# Data for the Ontario Restaurant Closures Experiment[*]

Simulation Methods and Parameterization Research

Lorena Almaraz De La Garza, Amy Farrow, and Kumalasari Sondjaja

20 February 2021

**Abstract**

This document explains how data for the Ontario restaurant closures experiment was simulated, including the research that was used for parameterization and the methods used to randomize data.

Analysis for this project uses the R statistical programming language (R Core Team 2020), and more specifically, the `tidyverse` package for data manipulation (Wickham et al. 2019). Because the data is managed using R Projects, `here` is used to reference file locations (Müller 2020). `kableExtra` formats tables (Zhu 2020). `bookdown` is used to format the report (Xie 2020).

First, we read in the skeleton for the survey data. This is what we would hypothetically know about a restaurant before they took the survey. We keep only the ID numbers, the group, and the last 3 characters of the postal code, which the survey-taker will have to enter to verify identity. All other variables, like name and full address, are only kept in the only frame for surveying, which contains no private information.

```r
# read in the data
# NOTE: scripts 01_scrape_health_depts,
# 02_stratified_cluster_sampling,
# 03_clean_restaurant_list_csvs,
# 04_restaurant_survey_frame,
# and 05_import_yelp_data must have been run already

table_for_surveys <- read_csv(here::here("outputs/data/table_for_surveys.csv"))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   name = col_character(),
##   address = col_character(),
##   unit = col_character(),
##   group = col_character(),
##   verify = col_character(),
##   ID = col_double()
## )
```

```r
survey_1 <- table_for_surveys %>%
  select(ID, group, verify) %>%
  arrange(ID)

# count the numbers for treat, control, and total
num_treat <- first(count(survey_1 %>%
                           filter(group == "treatment")))
```

[*] Code and data are available at: github.com/amycfarrow/ontariorestaurantclosuresexperiment.

```r
num_control <- first(count(survey_1 %>%
                             filter(group == "control")))

num_total <- first(count(survey_1))

survey_2 <- survey_1

survey_1
```

```
## # A tibble: 2,006 x 3
##       ID group     verify
##    <dbl> <chr>     <chr>
##  1     1 control   1E4
##  2     2 treatment 0N
##  3     3 control   6C7
##  4     4 control   1B8
##  5     5 control   1B0
##  6     6 control   2J9
##  7     7 treatment 3Y3
##  8     8 treatment 1K0
##  9     9 control   0H4
## 10    10 treatment 2J4
## # ... with 1,996 more rows
```

From there, data was simulated for all questions asked on the survey.

# 1 Type of service provided

First, some research:

```r
# Read in data
yelp_data <- read_csv(here("inputs/data/yelp_restaurants_ontario.csv")) %>%
  na.omit() # Remove all NAs
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   name = col_character(),
##   address = col_character(),
##   city = col_character(),
##   postal_code = col_character(),
##   latitude = col_double(),
##   longitude = col_double(),
##   stars = col_double(),
##   is_open = col_double(),
##   RestaurantsPriceRange2 = col_character(),
##   RestaurantsTakeOut = col_logical(),
##   RestaurantsDelivery = col_character(),
##   RestaurantsTableService = col_logical(),
##   OutdoorSeating = col_logical(),
##   categories = col_character(),
##   business_id = col_character()
## )
```

```
total_restaurants <- count(yelp_data)


# Check proportion of RestaurantsTakeOut
takeout <- yelp_data %>%
  select(RestaurantsTakeOut)
summary(takeout)
```

```
##  RestaurantsTakeOut
##  Mode :logical
##  FALSE:276
##  TRUE :4120
```

```
takeout_percentage <- takeout %>%
  group_by(RestaurantsTakeOut) %>%
  summarize(count = n()) %>%
  mutate(percentage = count/sum(count))
```

```
## ‘summarise()‘ ungrouping output (override with ‘.groups‘ argument)
```

```
takeout_percentage
```

```
## # A tibble: 2 x 3
##   RestaurantsTakeOut count percentage
##   <lgl>              <int>      <dbl>
## 1 FALSE                276     0.0628
## 2 TRUE                4120     0.937
```

```
# Check proportion of RestaurantsTableService
tableserv <- yelp_data %>%
  select(RestaurantsTableService)
summary(tableserv)
```

```
##  RestaurantsTableService
##  Mode :logical
##  FALSE:1448
##  TRUE :2948
```

```
tableserv_percentage <- tableserv %>%
  group_by(RestaurantsTableService) %>%
  summarize(count = n()) %>%
  mutate(percentage = count/sum(count))
```

```
## ‘summarise()‘ ungrouping output (override with ‘.groups‘ argument)
```

```
tableserv_percentage
```

```
## # A tibble: 2 x 3
##   RestaurantsTableService count percentage
##   <lgl>                   <int>      <dbl>
## 1 FALSE                    1448      0.329
## 2 TRUE                     2948      0.671
```

```
# Calculate invalid responses (neither dine-in nor takeout)
invalid <- yelp_data %>%
  filter(RestaurantsTableService == FALSE & RestaurantsTakeOut == FALSE) %>%
  count()

```

```r
valid_restaurants <- total_restaurants-invalid

# Calculate service type percentages
takeout_only <- yelp_data %>%
  filter(RestaurantsTableService == FALSE & RestaurantsTakeOut == TRUE) %>%
  count()/valid_restaurants

dinein_only <- yelp_data %>%
  filter(RestaurantsTableService == TRUE & RestaurantsTakeOut == FALSE) %>%
  count()/valid_restaurants

both <- yelp_data %>%
  filter(RestaurantsTableService == TRUE & RestaurantsTakeOut == TRUE) %>%
  count()/valid_restaurants

service_percentage<- data.frame(
  service = c("takeout", "dinein", "both"),
  percentage = c(takeout_only$n, dinein_only$n, both$n), options(digits = 2)
)

 service_percentage
```

```
##   service percentage digits
## 1 takeout      0.326      7
## 2  dinein      0.058      7
## 3    both      0.616      7
```

According to Ontario restaurant data from Yelp ("Download Yelp Dataset" 2019), approximately 6% of restaurants offer dine in service, 32.5% offer takeout, and 61.5% offer both.

