# Individual Characteristics and Geographic Background Shape Fertility Rate

Group 43

16/10/2020

**Abstract**

The ageing population and the low fertility rate have been two significant issues faced by almost all developed nations over the past decades, interacting with each other and together posing a challenge to the major economies worldwide (Jones, 2020). Canada, one of the developed nations, also faced a rather severe decline in birthrate, as shown by the fact that the total fertility rate decreased from 1.68 children per woman to 1.54 in 2016, the lowest level observed since 2003 (Government of Canada, 2018). Therefore, we explored the dataset of "General Social Survey on Family" conducted in the year of 2017, hoping to gain more insights into the current situation of the declining fertility rate in Canada. Adopting a logistic regression model that reveals how some relevant characteristics of respondents influence their fertility intention, we found that age, education, gender, birthplace province, and region of residence shape their likelihood of entering into parenthood.

## 1. Introduction

Previous studies have shown that the rate of increase in Canada's fertility rate has remained negative since 2011 (Canada Fertility Rate 1950-2020), amplifying the negative impact of the ageing population on economic growth (Dartford, 2020). Unfortunately, the current health crisis of COVID-19 also negatively affects fertility intentions: "people are choosing, in large part, to delay, defer or just not have a child or additional children at this time" because of the uncertainty of the pandemic, according to the CEO of the Vanier Institute for the Family (Weikle, 2020). Several future negative consequences are associated with a low fertility rate. Under the estimation that people over 65 years old will account for more than one-quarter of the Canadian population, the tax burden will mostly fall onto young working-age Canadians – an issue will be further aggravated if the current fertility rate persists. With that said, adopting a proactive population policy that promotes optimal stable population became increasingly crucial to enable Canada to achieve prosperous sustainability (Record, 2019). Therefore, we limited our geographic focus to Canada, hoping to discover how some individual characteristics, such as age, income, education and so on, affect the probability of choosing to give birth. Our final goal is to advise relevant Canadian agencies, based on our analysis, on targeting a specific audience group that currently has the lowest fertility intention. Hoping to present our findings more methodically, we initially commented on the 2017 GSS study in which our analysis was based. Afterwards, we detailed the model used to reveal the relationships between the variable in interest – the fertility intention – and other independent variables aforementioned. After specifying the backgrounds, here comes the report's backbone: we presented the model results and interpreted the results with visualizations. Specifically, we founded that fertility intention tends to increase with education level but decrease with age. Moreover, people's birthplace and region of residence also shape their fertility intentions. Lastly, to add depth to our analysis, we outlined some weaknesses of our research and offered some future steps that can be taken to polish our findings further.

# 2. Data

## 2.1 The 2017 GSS study

### 2.1.1 General Background (Key features, Strengths and Weaknesses)

The General Social Survey (GSS) program, which was initially established in 1985, has the objectives to observe changes in Canadians' living standards and well-being and to gather information on specific social policies. One exciting feature of the GSS is that it focuses on distinct center topics in different years, ranging from health to social engagement. This report will focus on the 2017 General Social Survey (GSS), which concerns the role that family plays in people's lives. The 2017 GSS is designed with a cross-sectional nature – information recorded in the survey is measured by observing many individuals and households from February 2nd to November 30th 2017. Covering a large number of aspects of respondents' family information, the 2017 GSS is a vault containing virtually all information needed by scholars to investigate topics concerning families. Specifically, the 2017 GSS recorded fourteen sections, with each concerning a major aspect of the respondents, ranging from family origins, fertility intentions to subjective well-being. Numerous survey questions and answers are stored under each section. However, the precise nature of the GSS program can hurt the data accuracy because non-responses tend to increase with the length of the survey questionnaire. Moreover, all respondents are interviewed via telephone. Therefore, households without telephones are excluded from the target population, leading to an issue that the target population may deviate from the true population, that is all people who are over 15 years old in Canada with certain exceptions that will be discussed below.

### 2.1.2 Methodology

The target population – the population intended to be studied – under the 2017 General Social Survey contains all people who are over 15 years old in Canada (excludes residents of the Yukon, Northwest Territories and Nunavut and full-time residents of institutions). To study the target population, a sample frame - the source material from which a sample can be drawn – is required. In the 2017 GSS, the frame is a combination of the lists of telephone numbers (landline and cellular) in use available to Statistical Canada and lists of all dwellings within the ten provinces. After specifying an appropriate frame, a sample with a pre-determined size can be selected. Therefore, the sample in the 2017 GSS includes 20,602 individuals who have the characteristics mentioned above and responded to the questionnaire. Notice that all respondents are interviewed via telephone.

The 2017 GSS implemented a two-stage sampling method. Specifically, the survey designer first implemented stratified sampling and then conducted simple random sampling without replacement. Stratified sampling is a sampling method that divides the population into subgroups, which are also known as strata. Under the 2017 GSS, strata are designed under geographical consideration. The majority of densely inhabited Census Metropolitan Areas (CMAs) are each treated as a different stratum. The non-CMA areas of each of the ten provinces further formed ten strata. As a result, twenty-seven strata are created in total. Afterwards, individuals are randomly selected within each stratum.

Adopting a two-stage sampling is, in fact, a double-edged sword. On one side, a multistage sampling offers flexibility in that researchers can incorporate different sampling methods together. Moreover, researchers can reach the desired type of or size of groups with multistage sampling. On the other side, one can argue that flexibility brings arbitrariness (Verial, 2018). Namely, a multistage sampling involves more subjectivity. For instance, under the 2017 GSS, each stratum is constructed under geographic consideration. However, no detailed reasoning is given to back up such a stratification. Therefore, we cannot rule out the possibility that the survey designer chose geographic areas to form strata just for convenience purposes, an example of non-random sampling. Notice that non-random sampling refers to a process in which a sample is not drawn by random chance, leading to potential biases (Hayes, 2008).

### 2.1.3 Questionnaire Design

The questionnaire under the 2017 General Social Survey has fourteen general topics, covering the respondents' family origin, conjugal history, fertility intentions and many more. Each broad topic contains a mass of questions so that data analysts can pick data in interest to conduct corresponding studies. However, as mentioned above, detailed coverage results in a lengthy questionnaire format can potentially lead to the non-response bias. To address the non-response bias, questionnaire designers implemented several measures, which we will discuss in detail in the next section.

### 2.1.4 Changes Since the Last GSS in 2015

To increase the GSS study's accuracy, professionals have modified the survey design compared to the last GSS vision conducted in 2015. Specifically, with the improvement of digital communication, the usage of fixed-line calls shrinks. Therefore, the new frame under the 2017 GSS study is modified to include "cell phone-only" home, a growing population but not covered by the previous random digital dialing (RDD) frames[1]. Moreover, in the case where multiple contact numbers are in the record, interviewers are instructed to follow a priority system in which a fixed phone number has the highest priority, followed by the mobile phone number. Furthermore, interviewers no longer directly ask respondents about their income, which is generally considered a personal issue. Instead, conductors of the GSS study used tax information to estimate each respondent's income. By doing such, the survey designers not only increase the data accuracy but also decrease the non-response rate.

