

What variable affects the number of children in the household

Zean Hua, Ziqi Ling, Shijie Sun, Jiayi Zeng

2020/10/18

Abstract

The Canadian aging phenomenon has been gradually changing the Canadian population structure in recent decades. The structure change is mainly caused by which “Canadian fertility is below the replacement level.” [“Section 2: Population by Age and Sex.” Annual Demographic Estimates: Canada, Provinces and Territories. 27 Nov. 2015. Web. 16 Oct. 2020] In this analysis, we explore Canadian fertility along with the reasons behind the low fertility rate with the data from 2017 Canadian Canada’s General Social Survey (GSS) from the perspective of the respondent’s happiness level, education level, income, and residential province. This paper also includes data approach methodology, the statistical model interpretation, results in the analysis as well as discussion about results’ weakness, improvements and implication in reality.

Introduction

Background and Introduction to the Analysis

At first, The reason why we are used this data because our group was particularly interested in this data. This data seems very simple, but there is a lot of content behind it that we should think about. Besides, we believe that having a child is related to the child’s parents and two families. Canada is a multicultural country; under such an environment, people become especially important. The number of children in the family can predict Canada’s population future trend, but these are the directions for our future research. This report uses education, family income, feeling about life, age, and province to predict the total number of children in each household. The variables we look for are related to the number of children in the household. For example, family income, will income difference makes people more or less willing to have children? Will different provinces have different benefits to stimulate population growth? According to the Handbook of Population and Family Economics, “demand models predict, and empirical studies confirm that an increase in women’s wages and education have a more negative impact on fertility”.[Schultz, T. Paul. “Chapter 8 Demand for Children in Low-Income Countries.” Handbook of Population and Family Economics. Elsevier, 06 Apr. 2005. Web. 14 Oct. 2020.] Does this result still hold if we review the number of children respondents have with GSS data and statistical models instead of economic models?

Through model analysis and data interpretation, we found that it's universal to have two children in Canada. There is quite a proportion of households that have no child. From the distribution of the total number of children respondents have, there are only 1.79% of respondents have six to seven children. Based on the data, we predict that the number of Canadian citizens who have six to seven children is no more than 1.79% approximately. To explore the reasons behind the low fertility rate, we further analyze the life satisfaction level, education level, income level and residential province. A majority of Canadian citizens are satisfied with their life. High satisfaction reveals Canadian well-considerate social benefits provided to citizens, but it does set a burden to encourage people to change their lifestyle. Based on the distribution of life satisfaction level, we estimate that around 72% of Canadian citizens are eighty percent and more satisfied with their life. It is quite difficult for the government to encourage citizens to have more children since adding one child will significantly alter the lifestyle of a family. To maximize the effect upon stimulating fertility rate, the government can start with the citizens who is high school degree and low income. From the distribution of education level, we have found that respondents who are high school degreed are the largest population group under the category of no child, two children and three children. The government can focus on high-school-degreed citizens if we want to solve the aging phenomena sooner. In terms of income level, we predict that the citizens who are from the middle class tend to bear more children in their households. The middle class is financially sufficient and educational adequate. If the government provides a more customized and detailed child benefit and women healthcare to the middle-class family, the productivity to increase fertility rate may be appreciable. To better help the government target population, we provide residential provinces distribution. Based on the residential province distribution, there is 27.3% of respondents from Ontario. According to the statistical data, we conclude that the Ontario population shall be the government targeted population because the largest proportion of our respondents are from Ontario. If the government chiefly perfects Ontario residents' social benefits, which are about child benefits and women's healthcare benefits. Based on this analysis, we suggest the government target middle-class, middle-satisfied, high-school-degreed Canadians to effectively increase the Canadian fertility rate.

Weaknesses

Our research database is from Canada's General Social Survey (GSS) program. The GSS program collects data through telephone interviews and internet questionnaires, along with the implication of Computer Assisted Telephone Interviewing (CATI). This data approach has advantages as well as weaknesses. The cost is low, but the coverage of respondents is limited. The use of the internet and telephone leads our data to not be able to concentrate on certain groups, which are elder and low-income. Although the number of cell-phone-only households has increased since cellphone gradually became everyone's necessity product, we still cannot ignore that there is a certain percentage of Canadians who do not have a cell phone. By collecting data from the internet and telephone, the samples collected have excluded certain population groups that need the society's extra attention the most.

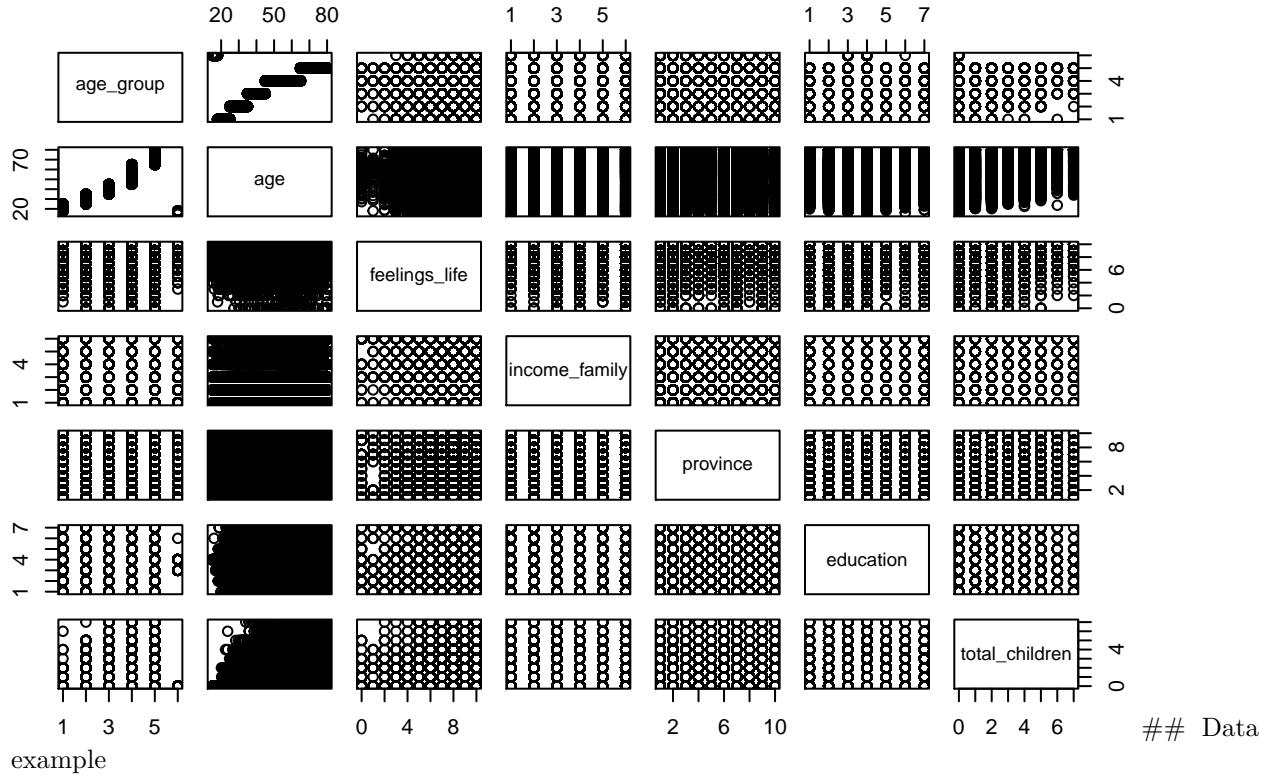
Another weakness of prediction is non-response bias. This may cause the survey's length and time-consumption. In our response, there is a large percentage of respondents answer valid skip. The non-response bias drops our valid response rate. Statistics Canada enhance the valid response rate by reviewing the respondents who did not provide a valid answer when they first responded to the survey. In 2017, the response rate reached over fifty percent.

With this analysis, we can better understand the reason behind the Canadian low fertility rate. From the distribution of respondents' residential province, this paper helps the Canadian federal and provincial governments better target their direction to improve childbearing benefit and education benefit. Our analysis can help determine the question of whether the Canadian government shall raise the Canadian Child Benefit (CCB) since aging phenomena has been a social concern over the years. If aging phenomena cannot be alleviated, the Canadian economy will be stroked sooner or later.

Data

The raw data

```
## [1] 20602     81
```



```
##   age_group age feelings_life      income_family province
## 1    45-65  45.65  52.7          8 $25,000 to $49,999 Quebec
## 2    45-65  45.65  51.1         10 $75,000 to $99,999 Manitoba
## 3    45-65  45.65  63.6         8 $75,000 to $99,999 Ontario
## 4    over 65 65.00  80.0        10 $100,000 to $ 124,999 Alberta
## 5    25-35  25.35  28.0         8 $50,000 to $74,999 Quebec
## 6    45-65  45.65  63.0         9 $50,000 to $74,999 Quebec
##                                         education total_children
## 1 High school diploma or a high school equivalency certificate           1
## 2                               Trade certificate or diploma                   5
## 3             Bachelor's degree (e.g. B.A., B.Sc., LL.B.)                  5
## 4 High school diploma or a high school equivalency certificate           1
## 5 College, CEGEP or other non-university certificate or di...              0
## 6 High school diploma or a high school equivalency certificate           2
```

Source of the Data

Our dataset is based on GSS 2017 and cleaned by Rohan Alexander and Sam Caetano. GSS is a General Social Survey which focuses on the families in Canada. The data of this paper contains total_children, feelings_right, education, income_family and province. The next paragraph will further explain the five variables. The data is originally collected and published by Statistics Canada. The 2017 GSS collected by Statistics Canada is our dataset. The survey is conducted from February to November in 2017 with a sample survey that is a cross-sectional design.

Variables in our Selected Dataset

The selected variables we used in our dataset are total_children, feelings_life, education, income_family and province. The variables are renamed after Rohan Alexander and Sam Caetano clean the code. The definition of each variable is explained below.

total_children: It originally called TOTCHDC. The variable measures the number of children in each households of our respondents.

feelings_life: It is originally called SLM_01. It is the variable to measure the life satisfaction level.

Education: It is originally called ESC1_01. It is variable to measure respondents' education level (whether respondents are currently working towards which education degree and diploma).

income_family: It is originally called TTLINCG2. The variable categorizes people into different income and aims to find out which income status the respondent is from.

province: It is originally called PRV. This variable indicates which province respondents is from.

Population and Sample, Methodology and Approach used to Collect and Process Data

The target respondents of GSS2017 include Canadian citizens who are aged 15 and above, excluding full-time institution workers and respondents in Yukon, Northwest Territories, Nunavut. The respondents are selected from the target population. Our Sampling population provide responses based on questionnaires. Statistics Canada collects data by distributing questionnaires to the sampling population. The survey is divided into fifteen different sections: entry component, respondent's date of birth, family origins, leaving the parental home, conjugal history, intentions and reasons to form a union, respondent's children, and other eight questions. The sampling strategy that Statistics Canada used is simple random sampling without replacement. Such a method allows interviewers to complete the questionnaire in numerous attempts with our randomly selected respondents. The survey frame uses two different components. One component is the listed telephone number in use (landline and cellular) available to Statistics Canada from various sources (for example, telephone companies and census of the population). The other component is the address register (AR), which is the list of dwellings within ten provinces in Canada. AR is the method that is used to group the telephone numbers, which are associated with the same valid addresses together. This approach to the respondents, created in 2013, combines telephone numbers (landline and cellular) with Statistics Canada Register and collects data via computer-assisted telephone interviews (CATI). Thus, the respondents are reached throughout phone calls with the choice of the language spoken. This sampling approach performs well in response coverage. However, it is time-consuming since there are millions of telephone numbers and addresses across ten provinces. Besides, this survey method needs less labour and the cost to hire people conducting telephone interviews will be a relatively large amount, but much cheaper than the traditional way of survey sampling. Other than the labour cost, Statistics Canada needs to attain telephone numbers from communication companies (Bell, etc.). The access fee will be costly, as well.