## 1.1 Survey 1

```r
# randomly assign service_provided using ratios above

set.seed(19)
survey_1 <- survey_1 %>%
  mutate(service_type = sample(x = c("dinein",
                                     "both",
                                     "takeout"),

      size = num_total,
      replace = TRUE,
      prob = c(0.06, 0.615, 0.325)
      ))

survey_1
```

```
## # A tibble: 2,006 x 4
##       ID group     verify service_type
##    <dbl> <chr>     <chr>  <chr>
##  1     1 control   1E4    both
##  2     2 treatment ON     both
##  3     3 control   6C7    takeout
##  4     4 control   1B8    both
##  5     5 control   1B0    both
##  6     6 control   2J9    both
##  7     7 treatment 3Y3    both
##  8     8 treatment 1K0    both
##  9     9 control   0H4    takeout
## 10    10 treatment 2J4    takeout
## # ... with 1,996 more rows
```

### 1.2 Survey 2

```
# randomly assign service_provided using ratios above

set.seed(19)
survey_2 <- survey_2 %>%
  mutate(service_type = sample(x = c("dinein",
                                     "both",
                                     "takeout"),

      size = num_total,
      replace = TRUE,
      prob = c(0.06, 0.615, 0.325)
      ))

survey_2
```

```
## # A tibble: 2,006 x 4
##       ID group     verify service_type
##    <dbl> <chr>     <chr>  <chr>
##  1     1 control   1E4    both
##  2     2 treatment 0N     both
##  3     3 control   6C7    takeout
##  4     4 control   1B8    both
##  5     5 control   1B0    both
##  6     6 control   2J9    both
##  7     7 treatment 3Y3    both
##  8     8 treatment 1K0    both
##  9     9 control   0H4    takeout
## 10    10 treatment 2J4    takeout
## # ... with 1,996 more rows
```

## 2 Demographic traits of owners

0.5% of small to medium enterprises in Canada are owned by persons with disabilities ("SME Profile: Ownership Demographics Statistics" 2020).

### 2.1 Survey 1

```
# randomly assign disability status using ratios above

set.seed(19893)
survey_1 <- survey_1 %>%
  mutate(disability = sample(c("yes", "no", "nonanswer"),
                             size = num_total,
                             prob = c(0.005, .985, 0.01),
                             replace = TRUE))
```

```
survey_1
```

```
## # A tibble: 2,006 x 5
##        ID group     verify service_type disability
##     <dbl> <chr>     <chr>  <chr>        <chr>
##  1      1 control   1E4    both         no
##  2      2 treatment 0N     both         no
##  3      3 control   6C7    takeout      no
##  4      4 control   1B8    both         no
##  5      5 control   1B0    both         no
##  6      6 control   2J9    both         no
##  7      7 treatment 3Y3    both         no
##  8      8 treatment 1K0    both         no
##  9      9 control   0H4    takeout      no
## 10     10 treatment 2J4    takeout      no
## # ... with 1,996 more rows
```

## 2.2 Survey 2

```
# randomly assign disability status using ratios above

set.seed(19893)
survey_2 <- survey_2 %>%
  mutate(disability = sample(c("yes", "no", "nonanswer"),
                             size = num_total,
                             prob = c(0.005, .985, 0.01),
                             replace = TRUE))

survey_2
```

```
## # A tibble: 2,006 x 5
##        ID group     verify service_type disability
##     <dbl> <chr>     <chr>  <chr>        <chr>
##  1      1 control   1E4    both         no
##  2      2 treatment 0N     both         no
##  3      3 control   6C7    takeout      no
##  4      4 control   1B8    both         no
##  5      5 control   1B0    both         no
##  6      6 control   2J9    both         no
##  7      7 treatment 3Y3    both         no
##  8      8 treatment 1K0    both         no
##  9      9 control   0H4    takeout      no
## 10     10 treatment 2J4    takeout      no
## # ... with 1,996 more rows
```