### 2.1.5 Dealing With Non-responses

The potential issue of non-response can significantly impact the accuracy of the final analysis. More specifically, with a large number of non-responses, the sample may deviate from the actual composition of the population intended to be studied, thus making a final finding groundless. Non-response occurs because of various reasons: poor survey designs with embarrassing questions asked, unsuitable delivery methods, prolix formats, etc. Deeply knowing the negative consequence of non-responses, the designers of the 2017 GSS implemented several measures that are discussed below:

1. Experts from Statistics Canada provide professional training to interviewers. Therefore, interviewers are familiar with interviewing techniques using CATI[2], survey knowledge and process.
2. Interviewers are provided with a manual to follow when calling the selected respondents.
3. Interviewers are instructed to explain the importance of GSS to the selected respondents, encouraging the selected respondents to participate in the survey.
4. Interviewers are instructed to reach the selected respondents up to two more times if they refused to participate in the first attempt.
5. The option of call back later at a convenient time is available to maximize the likelihood of having respondents participate in the survey.
6. Data experts impute missing data for critical variables based on respondents' answers in other survey questions (please see the next section for a detailed discussion).

Moreover, the 2017 GSS study also incorporates a three-stage adjustment system in dealing with non-responses[3]. In the first stage, adjustments apply to households for which no information is available. Afterwards, the GSS designers utilize the information available to Statistics Canada to model households' propensity to respond. Lastly, adjustments apply to the households for which only incomplete information

[1] A method for selecting people for involvement in telephone statistical surveys by generating telephone numbers at random

[2] Computer-assisted telephone interviewing (CATI) is a telephone surveying technique in which the interviewer follows a script provided by a software application

[3] Non-responding telephone numbers were grouped into three types: those with some auxiliary information available (in particular, a complete roster of household members), those with auxiliary information from various sources available to Statistics Canada and those with no auxiliary information.

is available. Using incomplete information, GSS designers again try to model households' propensity to respond. After the three-stage adjustment, non-responding telephone numbers are then dropped.

#### 2.1.6 Data Imputation

Data quality is essentially the foundation of performing any future statistical analysis. That said, missing or erroneous data is the biggest obstacle faced by survey designer and data analyst. The GSS program has tried its best to ensure the completeness and accuracy of data input derived from interviews with selected respondents. For instance, GSS designers adopted a data imputing method. Specifically, if some critical information is missing, such as sex and age, then an imputed answer based on respondents' answers in other relevant questions will be used as a substitute. Therefore, the edited data set of GSS contains no missing values for key variables. Similarly, in the case where respondents failed to provide information regarding the timing that they experienced a particular event, other relevant information, such as the date of the survey, age of the respondents, would be used to impute the best estimate. Adopting the data imputing method aims to provide future statisticians who will base their analysis on the GSS a complete data set. However, negative drawbacks are associated with such conduct, which we will discuss in-depth in the weaknesses section.

#### 2.1.7 Weight Adjustment

Weight adjustments are necessary when units are sampled with unequal probability (Lumley,2004) so that population estimates are consistent with the corresponding projected population counts. For instance, male respondents between 15 to 24 years old account for 3.7% of the sample while 7.5% population are males who are between 15 to 24 years old. Therefore, It is clear that the unweighted sample is not representative of the population we are looking at. With that said, several adjustments are made under the 2017 GSS study. First, given that GSS utilized stratification method and differences exist among strata, a stratum adjustment is introduced by adjusting the person weights for records within each geographic stratum. Moreover, the GSS also incorporates a province–age-sex adjustment to make sure the person weights match the projected province–age-sex distribution. Detailed formulas regarding the adjustments are beyond this report's scope and thus are included in the appendix.

### 2.2.2 Detailed data analysis

1. Age
   Age initially indicates the respondents' age when the survey was conducted, ranging from 15-year-old to 75-year-old. Please see figure below for the age distribution of the GSS respondents. For the convenience of comparing the age effect on the fertility rate, we performed a centering technique on the variable of age. Specifically, we subtract the average from each respondent's age and record the difference as a new variable name "ageC." Notice that the age distribution is highly right-skewed, meaning that we have lots of young and middle-aged respondents but a small number of elders. Specifically, the median age is 35 years old and 75% of all respondents are under 43 years old.

On the other hand, according to plot below, the age spread with respect to the binary outcome results (i.e. whether the respondents have fertile intention or not), follows an interesting curvy shape, providing us a hint in terms of model selection, which will be covered later in this report.

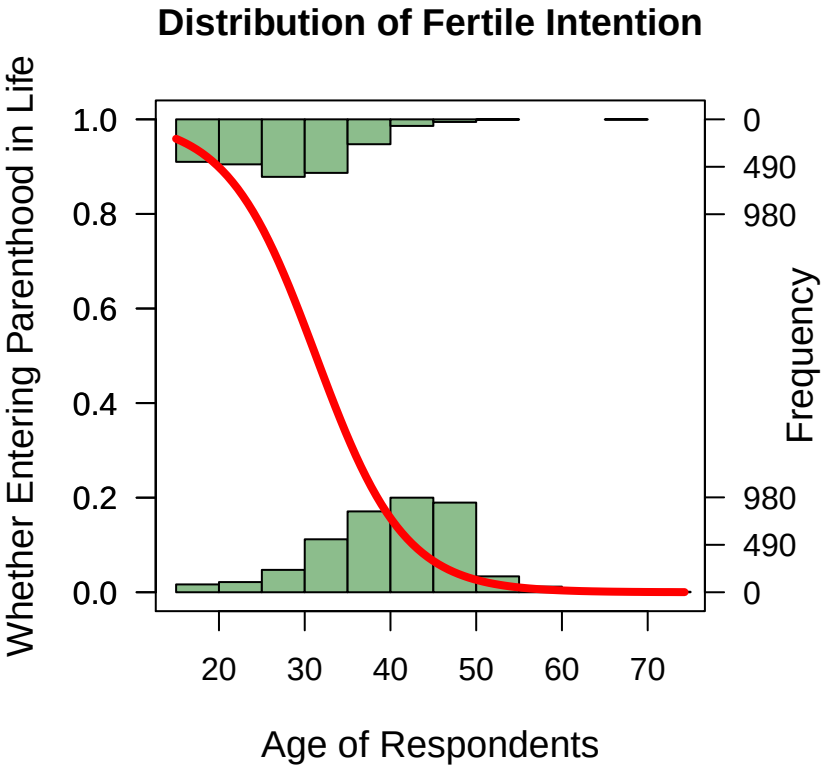



Figure 1: Logistic Plot with Age Spread

2. Education
   Education indicates the highest level of degree that the respondent obtained. Therefore, it is a categorical variable that can take on eight values: "High school diploma or a high school equivalency certificate", "Trade certificate or diploma", "Bachelor's degree (e.g. B.A., B.Sc., LL.B.)", "College, CEGEP or other non-university certificate or di...", "Less than high school diploma or its equivalent", "University certificate or diploma below the bachelor's level", "University certificate, diploma or degree above the bach...". For analyzing purposes, we removed the respondents with missing information on educational background. Below is a summary table that details the number of respondents for each education level. According to the table, we saw that respondents with a college degree or other non-university certificate formed the largest sub-group within our sample, followed by respondents with a high school degree and respondents with a bachelor's degree.