The Key Features, Strengths and Weaknesses about the Survey.

The key feature of the survey is its distribution method. The procedure to collect the data is completed by answering phone calls with computers' assistance. This is different from the traditional way of sampling by distributing printed questionnaires. Through such a distribution method, an advantage is a wide coverage over the target population and sample randomization. Everyone from our target population is possibly selected as one sample of GSS.

The weaknesses of this survey are the following. First, as respondents are able to complete the survey within a reasonable range of times while their former answers will not be replaced, the true number of sample size is underestimated. The sample size is around twenty thousand (without the replacement of the situation of answering the survey many times), which is not large enough to represent and predict a country's situation. Secondly, there is non-response bias, as the survey gives out to the respondents. One possible reason behind the non-response is the interview's length. If the interview is quite time-consuming with a long question, selected respondents may not have enough patience and time to complete all. In order to improve the response rate, Statistics Canada starts to re-contact respondents who did not complete the interview and explain the interview's importance to Canadian citizens and government. For respondents who are unavailable to answer the interview, Statistics Canada provides appointments as an option. Gradually, the Canadian response rate of GSS in 2017 has reached more than 50%. There are also adjustments applied to the non-responding telephone numbers; non-responding telephone numbers are dropped after they are weighted by responding to telephone numbers.

Model

Variance Component Model

Variance components model is a multilevel model without explanatory variable which could be used explain the variation in the number of children can be attributed to the provinces in which the respondents live, it has a form as following:

(Null Model in R code):

$$ChildrenNo_{ij} = \beta_0 + \mu_j + \epsilon_{ij}$$

where $ChildrenNo_{ij}$ is the total number of children for i th respondent living in j th province, μ_j is the deviation of BMI values of j th province from average, ϵ_{ij} is the error of modeling for the i th respondent living in j th province.

For testing the significance of a Multilevel Model, we compare it with a linear model only with an intercept term

(single model in R code):

$$ChildrenNo_{ij} = \beta_0 + \epsilon_{ij}$$

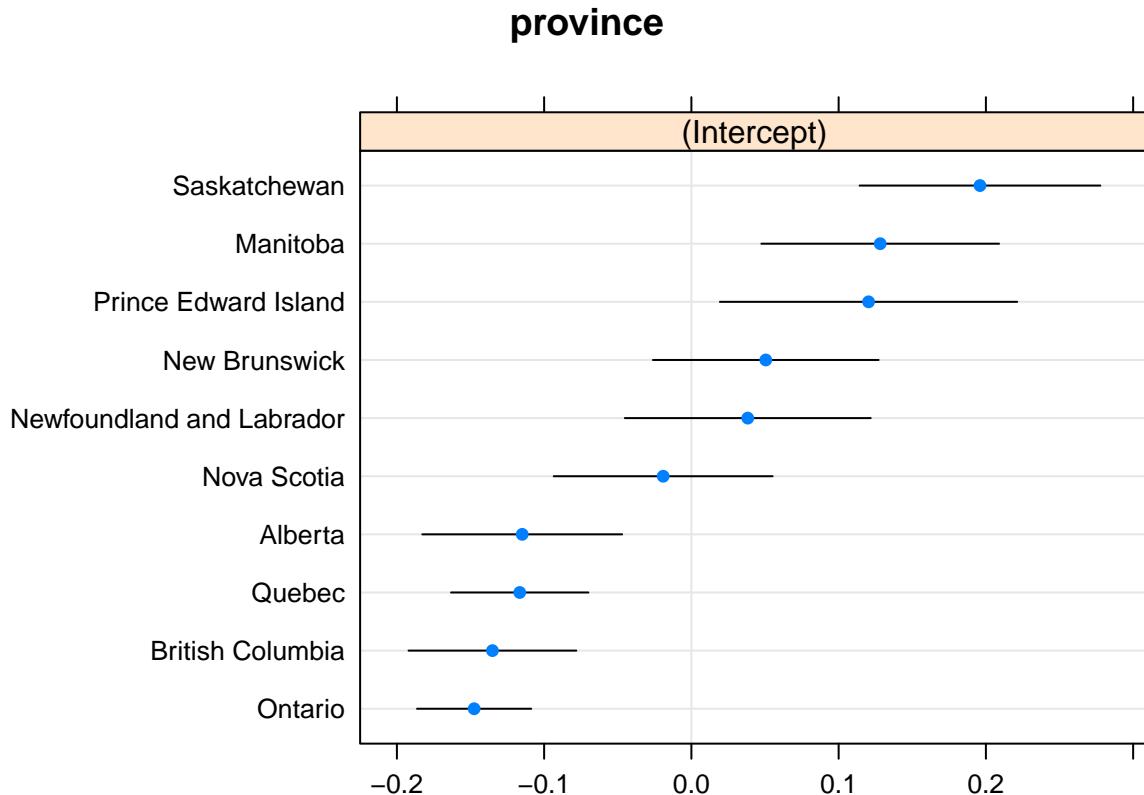
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ (1 | province)
##   Data: df_new
##
##      AIC      BIC  logLik deviance df.resid
##  72647.2  72670.9 -36320.6  72641.2     20036
##
## Scaled residuals:
##    Min     1Q  Median     3Q    Max
## -1.3026 -1.0707  0.1458  0.7227  3.6549
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   province (Intercept) 0.01544  0.1243
##   Residual            2.19427  1.4813
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.73356   0.04119  42.09
##
## Data: df_new
## Models:
## singlmodel: total_children ~ 1
## nullmodel: total_children ~ (1 | province)
##           npar  AIC  BIC logLik deviance Chisq Df Pr(>Chisq)
## singlmodel    2 72737 72753 -36366     72733
## nullmodel     3 72647 72671 -36321     72641 91.58  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

## $province
##                               (Intercept)
## Alberta                  -0.11488617
## British Columbia          -0.13511077
## Manitoba                 0.12822162
## New Brunswick             0.05051971
## Newfoundland and Labrador 0.03828181
## Nova Scotia               -0.01916776
## Ontario                  -0.14755168
## Prince Edward Island      0.12030719
## Quebec                   -0.11658388
## Saskatchewan             0.19596993
##
## with conditional variances for "province"

```

```
## $province
```



The testing results shows that the multilevel model does show a significant improvement in model fitting than a single level model.

Random Intercept Models

To explore the difference between provinces in total number of children after controlling on fixed level effects, we perform Random Intercept Model by adding variables into Variance Components model. The random intercept model with a single fixed effect has a form:

(ri1 model in R code):

$$ChildrenNo_{ij} = \beta_0 + \beta_1 Feeling + \mu_j + \epsilon_{ij}$$

We are interested about the question that "Do differences in total number of children between provinces remain after controlling for feeling of life?".

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ feelings_life + (1 | province)
##   Data: df_new
##
##      AIC      BIC  logLik deviance df.resid
## 72506.0 72537.6 -36249.0 72498.0     20035
##
## Scaled residuals:
##    Min     1Q  Median     3Q    Max
## -1.4009 -1.0297  0.1148  0.6734  3.8537
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   province (Intercept) 0.01457  0.1207
##   Residual            2.17868  1.4760
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 1.115059  0.065321 17.07
## feelings_life 0.076193  0.006354 11.99
##
## Correlation of Fixed Effects:
##          (Intr)
## feelings_lf -0.789
```

The coefficient for feeling is 0.076 and can be interpret as follows people with higher feeling score tend to have more children and the t-value for this coefficient is statistically significant. Compared to the variance components model, the intercept has hardly changed, nor has the level 2 variance. This suggests that the answer to the question, "Do differences in total number of children between provinces remain after controlling for feeling of life?" The answer is "Yes".

In order to check if the inclusion of an explanatory variable improves model fit, we compare this random fixed effect model (ri1) with a linear model and conduct a ANOVA testing between these two models

(sl1 model in R code):

$$ChildrenNo_{ij} = \beta_0 + \beta_1 Feeling + \epsilon_{ij}$$

```
# testing
sl1 <- lm (total_children ~ 1 + feelings_life, data = df_new)
summary (sl1)
```

```

## 
## Call:
## lm(formula = total_children ~ 1 + feelings_life, data = df_new)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.8201 -1.5868  0.1799  1.1799  5.8021 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.042387  0.052590 19.82   <2e-16 ***
## feelings_life 0.077767  0.006368 12.21   <2e-16 ***  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 1.48 on 20037 degrees of freedom
## Multiple R-squared:  0.007388, Adjusted R-squared:  0.007338 
## F-statistic: 149.1 on 1 and 20037 DF, p-value: < 2.2e-16

```

```
anova (ri1, sl1)
```

```

## Data: df_new
## Models:
## sl1: total_children ~ 1 + feelings_life
## ri1: total_children ~ feelings_life + (1 | province)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)    
## sl1     3 72590 72614 -36292     72584
## ri1     4 72506 72538 -36249     72498 86.254  1 < 2.2e-16 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

```

ri1 is better for fitting data

Checking if the inclusion of an explanatory variable improves model fit

```
anova (ri1, nullmodel)
```

```

## Data: df_new
## Models:
## nullmodel: total_children ~ (1 | province)
## ri1: total_children ~ feelings_life + (1 | province)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)    
## nullmodel 3 72647 72671 -36321     72641
## ri1        4 72506 72538 -36249     72498 143.27  1 < 2.2e-16 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

```

ri1 is better for fitting data

The results show that the random fixed effect model is better for fitting data

Similarly, we can examine the fixed effect of education and family income

"Do differences in total number of children between provinces remain after controlling on Education?"

```

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ education + (1 | province)
##   Data: df_new
##
##      AIC      BIC logLik deviance df.resid
## 72332.9 72404.0 -36157.4 72314.9     20030
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.5362 -0.9667  0.0997  0.5474  3.8488
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   province (Intercept) 0.01188  0.109
##   Residual            2.15906  1.469
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##   Estimate Std. Error t value
## (Intercept) 1.47525 0.04270 34.551
## educationCollege, CEGEP or other non-university certificate or di... 0.22082
## educationHigh school diploma or a high school equivalency certificate 0.26604
## educationLess than high school diploma or its equivalent          0.60811
## educationTrade certificate or diploma                            0.39118
## educationUniversity certificate or diploma below the bachelor's level 0.31177
## educationUniversity certificate, diploma or degree above the bach... 0.05293
## (Intercept) 0.04270
## educationCollege, CEGEP or other non-university certificate or di... 0.03253
## educationHigh school diploma or a high school equivalency certificate 0.03212
## educationLess than high school diploma or its equivalent          0.03628
## educationTrade certificate or diploma                            0.04554
## educationUniversity certificate or diploma below the bachelor's level 0.05964
## educationUniversity certificate, diploma or degree above the bach... 0.04204
## (Intercept) 34.551
## educationCollege, CEGEP or other non-university certificate or di... 6.789
## educationHigh school diploma or a high school equivalency certificate 8.283
## educationLess than high school diploma or its equivalent          16.760
## educationTrade certificate or diploma                            8.590
## educationUniversity certificate or diploma below the bachelor's level 5.228
## educationUniversity certificate, diploma or degree above the bach... 1.259
##
## Correlation of Fixed Effects:
##   (Intr) eCooncod esdoahsec ethsdoie edTcod ecodbtbl
## eC,CEooncod -0.422
## edHsdoahsec -0.430  0.555
## edcLthsdoie -0.384  0.493   0.499
## edctnTrdcod -0.308  0.392   0.399   0.359
## edUcodbtb'l -0.232  0.298   0.304   0.270   0.216
## eUc,dodatb. -0.319  0.423   0.427   0.377   0.299   0.230

```

The coefficient for education varies on different levles and can be interpret as follows people with higher level of education backgroud tend to have less children (for instance, the coefficient for people with diploma

or degree above the bachelor has the smallest coefficient = 0.05) while people with insufficient education experience are more likely having more children (for example, people with education less than high school).