Some research for racial and ethnic minorities and women in Canada and in business:

```
### Load census dataset by Health Regions ###
### Download link:
### https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/
### download-telecharger/comp/GetFile.cfm?Lang=E&FILETYPE=CSV&GEONO=058
census_2016 <- read.csv(here("inputs", "data", "census_2016.csv"))
census_2016 <- janitor::clean_names(census_2016)
```

```r
# Narrowed down categories
demographic_info <- c(1, 8, 1324, 1290, 1917)


### Function to get data from specific regions and rows ###

get_region_data <- function(y){
  census_2016 %>%
    filter(geo_code_por == y,
           member_id_profile_of_health_regions_2247 %in% demographic_info) %>%
    select(dim_profile_of_health_regions_2247,
           member_id_profile_of_health_regions_2247,
           dim_sex_3_member_id_1_total_sex,
           dim_sex_3_member_id_2_male,
           dim_sex_3_member_id_3_female)
}

### Get data from the randomly selected Health Regions ###
# GEO_CODE indicates public health regions, taken from Census 2016 links:
# https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/
# search-recherche/lst/results-resultats.cfm?Lang=E&TABID=1&G=1&Geo1=&Code1=
# &Geo2=&Code2=&GEOCODE=35&type=0

ontario <- get_region_data(35)

haliburton <- get_region_data(3535)
hamilton <- get_region_data(3537)
algoma <- get_region_data(3526)
simcoe_muskoka <- get_region_data(3560)
timiskaming <- get_region_data(3563)
windsor_essex <- get_region_data(3568)

northwest <- get_region_data(3549)
# Southwestern was created by almagamating oxford and elgin-st. thomas units:
oxford <- get_region_data(3552)
elgin <- get_region_data(3531)
waterloo <- get_region_data(3565)
durham <- get_region_data(3530)
sudbury <- get_region_data(3561)
brant <- get_region_data(3527)


### Put together demographic info in one table ###

populationss <- c("Total_Population",
                  "Indigenous_Population_25%_sample",
                  "Visible_Minority_25%_sample",
                  "Accommodation_and_Food_Services_25%_sample",
                  "Total_Population_Women")

get_pop_info <- function(x){
  c(as.numeric(x$dim_sex_3_member_id_1_total_sex
               [x$dim_profile_of_health_regions_2247 == "Population, 2016"]),
    as.numeric(x$dim_sex_3_member_id_1_total_sex
               [x$dim_profile_of_health_regions_2247 == "Aboriginal identity"]),
    as.numeric(x$dim_sex_3_member_id_1_total_sex
               [x$dim_profile_of_health_regions_2247 ==
                    "Total visible minority population"]),
    as.numeric(x$dim_sex_3_member_id_1_total_sex
               [x$dim_profile_of_health_regions_2247 ==
                    "72 Accommodation and food services"]),
    as.numeric(x$dim_sex_3_member_id_3_female
               [x$dim_profile_of_health_regions_2247 ==
                    "Total - Age groups and average age of the population - 100% data"]))
}

ontario_pop <- get_pop_info(ontario)

haliburton_pop <- get_pop_info(haliburton)
algoma_pop <- get_pop_info(algoma)
hamilton_pop <- get_pop_info(hamilton)
windsor_essex_pop <- get_pop_info(windsor_essex)
simcoe_muskoka_pop <- get_pop_info(simcoe_muskoka)
timiskaming_pop <- get_pop_info(timiskaming)

brant_pop <- get_pop_info(brant)
sudbury_pop <- get_pop_info(sudbury)
#southwestern_pop <- get_pop_info(southwestern)
oxford_pop <- get_pop_info(oxford)
elgin_pop <- get_pop_info(elgin)
northwest_pop <- get_pop_info(northwest)
waterloo_pop <- get_pop_info(waterloo)
durham_pop <- get_pop_info(durham)


populations <- bind_cols(populationss,
                         ontario_pop,
                         haliburton_pop,
                         algoma_pop,
                         hamilton_pop,
                         windsor_essex_pop,
                         simcoe_muskoka_pop,
                         timiskaming_pop,
                         brant_pop,
                         sudbury_pop,
                         oxford_pop,
                         elgin_pop,
                         northwest_pop,
                         waterloo_pop,
                         durham_pop)
```

```
## New names:
## * NA -> ...1
## * NA -> ...2
## * NA -> ...3
## * NA -> ...4
## * NA -> ...5
## * ...
```

```r
colnames(populations) <- c("Info", "Ontario", "Haliburton", "Algoma",
                           "Hamilton", "Windsor_Essex", "Simcoe_Muskoka",
                           "Timiskaming", "Brant", "Sudbury", "Oxford",
                           "Elgin", "Northwest", "Waterloo", "Durham")

populations <- populations %>%
  mutate(Southwestern = Oxford + Elgin) %>%
  mutate(total_treat =
           Haliburton + Algoma + Hamilton + Windsor_Essex + Simcoe_Muskoka + Timiskaming,
         total_control =
           Brant + Sudbury + Northwest + Waterloo + Durham + Southwestern) %>%
  select(-Oxford, -Elgin)

populations
```

```
## # A tibble: 5 x 16
##   Info  Ontario Haliburton Algoma Hamilton Windsor_Essex Simcoe_Muskoka
##   <chr>   <dbl>      <dbl>  <dbl>    <dbl>         <dbl>          <dbl>
## 1 Tota~  1.34e7     179083 113084   536917        398953         540249
## 2 Indi~  3.74e5       4795  15365    12135          9870          24110
## 3 Visi~  3.89e6       4655   2580   100055         70725          35055
## 4 Acco~  4.78e5       5545   4305    18325         14280          21600
## 5 Tota~  6.89e6      90755  57815   274390        202695         274225
## # ... with 9 more variables: Timiskaming <dbl>, Brant <dbl>, Sudbury <dbl>,
## #   Northwest <dbl>, Waterloo <dbl>, Durham <dbl>, Southwestern <dbl>,
## #   total_treat <dbl>, total_control <dbl>
```

```r
# Split the population info into our treatment and control groups.

populations_split <- matrix(ncol=4, nrow=5)
populations_split[,1] <- c("Total_Population",
                           "Indigenous_Population_25%_sample",
                           "Visible_Minority_25%_sample", "
                           Accommodation_and_Food_Services_25%_sample",
                           "Total_Population_Women")
populations_split[,2] <- ontario_pop

total_pop_treatment <- sum(populations[1,c(3:8)])
indigenous_treatment <- sum(populations[2,c(3:8)])
minority_treatment <- sum(populations[3,c(3:8)])
food_services_treatment <- sum(populations[4,c(3:8)])
women_treatment <- sum(populations[5,c(3:8)])

treatment <- c(total_pop_treatment,
               indigenous_treatment,
               minority_treatment,
               food_services_treatment,
               women_treatment)
populations_split[,3] <- treatment

total_pop_control <- sum(populations[1,c(9:14)])
indigenous_control <- sum(populations[2,c(9:14)])
minority_control <- sum(populations[3,c(9:14)])
```

```
food_services_control <- sum(populations[4,c(9:14)])
women_control <- sum(populations[5,c(9:14)])

control <- c(total_pop_control,
             indigenous_control,
             minority_control,
             food_services_control,
             women_control)
populations_split[,4] <- control

populations_split <- as.data.frame(populations_split)
colnames(populations_split) <- c("Info", "Ontario", "Treatment", "Control")

populations_split
```

```
##                                                                       Info
## 1                                                         Total_Population
## 2                                         Indigenous_Population_25%_sample
## 3                                              Visible_Minority_25%_sample
## 4 \n                            Accommodation_and_Food_Services_25%_sample
## 5                                                   Total_Population_Women
##    Ontario Treatment Control
## 1 13448494   1801335 1788702
## 2   374395     69035   84675
## 3  3885585    213570  297250
## 4   478070     64995   60550
## 5  6889105    916560  910715
```

```
### Get percentage proportions of demographic groups of interest

populations_split_percentage <- matrix(ncol=4, nrow=5)

for(i in 2:length(populations_split)){
  for(j in 1:5){
    populations_split_percentage[j,i] <-
      round(as.numeric(populations_split[j,i])/as.numeric(populations_split[1,i]),
            digits=3)
  }
}

#populations_percentage[1, 2:4] <- c(1, 1, 1)
populations_split_percentage[,1] <- c("Total_Population",
                                      "Indigenous_Population_25%_sample",
                                      "Visible_Minority_25%_sample",
                                      "Accommodation_and_Food_Services_25%_sample",
                                      "Total_Population_Women")

populations_split_percentage <- as.data.frame(populations_split_percentage)
colnames(populations_split_percentage) <- c("Info",
                                            "Ontario",
                                            "Treatment",
                                            "Control")

populations_split_percentage
```

```
##                                          Info Ontario Treatment Control
## 1                             Total_Population       1         1       1
## 2             Indigenous_Population_25%_sample   0.028     0.038   0.047
## 3                  Visible_Minority_25%_sample   0.289     0.119   0.166
## 4 Accommodation_and_Food_Services_25%_sample   0.036     0.036   0.034
## 5                       Total_Population_Women   0.512     0.509   0.509
```

