Study of total number of children of Canadian people

Qingyue Gao, Jiarong Ye, Xin Zhao

October 19, 2020

Abstract

For decades, Canadian fertility rate had been following the global trend of rapid decline. The younger population leaned toward having fewer or even no children, challenging the traditional belief that children were a crucial component of a happy life. In this paper, we analyzed the 2017 General Social Survey (GSS) with the topic of "Family". Using the multiple linear regression model, we discovered that the survey data supported our hypothesis that there was a positive correlation between the age, the life satisfactory level and the number of children. We also discovered that there was a negative correlation between the education level of the responders and their number of children. Our analysis helped us conceptualize how different factors can affect a person's number of children and opinion on life. However, we had to emphasize that our results only show correlations, not causations.

Keywords: Canada, fertility, children, happiness, education, age

Introduction

The peak of the baby boom in Canada was reached in 1959 with an average of 3.94 children per woman - and then, it's all downhill from here. Over decades, many researchers have been studying on the global decline of fertility rate in modern society. Some of the historical events that could have major impacts on fertility rate included the drastic decline in child mortality, the increased representation of women in higher education and paid workforce, as well as the readily more effective and available contraception methods.

But how did this decline of fertility rate affect people's life quality? Traditional culture believed that parenthood is an invaluable component of a successful life, whereas the younger generations held much less taboo against a smaller family or even childlessness. Indeed, many researches on this topic have reached ambiguous or even contradictory results, with many of them agreeing that whether having children positively correlates with life-satisfaction highly depends on factors such as the income level, the age of parenthood, and the available welfare system.

In this report, we intended to focus on the Canadian population and to analyze the factors that affect the number of children a person has as well as his/her happiness level, so that we can learn about what variables are correlated to the fertility decline in Canada and how this decline might affect the well-beings of the population. The dataset used in this report was acquired from GSS (2017). The General Social Survey (GSS) program consists of a series of independent, annual, voluntary, and cross-sectional surveys, with the topic of 2017 being "Family". Detailed information about this dataset can be found in the Data section.

We first analyzed the basic trends of the number of children owned within the sample population, as well as the correlation between age and number of children owned. To quantify the strength of the correlation, we employed a multiple linear regression model to study how age, feelings about life, and education level of a person each correlates with the number of children that person has.

Data

Our data was chosen from the 2017 General Social Survey. The GSS gathers data to monitor changes in the living condition of the Canadian population, and to provide information on social policy issues of current or emerging

interest. For instance, the 2017 survey gives insights to concerns including the number of families in Canada, their respective characteristics and socio-economic conditions such as feeling about life, and so forth, with classification variables such as age, sex, education and income to help delineate population groups. The data is therefore deemed appropriate for this analysis.

Target population includes all persons 15 years of age and older in Canada, excluding residents of the Yukon, Northwest Territories, and Nunavut and full-time residents of institutions. Sampling Frame comprises two components: lists of telephone numbers in use (both landline and cellular) available to Statistics Canada from various sources (telephone companies, Census of population, etc.), and the Address Register (AR): list of all dwellings within the ten provinces chosen in the target population. The combination ensures a good coverage of all households with telephone numbers.

To perform sampling, each of the ten provinces were divided into strata (i.e. geographic areas), then a simple random sample without replacement of records was performed in each stratum. The target sample size is 20,000, while the actual number of respondents was 20,602. 91.8% of the selected telephone numbers reached eligible households (a household has to include at least one person 15 years of age or older, to be considered eligible).

Survey is performed via computer assisted telephone interviews (CATI). Those who at first refused to participate were re-contacted up to two more times to encourage participation. For the timing inconvenience of the call, an appointment was arranged to call back at a more convenient time. For cases in which there was no one home, numerous call backs were made. The overall response rate for the 2017 GSS was 52.4%.

Potential drawback includes non-sampling errors due to non-response in the survey, in the form of both partial non-response and total non-response. In most cases, partial non-response occurs when the respondent does not understand or misinterprets a question, refuses to answer a question, or can not recall the requested information.