Compared to the variance components model, the intercept has hardly changed, nor has the level 2 variance. This suggests that the answer to the question, "Do differences in total number of children between provinces remain after controlling for education?" The answer is "Yes".

```
##
## Call:
## lm(formula = total_children ~ 1 + education, data = df_new)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -2.0396 -1.4148  0.1810  0.9604  5.5852 
##
## Coefficients:
##                               Estimate
## (Intercept)                  1.41479
## educationCollege, CEGEP or other non-university certificate or di... 0.22895
## educationHigh school diploma or a high school equivalency certificate 0.27949
## educationLess than high school diploma or its equivalent            0.62483
## educationTrade certificate or diploma                                0.40426
## educationUniversity certificate or diploma below the bachelor's level 0.32524
## educationUniversity certificate, diploma or degree above the bach... 0.04616
##                               Std. Error
## (Intercept)                  0.02411
## educationCollege, CEGEP or other non-university certificate or di... 0.03257
## educationHigh school diploma or a high school equivalency certificate 0.03216
## educationLess than high school diploma or its equivalent            0.03619
## educationTrade certificate or diploma                                0.04535
## educationUniversity certificate or diploma below the bachelor's level 0.05971
## educationUniversity certificate, diploma or degree above the bach... 0.04213
##                               t value
## (Intercept)                  58.681
## educationCollege, CEGEP or other non-university certificate or di... 7.029
## educationHigh school diploma or a high school equivalency certificate 8.692
## educationLess than high school diploma or its equivalent            17.265
## educationTrade certificate or diploma                                8.913
## educationUniversity certificate or diploma below the bachelor's level 5.447
## educationUniversity certificate, diploma or degree above the bach... 1.096
##                               Pr(>|t|) 
## (Intercept)                  < 2e-16 
## educationCollege, CEGEP or other non-university certificate or di... 2.15e-12 
## educationHigh school diploma or a high school equivalency certificate < 2e-16 
## educationLess than high school diploma or its equivalent            < 2e-16 
## educationTrade certificate or diploma                                < 2e-16 
## educationUniversity certificate or diploma below the bachelor's level 5.19e-08 
## educationUniversity certificate, diploma or degree above the bach... 0.273
##                               ***
## (Intercept)                  ***
## educationCollege, CEGEP or other non-university certificate or di... ***
## educationHigh school diploma or a high school equivalency certificate ***
## educationLess than high school diploma or its equivalent            ***
## educationTrade certificate or diploma                                ***
## educationUniversity certificate or diploma below the bachelor's level ***
```

```

## educationUniversity certificate, diploma or degree above the bach...
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.473 on 20032 degrees of freedom
## Multiple R-squared: 0.01744, Adjusted R-squared: 0.01715
## F-statistic: 59.27 on 6 and 20032 DF, p-value: < 2.2e-16

## Data: df_new
## Models:
## sl2: total_children ~ 1 + education
## ri2: total_children ~ education + (1 | province)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## sl2     8 72396 72459 -36190    72380
## ri2     9 72333 72404 -36157    72315 65.284 1  6.484e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Data: df_new
## Models:
## nullmodel: total_children ~ (1 | province)
## ri2: total_children ~ education + (1 | province)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## nullmodel  3 72647 72671 -36321    72641
## ri2        9 72333 72404 -36157    72315 326.33 6  < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Now we look into a random effect model which contains the combined fixed effects of feeling of life and education backgroud.

(ri12 model in R code)

$$ChildrenNo_{ij} = \beta_0 + \beta_1 Feeling + \beta_2 Education + \mu_j + \epsilon_{ij}$$

```

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ feelings_life + education + (1 | province)
##   Data: df_new
##
##      AIC      BIC   logLik deviance df.resid
##  72176.4  72255.4 -36078.2  72156.4     20029
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.6427 -0.9549  0.0763  0.5874  3.9605
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   province (Intercept) 0.01107  0.1052
##   Residual            2.14210  1.4636
##   Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##
```

Estimate

```

## (Intercept) 0.822338
## feelings_life 0.079571
## educationCollege, CEGEP or other non-university certificate or di... 0.228384
## educationHigh school diploma or a high school equivalency certificate 0.279726
## educationLess than high school diploma or its equivalent 0.618573
## educationTrade certificate or diploma 0.399119
## educationUniversity certificate or diploma below the bachelor's level 0.308589
## educationUniversity certificate, diploma or degree above the bach... 0.049846
##
## Std. Error
## (Intercept) 0.066424
## feelings_life 0.006307
## educationCollege, CEGEP or other non-university certificate or di... 0.032405
## educationHigh school diploma or a high school equivalency certificate 0.032011
## educationLess than high school diploma or its equivalent 0.036149
## educationTrade certificate or diploma 0.045365
## educationUniversity certificate or diploma below the bachelor's level 0.059401
## educationUniversity certificate, diploma or degree above the bach... 0.041880
##
## t value
## (Intercept) 12.380
## feelings_life 12.617
## educationCollege, CEGEP or other non-university certificate or di... 7.048
## educationHigh school diploma or a high school equivalency certificate 8.738
## educationLess than high school diploma or its equivalent 17.112
## educationTrade certificate or diploma 8.798
## educationUniversity certificate or diploma below the bachelor's level 5.195
## educationUniversity certificate, diploma or degree above the bach... 1.190
##
## Correlation of Fixed Effects:
## (Intr) flngs_ eCooncod esdoahsec ethsdoie edTcod ecodbtbl
## feelings_lf -0.779
## eC,CEooncod -0.285  0.018
## edHsdoahsec -0.301  0.034  0.555
## edcLthsdoie -0.264  0.023  0.493   0.500
## edctnTrdcod -0.208  0.014  0.392   0.399   0.359
## edUcodbtb'l -0.145 -0.004  0.298   0.303   0.270   0.216
## eUc,dodatb. -0.200 -0.006  0.423   0.427   0.377   0.299   0.230

## Data: df_new
## Models:
## ri1: total_children ~ feelings_life + (1 | province)
## ri12: total_children ~ feelings_life + education + (1 | province)
##      npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)
## ri1     4 72506 72538 -36249    72498
## ri12   10 72176 72255 -36078    72156 341.61  6 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Data: df_new
## Models:
## ri2: total_children ~ education + (1 | province)
## ri12: total_children ~ feelings_life + education + (1 | province)
##      npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)
## ri2     9 72333 72404 -36157    72315
## ri12   10 72176 72255 -36078    72156 158.55  1 < 2.2e-16 ***

```

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The results of ANOVA indicate that later model combined Feeling and Education is more appropriate for modeling than previous model only contains singel fixed effect.

For the effect of family income,

(ri3 model in R code)

$$ChildrenNo_{ij} = \beta_0 + \beta_3 Income + \mu_j + \epsilon_{ij}$$

```

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ income_family + (1 | province)
##   Data: df_new
##
##      AIC      BIC logLik deviance df.resid
## 72644.0 72707.2 -36314.0 72628.0     20031
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.3394 -1.0647  0.1455  0.7150  3.7049
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   province (Intercept) 0.01547  0.1244
##   Residual            2.19281  1.4808
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                 1.73925  0.05128 33.919
## income_family$125,000 and more 0.00657  0.03891  0.169
## income_family$25,000 to $49,999 0.04872  0.03955  1.232
## income_family$50,000 to $74,999 -0.02770  0.04062 -0.682
## income_family$75,000 to $99,999 -0.01440  0.04251 -0.339
## income_familyLess than $25,000 -0.07804  0.04313 -1.809
##
## Correlation of Fixed Effects:
##          (Intr) i_$1am i_$2t$ i_$5t$ i_$7t$
## i_$125,00am -0.518
## i_$25,000t$ -0.514  0.672
## i_$50,000t$ -0.499  0.655  0.647
## i_$75,000t$ -0.477  0.626  0.618  0.601
## i_Lt$25,000 -0.470  0.616  0.610  0.593  0.567

```

The outcome of this multilevel model is ambiguous to explain the effects of familiy income on total number of children, since the t value for coefficients are not all highly significant. However, if we add this term into ri12 model we find out that low income familiy are more likely to have more children when it compares to the rich familiy.

Combined all three fixed effects: Feeling, Education and Income

(ri123 model in R code)

$$ChildrenNo_{ij} = \beta_0 + \beta_1 Feeling + \beta_2 Education + \beta_3 Income + \mu_j + \epsilon_{ij}$$

```

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ feelings_life + education + income_family +
##           (1 | province)
## Data: df_new
##
##      AIC      BIC logLik deviance df.resid
## 72160.3 72278.9 -36065.2 72130.3     20024
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.6929 -0.9416  0.0704  0.5761  3.9658
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   province (Intercept) 0.01115  0.1056
##   Residual            2.13931  1.4626
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##                                         Estimate
## (Intercept)                           0.877181
## feelings_life                          0.074258
## educationCollege, CEGEP or other non-university certificate or di... 0.240286
## educationHigh school diploma or a high school equivalency certificate 0.301017
## educationLess than high school diploma or its equivalent             0.656927
## educationTrade certificate or diploma                                0.412645
## educationUniversity certificate or diploma below the bachelor's level 0.315672
## educationUniversity certificate, diploma or degree above the bach... 0.043909
## income_family$125,000 and more                               0.033512
## income_family$25,000 to $49,999                            -0.013913
## income_family$50,000 to $74,999                            -0.052908
## income_family$75,000 to $99,999                            -0.011154
## income_familyLess than $25,000                            -0.147826
##                                         Std. Error
## (Intercept)                           0.074209
## feelings_life                          0.006414
## educationCollege, CEGEP or other non-university certificate or di... 0.032596
## educationHigh school diploma or a high school equivalency certificate 0.032462
## educationLess than high school diploma or its equivalent             0.037389
## educationTrade certificate or diploma                                0.045528
## educationUniversity certificate or diploma below the bachelor's level 0.059407
## educationUniversity certificate, diploma or degree above the bach... 0.041902
## income_family$125,000 and more                               0.038538
## income_family$25,000 to $49,999                            0.039573
## income_family$50,000 to $74,999                            0.040239
## income_family$75,000 to $99,999                            0.042012
## income_familyLess than $25,000                            0.043752
##                                         t value
## (Intercept)                         11.820
## feelings_life                        11.578
## educationCollege, CEGEP or other non-university certificate or di... 7.372
## educationHigh school diploma or a high school equivalency certificate 9.273
## educationLess than high school diploma or its equivalent            17.570
## educationTrade certificate or diploma                                9.064

```