We established expected percentages for the treatment and control regions based on census data ("Census Profile, 2016 Census" 2017).

15.6% of business owners are women in Canada ("SME Profile: Ownership Demographics Statistics" 2020).

Visible minorities are 25% of the population ("Number and Proportion of Visible Minority Population in Canada, 1981 to 2036" 2017).

12.2% of business owners are visible minorities ("SME Profile: Ownership Demographics Statistics" 2020).

Therefore, visible minorities are 0.49 as likely to own a business. Therefore, we can estimate that 0.058 of restaurant owners are a visible minority in the treatment group, and 0.081 in the control group.

Indigenous people have 3.7% service business ownership as opposed to 15.3% in the reference population ("Table a-1 Counts of Businesses Per 1,000 Residents by Province/Territory and Industry" 2019). They are 0.24 times as likely to own a service business in Ontario. Therefore, we can estimate that 0.009 of business owners in the treatment group are Indigenous, and 0.011 in the control group.

## 2.3 Survey 1

```r
# Randomly assign racial and ethnic minorities, and women, using the ratios above:

set.seed(19893)
survey_1 <- survey_1 %>%
  mutate(woman = sample(c("yes","no","nonanswer"),
                        size = num_total,
                        prob = c(0.154, 0.836, 0.01),
                        replace = TRUE))

set.seed(19893)
survey_1 <- bind_rows(
  survey_1 %>%
    filter(group == "treatment") %>%
    mutate(indigenous = sample(c("yes", "no", "nonanswer"),
                               size = num_treat,
                               prob = c(0.009, 0.981, 0.01),
                               replace = TRUE))
  ,
  survey_1 %>%
    filter(group == "control") %>%
   mutate(indigenous = sample(c("yes", "no", "nonanswer"),
                              size = num_control,
                              prob = c(0.011, 0.979, 0.01),
                              replace = TRUE))
)

set.seed(19893)
survey_1 <- bind_rows(
  survey_1 %>%
    filter(group == "treatment", indigenous != "yes") %>%
    mutate(visible = sample(c("yes", "no", "nonanswer"),
                               size = first(count(survey_1 %>%
                                                    filter(group == "treatment",
                                                            indigenous != "yes"))),
                               prob = c(0.057, 0.933, 0.01),
                            replace = TRUE))
  ,
  survey_1 %>%
    filter(group == "control", indigenous != "yes") %>%
   mutate(visible = sample(c("yes", "no", "nonanswer"),
                              size = first(count(survey_1 %>%
                                                   filter(group == "control",
                                                           indigenous != "yes"))),
                              prob = c(0.08, 0.91, 0.01),
                           replace = TRUE))
  ,
  survey_1 %>%
    filter(indigenous == "yes") %>%
    mutate(visible = "no")
)

survey_1
```

```
## # A tibble: 2,006 x 8
##        ID group     verify service_type disability woman indigenous visible
##     <dbl> <chr>     <chr>  <chr>        <chr>      <chr> <chr>      <chr>
##  1      2 treatment ON     both         no         no    no         no
##  2      7 treatment 3Y3    both         no         no    no         no
##  3      8 treatment 1K0    both         no         no    no         no
##  4     10 treatment 2J4    takeout      no         no    no         no
##  5     12 treatment 1E5    both         no         no    no         yes
##  6     18 treatment 1S0    both         no         yes   no         no
##  7     20 treatment 5S6    takeout      no         yes   no         no
##  8     22 treatment 7K6    dinein       no         no    no         no
##  9     24 treatment 2G3    takeout      no         no    no         no
## 10     25 treatment 6S4    takeout      no         no    no         no
## # ... with 1,996 more rows
```

