_on_int <- sample.int(nrow(data_on), size = n_on, replace = FALSE)
sample_on <- data_on[sample_on_int,]
sample_on$exp <- ifelse(sample_on$Rent > 1000, 1,0)


data_off <- read.csv("offcampus.csv")
data_off <- data_off[data_off$Family == "No", ]
data_off <- data_off[data_off$Rent != "0", ]
sample_off_int <- sample.int(nrow(data_off), size = n_off, replace = FALSE)
sample_off <- data_off[sample_off_int,]
sample_off$exp <- ifelse(sample_off$Rent > 1000, 1,0)


avg_str <- (n_on/n)*mean(sample_on$Rent)+(n_off/n)*mean(sample_off$Rent)
avg_str
se_str <- sqrt((N_on/N)^(2)*var(sample_on$Rent)/n_on
+ (N_off/N)^(2)*var(sample_off$Rent)/n_off)
se_str
CI_str <- c(avg_str - 1.96*se_str, avg_str + 1.96*se_str)
CI_str


prop_str <- (n_on/n)*mean(sample_on$exp)+(n_off/n)*mean(sample_off$exp)
prop_str
se_prop_str <- sqrt((N_on/N)^(2)*mean(sample_on$exp)*
(1-mean(sample_on$exp))/n_on
```

$+ (\text{N\_{\bf off}/N})\,\hat{}\,(2)*{\bf mean}({\bf sample\_off\$exp})*(1-{\bf mean}({\bf sample\_off\$exp}))/{\bf n\_off})$

$\text{se\_{\bf prop}\_str}$

$\text{CI\_{\bf prop}\_str} \leftarrow {\bf c}({\bf prop}\_str - 1.96*{\bf se\_prop}\_str,\ {\bf prop}\_str + 1.96*{\bf se\_prop}\_str)$

$\text{CI\_{\bf prop}\_str}$

STAT344 Sample Survey Group Project

Part I: Simple Random Sample

* Required

1. Are you living with your family?   *

   Mark only one oval.

   ◯ Yes

   ◯ No

2. Are you living on campus?   *

   Mark only one oval.

   ◯ Yes

   ◯ No

3. If on-campus, which residence are you currently living?

   _____

4. If off-campus, which part of Vancouver are you currently living?

   _____

5. How much rent are you paying per month?   *

   _____

STAT344 Group Project (Part 2A)

Stratified Sampling (On-campus)

* Required

1. Are you living with your family? *

   Mark only one oval.

   ( ) Yes

   ( ) No

2. Are you living on campus? *

   Mark only one oval.

   ( ) Yes

   ( ) No

3. If on-campus, which residence are you currently living?

   _____

4. If off-campus, which part of Vancouver are you currently living?

   _____

5. How much rent are you paying per month? *

   _____

STAT344 Group Project (Part 2B)

Stratified Sampling (Off-campus)

* **Required**

1. Are you living with your family? *

    Mark only one oval.

    ◯ Yes

    ◯ No

2. Are you living on campus? *

    Mark only one oval.

    ◯ Yes

    ◯ No

3. If on-campus, which residence are you currently living?

    _____

4. If off-campus, which part of Vancouver are you currently living?

    _____

5. How much rent are you paying per month? *

    _____