```

## educationUniversity certificate or diploma below the bachelor's level      5.314
## educationUniversity certificate, diploma or degree above the bach...       1.048
## income_family$125,000 and more                                         0.870
## income_family$25,000 to $49,999                                         -0.352
## income_family$50,000 to $74,999                                         -1.315
## income_family$75,000 to $99,999                                         -0.266
## income_familyLess than $25,000                                         -3.379

## Data: df_new
## Models:
## ri12: total_children ~ feelings_life + education + (1 | province)
## ri123: total_children ~ feelings_life + education + income_family +
## ri123:          (1 | province)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## ri12     10 72176 72255 -36078    72156
## ri123    15 72160 72279 -36065    72130 26.039  5  8.771e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Random Slope Model

In random intercept models, we assumed that the fixed effects were the same for all province. In this section, we check the impacts of interested variables within provinces by adding random term μ_{ij} to the coefficient of explanatory variable. Let us consider the impact of explanatory variable Feeling of life within in different provinces as an example. The random slope model has a form:

(rs.gen1 model in R code)

$$ChildrenNo_{ij} = \beta_0 + (\beta_1 + \mu_{1ij})Feeling + \mu_j + \epsilon_{ij}$$

```

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ feelings_life + education + income_family +
##           (1 + feelings_life | province)
## Data: df_new
##
##      AIC      BIC logLik deviance df.resid
## 72164.3 72298.7 -36065.1  72130.3     20022
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -1.6970 -0.9417  0.0703  0.5763  3.9659
##
## Random effects:
##   Groups   Name        Variance Std.Dev. Corr
##   province (Intercept) 9.038e-03 0.095066
##           feelings_life 1.640e-06 0.001281 1.00
##   Residual            2.139e+00 1.462639
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##                                         Estimate
## (Intercept)                               0.872099
## feelings_life                            0.074863
## educationCollege, CEGEP or other non-university certificate or di... 0.240331

```

```

## educationHigh school diploma or a high school equivalency certificate 0.301010
## educationLess than high school diploma or its equivalent 0.657020
## educationTrade certificate or diploma 0.412613
## educationUniversity certificate or diploma below the bachelor's level 0.315623
## educationUniversity certificate, diploma or degree above the bach... 0.043954
## income_family$125,000 and more 0.033477
## income_family$25,000 to $49,999 -0.013822
## income_family$50,000 to $74,999 -0.052906
## income_family$75,000 to $99,999 -0.011159
## income_familyLess than $25,000 -0.147786
##
## Std. Error
## (Intercept) 0.072697
## feelings_life 0.006425
## educationCollege, CEGEP or other non-university certificate or di... 0.032596
## educationHigh school diploma or a high school equivalency certificate 0.032462
## educationLess than high school diploma or its equivalent 0.037389
## educationTrade certificate or diploma 0.045527
## educationUniversity certificate or diploma below the bachelor's level 0.059407
## educationUniversity certificate, diploma or degree above the bach... 0.041902
## income_family$125,000 and more 0.038539
## income_family$25,000 to $49,999 0.039572
## income_family$50,000 to $74,999 0.040239
## income_family$75,000 to $99,999 0.042012
## income_familyLess than $25,000 0.043752
##
## t value
## (Intercept) 11.996
## feelings_life 11.652
## educationCollege, CEGEP or other non-university certificate or di... 7.373
## educationHigh school diploma or a high school equivalency certificate 9.273
## educationLess than high school diploma or its equivalent 17.573
## educationTrade certificate or diploma 9.063
## educationUniversity certificate or diploma below the bachelor's level 5.313
## educationUniversity certificate, diploma or degree above the bach... 1.049
## income_family$125,000 and more 0.869
## income_family$25,000 to $49,999 -0.349
## income_family$50,000 to $74,999 -1.315
## income_family$75,000 to $99,999 -0.266
## income_familyLess than $25,000 -3.378
## convergence code: 0
## boundary (singular) fit: see ?isSingular

## Data: df_new
## Models:
## ri123: total_children ~ feelings_life + education + income_family +
## ri123: (1 | province)
## rs.gen1: total_children ~ feelings_life + education + income_family +
## rs.gen1: (1 + feelings_life | province)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## ri123     15 72160 72279 -36065    72130
## rs.gen1   17 72164 72299 -36065    72130  0.0283  2     0.9859

```

Since the correlation between a province random intercept and the same province random effect for feeling is positively equal to 1. In this case, it might suggest that a province which is towards the upper end of the distributions for intercepts will also be relatively high up in the distribution of random effects for people

with higher score in feeling life. However, we notice that the effect of province does not hold by and the testing results shows that rs.gen1 is not perform as well as the random intercept model ri123. Therefore, the impact of feeling life score does not vary between provinces.

Similarly, we check whether the impact of Education varies between provinces.

```
rs.gen2 <- lmer(total_children ~ feelings_life + education + income_family + (1 + education|province), data = df_new)

## boundary (singular) fit: see ?isSingular

summary(rs.gen2)

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ feelings_life + education + income_family +
##           (1 + education | province)
## Data: df_new
##
##      AIC      BIC  logLik deviance df.resid
## 72199.6 72531.6 -36057.8 72115.6     19997
## 
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.7759 -0.9505  0.0662  0.5644  3.9409
## 
## Random effects:
## Groups   Name
## province (Intercept)
##   educationCollege, CEGEP or other non-university certificate or diploma
##   educationHigh school diploma or a high school equivalency certificate
##   educationLess than high school diploma or its equivalent
##   educationTrade certificate or diploma
##   educationUniversity certificate or diploma below the bachelor's level
##   educationUniversity certificate, diploma or degree above the bachelor's level
## Residual
## Variance Std.Dev. Corr
## 0.001833 0.04281
## 0.005602 0.07484  0.29
## 0.007474 0.08645  0.46  0.98
## 0.019095 0.13818  0.78  0.82  0.91
## 0.011538 0.10741  -0.06 0.87  0.80  0.56
## 0.020315 0.14253  0.46  0.83  0.85  0.85  0.83
## 0.004229 0.06503  0.64  0.91  0.97  0.98  0.70  0.90
## 2.137019 1.46185
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##                                         Estimate
## (Intercept)                                0.84192
## feelings_life                               0.07415
## educationCollege, CEGEP or other non-university certificate or diploma 0.27238
## educationHigh school diploma or a high school equivalency certificate 0.34191
## educationLess than high school diploma or its equivalent            0.71860
## educationTrade certificate or diploma                         0.46171
```

```

## educationUniversity certificate or diploma below the bachelor's level 0.37948
## educationUniversity certificate, diploma or degree above the bach... 0.07943
## income_family$125,000 and more 0.03316
## income_family$25,000 to $49,999 -0.01458
## income_family$50,000 to $74,999 -0.05326
## income_family$75,000 to $99,999 -0.01247
## income_familyLess than $25,000 -0.15127
##
## Std. Error
## (Intercept) 0.06735
## feelings_life 0.00641
## educationCollege, CEGEP or other non-university certificate or di... 0.04126
## educationHigh school diploma or a high school equivalency certificate 0.04316
## educationLess than high school diploma or its equivalent 0.05790
## educationTrade certificate or diploma 0.05843
## educationUniversity certificate or diploma below the bachelor's level 0.07593
## educationUniversity certificate, diploma or degree above the bach... 0.04697
## income_family$125,000 and more 0.03852
## income_family$25,000 to $49,999 0.03955
## income_family$50,000 to $74,999 0.04022
## income_family$75,000 to $99,999 0.04199
## income_familyLess than $25,000 0.04373
##
## t value
## (Intercept) 12.500
## feelings_life 11.567
## educationCollege, CEGEP or other non-university certificate or di... 6.602
## educationHigh school diploma or a high school equivalency certificate 7.922
## educationLess than high school diploma or its equivalent 12.412
## educationTrade certificate or diploma 7.902
## educationUniversity certificate or diploma below the bachelor's level 4.998
## educationUniversity certificate, diploma or degree above the bach... 1.691
## income_family$125,000 and more 0.861
## income_family$25,000 to $49,999 -0.368
## income_family$50,000 to $74,999 -1.324
## income_family$75,000 to $99,999 -0.297
## income_familyLess than $25,000 -3.459

```

```

##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(x, correlation=TRUE)  or
##      vcov(x)      if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular

```

```
anova (rs.gen2, ri123)
```

```

## Data: df_new
## Models:
## ri123: total_children ~ feelings_life + education + income_family +
## ri123:      (1 | province)
## rs.gen2: total_children ~ feelings_life + education + income_family +
## rs.gen2:      (1 + education | province)
##      npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)

```

```

## ri123      15 72160 72279 -36065      72130
## rs.gen2     42 72200 72532 -36058      72116 14.757 27      0.9727

```

It implies that impact of education score on total number of children does not vary between provinces.

```

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: total_children ~ feelings_life + education + income_family +
##           (1 + income_family | province)
## Data: df_new
##
##          AIC      BIC logLik deviance df.resid
## 72164.7 72441.4 -36047.3 72094.7    20004
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.7095 -0.9439  0.0740  0.5724  4.0111
##
## Random effects:
##   Groups   Name        Variance Std.Dev. Corr
##   province (Intercept) 0.001706 0.04131
##   income_family$125,000 and more 0.001741 0.04172  0.87
##   income_family$25,000 to $49,999 0.021466 0.14651 -0.15  0.34
##   income_family$50,000 to $74,999 0.007309 0.08549 -0.02  0.43  0.97
##   income_family$75,000 to $99,999 0.010563 0.10277  0.07  0.54  0.97
##   income_familyLess than $25,000 0.056026 0.23670 -0.02  0.33  0.83
##   Residual                  2.133641 1.46070
##
## 
## 
## 
## 
## 
## 
## 
## 
## 
## 0.95
## 0.93  0.77
##
## Number of obs: 20039, groups: province, 10
##
## Fixed effects:
##                               Estimate
## (Intercept)                0.839500
## feelings_life               0.074097
## educationCollege, CEGEP or other non-university certificate or di... 0.237731
## educationHigh school diploma or a high school equivalency certificate 0.298969
## educationLess than high school diploma or its equivalent            0.649050
## educationTrade certificate or diploma                                0.413968
## educationUniversity certificate or diploma below the bachelor's level 0.319749
## educationUniversity certificate, diploma or degree above the bach... 0.041357
## income_family$125,000 and more                           0.048748
## income_family$25,000 to $49,999                          0.042525
## income_family$50,000 to $74,999                         -0.015971
## income_family$75,000 to $99,999                         0.025518
## income_familyLess than $25,000                         -0.032830
##                               Std. Error
## (Intercept)                0.067290
## feelings_life               0.006406