Table 1: Educational Background of the Sample Respondents

| education | n |
|---|---|
| Bachelor's degree (e.g. B.A., B.Sc., LL.B.) | 1382 |
| College, CEGEP or other non-university certificate or di... | 1634 |
| High school diploma or a high school equivalency certificate | 1463 |
| Less than high school diploma or its equivalent | 638 |
| Trade certificate or diploma | 540 |
| University certificate or diploma below the bachelor's level | 185 |
| University certificate, diploma or degree above the bach... | 515 |

3. Region of residence
   Region of residence indicates the region in which the respondent residences. With that said, it is a categorical variable that can take on five values: "Atlantic region", "Quebec", "Ontario", "Prairie region", and "British Columbia". Below is a summary table that details the number of respondents in each region of residence. According to figure below, we observed that not many respondents from British Columbia are included in the sample while the other four regions have roughly the same number of respondents.

Please see figure below to have a better sense of the composition of the respondents in terms of their region of residence

4. Personal income
   Personal income indicates the respondent's income, which is a categorical variable that can take on sixe values:"$25,000 to $49,999", "Less than $25,000", "$50,000 to $74,999", "$125,000 and more", "$75,000 to $99,999", and"$100,000 to $124,999". Below is a summary table that details the number of respondents in each income group. According to the plots, 1880 respondents in our data set earn less than $25,000 per year, forming the largest sub-group in terms of income information. On the other hand, there are only 269 respondents in our data set making $125,000 and over every year. To conclude, we have more respondents with low or medium income than wealthy respondents in the data set, which is consistent with the distribution of income in the population.

Table 2: Personal Income of the Respondents

| income_respondent | n |
|---|---|
| $100,000 to $ 124,999 | 332 |
| $125,000 and more | 269 |
| $25,000 to $49,999 | 1741 |
| $50,000 to $74,999 | 1326 |
| $75,000 to $99,999 | 809 |
| Less than $25,000 | 1880 |

5. Birthplace province
   Birthplace province indicates the region in which the respondent was born. With that said, it is a categorical variable that can take on ten values:"Newfoundland and Labrador", "Prince Edward Island", "Nova Scotia", "New Brunswick", "Quebec", "Ontario", "Manitoba", "Saskatchewan", "Alberta", "British Columbia", "Yukon / Northwest Territories / Nunavut", respondents lacking records of birthplace recorded were excluded. Below is a summary table that details the number of respondents born in each province. According to table, we saw that a large number of respondents are from Ontario and Quebec, 1705 and 1316 respectively. By contrast, only four respondents are coming from Yukon/Northwest Territories/Nunavut. Other provinces each contribute roughly the same number of respondents - 400 to 500 respondents for each province.

Table 3: Birth Place Orgins of the Respondents

| place_birth_province | n |
|---|---|
| Alberta | 513 |
| British Columbia | 569 |
| Manitoba | 371 |
| New Brunswick | 415 |
| Newfoundland and Labrador | 451 |

| place_birth_province | n |
|---|---|
| Nova Scotia | 425 |
| Ontario | 1705 |
| Prince Edward Island | 194 |
| Quebec | 1316 |
| Saskatchewan | 394 |
| Yukon / Northwest Territories / Nunavut | 4 |

6. Sex
   Sex is a categorical variable that can take on two values: male and female. For analyzing purposes, we transformed gender as a dummy variable, a variable can only be either 0 or 1. In our analysis, gender is 1 if the respondent is male or 0 otherwise. Approximately 55% of respondents in our data are female. The gender distribution is relatively balanced among the respondents, as there are 3019 male respondents and 3338 female respondents.

# 3. Model

## 3.1 Choice of a Logistic Regression

A logistic regression model (GLM) indicates a linear regression relationship between a binary response and one or more covariates, which are also called independent variables. The purpose of the model is to explore the relationship between a binary outcome response and a bunch of potentially influential factors. Through data cleaning, we have transformed the variable of fertility intention into a binary response, with 1 indicates that the respondents is planning to enter into parenthood and 0 otherwise. Therefore, a logistic regression model is a perfect fit given that we are looking for how some individual characteristics can influence respondents' fertility intention, which is a binary response.

As discussed in the previous section, we settled on six variables, including four categorical variables, one dummy variable of gender, and one numeric variable of age. For categorical or dummy variables, we do not consider the covariates' coefficients if these covariates do not match the respondent's situation. With that in mind, utilizing a logistic model, we can investigate how people in different "groups" react to the question of "do you currently have a fertility intention?" For instance, if we use people born in Ontario as a reference group, we can derive the difference of log odds between respondents born in Ontario and those born in Quebec through the variable coefficient that indicates whether the respondent was born in Quebec. For numeric variables, we can observe how the log odds change regarding the movement of the numeric variable, that is, how the log odds of having fertility intention respond to the changes in the respondents' age in our case. To conclude, a logistic regression model is indeed a perfect choice given that we are interested in investigating how individual characteristics of respendents shape their fertility intention.

## 3.2 Model Construction

Given that we are interested in testing whether the respondents' age, gender, economic status, education level, region of residence, and the birth place would influence their fertility intention, we used GSS data to construct a generalized linear model in a logistic set up.

The corresponding statistical model is shown below:

$$Y_i \sim Binomial(N_i, p_i)$$
$$log(\frac{p_i}{1-p_i}) = logit(p_i) = X_i\beta$$

Where,

- $N_i$ is the total sample size
- $Y_i$ is the number of event in interest happening - the number of respondents with parenthood plan sometimes in their life
- $p_i$ is the probability of the respondent expressing a birth intention
- $X_i$ is the vector of covariates (i.e. age, gender, personal income, birth place, region, and education level)
- $\beta$ represents the coefficient vector corresponding to each covariate

The model construction process involves a two-step procedure:

1. Conduct a regression associating fertility intention to respondents' age, gender, education, economic status, region of residence and region of birth. As shown below, for individual $i$, let $y_i$ equals to 1 if he or she plans to enter parenthood sometimes in life, and 0 otherwise.

$$Pr(y_i = 1) = \text{logit}^{-1}\left(\beta_{a[i]}^{age} + \beta_{e[i]}^{educ} + \beta_{b[i]}^{birthplace} + \beta_{r[i]}^{region} + \beta_{inc[i]}^{income} + \beta_{g[i]}^{gender}\right)$$

where the $\beta's$ are age-group, education, region of residence, personal income, birth province and gender, respectively. The notation $a[i]$ refers to the age-group $a$ to which respondent $i$ belongs.

2. Apply the estimation of the logistic regression to the general population weights in age, personal income, education level, region of residence from General Social Survey through post-stratifying.