## 2.4 Survey 2

```
# Randomly assign racial and ethnic minorities, and women, using the ratios above:

set.seed(19893)
survey_2 <- survey_2 %>%
  mutate(woman = sample(c("yes","no","nonanswer"),
                        size = num_total,
                        prob = c(0.154, 0.836, 0.01),
                        replace = TRUE))

set.seed(19893)
survey_2 <- bind_rows(
  survey_2 %>%
    filter(group == "treatment") %>%
    mutate(indigenous = sample(c("yes", "no", "nonanswer"),
                               size = num_treat,
                               prob = c(0.009, 0.981, 0.01),
                               replace = TRUE))
  ,
  survey_2 %>%
    filter(group == "control") %>%
   mutate(indigenous = sample(c("yes", "no", "nonanswer"),
                              size = num_control,
                              prob = c(0.011, 0.979, 0.01),
                              replace = TRUE))
)

set.seed(19893)
survey_2 <- bind_rows(
  survey_2 %>%
    filter(group == "treatment", indigenous != "yes") %>%
    mutate(visible = sample(c("yes", "no", "nonanswer"),
                               size = first(count(survey_2 %>%
                                                    filter(group == "treatment",
                                                           indigenous != "yes"))),
                               prob = c(0.057, 0.933, 0.01),
                            replace = TRUE))
  ,
  survey_2 %>%
    filter(group == "control", indigenous != "yes") %>%
   mutate(visible = sample(c("yes", "no", "nonanswer"),
                              size = first(count(survey_2 %>%
                                                   filter(group == "control",
                                                          indigenous != "yes"))),
                              prob = c(0.08, 0.91, 0.01), replace = TRUE))
  ,
  survey_1 %>%
    filter(indigenous == "yes") %>%
    mutate(visible = "no")
)

survey_2
```

```
## # A tibble: 2,006 x 8
##       ID group     verify service_type disability woman indigenous visible
##    <dbl> <chr>     <chr>  <chr>        <chr>      <chr> <chr>      <chr>
##  1     2 treatment ON     both         no         no    no         no
##  2     7 treatment 3Y3    both         no         no    no         no
##  3     8 treatment 1K0    both         no         no    no         no
##  4    10 treatment 2J4    takeout      no         no    no         no
##  5    12 treatment 1E5    both         no         no    no         yes
##  6    18 treatment 1S0    both         no         yes   no         no
##  7    20 treatment 5S6    takeout      no         yes   no         no
##  8    22 treatment 7K6    dinein       no         no    no         no
##  9    24 treatment 2G3    takeout      no         no    no         no
## 10    25 treatment 6S4    takeout      no         no    no         no
## # ... with 1,996 more rows
```

# 3 Shutdowns

# 4 Survey 2

43.2% of restaurants shut down temporarily due to the pandemic ("Impact of Covid-19 on Business or Organization Status, by Business Characteristics" 2021).

3% closed of restaurants in the US closed permanently between March and June (Sparks 2020). This period was about six times as long as our two week closure.

```
# Randomly assign shutdown using the ratios above only for the treatment group:

set.seed(19893)
survey_2 <- bind_rows(
  survey_2 %>%
    filter(group == "treatment") %>%
    mutate(shutdown = sample(c("none", "temporary", "permanent"),
                             size = num_treat,
                             prob = c(0.558, 0.432, 0.01),
                             replace = TRUE))
  ,
  survey_2 %>%
    filter(group == "control") %>%
    mutate(shutdown = sample(c("none", "temporary", "permanent"),
                             size = num_control,
                             prob = c(.99, 0.005, 0.005),
                             replace = TRUE))
) %>%
  arrange(ID)

survey_2
```

```
## # A tibble: 2,006 x 9
##       ID group  verify service_type disability woman indigenous visible shutdown
##    <dbl> <chr>  <chr>  <chr>        <chr>      <chr> <chr>      <chr>   <chr>
##  1     1 contr~ 1E4    both         no         no    no         no      none
##  2     2 treat~ 0N     both         no         no    no         no      tempora~
##  3     3 contr~ 6C7    takeout      no         yes   no         no      none
##  4     4 contr~ 1B8    both         no         yes   no         no      none
##  5     5 contr~ 1B0    both         no         yes   no         no      none
##  6     6 contr~ 2J9    both         no         no    no         no      none
##  7     7 treat~ 3Y3    both         no         no    no         no      none
##  8     8 treat~ 1K0    both         no         no    no         no      tempora~
##  9     9 contr~ 0H4    takeout      no         no    no         no      none
## 10    10 treat~ 2J4    takeout      no         no    no         no      tempora~
## # ... with 1,996 more rows
```

# 5 Revenue

## 5.1 Survey 1

In 2019, there were 25,836 restaurants and eating-places that were employers ("Canadian Business Counts, with Employees, December 2019" 2021) and 6,968 restaurants that were not employers ("Canadian Business Counts, Without Employees, December 2019" 2021) in Canada, for a total of 32,804 restaurants.

In the same year, full-service restaurants had $13,456,600,000 in revenue, and limited-service eating places had $14,082,700,000 in revenue ("Food Services and Drinking Places, Summary Statistics" 2021).

From this, we can calculate an estimated $69,959 average revenue per month.

From an analysis of Kaggle restaurant data ("Kaggle Restaurant Revenue Prediction" 2019), we can see that restaurant revenue looks like an F distribution. The distribution rf(n, df1, df2, ncp) = rf(num_rest, 10, 5) gives an approximate shape. Using the f-distribution formula, we know the distribution has a mean of 10/9, so we would multiply by 62,963 to get the mean to 69,959.

```
# Randomly sample from the scaled F distribution to generate restaurant revenues:

set.seed(19894)
survey_1 <- survey_1 %>%
  mutate(revenue = 62963 * rf(num_total,20,20))

survey_1
```

```
## # A tibble: 2,006 x 9
##       ID group   verify service_type disability woman indigenous visible revenue
##     <dbl> <chr>   <chr>  <chr>        <chr>      <chr> <chr>      <chr>     <dbl>
##  1     2 treatm~ ON     both         no         no    no         no        87879.
##  2     7 treatm~ 3Y3    both         no         no    no         no        47324.
##  3     8 treatm~ 1K0    both         no         no    no         no        52722.
##  4    10 treatm~ 2J4    takeout      no         no    no         no       108558.
##  5    12 treatm~ 1E5    both         no         no    no         yes       73843.
##  6    18 treatm~ 1S0    both         no         yes   no         no       106363.
##  7    20 treatm~ 5S6    takeout      no         yes   no         no        81652.
##  8    22 treatm~ 7K6    dinein       no         no    no         no        80808.
##  9    24 treatm~ 2G3    takeout      no         no    no         no        58740.
## 10    25 treatm~ 6S4    takeout      no         no    no         no        35745.
## # ... with 1,996 more rows
```