```

```

## educationCollege, CEGEP or other non-university certificate or di... 0.032563
## educationHigh school diploma or a high school equivalency certificate 0.032434
## educationLess than high school diploma or its equivalent 0.037385
## educationTrade certificate or diploma 0.045455
## educationUniversity certificate or diploma below the bachelor's level 0.059334
## educationUniversity certificate, diploma or degree above the bach... 0.041849
## income_family$125,000 and more 0.041007
## income_family$25,000 to $49,999 0.062104
## income_family$50,000 to $74,999 0.048768
## income_family$75,000 to $99,999 0.053688
## income_familyLess than $25,000 0.088290
##
t value
## (Intercept) 12.476
## feelings_life 11.567
## educationCollege, CEGEP or other non-university certificate or di... 7.301
## educationHigh school diploma or a high school equivalency certificate 9.218
## educationLess than high school diploma or its equivalent 17.361
## educationTrade certificate or diploma 9.107
## educationUniversity certificate or diploma below the bachelor's level 5.389
## educationUniversity certificate, diploma or degree above the bach... 0.988
## income_family$125,000 and more 1.189
## income_family$25,000 to $49,999 0.685
## income_family$50,000 to $74,999 -0.327
## income_family$75,000 to $99,999 0.475
## income_familyLess than $25,000 -0.372
## convergence code: 0
## boundary (singular) fit: see ?isSingular

## Data: df_new
## Models:
## ri123: total_children ~ feelings_life + education + income_family +
## ri123: (1 | province)
## rs.gen3: total_children ~ feelings_life + education + income_family +
## rs.gen3: (1 + income_family | province)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## ri123     15 72160 72279 -36065    72130
## rs.gen3   35 72165 72441 -36047    72095 35.632 20     0.01699 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## $province
##                               (Intercept) income_family$125,000 and more
## Alberta             -0.019421584          -0.036019270
## British Columbia    -0.015500396          -0.032115674
## Manitoba            0.014951514           0.036375476
## New Brunswick       0.002208183           0.008686142
## Newfoundland and Labrador -0.051151102        -0.035304977
## Nova Scotia          0.015810404           0.010054274
## Ontario              -0.042992634          -0.038413209
## Prince Edward Island 0.016775881           0.030637598
## Quebec               0.025472350          -0.011420368
## Saskatchewan         0.053847385           0.067520008
##                               income_family$25,000 to $49,999
## Alberta                  -0.12453850

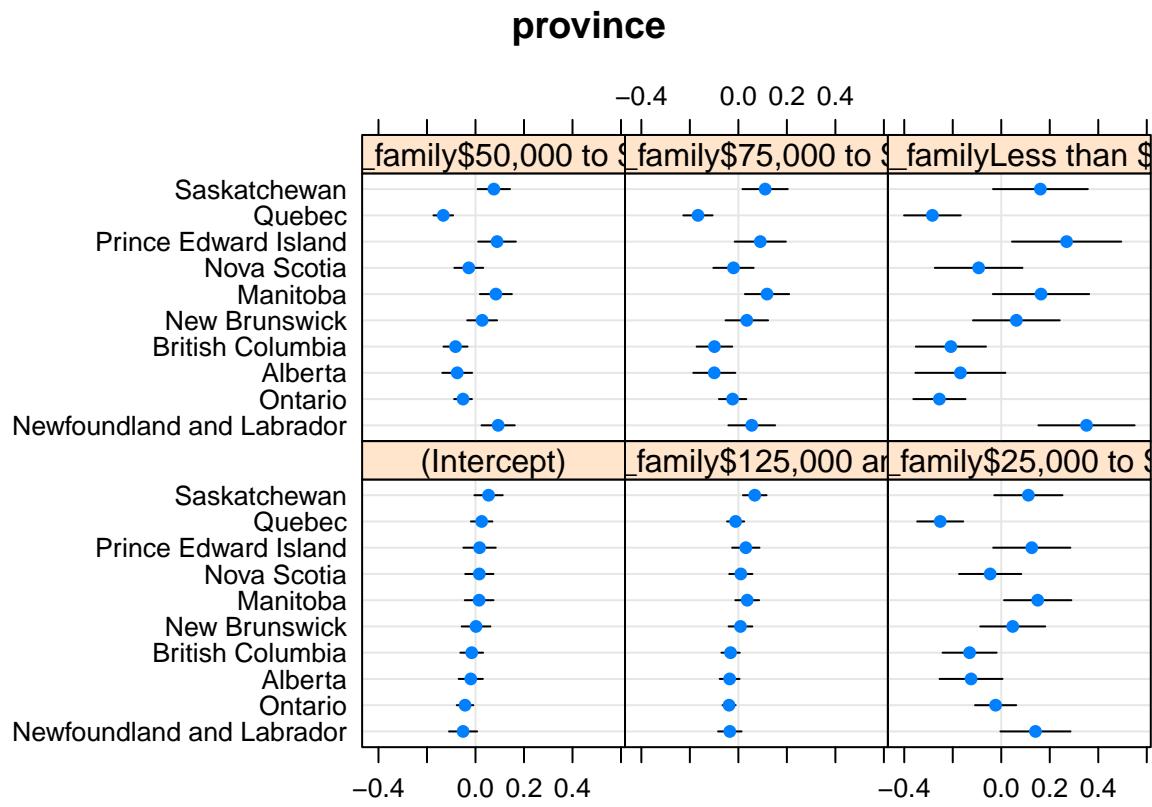
```

```

## British Columbia           -0.13067993
## Manitoba                  0.15027923
## New Brunswick              0.04725093
## Newfoundland and Labrador 0.14102930
## Nova Scotia                -0.04582302
## Ontario                    -0.02334612
## Prince Edward Island       0.12576258
## Quebec                     -0.25175574
## Saskatchewan               0.11182126
##                               income_family$50,000 to $74,999
## Alberta                    -0.07537821
## British Columbia            -0.08233617
## Manitoba                   0.08411338
## New Brunswick               0.02742433
## Newfoundland and Labrador  0.09339775
## Nova Scotia                 -0.02768386
## Ontario                     -0.05131225
## Prince Edward Island        0.08888938
## Quebec                      -0.13286105
## Saskatchewan               0.07574670
##                               income_family$75,000 to $99,999
## Alberta                    -0.09939858
## British Columbia            -0.09879527
## Manitoba                   0.11810530
## New Brunswick               0.03474188
## Newfoundland and Labrador  0.05515878
## Nova Scotia                 -0.02015230
## Ontario                     -0.02378584
## Prince Edward Island        0.09042648
## Quebec                      -0.16664021
## Saskatchewan               0.11033976
##                               income_familyLess than $25,000
## Alberta                    -0.16848463
## British Columbia            -0.20770536
## Manitoba                   0.16387110
## New Brunswick               0.06225396
## Newfoundland and Labrador  0.35147242
## Nova Scotia                 -0.09320173
## Ontario                     -0.25507533
## Prince Edward Island        0.26940198
## Quebec                      -0.28393687
## Saskatchewan               0.16140446
##
## with conditional variances for "province"

## $province

```

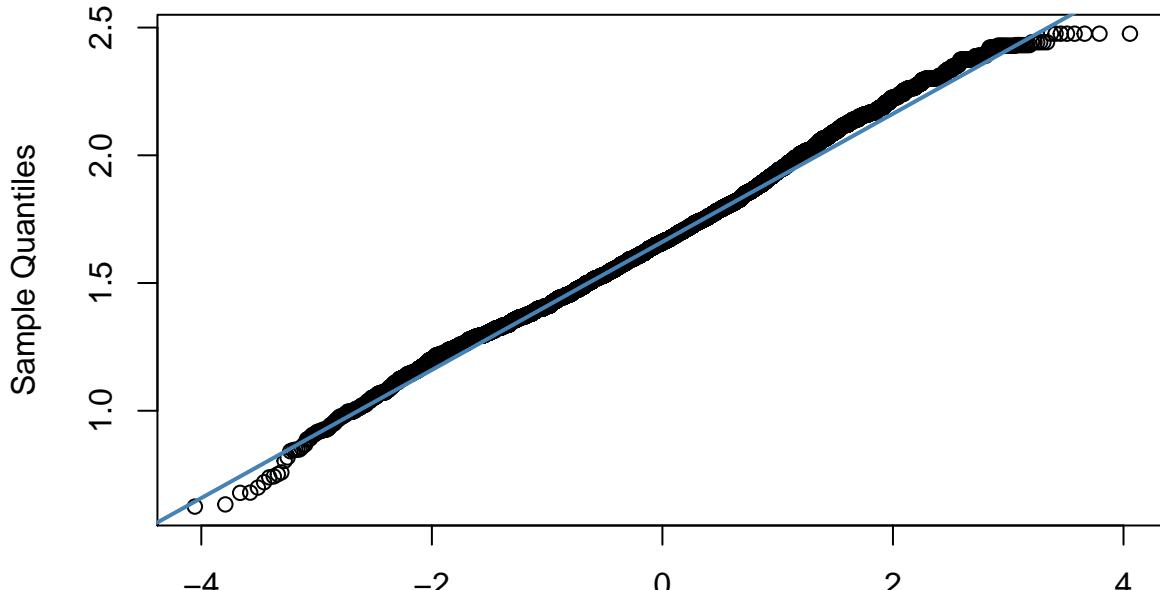


The results shows that the impact of family income on number of children does vary between provinces.

Model Diagnose

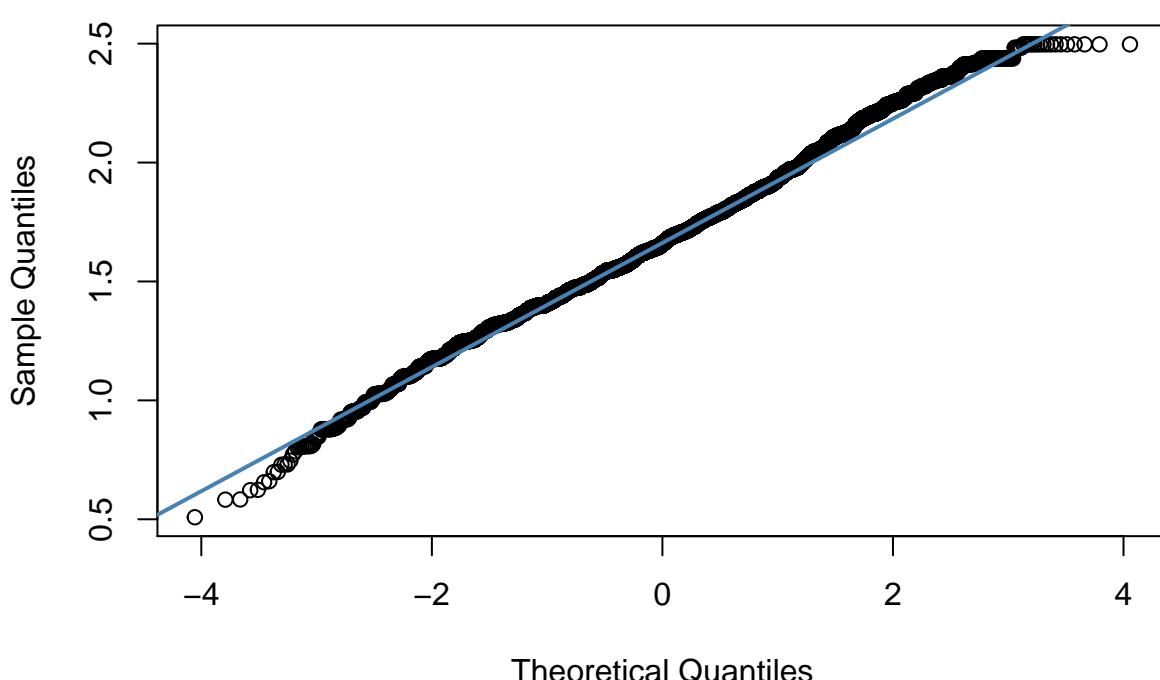
The diagnostic plots for the random intercept model (with all three variables: Feeling, Income and Education) and the random slope model with random slope of Education are depicted in the following figures. Basically, the normality assumption could be reasonably satisfied in either of these two models since most of the dots are affiliated to the straight line in the Normal Q-Q plot.