Model

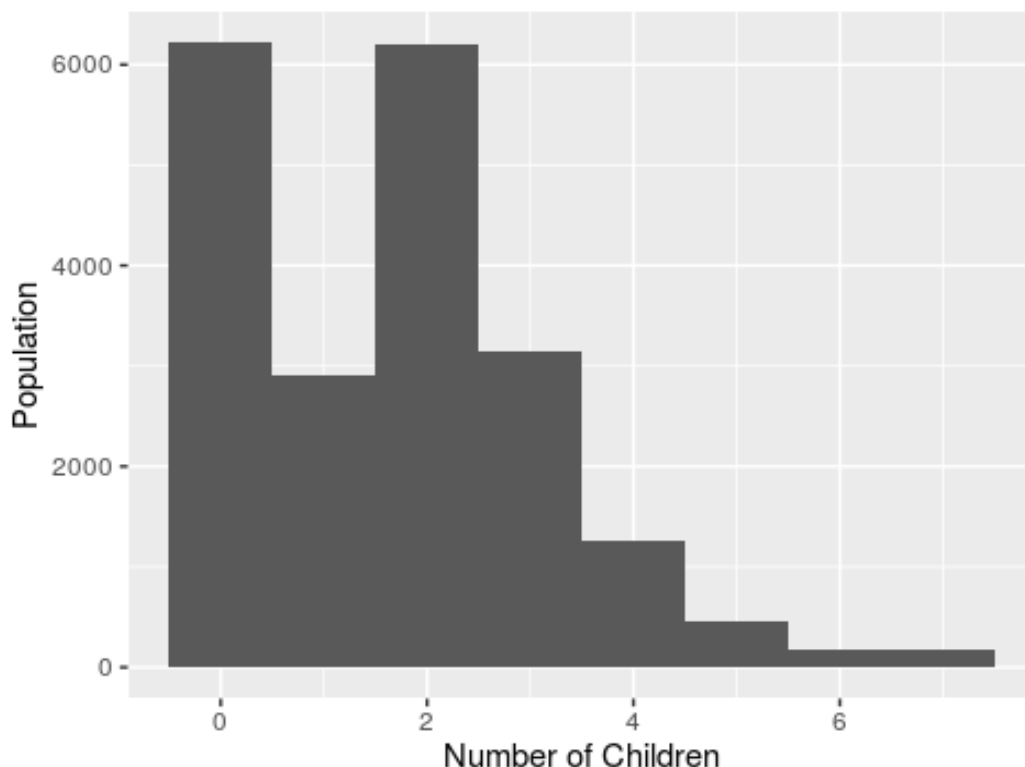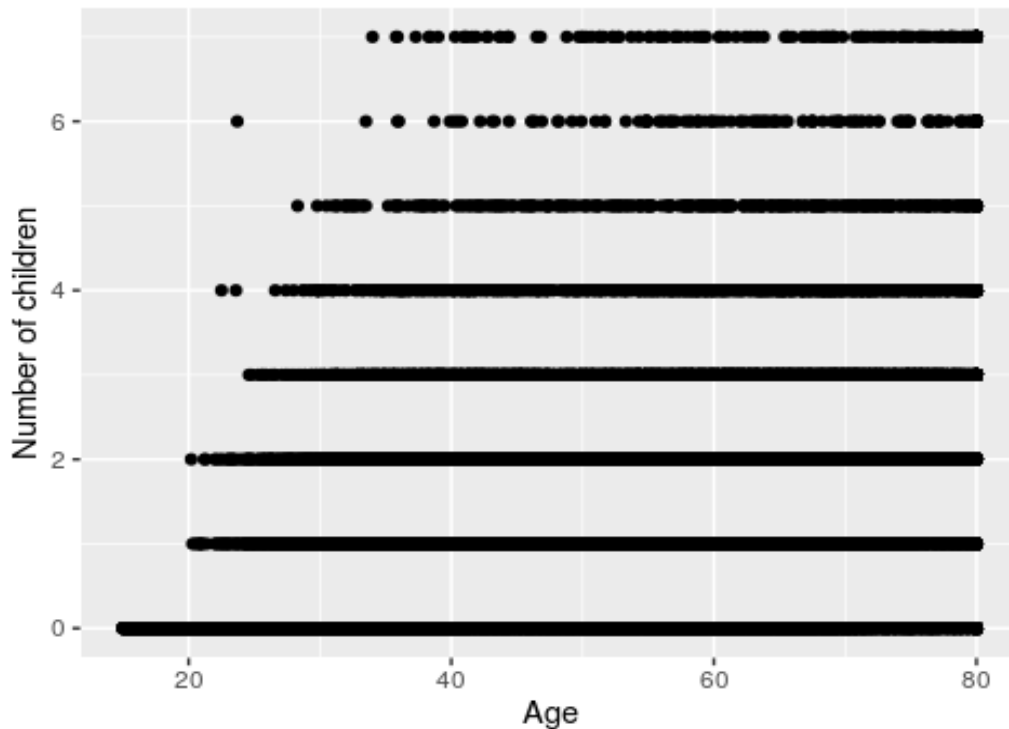Figure1: Histogram shows the the survey population grouped by the

