Apply Linear Regression Model to analyze the number of children someone has

Junming Zhang

Saturday, October 17, 2020

Abstract

Having children is one of the key factors of a family, and children play an important role in family relationship. Also, birth rate influences the future of a community because they will be the labour force of the community, which has an impact on the economical and social development, and even existence of the community. Therefore, it is neccessary to have an indicator to suggest what factors influence the number of children a family has, and so the government can make policies to control birth rate based on the indicator. We tried to investigate such a factor by the dataset **Canadian General Social Survey (GSS)** (citation 5) of year 2017, in which had some attributes like number of childrens one person has and some other features of a person which may help (total_children, marital_status, education, partner_education, income_family, self_rated_health and self_rated_mental_health). To generate such an indicator, we built a linear regression model with the dataset, which predicts the number of children one may have with some attributes I searched before which may help.

Introduction

Our goal is to find a model to predict how many children one may have based on attributes may have strong affects, and test how strongly are these attributes correlated to the number of children one may have. To create such a model, we use the dataset **Canadian General Social Survey (GSS)** from year 2017, which has attributes that may contribute to the our research goal, like total_children, marital_status, education, partner_education, income_family, self_rated_health and self_rated_mental_health, and we built our model with these attributes. Total chilren is the attribute we are interested in. Education and partner_education are important because people who are entolled in the tertiary education tend to postpone their marriage and have fewer children (citation 1). We consider income_family because income_family is a mirror of the economy development, which is related to birth rate, for instance, economic depression may mean low fertility (Pobric & Robinson, 2015). We look into marital_status, because the type of partnership may contribute to birth rate, for example, those who get married may have more children than those who cohabit (Martinez, Daniels, & Chandra, 2012). Finally but still important, the health status impacts the number of children on may have, for example, countries with higher HDI (Human Development Index, which involves life expectancy, education, and per capita income) may have lower fertility rate, which is reflected by the citation 4, countries with lower HDI have higher fertility rate and vice versa. Since health is related to HDI for life expectancy, we put the self_rated_health and self_rated_mental_health into the model. Therefore, we create a linear regression model with these attributes to predict the number of children one may have, and analyze how is our prediction related to these attributes, and how strong is our interest and these attributes correlate.

Data

The dataset is obtained from **Canadian General Social Survey (GSS)** of year 2017, it contains all the attributes I listed in the Introduction section that are used to build the model. To make the dataset, they use a questionnaire and interview the respondents on phone call (Beaupré, 2020). A brief outline of the questionnaire is following (Beaupré, 2020):

- 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
- Fertility intentions
- Maternity/parental leave
- Organization and decision making within the household
- Arrangements and financial support after a separation/divorce
- Labour market new and education
- Health and subjective well-being
- Characteristics of respondent's dwelling
- Characteristics of respondent of spouse/partner

The questionnaire to build the dataset was delivered by telephone (Beaupré, 2020), and this questionnaire was helpful because it covers many details on the personal conditions of the respondent, like health and education, for which there is significant proof that influences the fertility rate. However, the previous nationalities of respondents are ignored, which may also introduce errors in the result, because some countries may have special cultures and religions that affect the fertility rate. There are also pros and cons for collecting data by phone. The benefit is that, since the most people have their own telephones today, so it is easy to connect and the data can be collected with lower costs. However, some people may not respond to the phone call, which leads to the non-participation error.

There are 81 variables/attributes and 20602 observations in the dataset. The variables generally cover many aspects about the living conditions and the personal conditions of the interviewee, which may suggest our interest, and we tested some of them which are possibly helpful according to the documents and references we found, and investigate the correlation between the variables in our scope and our interest. And also, the dataset has a large number of observations with respect to the place where the data were collected, and thus this makes the results (in Canada) can be found from the dataset more representative. However, since the dataset is only limited to one country (Canada), the variables do not reflect other factors may also have impacts but not suitable for just one country, like policy, war or peace, natural conditions, and so hard to reflect worldwide facts.

The data are collected with the stratified random sampling (simple random sampling without replacement in the stratum) method (Beaupré, 2020), a probability sampling approach. The target population for the dataset included all persons 15 years of age and older in Canada, excluding: 1. Residents of the Yukon, Northwest Territories, and Nunavut; and 2. Full-time residents of institutions (Beaupré, 2020). The frame of the survey is 1. Lists of telephone numbers in use (both landline and cellular) available to Statistics Canada from various sources (telephone companies, Census of population, etc.); and 2. The Address Register (AR): List of all dwellings within the ten provinces. The probability sampling method (collection approach for this dataset) decreases errors like generalization and more representative for the whole population. However,

there are some drawbacks of the dataset from both non-sampling error. The non-sampling error is mainly from (patial or total) non-participation. This is handled by adjusting the weights to less for non-participation cases (Beaupré, 2020).