Normal Q-Q Plot



Theoretical Quantiles

Normal Q-Q Plot



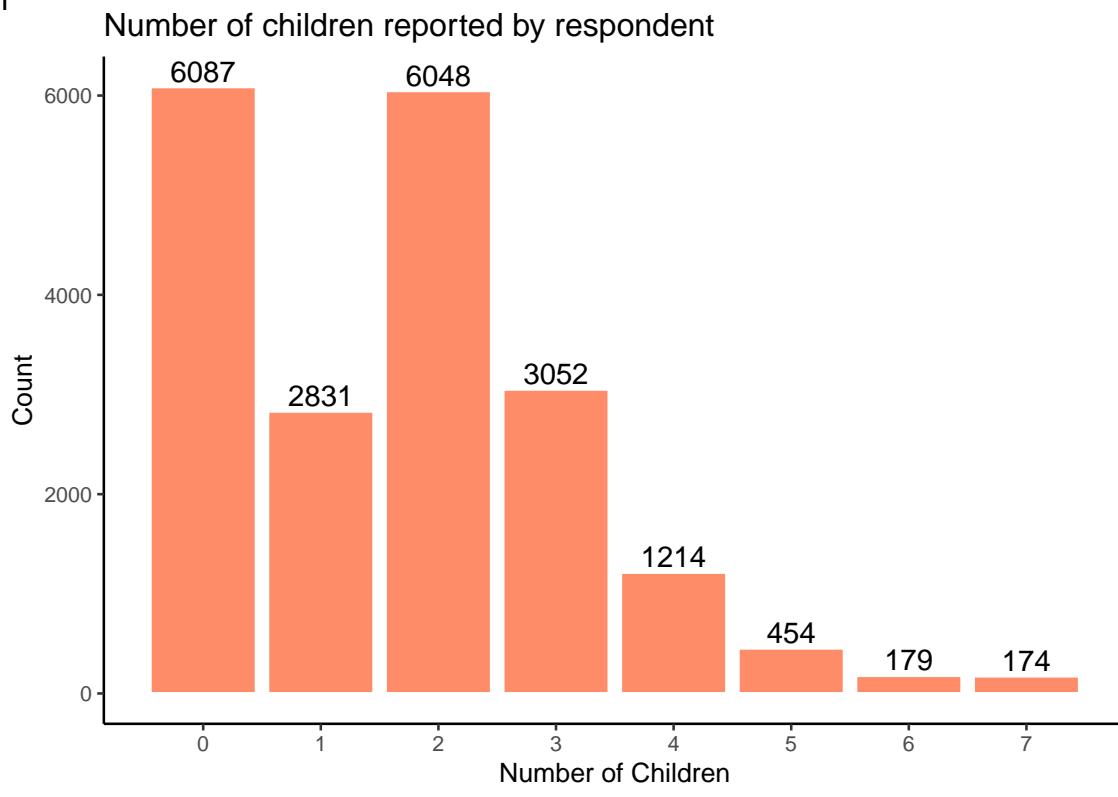
Theoretical Quantiles

Disscussion of strength &weakness

The superiority of the multilevel regression model is obvious: it is more reasonable for inference on data that are clustered in groups (hierarchical data), comparing to an ordinary single fixed effect model (e.g. standard ordinary least squares model). Instead of estimating separate intercept/slope in multiple groups, multilevel regression model can minimize the number of estimated parameters: the mean and variance of the intercepts, or the average slope along with its variance across the groups. Generally, a multilevel regression model is capable of dealing with data that is hierarchically structured, modeling variation in the social world, and conducting explicitly tests for differences in parameters by groups. However, it has a number of potential weaknesses. The main concern that we might encounter is that we assume that the means of multiple levels are samples from a normal distribution, for instance, within each province, the total number of children follows a normal distribution with a provincial specified mean β_{0j} , and a variance σ^2 . Moreover, we assume that the provincial-specified means vary according to a normal distribution, then, heterogeneity in responses to lower-level variables may not be hold since the groupwise heterogeneity in the mean are controlled. Additionally, a multilevel regression model is a complicated model which is more likely to generate computational problems and consume our attention to understand the results than an ordinary regression model.

Results

Figure1



Source : General Social Survey, Families, 2017.

Figure 1 demonstrates the total number of children respondents have a cap of seven in a bar plot. A total of 20602 samples was collected with 20039 valid responses. The plot is right-skewed. The left skewness shows that a majority of respondents have fewer children. The two largest proportions are respondents who have no child and respondents who have two children. There are 6087 respondents who have no child, while 6048 respondents have two children. The least proportion shown in the graph is the respondents who have seven children, with a number of 174. There are 30.3% of respondents have no child; 14.1% of respondents have one child; 30.2% of Respondents have two children; 15.2% of respondents have three children; 6.1% of respondents have four children; 2.3% of respondents have five children; 0.90% respondents have six children; 0.90% respondents have seven children.

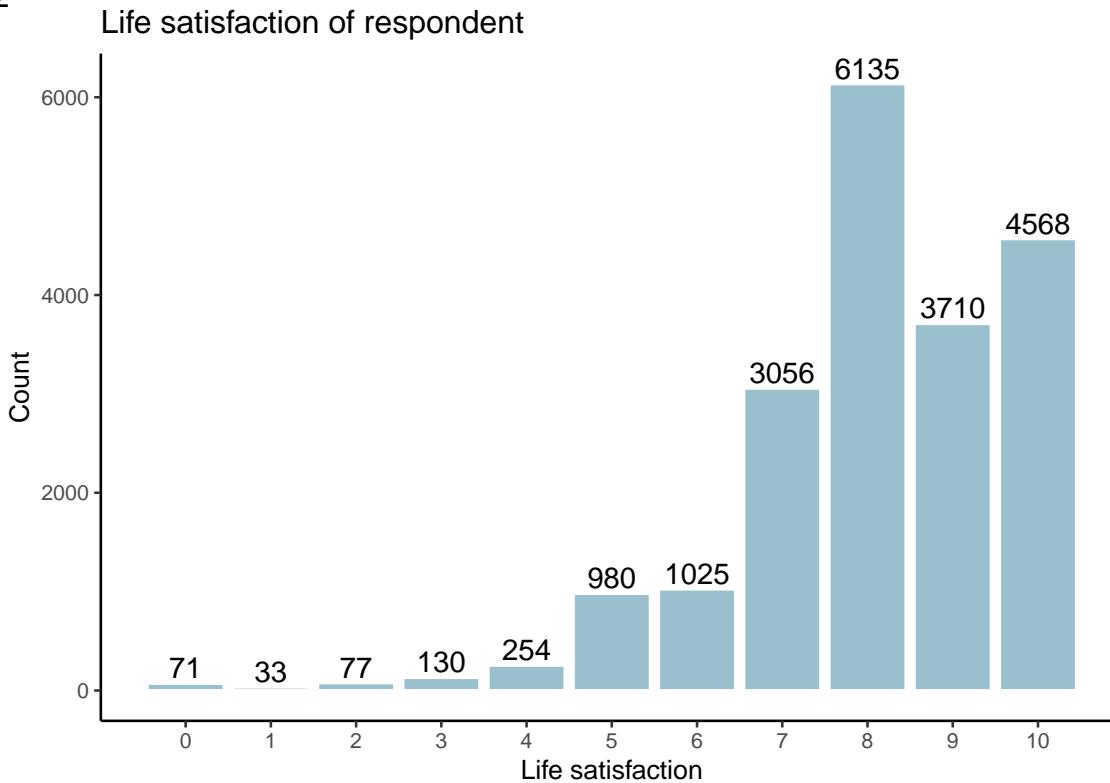
```

## # A tibble: 6 x 6
##   age_group     n   Mean Median    SD   freq
##   <chr>     <int>  <dbl>  <dbl> <dbl>  <dbl>
## 1 18-25      1075  0.0958     0  0.426  0.0536
## 2 25-35      2739  0.773     0  1.07   0.137
## 3 35-45      3116  1.64      2  1.31   0.156
## 4 45-65      6985  1.81      2  1.35   0.349
## 5 over 65    5705  2.38      2  1.55   0.285
## 6 under 18    419   0          0  0       0.0209

```

For the first table, we divided the respondent into different age groups and created a function for summarizing some statistics properties, such as mean and median, of the total number of children in their household for the different age groups. As we can see in table 1, the mean of the total number of children for the people who are over 65 is around 2.38, which means there are at least two children on average for the respondent who is older than 65. As for the respondent who is under 18, we know that the mean and median are both 0, which means that for the respondent who is under 18, they do not have children. This data comply with common sense.

Figure2



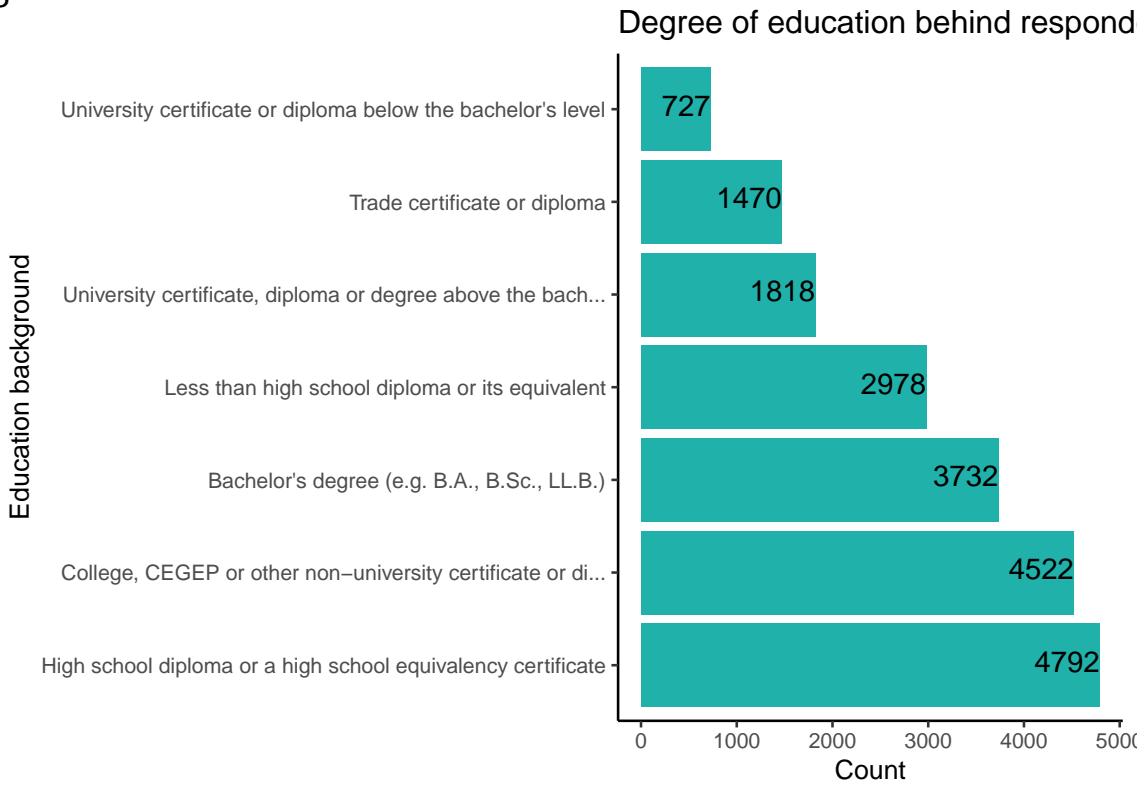
Source : General Social Survey, Families, 2017.