Figure 2: Scatterplot showing the relationship between age and

```
##
## Call:
## lm(formula = total_children ~ age + feelings_life + education,
##     data = bb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9818 -0.8246 -0.1591  0.7890  6.0508
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.5427849  0.0589935  -9.201  < 2e-16 ***
## age             0.0368320  0.0005318  69.259  < 2e-16 ***
## feelings_life   0.0578064  0.0056971  10.147  < 2e-16 ***
## education      -0.2025739  0.0264768  -7.651 2.08e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.322 on 20035 degrees of freedom
##    (222 observations deleted due to missingness)
## Multiple R-squared:  0.2079, Adjusted R-squared:  0.2078
## F-statistic:  1753 on 3 and 20035 DF,  p-value: < 2.2e-16
```

Multiple linear regression model is implemented to find the relationship between total number of children, age, feelings about life, and gender (how the latter three variables influence the number of children).

$$Total.number.of.children = \beta_0 + \beta_1 \times age + \beta_2 \times Feelings.about.life + \beta_3 \times Education + \epsilon_i$$

Multiple linear regression model analyses relationships between multiple explanatory variables and the response variable, as numerous factors influence the total number of children that a respondent has, this model is appropriate for the situation. The model runs with R.

- β_0 is the estimated coefficients of intercept.

- β_1 is the estimated coefficients of slope of the age (explanatory variable).

- β_2 is the estimated coefficients of slope of the feelings about life (explanatory variable).

- β_3 is the estimated coefficients of slope of the education (explanatory variable).

- The unobservable error component ϵ accounts for the failure of data to lie on the straight line and represents the difference between the true and observed realization of the response variable.

Total number of children is defined as the total number of children reported by respondent, and it is the response variable (numerical variable). The number is capped at 7 children and more. Among 20,602 reponses, 10 respond "Don't know", 8 respond "Refusal", and 1 response is stated, therefore the data is sufficient to clean and carry out analysis.

Age is defined as the age of respondent at time of the survey interview, and it is an explanatory variable (numerical variable). This variable is capped at 80 years. Age is used instead of age group as the analysis tries to find the relationship between age and total number of children, while age gives more insight to the relationship than age groups. Among 20,602 responses, 19,421 reported age figures between 15 and 79, 1,181 reported 80 years and over. The data is sufficient to use.

Feelings about life is defined as: the feelings about life as a whole of the of respondent at time of the survey interview. The survey question is "Using a scale of 0 to 10 where 0 means 'Very dissatisfied' and 10 means 'Very satisfied', how do you feel about your life as a whole right now?" (feelings about life may influence people's willingness of giving birth, adoption, and other decisions on children parenting).