## 3.3 Model Construction and Survey Design of the 2017 GSS

Given that the 2017 GSS study was conducted under a two-stage sampling method involving stratification and simple random sampling without replacement, we modified our model construction accordingly. As discussed in the previous section, each stratum is designed under geographical consideration under the 2017 GSS study. The majority of densely inhabited Census Metropolitan Areas (CMAs) are each treated as a different stratum. The non-CMA areas of each of the ten provinces further formed ten strata. As a result, twenty-seven strata are created in total. To reflect such a stratification, we utilized the survey package to make our model consistent with the GSS study. However, we instead only introduced ten strata; each represents an individual province because the original data set does not contain a variable that links each observation with the stratum it belongs. With that said, our model might not entirely reflect the survey method of the GSS study, but it is our best try with the data set we currently have in hand. To address that, we manually computed the population figures in 27 strata (see Appendix) based on the information available on the official website of Statistic Canada, hoping that we can find a way to use them as a benchmark to correct our stratification methods incorporated in the model in future studies.

## 3.4 Modelling Software

Analysis was conducted in R (R Core Team, 2019) and our logistic regression model were produced using the built-in function svyglm in R.

### 3.5 Assumptions of Logistic Regression

Running a logistic regression needs to fulfill five underlying assumptions:

1. Binary dependent variable
   In our case, the fertility intention is coded with "Yes, will consider having children sometimes in life" and "No, will not consider having children." Therefore, this requirement is passed.

2. Independent observations
   Given that each observation is a respondent who is over 15 years old and live in Canada and the same respondent will not be surveyed twice, this requirement is passed.

3. Linearity between covariates and the log odds
   The model assumes the log odds of the event in interest is linearly related to its predictors.

4. No multi-collinearity among the independent variables
   There should be no obvious correlations between the covariates of interests.

5. Large sample size
   A considerable size of data set with frequent measurements is indispensable for conducting a logistic regression model. The 2017 GSS study has a sample size of more than 20000 respondents. Therefore, this requirement is passed.

*Note:logistic regressions do not require a linear relationship between the outcome variables and its predictors. Therefore, the residuals do not need to follow a normal distribution; namely, data is not subject to the homogeneity of variance.

### 3.6 Alternative model (The Probit Model):

The Binary Response model is applicable when the dependent variable is a dummy variable; thus, the OLS method becomes unsuitable as it does not guarantee the value of a dependent variable to be within the unit interval.
Furthermore, the partial effect under the OLS method is constant by construction, while the partial effect on a binary dependent variable may not be constant. Given that we are interested in how fertility intention, a variable that can only take on a value of 1 if the respondent expressed a fertility intention and 0 otherwise, a binary response model is the ideal choice. The broad category of the binary response model contains two standard practices that researchers frequently use: the probit model, which uses the cumulative distribution fuction of the standard normal distribution, and the logistic model, which uses the cumulative distribution fuction of the logistic distribution. Both the probit model and the logistic model take a dummy variable as the dependent variable. The only difference is that the logit model uses a logit link function $f(\bar{y}) = ln(\frac{p}{1-p})$, while the probit model uses an inverse normal link function $f(\bar{y}) = \Phi^{-1}(p)$. Empirical evidence has shown that the difference in the model's overall results is usually slight to non-existent (Grace-Martin et al., 2020). According to a study conducted by Chen and Tsurumi in 2010, there are five criteria[4] to choose between the logit and the probit model; however, they also conclude that these five criteria do not return a decisive suggestion regarding the model choice for the majority of the time (Giles, 1970).
Therefore, the choice between the logit model and the probit model comes to depend on interpretation. The logistic model tends to offer a more intuitive interpretation of the coefficients because we can always

[4]1. The deviance information criterion (DIC). The predictive deviance information criterion (PDIC). 3. The unweighted sum of squared errors (USSE). 4. The weighted sum of squared errors (WSSE). 5.Akaike's information criterion (AIC).

transform log-odds into odds ratios through some easy mathematical operations. However, the probit model does not provide such an advantage because of its inverse normal link function mentioned above.

# 4. Result

### 4.1 Interpreting Logsitic Results

Table 4: Coefficient Table of Regression Linear Model

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -5.30 | 0.32 | -16.82 | 0.00 |
| place_birth_provinceBritish Columbia | 0.00 | 0.21 | 0.01 | 1.00 |
| place_birth_provinceManitoba | 0.12 | 0.20 | 0.62 | 0.54 |
| place_birth_provinceNew Brunswick | 0.59 | 0.29 | 2.03 | 0.04 |
| place_birth_provinceNewfoundland and Labrador | 0.40 | 0.28 | 1.43 | 0.15 |
| place_birth_provinceNova Scotia | 0.34 | 0.26 | 1.33 | 0.18 |
| place_birth_provinceOntario | 0.33 | 0.20 | 1.61 | 0.11 |
| place_birth_provincePrince Edward Island | 0.12 | 0.41 | 0.29 | 0.77 |
| place_birth_provinceQuebec | 0.38 | 0.28 | 1.38 | 0.17 |
| place_birth_provinceSaskatchewan | 0.07 | 0.19 | 0.35 | 0.73 |
| place_birth_provinceYukon / Northwest Territories / Nunavut | -10.19 | 0.68 | -15.09 | 0.00 |
| regionBritish Columbia | 0.67 | 0.23 | 2.92 | 0.00 |
| regionOntario | 0.38 | 0.21 | 1.79 | 0.07 |
| regionPrairie region | 0.51 | 0.21 | 2.43 | 0.01 |
| regionQuebec | 0.47 | 0.27 | 1.75 | 0.08 |
| ageC | -0.21 | 0.01 | -30.52 | 0.00 |
| educationCollege, CEGEP or other non-university certificate or di... | -0.28 | 0.11 | -2.56 | 0.01 |
| educationHigh school diploma or a high school equivalency certificate | -0.64 | 0.13 | -5.03 | 0.00 |
| educationLess than high school diploma or its equivalent | -1.58 | 0.19 | -8.41 | 0.00 |
| educationTrade certificate or diploma | -0.32 | 0.15 | -2.08 | 0.04 |
| educationUniversity certificate or diploma below the bachelor's level | 0.22 | 0.20 | 1.07 | 0.28 |
| educationUniversity certificate, diploma or degree above the bach... | 0.39 | 0.14 | 2.81 | 0.01 |
| income_respondent$125,000 and more | -0.08 | 0.25 | -0.31 | 0.76 |
| income_respondent$25,000 to $49,999 | 0.35 | 0.19 | 1.84 | 0.07 |
| income_respondent$50,000 to $74,999 | 0.13 | 0.19 | 0.69 | 0.49 |
| income_respondent$75,000 to $99,999 | 0.13 | 0.19 | 0.66 | 0.51 |
| income_respondentLess than $25,000 | 0.17 | 0.20 | 0.86 | 0.39 |
| is_male | 0.73 | 0.08 | 8.78 | 0.00 |