## 5.2 Survey 2

First, the revenues from the first survey will have a randomized factor (normal distribution with a mean of 1 and a standard distribution of 0.1) to add some random variance between months.

Second, places that are closed permanently will be assumed to experience -75% average revenue, and temporary shutdowns will be assumed to experience -50% revenue.

Revenues went to -35% during the first lockdown wave in March (Dixon 2020). We will assume that these numbers are accurate for a dine-in/takeout restaurant that is experiencing a takeout-only closure. They are closed 14 out of 31 days, so we will assume average losses of -17%.

We will assume that takeout-only restaurants will not have losses on average, and dine-in only places will have -100% revenue on average for the days they are closed. With closures 14 out of 31 days in the month, dine-in only will have average losses of -45%.

All businesses had a median decrease of -15%, but visible minority-owned businesses had a median decrease of -25% ("Business or Organization Revenue from August 2020 Compared with August 2019, by Business Characteristics" 2021). Therefore, an additional -12% will be applied to visible minority-owned restaurants.

There seems to be no significant difference in business losses for disabled owners or female business owners ("Business or Organization Revenue from August 2020 Compared with August 2019, by Business Characteristics" 2021).

```
# Randomly sample from the scaled F distribution to generate restaurant revenues, and apply a random nu
```

```r
set.seed(19894)
survey_2 <- survey_2 %>%
  mutate(revenue = 62963 * rf(num_total,20,20)) %>%
  mutate(factor = abs(rnorm(num_total, 1, .1))) %>%
  mutate(revenue = factor * revenue) %>%
  select(-factor)

# Reduce the treatment revenues to those that shutdown:

survey_2 <-
  bind_rows(
    survey_2 %>%
      filter(shutdown == "none") %>%
      mutate(revenue = revenue)
    ,
    survey_2 %>%
      filter(shutdown == "temporary") %>%
      mutate(revenue = 0.50 * revenue)
    ,
    survey_2 %>%
      filter(shutdown == "permanent") %>%
      mutate(revenue = 0.25 * revenue)
  ) %>%
  arrange(ID)

# reduce the treatment revenues for those with more dine-in and less take-out service:

survey_2 <-
  bind_rows(
    survey_2 %>%
      filter(service_type == "dinein" & group == "treatment") %>%
      mutate(revenue = 0.55 * revenue)
    ,
    survey_2 %>%
      filter(service_type == "both" & group == "treatment") %>%
      mutate(revenue = 0.83 * revenue)
    ,
    survey_2 %>%
      filter(service_type == "takeout" & group == "treatment") %>%
      mutate(revenue = revenue)
    ,
    survey_2 %>%
      filter(group == "control")
  ) %>%
  arrange(ID)

# reduce the treatment revenues further for racial and ethnic minority owned businesses:

survey_2 <-
  bind_rows(
    survey_2 %>%
      filter((visible == "yes" | indigenous == "yes")
             & group == "treatment") %>%
      mutate(revenue = 0.88 * revenue)
    ,
    survey_2 %>%
      filter(visible == "no" & group == "treatment")
    ,
    survey_2 %>%
      filter(group == "control")
  )

survey_2
```

```
## # A tibble: 2,006 x 10
##       ID group verify service_type disability woman indigenous visible shutdown
##    <dbl> <chr> <chr>  <chr>        <chr>      <chr> <chr>      <chr>   <chr>
##  1    12 trea~ 1E5    both         no         no    no         yes     tempora~
##  2    65 trea~ 4R6    takeout      no         no    yes        no      tempora~
##  3   116 trea~ 1N4    both         no         no    yes        no      tempora~
##  4   140 trea~ 1B0    takeout      nonanswer  nona~ no         yes     tempora~
##  5   147 trea~ 0A1    takeout      no         no    no         yes     tempora~
##  6   186 trea~ 1M0    both         no         no    no         yes     tempora~
##  7   260 trea~ 1L0    both         no         no    no         yes     tempora~
##  8   278 trea~ 4K1    takeout      no         no    yes        no      none
##  9   279 trea~ 0N     takeout      no         no    yes        no      none
## 10   302 trea~ 4B7    both         no         no    no         yes     tempora~
## # ... with 1,996 more rows, and 1 more variable: revenue <dbl>
```

# 6 Number of employees

$82,000 in yearly sales (gross revenue) per full-time equivalent employee ("2019 Restaurant Industry Factbook" 2019)

Therefore 6,833 monthly revenue per full-time equivalent employee.

In January 2021, there were 2187300 full time sales and service employees in Canada and 1164000 part time sales and service employees ("Average Usual Hours and Wages by Selected Characteristics, Monthly, Unadjusted for Seasonality (X 1,000)" 2021). Therefore there are 1.9 full time employees per part time employee. Part time employees in service and sales averaged 16.6 hours per week. We will assume that 2 PTE = 1 FTE, or 1 PTE = 0.5 FTE. Therefore, for every 1.9 + 1 employees, there are 1.9 + 1*0.5 FTE. For every 2.9 employees, there are 2.4 FTE.