Education is an explanatory variable (binary categorical variable) with two responses "Less than high school diploma or its equivalent"(reports 0 in the model) or otherwise(reports 1 in the model). It is defined as: the highest certificate, diploma or degree that the respondent have completed at the time of the interview (attaining high school diploma and above may have significant influence over people's willingness of giving birth, adoption, and other decisions on children parenting, as high school and higher education cover important stage of sexual maturity and mental maturity of people, comparing the two groups with different education level will give insight to the relationship between higher education and total number of children).

Results

Histogram in figure 1 is right skewed, showing a general trend that the population size decreases as the number of children increases. The distribution is bimodal, which means most people have either no child or only 2 children.

Scatter plot in figure 2 shows a weak, positive, nonlinear pattern. The plot doesn't necessarily suggest a linear relationship; however, the positive pattern does show that people with higher age have relatively more children than the younger people. The outlier observed at x-axis of 24 and y-axis of 6 implies that a 24-year-old individual has 6 children, which is an usual data point within the sample (an outlier is an unusual value that is far from other data points).

From the first table, we can get equation: (education=1 if high school diploma or greater, education=0 if less than high school diploma)

$$Total.num\widehat{ber.of}.children = -0.5428 + 0.0368 \times age + 0.0578 \times Feelings.about.life - 0.2028 \times Education$$

The positive slope coefficient for both age and feelings about life ($\widehat{\beta_1}$ and $\widehat{\beta_2}$) suggests that the number of children increases with one's age and increases with one's feelings about life (people having better feelings about life would have relatively more children).

Education level has influence over the total number of children and is incorporated in the model as a categorical variable with the coefficient for education ($\widehat{\beta_3}$) being negative, suggesting that people in the group of high school education level on average have less children than those in group of not attaining high school education level.

The interpretation of the model is: when age increases by one unit, the average value of the total number of children increases by 0.0368 for both groups (people with high education level and people with low education level); when feeling about life increases by one unit, the average value of the total number of children increases by 0.0578 for both groups (people with high education level and people with low education level); the difference in average value of the total number of children between two groups (people with high education level and people with low education level) is 0.2025.

The standard error of age is 0.0005, the standard error of feeling about life is 0.0057, and the standard error of education is 0.0264. The standard errors for all three variables are small, which indicates that the sample mean is a considerably accurate reflection of the actual population mean. The p-value for the three variables, age, feelings about life and education are less than 0.05, so it rejects the null hypothesis. Therefore the parameter of slope, $\widehat{\beta_1}, \widehat{\beta_2}$ and $\widehat{\beta_3}$ are not 0 and there is a certain correlation between the total number of children and these three variables.

Discussion

The right skewed, bimodal histogram suggests that the average number of children of the population is between 1.5 and 2, with most people having either zero or two children. The large number of people having zero children is expected, as one of the weaknesses of our research is that the dataset surveyed people above 15 of age, there is no indication of the marital state of the population. It is rather surprising that the number of people having 2 children rival those having none, suggesting that having two children is not the norm in most families.

Availability of contraception enables people to have free will of choice in family planning, therefore children are given birth with careful consideration rather than unplanned pregnancy. The analysis is then able to hypothesize that having 2 children is considered appropriate or ideal within most of the Canadian families.

The results from the multiple linear regression model support that there exists statistically significant correlation between age, satisfaction about life, education level, and number of children. Among these variables, age and satisfaction about life are positively correlated with the number of children, while the education level is negatively correlated with the number of children; the negative correlation between education and number of children is also significantly stronger than the other positive correlations.

The positive correlation between age and number of children is of no surprise, given that older people have more time to "accumulate" children. The negative correlation between education and number of children is an interesting topic for further discussion. Recall that for this study, we divided education level into two categories, with people with high school or below degree as having "low" education level, and those with more than high school diploma as having "high" education.