Table 5: 95% Confidence Interval of Logistic Regression Model Coefficients

| | 2.5 % | 97.5 % |
|---|---|---|
| (Intercept) | -5.92 | -4.68 |
| place_birth_provinceBritish Columbia | -0.40 | 0.41 |
| place_birth_provinceManitoba | -0.27 | 0.51 |
| place_birth_provinceNew Brunswick | 0.02 | 1.17 |
| place_birth_provinceNewfoundland and Labrador | -0.15 | 0.96 |
| place_birth_provinceNova Scotia | -0.16 | 0.85 |
| place_birth_provinceOntario | -0.07 | 0.73 |
| place_birth_provincePrince Edward Island | -0.68 | 0.92 |
| place_birth_provinceQuebec | -0.16 | 0.92 |
| place_birth_provinceSaskatchewan | -0.30 | 0.43 |
| place_birth_provinceYukon / Northwest Territories / Nunavut | -11.52 | -8.87 |
| regionBritish Columbia | 0.22 | 1.12 |
| regionOntario | -0.04 | 0.80 |
| regionPrairie region | 0.10 | 0.91 |
| regionQuebec | -0.06 | 1.00 |
| ageC | -0.23 | -0.20 |
| educationCollege, CEGEP or other non-university certificate or di... | -0.49 | -0.07 |
| educationHigh school diploma or a high school equivalency certificate | -0.88 | -0.39 |
| educationLess than high school diploma or its equivalent | -1.94 | -1.21 |
| educationTrade certificate or diploma | -0.62 | -0.02 |
| educationUniversity certificate or diploma below the bachelor's level | -0.18 | 0.61 |
| educationUniversity certificate, diploma or degree above the bach... | 0.12 | 0.66 |
| income_respondent$125,000 and more | -0.56 | 0.41 |
| income_respondent$25,000 to $49,999 | -0.02 | 0.72 |
| income_respondent$50,000 to $74,999 | -0.24 | 0.49 |
| income_respondent$75,000 to $99,999 | -0.25 | 0.51 |
| income_respondentLess than $25,000 | -0.22 | 0.56 |
| is_male | 0.57 | 0.89 |

According to coefficient table above, we could obtain the "baseline probability" from our fertile intention estimation model. The baseline probability is the estimated childbearing probability of Alberta female respondents who held a bachelor's degree and currently reside in Atlantic region, with personal earning ranges from $10000 to $12499. We estimate this probability to be 0.49% . We are 95% confident that this probability would be between 0.27% and 0.92%.

The age effect on the odds of having children was statistically significant: aging post a negative influence on the odds of having children. As the respondents grow older, they are less likely to plan for childbearing. For one year of age older than the average (i.e. 52-year-old), the respondent's log odds of childbearing would decrease by 0.21 as the respondent is one year older than the average.

Now, we shift the attention to gender difference. When the gender of the respondent is male, the odds of him having a fertility intention would increase approximately 1 time in comparison with female respondents with the same economic situation, education level, region of residence and birthplace. The estimated odds are considered to be statistically supported ($p<0.005$).

For respondents with various regions of residence areas, the odds of them entering parenthood would differ as well. British Columbia residents have significantly 1.95 times higher odds of childbearing than the ones with region in Alberta province. Our regression model also reveals that this increased percentage would be 66.5% in fertile odds when comparing Prairie residents with Atlantic region residents.

Noticeably, respondents born in Yukon province or Northwest Territories have 10.19 log odds of childbearing

times lower compared to the baseline group with Alberta as birthplace. If the respondents' birth origin change from Alberta province to New Brunswick, his or her odds of having a fertile plan in life will increase by 80%.

However, there is no enough statistical evidence supporting the hypothesis that the respondents with various economic status would have different odds in childbearing (p's >0.05). Large p-values indicate that the personal income factor did not have a significant contribution in estimating the odds of having a fertile plan.

Last, for the respondents that have different educational backgrounds, the odds of their fertility intention would vary. For university graduates who hold degrees above the bachelors, they tend to have 1.47 times higher odds of embracing parental identity, compared to the baseline bachelor group. While the respondents holding only high school diploma tend to have 47.3% lower odds of having children.

## 4.2 Model Checks and Diagnostic Issues

### 4.2.1 Logistic Regression Model in General

The expressiveness of the logistic regression model is restricted, since its pre-assumption separates the outcome variable (i.e. respondents' intention to childbearing) into completely two levels. While in real life circumstances, the partial population does hold an absolute attitude on entering parenthood sometimes in life. Logistic regression shares its pros and cons, it provides probabilities of the individual will have children, which makes the model more informative compared to other classification models.

### 4.2.2 Convergence in Logistic Regression

Likelihood maximization algorithm to converge would fail for estimating logistic regression model, which is indicated with non-existence under our context (Allison, 2008).

### 4.2.3 Cross Validation

Table 6: Confusion Matrix of Repeated Cross Validation

| | FALSE | TRUE |
|---|---|---|
| 0 | 2747 | 433 |
| 1 | 539 | 1367 |

For the purpose of evaluating our logistic regression model on fertile intention, one of the most common ways is to assess its accuracy in predicting the outcome variable under different subsets of population data. The technique performed is K-fold cross validation, with partitioning the current survey dataset into equally sized subsets. For evaluating the existing logistic regression model, we conducted a repeated five-fold cross validation. One subset data (i.e. fold) would be selected as testing set, while our logistic regression would run through the remaining folds to provide a prediction of the testing data. The testing-and-training process is repeated for five times. The predicted outcomes were tracked and were compared with the testing sets with known true outcomes. The model's performances on each repeated trial are shown through a table defined as a confusion matrix. As shown in the confusion matrix above, the accuracy of the model's prediction on the respondents' attitude towards whether children will have is 80%, which proves the predicting power of our logistic regression model.

### 4.2.4 Linearity of Quantitative Variable in Logistic Model

```
## Don't know how to automatically pick scale for object of type svystat. Defaulting to continuous.
```

```
## `geom_smooth()` using formula 'y ~ x'
```

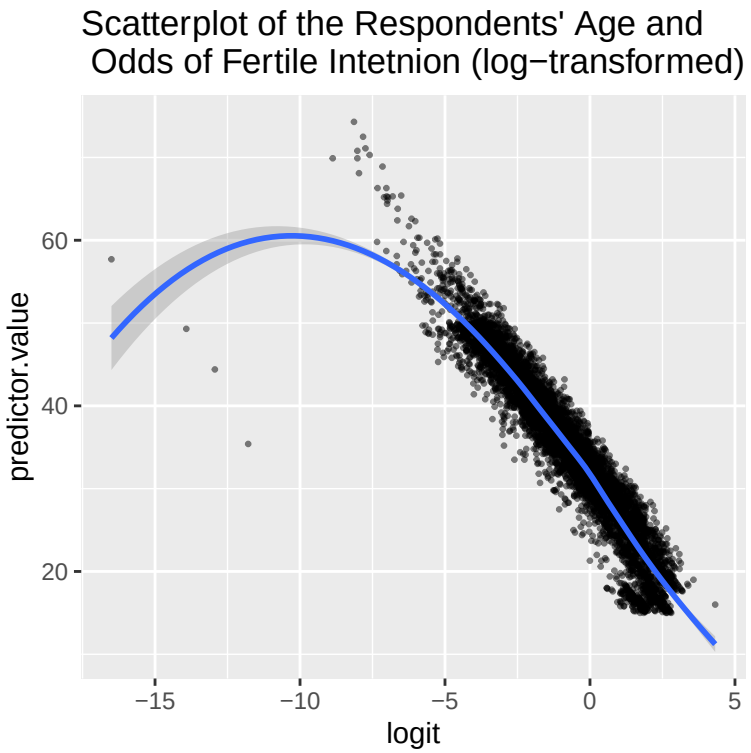



Figure 2: Linearly Association of Age and Fertile Intention Outcome (in logit scale).