Model

Here is the number of observations for each stratum (since the data are collected by stratified sampling without replacement), the stratum was divided based on the province the interviewee lived (Beaupré, 2020), we built our linear regression model based on the stratification below:

```
## # A tibble: 10 x 2
## # Groups:   province [10]
##    province                      n
##    <chr>                     <int>
##  1 Alberta                    1728
##  2 British Columbia           2522
##  3 Manitoba                   1192
##  4 New Brunswick              1337
##  5 Newfoundland and Labrador  1094
##  6 Nova Scotia                1425
##  7 Ontario                    5621
##  8 Prince Edward Island        708
##  9 Quebec                     3822
## 10 Saskatchewan               1153
```

```
## Warning in summary.lm(mysvylm): residual degrees of freedom in object suggest
## this is not an "lm" fit
```

```
##
## Call:
## svyglm(formula = total_children ~ as.factor(marital_status) +
##     as.factor(education) + as.factor(partner_education) + as.factor(income_family) +
##     as.factor(self_rated_health) + as.factor(self_rated_mental_health),
##     design = gss.design)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0609 -0.8203 -0.0475  0.7146  6.6155
##
## Coefficients:
##                                                                                  Estimate
## (Intercept)                                                                      2.025483
## as.factor(marital_status)Living common-law                                      -0.480726
## as.factor(marital_status)Married                                                 0.113549
## as.factor(marital_status)Separated                                               0.197398
## as.factor(marital_status)Single, never married                                  -1.454016
## as.factor(marital_status)Widowed                                                 0.304508
## as.factor(education)College, CEGEP or other non-university certificate or di...  0.096823
## as.factor(education)High school diploma or a high school equivalency certificate  0.267416
## as.factor(education)Less than high school diploma or its equivalent              0.555937
## as.factor(education)Trade certificate or diploma                                 0.262349
## as.factor(education)University certificate or diploma below the bachelor's level  0.152244
## as.factor(education)University certificate, diploma or degree above the bach...  -0.008789
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...  0.077587
## as.factor(partner_education)High school diploma or a high school equivalency certi...  0.218896
```

```
## as.factor(partner_education)Less than high school diploma or its equivalent               0.419406
## as.factor(partner_education)Trade certificate or diploma                                   0.195518
## as.factor(partner_education)University certificate or diploma below the bachelor's level  0.088903
## as.factor(partner_education)University certificate, diploma or degree above the ba...    -0.067516
## as.factor(income_family)$125,000 and more                                                  0.094280
## as.factor(income_family)$25,000 to $49,999                                                 0.011437
## as.factor(income_family)$50,000 to $74,999                                                -0.068498
## as.factor(income_family)$75,000 to $99,999                                                -0.043548
## as.factor(income_family)Less than $25,000                                                 -0.137966
## as.factor(self_rated_health)Excellent                                                     -1.183581
## as.factor(self_rated_health)Fair                                                          -1.058813
## as.factor(self_rated_health)Good                                                          -1.134601
## as.factor(self_rated_health)Poor                                                          -0.893970
## as.factor(self_rated_health)Very good                                                     -1.230207
## as.factor(self_rated_mental_health)Excellent                                               0.801803
## as.factor(self_rated_mental_health)Fair                                                    0.765855
## as.factor(self_rated_mental_health)Good                                                    0.829083
## as.factor(self_rated_mental_health)Poor                                                    0.625163
## as.factor(self_rated_mental_health)Very good                                               0.793638
##                                                                                         Std. Error
## (Intercept)                                                                               0.417433
## as.factor(marital_status)Living common-law                                                0.085221
## as.factor(marital_status)Married                                                          0.081236
## as.factor(marital_status)Separated                                                        0.155749
## as.factor(marital_status)Single, never married                                            0.091311
## as.factor(marital_status)Widowed                                                          0.163705
## as.factor(education)College, CEGEP or other non-university certificate or di...           0.035602
## as.factor(education)High school diploma or a high school equivalency certificate          0.037889
## as.factor(education)Less than high school diploma or its equivalent                       0.050373
## as.factor(education)Trade certificate or diploma                                          0.049467
## as.factor(education)University certificate or diploma below the bachelor's level         0.063015
## as.factor(education)University certificate, diploma or degree above the bach...          0.043441
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...    0.036804
## as.factor(partner_education)High school diploma or a high school equivalency certi...    0.037107
## as.factor(partner_education)Less than high school diploma or its equivalent               0.049773
## as.factor(partner_education)Trade certificate or diploma                                  0.050747
## as.factor(partner_education)University certificate or diploma below the bachelor's level  0.062557
## as.factor(partner_education)University certificate, diploma or degree above the ba...     0.043925
## as.factor(income_family)$125,000 and more                                                 0.037113
## as.factor(income_family)$25,000 to $49,999                                                0.044198
## as.factor(income_family)$50,000 to $74,999                                                0.041058
## as.factor(income_family)$75,000 to $99,999                                                0.041185
## as.factor(income_family)Less than $25,000                                                 0.066378
## as.factor(self_rated_health)Excellent                                                     0.284400
## as.factor(self_rated_health)Fair                                                          0.285916
## as.factor(self_rated_health)Good                                                          0.283810
## as.factor(self_rated_health)Poor                                                          0.292315
## as.factor(self_rated_health)Very good                                                     0.283753
## as.factor(self_rated_mental_health)Excellent                                              0.308172
## as.factor(self_rated_mental_health)Fair                                                   0.311638
## as.factor(self_rated_mental_health)Good                                                   0.307853
## as.factor(self_rated_mental_health)Poor                                                   0.331887
## as.factor(self_rated_mental_health)Very good                                              0.307905
##                                                                                          t value
```