People with higher education levels turn out to have significantly less children than people with low education levels, even though due to the sampling method of the survey the low education group contains some young highschool students that are not of legal age to have children yet. Possible explanations for this can be similar to how the increased participation of women in higher education greatly reduced the global fertility rate - better education and care leading to higher value of human capital per child, paired with a better understanding of modern birth control methods.

The correlation between feeling of life and number of children owned is also a topic that kindles many further discussions. Due to the nature of multiple linear regression model, or any non-experimental statistical models in general, we can only acquire information about the correlation between two variables, not their causation. The positive correlation coefficient and the small p-value suggests that people with higher opinion on life tend to have more children, but we cannot tell from the data whether their higher opinion on life encourages these people to have more children, or whether having more children raises the people's opinion on life. For all we know, there could be (and very like exists) other factors that play some determinant roles in both of these variables - for example, maybe richer people are more satisfied with their lives, while also happen to be better capable of affording large numbers of children.

Weaknesses

So here comes our first weakness. Because this is a statistical analysis on a survey, not a controlled environment, the results we acquire would only be able to show correlations and not causations. We have proved that there existed a positive correlation between content of life and number of children owned, but we cannot directly answer our questions on what factors contribute to the decreasing fertility rate and whether the fact of having more children alone will make you happier. At best, we can say that happier people also tend to have more children, and reason for which requires future studies to analyze.

There is also the problem of the sample population. The topic of the 2017 GSS is "family", not specifically "children", so the survey targeted all people that are 15 years or older in Canada. As a result, the sample population likely included a significant portion of underage people, which can be broadened to populations such as college students who were too young to ever consider the idea of having a child.

A weakness even more fundamental than population bias is that "feeling about life" is a very subjective concept, and is therefore evaluated in different ways among respondents. It is also subject to many temporary factors, where the respondent happened to "had a bad day". This inconsistency is shown in the result as the model has a R^2 value of 0.2079, suggesting that the model is not a good fit given the data.

Next Steps

In order to minimize the third-party factors such as family income and to study the sole effect of the number of children on a person's opinion on life quality, an interesting way demonstrated by Myrskyla and Margolis is to use panel data and study on how parental well-being changes as the children grow up and as they give birth to more children. Because the respondents are compared with themselves in different time frames, the confounding variable bias is relatively trivialized - although the subjectivity bias would still remain.

Talking about confounding variables, cultural opinions have significant influence over the number of children to give birth and when to give birth, while different cultures perceive the relationship between having children and happiness in very distinct ways. Since multiculturalism is promoted in Canada, a possible future direction is to incorporate cultural background into future analysis.

References

Statistics Canada, "General social survey (GSS) 2017: Cycle 31, Family". Retrieved on Oct. 2020. http://www.odesi.ca.

Caetano, Samantha-Jo, et al. "gss_cleaning.R" Rstudio Team. University of Toronto, 2020.

Roser, Max, et al. "Child and Infant Mortality." Our World in Data, 10 May 2013, ourworldindata.org/child-mortality.

Bailey, Martha J. "Fifty Years of Family Planning: New Evidence on the Long-Run Effects of Increasing Access to Contraception." Brookings Papers on Economic Activity, U.S. National Library of Medicine, 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC4203450/.

Kim, Jungho. "Female education and its impact on fertility." Ajou Univeristy, Korea. IZA World of Labor: 2016.228.

Government of Canada, Statistics Canada. "Fertility: Fewer Children, Older Moms." Government of Canada, Statistics Canada, 17 May 2018, www150.statcan.gc.ca/n1/pub/11-630-x/11-630-x2014002-eng.htm.

Leridon, Henri. "Demographic Effects of the Introduction of Steroid Contraception in Developed Countries." OUP Academic, Oxford University Press, 14 June 2006, academic.oup.com/humupd/article/12/5/603/778783.

Myrskylä, Mikko & Margolis, Rachel "Happiness: Before and After the Kids." MPIDR Working Paper WP 2012-013. Max Planck Institute for Demographic Research, Germany, Feb 2012.