To assess the linearity between our only numeric variable, age of the respondents, and the dependent variable of fertility intention, we drew a scatter plot to further visualize the situation. According to the plot above, we can see that most of the scatters form a approximately linear trend, suggesting that the assumption of a linearity between the continuous covariate and the log odds of the outcome variable was satisfied in our model.

#### 4.2.5 Multicolinearity Issues

Multicolinearity refers to the situation where the predictor variables in the logistic regression model are highly correlated. To assess this potential issue, variance inflation factors were computed using R, the result indicates two predictors, regions of residency and birth places of the respondents in the model are with relatively high correlation by the rule of thumb (VIF's >5). The solution to this problematic colinearity is through removing one of the concerned variable in the futrue model conduction.

# 5. Discussion:

## 5.1 Fertility Intention and Birthplace

For analyzing purposes, we will exclude respondents who failed to provide information regarding their birthplace. According to the regression result above, fertility intention is associated with the birthplace in which each respondent was born. Using Alberta as a base group, we saw that people who are born in New Brunswick have the highest fertility intention while respondents born in Yukon/Northwest Territories/Nunavut have the least fertility intention, holding all other factors (i.e. age, gender, income, etc.) constant. Some scholars

have argued that the low fertility intention within the three territories can be attributed to their underdeveloped local economies compared to the rest of Canada (Waddell, 2016). Specifically, people living in a less advanced economy may take the availability of future workplace opportunities into considerations when deciding on whether to enter into parenthood. Therefore, one potential solution to accelerate the birthrate in these three territories is to carry out policies concerning the local economic development.

On the other hand, we also saw that respondents born in Quebec are less likely to have fertility intentions compared to the rest of Canada. One potential explanation is the cultural difference in Quebec (United Nations Expert Group, 2017). Quebec, a former French colony, has a distinct cultural identity. Moreover, Quebecois are more gender-egalitarian compared to people from the rest of the country, as shown by the higher female labour participation rate since the "Quiet Revolution" in the late 1950s. With that said, one can argue that the low fertility intention for respondents born in Quebec is at least partially because female Quebecois devote more time in the workplace.

Last, according to our regression analysis, respondents from Ontario and the western provinces – British Columbia, Manitoba, Alberta, Saskatchewan, are more likely to express fertility intention. Such a result fits with our expectations. As indicated in many articles, the recent trend has been that aboriginal people are more likely to enter into parenthood because they want to pass on their unique cultural identity to the next generation (Kirkup, 2017). Ontario is, in fact, the province where the largest number of aboriginal people lived. Moreover, 60% of the entire aboriginal population lives in one of the four western provinces. Therefore, the high fertility intention in Ontario and the western provinces may be associated with their above-average population composition of aboriginal people.

## 5.2 Fertility Intention and Region of Residence

According to the regression, respondents who reside in the Atlantic region (a reference group) have the least fertility intention on average. In contrast, people currently living in Quebec are more likely to express fertility intention. Such a finding is rather interesting because we just discovered that born in Quebec tend to lower the people's willingness to enter into parenthood due to factors such as cultural issues. Now, we found out that living in Quebec has the opposite effect on people's fertility intentions. However, such a discrepancy can, in fact, be explained by the policies concerning giving birth in Quebec. Quebec has adopted a series of family-friendly policies to boost the fertility rate. Compared to the federal-level policy, the Quebec policies offer more extended maternity leave and higher paychecks during the maternity leave, cover more women, and even reserve some leave for fathers. Therefore, people who are living in Quebec, the actual benefit group of the family-friendly policies tend to express a higher fertility intention.

## 5.3 Fertility and Age

In our regression model, the younger the respondent is, the more likely that he or she expresses a fertility intention. Such a finding is not surprising, as a mass of online articles have indicated that the fertility rate is negatively affected by age (American College of Obstetricians and Gynecologists, 2014). However, we are not advising aged people to consider entering into parenthood or have additional children. In fact, we do recommend the opposite: the government should put more effort into influencing young people's birth-giving decision. According to a study conducted by Mikko Myrskyla and Andrew Fenelon, offspring born to mothers over 35 years old have worse outcomes in terms of mortality, height, obesity, the number of diagnosed conditions compared to that born to mothers aged between 25 to 34 years old (Myrskylä & Fenelon, 2012). ## 5.4 Fertility and Gender

Scholars have proposed that solving the low fertility rate requires paying attention to women's altitude (Golmakani, Fazeli, Taghipour, & Shakeri, 2015), which is consistent with our regression result. In our regression model, we discovered that males are more likely to express a fertility intention than females. Such a finding is what we expected before conducting the regression. Believe it or not, gender stereotypes still exist. In some cultures, people still consider that the responsibilities and duties of the family, including taking care of the baby, fall onto females. In contrast, males take on more professional responsibilities. With

that in mind, females need to consider more factors when deciding whether to enter into parenthood. For instance, they might need to sacrifice work time to the family and to the baby. Moreover, previous studies have shown that women are less likely to progress than male colleagues after childbirth (Smith, 2019). As a result, males are more likely to have a fertility intention as they do not need to worry about such things.

### 5.5 Fertility and Education

A Popular view has been that the fertility rate tends to decrease as the education level increases (Nairobi, Seoul and Torodi, 2019). However, we have seen a reverse relationship between fertility intention and education in our regression model. Specifically, people with a higher school diploma or less tend to express a lower fertility intention while the fertility intention is the strongest for people who hold a university degree or above. Such a finding is interesting, as it suggests that, after controlling age, income, gender, birthplace and region of residence, the fertility rate tends to move in the same direction as the education level, which seems contrary to the prevailing view. However, the prevailing view is derived from a global context, but our analysis has a geographical focus on Canada. Thus, the discrepancy is no longer a big surprise. Specifically, when extending the focus to a global context, researchers counter outliers that could potentially mess up the real relationship that in interest. For instance, an average woman in Niger, an impoverished country where the education system is not available, has seven children (Nairobi, Seoul and Torodi, 2019). By contrast, people in developed nations are generally educated and give birth to a more reasonable number of children. Therefore, countries like Niger can coercively turn the relationship between fertility rate and the education level into positive. With that in mind, given that our analysis has a geographic scope of Canada, our result is not contrary to the prevailing view. In fact, the topic of fertility and education is an interesting social study topic which all analysts found worth further investigation. Thus, we will conduct future studies that extensively focus on the relationship between fertility intention and education.