With a total of 20039 valid data collected, this bar plot illustrates a left-skewed respondents' life satisfaction level distribution. The largest proportion of respondents are 80% satisfied with their life while the least proportion of respondents are 1% satisfied with their life. There are 6135 respondents (30.6% of the total respondents) are eighty percent satisfied, and 33 respondents (0.16% of the total respondents) are one percent satisfied with their life. There are 22.89% of respondents who are 100% satisfied with their current life. Approximately 72% of respondents are satisfied with a percentage of eighty or higher. More than half of Canadians are predicted to be highly satisfied with their life.

```
## # A tibble: 11 x 6
##   feelings_life     n   Mean Median    SD   freq
##   <int> <int> <dbl> <dbl> <dbl> <dbl>
## 1 0      71   1.52    2  1.36 0.0035
## 2 1      33   1.24    0  1.44 0.0016
## 3 2      77   1.60    1  1.67 0.0038
## 4 3     130   1.39    1  1.52 0.0065
## 5 4     254   1.37    1  1.41 0.0127
## 6 5     980   1.67    2  1.52 0.0489
## 7 6    1025   1.47    1  1.48 0.0512
## 8 7    3056   1.45    1  1.45 0.152
## 9 8    6135   1.64    2  1.47 0.306
## 10 9   3710   1.68    2  1.44 0.185
## 11 10   4568   1.93    2  1.52 0.228
```

For this table, we can see with the different life satisfaction, the mean total number of children is different. The highest mean of the total number of children is 1.93 for the respondent who chooses 10 (very satisfied), and the lowest is 1.24 for the respondent who chooses 1(dissatisfied).

Figure3



Source: General Social Survey, Families, 2017

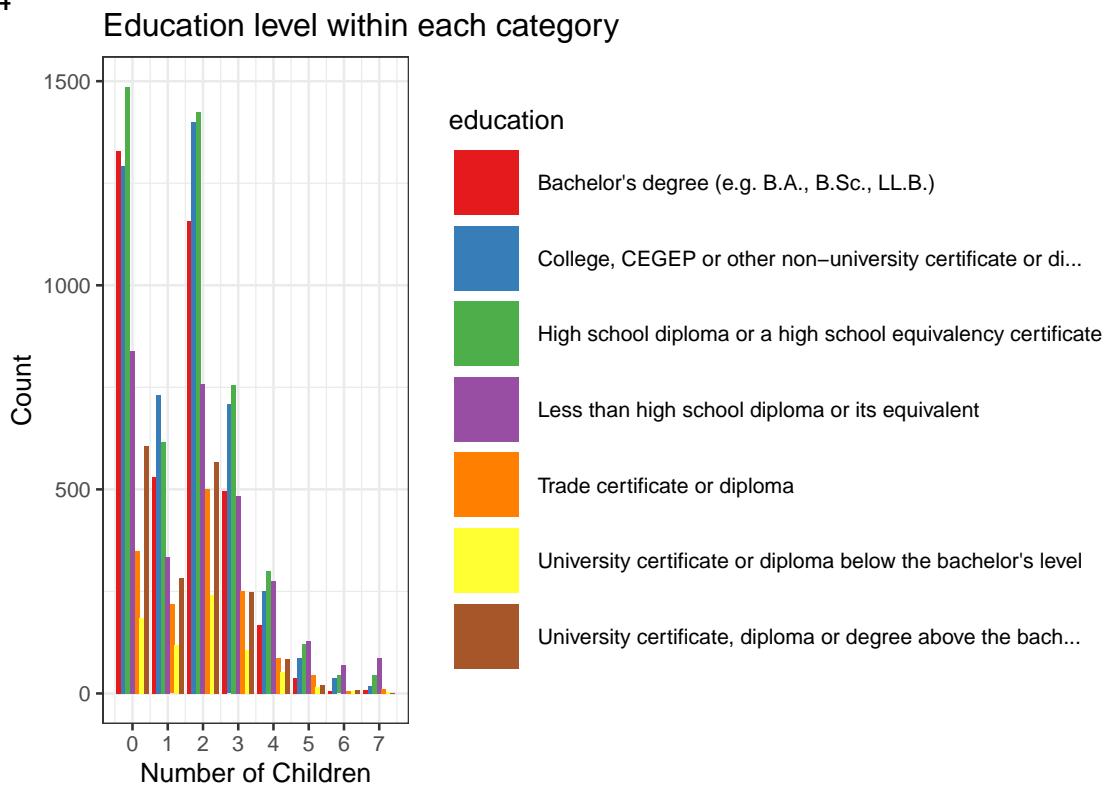
Figure 3 The graph is the respondents' education level distribution. With a total of 20039 collected samples, there are 4522 respondents who have a college degree and 3732 respondents who have bachelor's degrees. There are 1470 respondents who have a trade diploma and 727 respondents who have a university diploma below the bachelor's degree. What're more, around 4792 respondents complete their high school diploma, and with 2978 respondents less than high school diplomas.

```
## # A tibble: 7 x 6
##   education          n  Mean Median    SD freq
##   <chr>      <int> <dbl>  <dbl> <dbl> <dbl>
## 1 Bachelor's degree (e.g. B.A., B.Sc., LL.B.) 3732  1.41   2  1.31 0.186
## 2 College, CEGEP or other non-university certif~ 4522  1.64   2  1.40 0.226
## 3 High school diploma or a high school equivale~ 4792  1.69   2  1.52 0.239
## 4 Less than high school diploma or its equivale~ 2978  2.04   2  1.81 0.149
## 5 Trade certificate or diploma                 1470  1.82   2  1.43 0.0734
## 6 University certificate or diploma below the b~ 727   1.74   2  1.41 0.0363
## 7 University certificate, diploma or degree abo~ 1818  1.46   2  1.31 0.0907
```

Table 3

For this table, we can find that the mean of the total number of children is highest with 2.04 for the respondent whose education level is less than high school and the lowest mean is 1.41, which means for the respondent who is bachelor's degree, they are less likely to have children in this data.

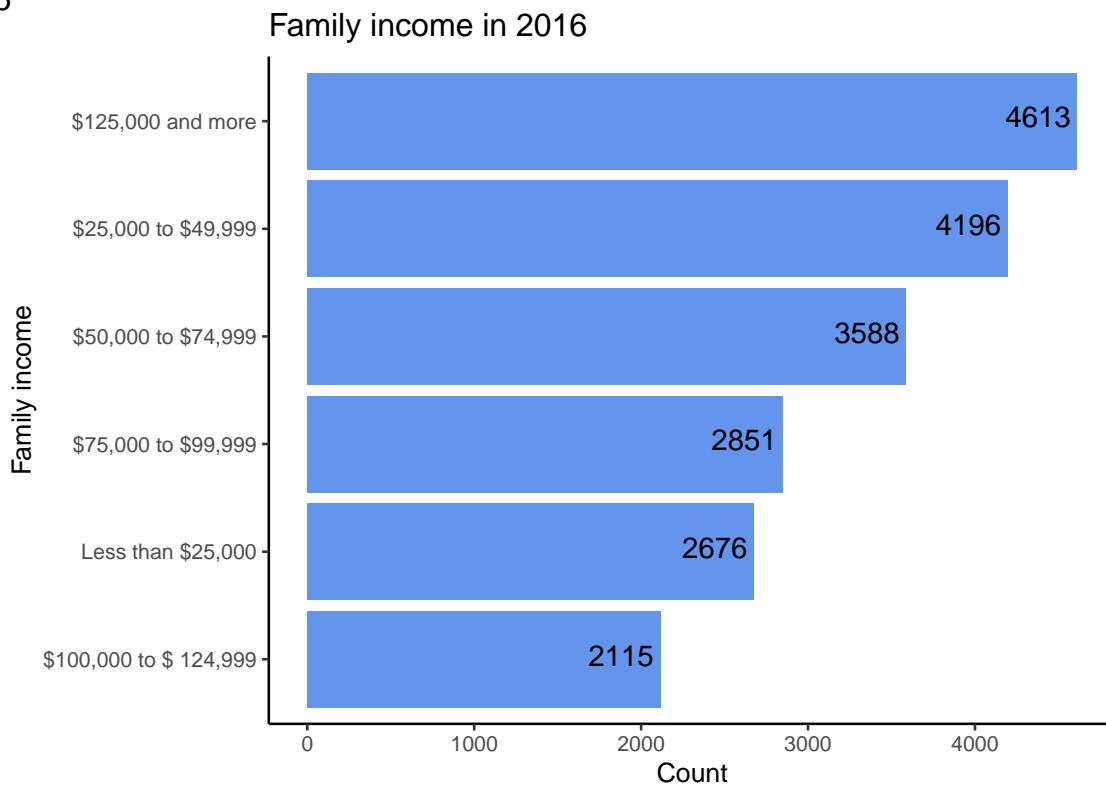
Figure 4



Source: General Social Survey, Families, 2017

Figure 4 This graph shows the education level within each category of the number of children respondents have. The largest proportion of respondents are respondents who have a high school degree. High school diploma respondents are the largest proportion of the four categories, which are the respondents are no child, two children, three children and four children. Besides that, for the respondents who have a college certificate, they are more likely to have one child under a different education level. For the respondent who has five or more children, respondents whose education level is less than a high school diploma occupies a heavy proportion.

Figure 5

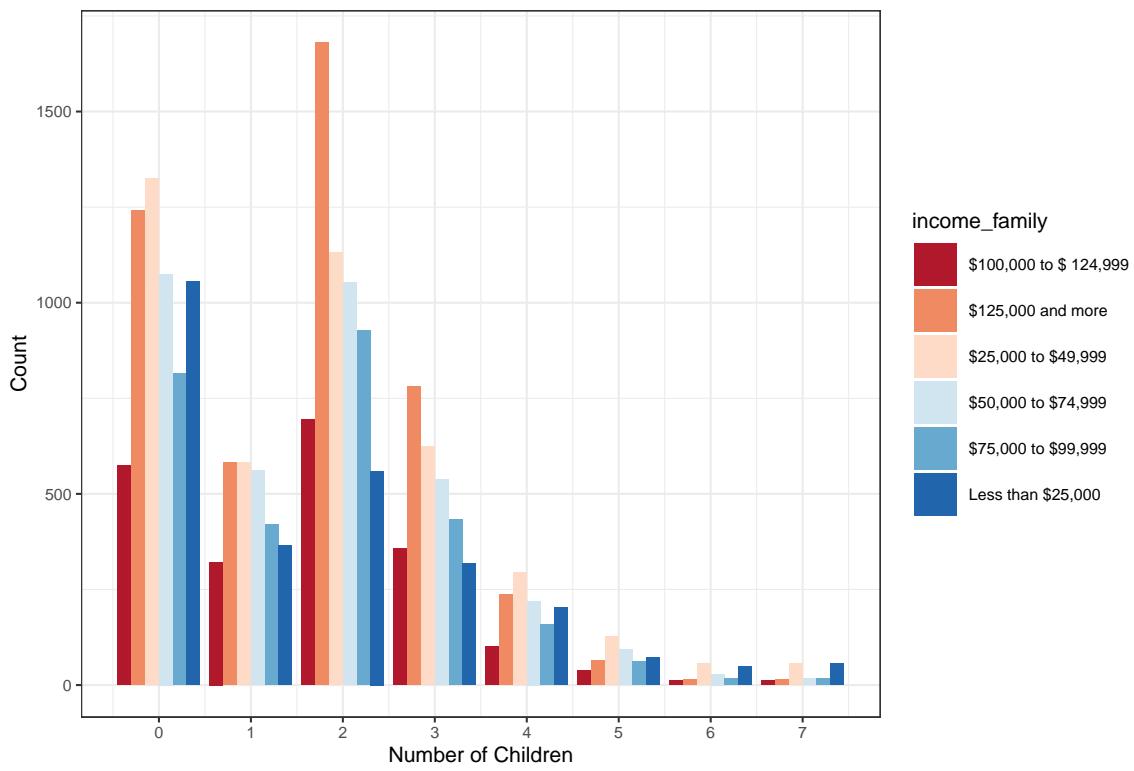


Source: T1FF 2016, Statistics Canada

Figure 5 With a total of 20039 valid responses, this is the distribution of respondents' family income. The largest proportion of respondents earn 125,000 dollars or more each year. There are 4701 respondents who annually earn 125,000 dollars or more, whereas 2154 respondents earn 100,000 dollars to 124,999 dollars. The distribution shows a right-skewed characteristic, which indicates that Canadian respondents earn a relatively high income to support their well-being.