2.4 FTE -> 1.9 FT + 1 PT.

$ 1.9 / (1.9 + 0.5) = 1.9 / 2.4 = 0.79 $

$ 1 / (1.9 + 0.5) = 1 / 2.4 = 0.42 $

## 6.1 Survey 1

```
# Use revenue to generate realistic numbers of employees for restaurants:

set.seed(19893)
survey_1 <- survey_1 %>%
  mutate(fte = revenue / 6833) %>%
  mutate(ft = round(fte * 0.79)) %>%
  mutate(pt = round(fte * 0.42)) %>%
  select(-fte)

survey_1 <- bind_rows(
  survey_1 %>% filter(ft < 1) %>%
    mutate(ft = 1),
  survey_1 %>% filter(ft >= 1)
)

survey_1
```

```
## # A tibble: 2,006 x 11
##       ID group verify service_type disability woman indigenous visible revenue
##    <dbl> <chr> <chr>  <chr>        <chr>      <chr> <chr>      <chr>     <dbl>
##  1     2 trea~ ON     both         no         no    no         no       87879.
##  2     7 trea~ 3Y3    both         no         no    no         no       47324.
##  3     8 trea~ 1K0    both         no         no    no         no       52722.
##  4    10 trea~ 2J4    takeout      no         no    no         no      108558.
##  5    12 trea~ 1E5    both         no         no    no         yes      73843.
##  6    18 trea~ 1S0    both         no         yes   no         no      106363.
##  7    20 trea~ 5S6    takeout      no         yes   no         no       81652.
##  8    22 trea~ 7K6    dinein       no         no    no         no       80808.
##  9    24 trea~ 2G3    takeout      no         no    no         no       58740.
## 10    25 trea~ 6S4    takeout      no         no    no         no       35745.
## # ... with 1,996 more rows, and 2 more variables: ft <dbl>, pt <dbl>
```

## 6.2 Survey 2

For the second survey, we will assume that treatment restaurants shifted towards part time random amounts–that is, the number of full time employees per part time employee dropped from 1.9. X is some number less than or equal to 1.9.

```
# Use revenue to generate realistic numbers of employees for restaurants:

set.seed(19893)
survey_2 <- survey_2 %>%
  mutate(fte = revenue / 6833) %>%
  mutate(ptincrease = 1.9 * (1 - abs(rnorm(num_total, 0, sd = 0.1)))) %>%
  mutate(ft = round(fte * (ptincrease / (ptincrease + 0.5)))) %>%
  mutate(pt = round(fte * (1 / (ptincrease + 0.5)))) %>%
  select(-fte, - ptincrease)

survey_2 <- bind_rows(
  survey_2 %>% filter(ft < 1) %>%
    mutate(ft = 1),
  survey_2 %>% filter(ft >= 1)
)

survey_2
```

```
## # A tibble: 2,006 x 12
##       ID group verify service_type disability woman indigenous visible shutdown
##    <dbl> <chr> <chr>  <chr>        <chr>      <chr> <chr>      <chr>   <chr>
##  1    12 trea~ 1E5    both         no         no    no         yes     tempora~
##  2    65 trea~ 4R6    takeout      no         no    yes        no      tempora~
##  3   116 trea~ 1N4    both         no         no    yes        no      tempora~
##  4   140 trea~ 1B0    takeout      nonanswer  nona~ no         yes     tempora~
##  5   147 trea~ 0A1    takeout      no         no    no         yes     tempora~
##  6   186 trea~ 1M0    both         no         no    no         yes     tempora~
##  7   260 trea~ 1L0    both         no         no    no         yes     tempora~
##  8   278 trea~ 4K1    takeout      no         no    yes        no      none
##  9   279 trea~ 0N     takeout      no         no    yes        no      none
## 10   302 trea~ 4B7    both         no         no    no         yes     tempora~
## # ... with 1,996 more rows, and 3 more variables: revenue <dbl>, ft <dbl>,
## #   pt <dbl>
```