## 6. Weaknesses and Next Steps

We have discovered that fertility intention tends to increase with education level but decrease with income and age, after controlling other factors. Moreover, people's birthplace and region of residence also shape their fertility intentions. However, it is impossible to offer a uniform suggestion across all areas about boosting fertility intention because each province has its unique cause of low ertility intention. For instance, Yukon, Northwest Territories and Nunavut have suffered a low birth rate highly because of their poor local economies. That said, local government in these three territories might have to boost the economic prosperity before addressing the low fertility rate. Moreover, We have learned that Quebec's family-friendly policy indeed has a positive effect on influencing people entering into parenthood. Therefore, other provincial governments might need to adopt such a system. However, how to choose the target audience of these policies in each province brings further headaches. Unfortunately, there isn't a single formula to copy and paste for local governments. For instance, we discovered that the aboriginal people living in British Columbia have already exhibited strong fertility intentions by conducting online research. Therefore, the provincial government in British Columbia should make its policies more appealing to the non-aboriginal population.

Although, we have seen that, according to our model, fertility intention is associated with other individual characteristics of the respondents. The underlying relationships we discovered may not be 100% correct. First, we conducted our analysis based on the data set of the 2017 GSS study. Therefore, the accuracy of the original data source significantly influences our model results. Unfortunately, the GSS study utilized an imputed data method to void missing values appear for critical variables, a method that we have discussed in detail previously. With that in mind, it is possible that some of the values we incorporated in our model are imputed and may not reflect the real situation of some respondents. If that is the case, the validity of relationships that we have discovered is open to doubt. Second, we did not follow the stratification specified in the GSS study when constructing our logistic model because there isn't a variable that allows us to correctly link those 27 strata with each respondent. Therefore, we have tried our best to instead use ten provinces as ten strata. Although such modification will not impact the variable's coefficients, it influences standard

errors in our model. As a result, our confidence interval estimation might not be accurate enough. Though unachievable with the data set that we currently have in hand, we still manually compute the population figures within each of the 27 strata according to the official website of Statistic Canada (see Appendix), hoping that we can revise our model construction in the future.

# 7. Acknowledgement

Special thanks to Professor Rohan Alexandar at University of Toronto,for we have used part of his code to do the analysis.

# 8. References

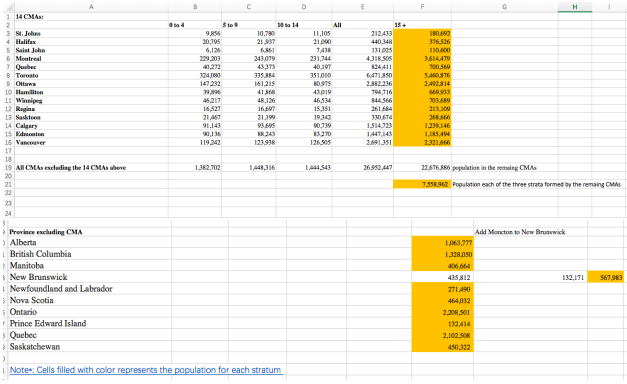

- American College of Obstetricians and Gynecologists. (2014, March). Female Age-Related Fertility Decline. Retrieved October 17, 2020, from https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2014/03/female-age-related-fertility-decline
- Assumptions of Logistic Regression. (n.d.). Retrieved October 16, 2020, from https://www.statisticssolutions.com/assumptions-of-logistic-regression/
- Atmathew. (2015, August 18). Evaluating Logistic Regression Models. Retrieved October 16, 2020, from https://www.r-bloggers.com/2015/08/evaluating-logistic-regression-models/
- Atmathew. (2015, August 18). Evaluating Logistic Regression Models. Retrieved October 18, 2020, from https://www.r-bloggers.com/2015/08/evaluating-logistic-regression-models/
- Canada Fertility Rate 1950-2020. (n.d.). Retrieved October 14, 2020, from https://www.macrotrends.net/countries/CAN/canada/fertility-rate
- Canada must keep an eye on the impact of our low fertility rate. (2019, May 27). Retrieved October 14, 2020, from https://www.therecord.com/opinion/letters-to-the-editors/2019/05/27/canada-must-keep-an-eye-on-the-impact-of-our-low-fertility-rate.html
- Cost of Raising a Child: The Key Elements to Consider. (2019, May 17). Retrieved October 17, 2020, from https://www.nbc.ca/personal/advice/budget/how-much-does-it-cost-to-raise-a-child.html
- Dartford, K. (2020, March 11). An ageing population: Europe's demographic crisis explained. Retrieved October 19, 2020, from https://www.euronews.com/2020/02/12/a-low-birth-rate-and-a-rapidly-ageing-population-europe-s-demographic-crisis-explained
- Giles, D. (1970, January 01). Choosing Between the Logit and Probit Models. Retrieved October 18, 2020, from https://davegiles.blogspot.com/2016/06/choosing-between-logit-and-probit-models.html
- Golmakani, N., Fazeli, E., Taghipour, A., & Shakeri, M. (2015). Relationship between gender role attitude and fertility rate in women referring to health centers in Mashhad in 2013. Retrieved October 17, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4387654/
- Government of Canada, S. (2018, June 11). Report on the Demographic Situation in CanadaFertility: Overview, 2012 to 2016. Retrieved October 14, 2020, from https://www150.statcan.gc.ca/n1/pub/91-209-x/2018001/article/54956-eng.htm
- Grace-Martin, K., Abdulfatah, Y., Keri, Chuol, G., Bgal, Brenneman, M., . . . Lele, U. (2020, September 29). The Difference Between Logistic and Probit Regression. Retrieved October 18, 2020, from https://www.theanalysisfactor.com/the-difference-between-logistic-and-probit-regression/
- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr
- Hayes, A. (2008, June 05). Sampling, Nonrandom. Retrieved October 17, 2020, from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405186407.wbiecs001
- Kirkup, K. (2017, October 25). High fertility rate, growing sense of self drives Canada's Indigenous population up, census finds. Retrieved October 16, 2020, from https://www.thestar.com/news/canada/2017/10/25/high-fertility-rate-growing-sense-of-self-drives-canadas-indigenous-population-up-census-finds.html