Figure 6

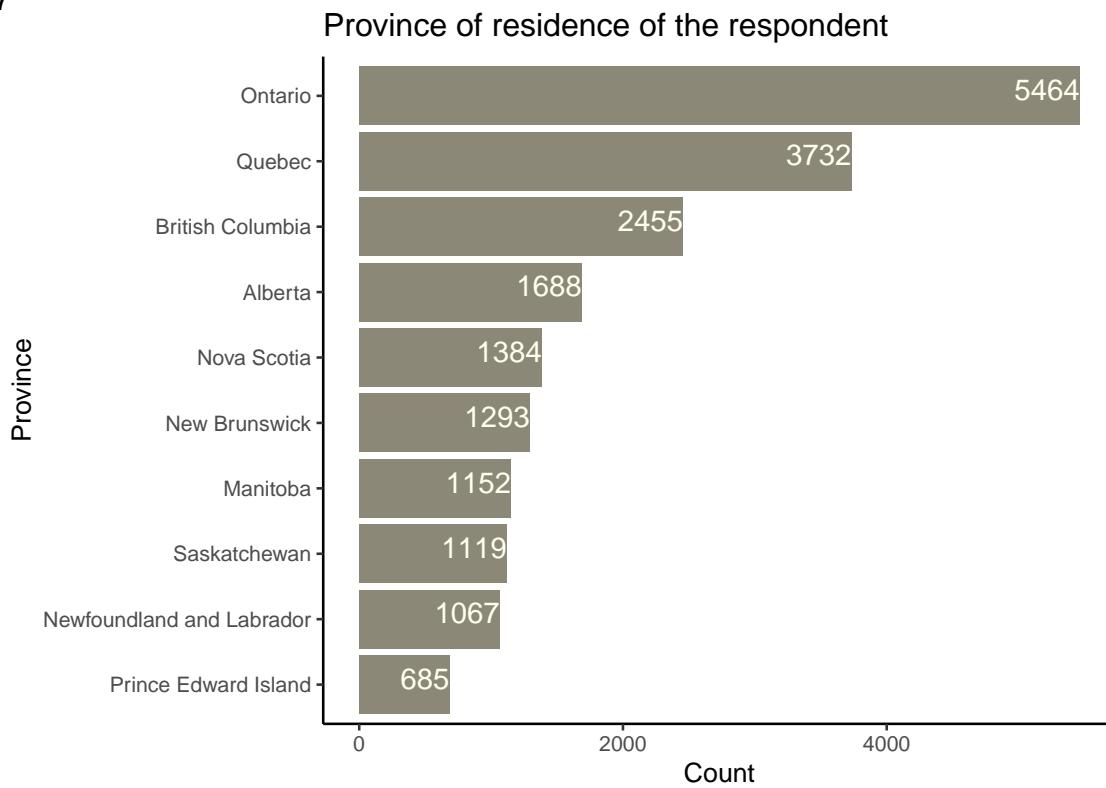
The number of different income categories in the number of children in the household



Source: General Social Survey, 2017.

Figure 6 This distribution shows the proportion of income levels within each number of children respondents have. The graph shows a detailed look into the childbearing favouritism in which respondents at different income levels are responded to. The outstanding feature of this graph is that respondents who earn 125,000 dollars and more are in the largest proportion under the category of two children. The respondents who earn 25,000 dollars to 49,999 dollars are in the largest proportion under the category that is not two children.

Figure 7



Source: General Social Survey, Families, 2017.

Figure 7 This is a distribution of respondents' residential province. A majority of respondents are from Ontario Province, with a number of 5616. With a total of 20583 collected samples, 27.3% respondents are from Ontario; 18.6% respondents are from Quebec; 12.3% respondents are from British Columbia; 8.40% respondents are Alberta; 6.90% respondents are from Nova Scotia; 6.49% respondents are from New Brunswick; 5.79% respondents are from Manitoba; 5.60% respondents are from Saskatchewan; 5.31% respondents are from Newfoundland and Labrador; 3.43% respondents are from Prince Edward Island.

Discussion

Figure 1 & Table 1

As we can know from the figure 1, the samples' households are common to have two kids, which is about 30% of the aggregates. The proportion of respondents whose is without any child is also around 30%. A relatively small proportion of respondents have more than three kids. The child distribution among the samples' households is closed to realistic phenomena. According to the census of 2016 in Canada, the number of couples living with at least one child is 51.1%. This is the lowest level on record. There is also an increase in the proportion of couples living without a child. Compared to 2011, the proportion increases by 7.2%. Such increase is relatively large. The trends in the share of couples living with and without children indicates the diversity of households and families in Canada. From Figure 1 and Table 1, we can find that most Canadian families are likely to have no children or have two children. This is a very interesting point; the reason why the number of one child is only half of that of no children or two children. In our opinion, for those respondents who do not have children, they are probably young and still in the early stages of their careers. Work is what they need most for their current situation. Secondly, it may be because some people do not want a child. This phenomenon is also widespread. Why is the respondent with two children more than twice as many as the respondent with one child? Having two children be more conducive to the growth of the child? As children grow up, it is indeed a delightful thing to have a sibling. Besides, we feel that for most Canadians, they have enough economic capital money to support their children lives and future education.

Figure 2 & Table 2

Most of our respondents are satisfied with their life. More than 70% of the aggregates have a high index of life satisfaction. The average of our samples' feelings_life is 7.88, which is relatively high. The variable, feelings_life, measures the samples' satisfaction. According to the World Happiness Report, Canadians' average index of happiness is about 7.32 in 2017. Since the sample size is not large enough, there is a gap between these the average of samples' feelings_life with the Canadians' average index of happiness. The samples do reflect the majority of the population.

Figure 3,4 & Table 3

From Figure 3, 4 and table 3, we can find that for most Canadians, they have at least complete a high school diploma, and whose education level is bachelor's degree and above bachelor's degree also account for a certain percentage. Will education level also affect the number of children in the household? We can imagine that it is difficult for those who currently focus on their studies to take care of their children. In this way, they probably will postpone marriage and childbearing. Additionally, we can find that those respondents with an education level less than a high school diploma (relatively low diplomas) are more willing to have more children. Those with higher education levels give birth to 4, or the number of children above four will be relatively small. What caused these unusual phenomena? Does this situation fit the real social world situation?

Figure 5 & 6

We think that households with low income tend to have more kids. The majority of the households with more than 3 kids are those families who have an annual income of less than \$50,000. However, higher-income families are common to have 0 or 2 kids in each household. According to figure5 and 6, we found that the two most income ranges are 125,000dollars and more and 25,000 dollars to 49,999 dollars. The gap between the two ranges is enormous, and income imbalance is also a problem for modern society. Through figure 6, we found that respondents with incomes ranging from 25,000 dollars to 49,999 dollars are more likely to have no children or have four or more children. What is the reason cause the gap between them? We assume that it is easier to obtain a high-paying job with a higher education level under the general social environment. Combining the content of this paragraph with the previous one, is it possible that those relatively low-income respondents have a relatively basic level of education, which is with common sense.

Conclusion

Through this model and graph, it is not difficult to find that the number of children in the household is not just as simple as what it looks like. This data comes from GSS, which is raw data. Rohan Alexander and Sam Caetano help us clean the data, and then we group by different age groups. We found that for this data, more than 50% of respondents are over 45 years old. These age differences may cause data bias. For example, older people will have more children than younger people, which meets the actual social situation. In general, our report does prove that some variables may cause differences in the number of children in Canadian families. However, some biases in the data will cause uncertainty. One of the most important findings is that people with lower income will be more likely to have a big family, and ill-educated respondents generally have more children than high-educated respondents. Under the social environment that people will receive higher incomes with a higher education background, Can we boldly assume that those with low incomes are people with low educational backgrounds? If so, there will be more concern for the future report.

Weaknesses

The first weakness is the inaccurate prediction due to the limited sample size. Since interviewers may review the respondent multiple times so that respondents fully complete the survey, the real sample size is underestimate. The Second weakness is the non-response bias. If the interview is quite time-consuming with long question, selected respondents may not have patience and time to complete all. The last and most important is that the respondent bias. Most of the age of respondents are over 45 which is higher than the mean of Canadian

Next Steps

To reduce questionnaires' shortcomings, we re-contact respondents who did not complete the interview the first time, explain questionnaires' importance and convince them to complete the interviews. For respondents who are unavailable to answer the interview, we provide appointments as an option. Gradually, the Canadian response rate of GSS in 2017 has reached more than 50%. This analysis gives government a clearer version to understand the reason behind Canadian low fertility rate and to further improve provincial benefits. With our results, we can provide it federal and provincial government so that they can solve the aging phenomena with a better understanding. Population aging is a direct negative effect on Canadian labor force and capital accumulation. To minimize the harm caused by population aging, Canadian government should encourage citizens on childbearing with a more customized Child Benefits to different income-level households. If government does better allocating resource, Canadian fertility rate may gradually increase.

References

- Alexander, R, & Caetano, S. (2020, October 7th). [Clean-up the 2017 GSS data obtained from the U of T library]. Unpublished raw data.
- Canada Happiness index - data, chart. (n.d.). Retrieved October 19, 2020, from <https://www.theglobaleconomy.com/Canada/happiness/>
- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Government of Canada, S. (2017, August 02). Families, households and marital status: Key results from the 2016 Census. Retrieved October 17, 2020, from <https://www150.statcan.gc.ca/n1/daily-quotidien/170802/dq170802a-eng.htm>
- Government of Canada, S. (2017, November 29). Education in Canada: Key results from the 2016 Census. Retrieved October 19, 2020, from <https://www150.statcan.gc.ca/n1/daily-quotidien/171129/dq171129a-eng.htm>
- Hadley Wickham, Romain Fran?ois, Lionel Henry and Kirill M?ller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. <https://CRAN.R-project.org/package=dplyr>
- Hadley Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- 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>
- Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5
- Statistics Canada. (2017). General Social Survey On Family (Cycle 31), 2017 - Canadian General Social Surveys (GSS).. [online] Available at: <https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/html/gss.htm>
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>.
- Xie Y (2019). “TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX Live.” *TUGboat*, 30–32. <http://tug.org/TUGboat/Contents/contents40-1.html>.
- Xie Y (2020). tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents. R package version 0.26, <https://github.com/yihui/tinytex>.
- Yihui Xie and J.J. Allaire and Garrett Grolemund (2018). R Markdown: The Definitive Guide. Chapman and Hall/CRC. ISBN 9781138359338. URL <https://bookdown.org/yihui/rmarkdown>.