# 7 Non-response

A professional mail survey with follow up found a non-response rate of 61% (Suominen et al. 2012). We will also assume that some people will mail in answers and fill in the form incompletely.

```
# Randomly select only a subset of all the generated responses to keep, simulating nonresponse:
# Also randomly select columns to delete in partial responses, and delete the data from those rows/colu

num_cols_1 = 11

set.seed(891)
survey_1 <- survey_1 %>%
  mutate(nonresponse = sample(c("yes","partial","no"),
                              size = num_total,
                              prob = c(0.6, 0.01, 0.39),
                              replace = TRUE)) %>%
  filter(nonresponse != "yes")

set.seed(891)
survey_1 <- survey_1 %>%
  mutate(which_partial = sample(c(4:num_cols_1),
                                size = first(count(survey_1)),
                                replace = TRUE))

for (i in c(1:as.numeric(count(survey_1)))){
  if (survey_1[i, "nonresponse"] == "partial") {
    survey_1[i, as.numeric(survey_1[i,"which_partial"])] <- NA
  }
}

survey_1 <- survey_1 %>%
  select(-nonresponse, -which_partial)

survey_1
```

```
## # A tibble: 784 x 11
##       ID group verify service_type disability woman indigenous visible revenue
##    <dbl> <chr> <chr>  <chr>        <chr>      <chr> <chr>      <chr>     <dbl>
##  1     2 trea~ 0N     both         no         no    no         no       87879.
##  2    12 trea~ 1E5    both         no         no    no         yes      73843.
##  3    22 trea~ 7K6    dinein       no         no    no         no       80808.
##  4    24 trea~ 2G3    takeout      no         no    no         no       58740.
##  5    27 trea~ 5K5    both         no         no    no         no       64689.
##  6    30 trea~ 2N2    both         no         no    no         no       75411.
##  7    51 trea~ 1A0    takeout      no         no    no         no      100323.
##  8    54 trea~ 1T7    both         no         no    no         no       44673.
##  9    64 trea~ 1K2    dinein       no         no    no         no      292752.
## 10    67 trea~ 3Y8    both         no         no    no         no       48750.
## # ... with 774 more rows, and 2 more variables: ft <dbl>, pt <dbl>
```

```
# Randomly select only a subset of all the generated responses to keep, simulating nonresponse:
# Also randomly select columns to delete in partial responses, and delete the data from those rows/colu

num_cols_2 = 12

set.seed(198)
survey_2 <- survey_2 %>%
  mutate(nonresponse = sample(c("yes","partial","no"),
                              size = num_total,
                              prob = c(0.6, 0.01, 0.39),
                              replace = TRUE)) %>%
  filter(nonresponse != "yes")

set.seed(198)
survey_2 <- survey_2 %>%
  mutate(which_partial = sample(c(4:num_cols_1),
                                size = first(count(survey_2)),
                                replace = TRUE))

for (i in c(1:as.numeric(count(survey_2)))){
  if (survey_2[i, "nonresponse"] == "partial") {
    survey_2[i, as.numeric(survey_2[i,"which_partial"])] <- NA
  }
}

survey_2 <- survey_2 %>%
  select(-nonresponse, -which_partial)

survey_2
```

```
## # A tibble: 795 x 12
##       ID group verify service_type disability woman indigenous visible shutdown
##    <dbl> <chr> <chr>  <chr>        <chr>      <chr> <chr>      <chr>   <chr>
##  1    12 trea~ 1E5    both         no         no    no         yes     tempora~
##  2   186 trea~ 1M0    both         no         no    no         yes     tempora~
##  3   302 trea~ 4B7    both         no         no    no         yes     tempora~
##  4   322 trea~ 4V1    both         no         no    no         yes     tempora~
##  5   385 trea~ 1L0    both         no         no    no         yes     tempora~
##  6   399 trea~ 1W0    both         no         no    no         yes     tempora~
##  7   437 trea~ 2M4    both         no         no    no         yes     tempora~
##  8   470 trea~ 1A0    takeout      no         no    no         yes     tempora~
##  9   541 trea~ T0N    both         no         no    no         yes     tempora~
## 10   651 trea~ 1P0    both         no         no    no         yes     tempora~
## # ... with 785 more rows, and 3 more variables: revenue <dbl>, ft <dbl>,
## #   pt <dbl>
```

```
# Save the two survey results tables:

write_csv(survey_1, here("outputs/data/survey_1.csv"))
```

```r
write_csv(survey_2, here("outputs/data/survey_2.csv"))
```

# References

“2019 Restaurant Industry Factbook.” 2019. National Restaurant Organization. https://restaurant.org/downloads/pdfs/research/soi/restaurant_industry_fact_sheet_2019.pdf.

“Average Usual Hours and Wages by Selected Characteristics, Monthly, Unadjusted for Seasonality (X 1,000).” 2021. Statistics Canada. https://doi.org/https://doi.org/10.25318/1410032001-eng.

“Business or Organization Revenue from August 2020 Compared with August 2019, by Business Characteristics.” 2021. Statistics Canada. https://doi.org/https://doi.org/10.25318/3310028101-eng.

“Canadian Business Counts, with Employees, December 2019.” 2021. Statistics Canada. https://doi.org/https://doi.org/10.25318/3310022201-eng.

“Canadian Business Counts, Without Employees, December 2019.” 2021. Statistics Canada. https://doi.org/https://doi.org/10.25318/3310022301-eng.

“Census Profile, 2016 Census.” 2017. Statistics Canada. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E.

Dixon, Vince. 2020. “By the Numbers: COVID-19’s Devastating Effect on the Restaurant Industry.” https://www.eater.com/2020/3/24/21184301/restaurant-industry-data-impact-covid-19-coronavirus.

“Download Yelp Dataset.” 2019. Yelp Inc. https://www.yelp.com/dataset/download.

“Food Services and Drinking Places, Summary Statistics.” 2021. Statistics Canada. https://doi.org/https://doi.org/10.25318/2110017101-eng.

“Impact of Covid-19 on Business or Organization Status, by Business Characteristics.” 2021. Statistics Canada. https://doi.org/https://doi.org/10.25318/3310027601-eng.

“Kaggle Restaurant Revenue Prediction.” 2019. http://www.mbel.io/2019/08/23/kaggle-restaurant-revenue-prediction/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

“Number and Proportion of Visible Minority Population in Canada, 1981 to 2036.” 2017. Statistics Canada. https://www.statcan.gc.ca/eng/dai/btd/othervisuals/other010.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

“SME Profile: Ownership Demographics Statistics.” 2020. Innovation, Science; Economic Development Canada. https://www.ic.gc.ca/eic/site/061.nsf/eng/h_03115.html.

Sparks, Hannah. 2020. “More Than Half of Us Restaurant Closures Are Permanent Due to Coronavirus: Study.” https://nypost.com/2020/06/30/more-than-half-of-us-restaurants-closed-due-to-covid-19-study/.

Suominen, Sakari, Karoliina Koskenvuo, Lauri Sillanmäki, Jussi Vahtera, Katariina Korkeila, Mika Kivimäki, Kari J Mattila, et al. 2012. “Non-Response in a Nationwide Follow-up Postal Survey in Finland: A Register-Based Mortality Analysis of Respondents and Non-Respondents of the Health and Social Support (Hessup) Study.” *BMJ Open* 2 (2). https://doi.org/10.1136/bmjopen-2011-000657.

“Table a-1 Counts of Businesses Per 1,000 Residents by Province/Territory and Industry.” 2019. Statistics Canada. https://www150.statcan.gc.ca/n1/pub/18-001-x/2019002/app-ann/app-ann-tbla1-eng.htm.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with Rmarkdown*. https://github.com/rstudio/bookdown.

Zhu, Hao. 2020. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.