- Lavallée, P. & Beaumont, J.-F. (2015), Why We Should Put Some Weight on Weights. Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach, Invited article, Retrieved from https://surveyinsights.org/?p=6255
- Long JA (2020). jtools: Analysis and Presentation of Social Scientific Data. R package version 2.1.0, <URL: https://cran.r-project.org/package=jtools>.
- Markham, K. (2020, February 03). Simple guide to confusion matrix terminology. Retrieved October 18, 2020, from https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/
- Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret
- Myrskylä, M., & Fenelon, A. (2012, November). Maternal age and offspring adult health: Evidence from the health and retirement study. Retrieved October 17, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3881604/
- Sam Firke (2020). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.0.1. https://CRAN.R-project.org/package=janitor
- Sing T, Sander O, Beerenwinkel N, Lengauer T (2005). "ROCR: visualizing classifier performance in R." *Bioinformatics*, *21*(20), 7881. <URL: http://rocr.bioinf.mpi-sb.mpg.de>.
- Smith, J. (2019, October 22). Women less likely to progress at work than male colleagues after childbirth. Retrieved October 17, 2020, from https://workplaceinsight.net/women-less-likely-to-progress-at-work-than-their-male-counterparts-following-childbirth/
- Stubben, C.J. and Milligan, B.G. 2007. Estimating and Analyzing Demographic Models Using the popbio Package in R. Journal of Statistical Software 22:11.
- T. Lumley (2010) Complex Surveys: A Guide to Analysis Using R. John Wiley and Sons.
- Technology, A. (n.d.). Computing in the Humanities and Social Sciences. Retrieved October 19, 2020, from http://www.chass.utoronto.ca/
- Thanks to education, global fertility could fall faster than expected. (n.d.). Retrieved October 17, 2020, from https://www.economist.com/international/2019/02/02/thanks-to-education-global-fertility-could-fall-faster-than-expected
- Thomas Lumley ,2004,Analysis of complex survey samples, Department of Biostatistics University of Washington
- United Nations Expert Group. (2017, November 2). Regional variations in fertility trends and policies in Canada [PDF]. New York: United Nations Expert Group.
- Waddell, S. (2016, February 15). Yukon has low fertility rate, aging population. Retrieved October 16, 2020, from https://www.whitehorsestar.com/News/yukon-has-low-fertility-rate-aging-population
- Weikle, B. (2020, August 03). The COVID-19 pandemic is expected to lower the birth rate. Here's why that matters | CBC News. Retrieved October 14, 2020, from https://www.cbc.ca/news/health/covid-19-birthrate-1.5670539
- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595
- (n.d.). Retrieved October 16, 2020, from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710013501
- (PDF) Convergence Failures in Logistic Regression. (n.d.). Retrieved October 18, 2020, from https://www.researchgate.net/publication/228813245_Convergence_Failures_in_Logistic_Regression


# 9. Appendix

Full code and data supporting this analysis is available at: https://github.com/DuoyiJoy/sta304Assignment3

Twenty-seven Strata:



| 14 CMAs: | | | | | | |
|---|---|---|---|---|---|---|
| | 0 to 4 | 5 to 9 | 10 to 14 | All | 15 + | |
| St. Johns | 9,856 | 10,780 | 11,105 | 212,433 | 180,692 | |
| Halifax | 20,795 | 21,937 | 21,090 | 440,348 | 376,526 | |
| Saint John | 6,126 | 6,861 | 7,438 | 131,025 | 110,601 | |
| Montreal | 229,203 | 243,079 | 231,744 | 4,318,505 | 3,614,479 | |
| Quebec | 40,272 | 43,373 | 40,197 | 824,411 | 700,569 | |
| Toronto | 324,080 | 335,884 | 351,010 | 6,471,850 | 5,460,876 | |
| Ottawa | 147,232 | 161,215 | 80,975 | 2,882,236 | 2,492,814 | |
| Hamilton | 39,896 | 41,868 | 43,019 | 794,716 | 669,933 | |
| Winnipeg | 46,217 | 48,126 | 46,534 | 844,566 | 703,689 | |
| Regina | 16,527 | 16,697 | 15,351 | 261,684 | 213,109 | |
| Sasktoon | 21,467 | 21,199 | 19,342 | 330,674 | 268,666 | |
| Calgary | 91,143 | 95,695 | 90,739 | 1,514,723 | 1,239,146 | |
| Edmonton | 90,136 | 88,243 | 83,270 | 1,447,143 | 1,185,494 | |
| Vancouver | 119,242 | 123,938 | 126,505 | 2,691,351 | 2,321,666 | |
| | | | | | | |
| All CMAs excluding the 14 CMAs above | 1,382,702 | 1,448,316 | 1,444,543 | 26,952,447 | 22,676,886 | population in the remaing CMAs |
| | | | | | 7,558,962 | Population each of the three strata formed by the remaing CMAs |

| Province excluding CMA | | Add Moncton to New Brunswick | |
|---|---|---|---|
| Alberta | 1,063,777 | | |
| British Columbia | 1,328,050 | | |
| Manitoba | 406,664 | | |
| New Brunswick | 435,812 | 132,171 | 567,983 |
| Newfoundland and Labrador | 271,490 | | |
| Nova Scotia | 464,032 | | |
| Ontario | 2,208,501 | | |
| Prince Edward Island | 132,414 | | |
| Quebec | 2,102,508 | | |
| Saskatchewan | 450,322 | | |

Note*: Cells filled with color represents the population for each stratum

| Province excluding CMA and Census agglomerarions | | | | | |
|---|---|---|---|---|---|
| Alberta | 50,862 | 52,412 | 52,618 | 755,102 | 599,210 |
| British Columbia | 26,072 | 28,777 | 28,717 | 573,877 | 490,311 |
| Manitoba | 29,758 | 29,826 | 28,427 | 384,927 | 296,916 |
| New Brunswick | 11,957 | 13,096 | 13,419 | 283,340 | 244,868 |
| Newfoundland and Labrador | 8,567 | 9,939 | 11,193 | 238,321 | 208,622 |
| Nova Scotia | 13,395 | 14,489 | 15,400 | 320,355 | 277,071 |
| Ontario | 76,818 | 78,388 | 78,407 | 1,449,544 | 1,215,931 |
| Prince Edward Island | 2,917 | 3,059 | 3,273 | 60,515 | 51,266 |
| Quebec | 77,814 | 83,008 | 79,043 | 1,569,706 | 1,329,841 |
| Saskatchewan | 27,299 | 28,321 | 28,080 | 399,077 | 315,377 |
| | | | | | |
| Add back Census agglomerarions | | | | | |
| Alberta | 38,494 | 38,705 | 37,984 | 579,750 | 464,567 |
| British Columbia | 46,018 | 50,255 | 50,643 | 984,655 | 837,739 |
| Manitoba | 10209 | 10521 | 9494 | 139972 | 109,748 |
| New Brunswick | 9777 | 10777 | 11131 | 222629 | 190,944 |
| Newfoundland and Labrador | 970 | 3320 | 3630 | 70788 | 62,868 |
| Nova Scotia | 3220 | 9844 | 10667 | 210692 | 186,961 |
| Ontario | 56003 | 57678 | 59153 | 1165404 | 992,570 |
| Prince Edward Island | 4538 | 5247 | 5499 | 96432 | 81,148 |
| Quebec | 41629 | 46057 | 45010 | 905363 | 772,667 |
| Saskatchewan | 10763 | 10613 | 10635 | 166956 | 134,945 |

Appendix A

$$StratumAdjustment = \frac{Projected\ Population\ Count\ for\ the\ Stratum}{Sum\ of\ the\ Person\ Weights\ for\ the\ Stratum}$$

$$StratumAdjustment = \frac{Projected\ Province\ -Age-Sex\ Group\ Population\ Count}{Sum\ of\ the\ Provice\ -Age-\ Sex\ Group\ Person\ Weight}$$