```
## (Intercept)                                                                    4.852
## as.factor(marital_status)Living common-law                                    -5.641
## as.factor(marital_status)Married                                               1.398
## as.factor(marital_status)Separated                                             1.267
## as.factor(marital_status)Single, never married                               -15.924
## as.factor(marital_status)Widowed                                               1.860
## as.factor(education)College, CEGEP or other non-university certificate or di...  2.720
## as.factor(education)High school diploma or a high school equivalency certificate  7.058
## as.factor(education)Less than high school diploma or its equivalent           11.036
## as.factor(education)Trade certificate or diploma                               5.304
## as.factor(education)University certificate or diploma below the bachelor's level  2.416
## as.factor(education)University certificate, diploma or degree above the bach...  -0.202
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...  2.108
## as.factor(partner_education)High school diploma or a high school equivalency certi...  5.899
## as.factor(partner_education)Less than high school diploma or its equivalent    8.426
## as.factor(partner_education)Trade certificate or diploma                       3.853
## as.factor(partner_education)University certificate or diploma below the bachelor's level  1.421
## as.factor(partner_education)University certificate, diploma or degree above the ba...  -1.537
## as.factor(income_family)$125,000 and more                                      2.540
## as.factor(income_family)$25,000 to $49,999                                     0.259
## as.factor(income_family)$50,000 to $74,999                                    -1.668
## as.factor(income_family)$75,000 to $99,999                                    -1.057
## as.factor(income_family)Less than $25,000                                     -2.078
## as.factor(self_rated_health)Excellent                                         -4.162
## as.factor(self_rated_health)Fair                                              -3.703
## as.factor(self_rated_health)Good                                              -3.998
## as.factor(self_rated_health)Poor                                              -3.058
## as.factor(self_rated_health)Very good                                         -4.335
## as.factor(self_rated_mental_health)Excellent                                   2.602
## as.factor(self_rated_mental_health)Fair                                        2.458
## as.factor(self_rated_mental_health)Good                                        2.693
## as.factor(self_rated_mental_health)Poor                                        1.884
## as.factor(self_rated_mental_health)Very good                                   2.578
##                                                                             Pr(>|t|)
## (Intercept)                                                                 1.24e-06
## as.factor(marital_status)Living common-law                                  1.73e-08
## as.factor(marital_status)Married                                            0.162211
## as.factor(marital_status)Separated                                          0.205032
## as.factor(marital_status)Single, never married                               < 2e-16
## as.factor(marital_status)Widowed                                            0.062896
## as.factor(education)College, CEGEP or other non-university certificate or di...  0.006545
## as.factor(education)High school diploma or a high school equivalency certificate  1.78e-12
## as.factor(education)Less than high school diploma or its equivalent           < 2e-16
## as.factor(education)Trade certificate or diploma                            1.16e-07
## as.factor(education)University certificate or diploma below the bachelor's level  0.015707
## as.factor(education)University certificate, diploma or degree above the bach...  0.839674
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...  0.035044
## as.factor(partner_education)High school diploma or a high school equivalency certi...  3.75e-09
## as.factor(partner_education)Less than high school diploma or its equivalent  < 2e-16
## as.factor(partner_education)Trade certificate or diploma                    0.000117
## as.factor(partner_education)University certificate or diploma below the bachelor's level 0.155302
## as.factor(partner_education)University certificate, diploma or degree above the ba...  0.124301
## as.factor(income_family)$125,000 and more                                   0.011088
## as.factor(income_family)$25,000 to $49,999                                  0.795820
```

```
## as.factor(income_family)$50,000 to $74,999                                    0.095274
## as.factor(income_family)$75,000 to $99,999                                    0.290369
## as.factor(income_family)Less than $25,000                                     0.037686
## as.factor(self_rated_health)Excellent                                         3.18e-05
## as.factor(self_rated_health)Fair                                              0.000214
## as.factor(self_rated_health)Good                                              6.43e-05
## as.factor(self_rated_health)Poor                                              0.002231
## as.factor(self_rated_health)Very good                                         1.47e-05
## as.factor(self_rated_mental_health)Excellent                                  0.009285
## as.factor(self_rated_mental_health)Fair                                       0.014004
## as.factor(self_rated_mental_health)Good                                       0.007089
## as.factor(self_rated_mental_health)Poor                                       0.059635
## as.factor(self_rated_mental_health)Very good                                  0.009962
##
## (Intercept)                                                                   ***
## as.factor(marital_status)Living common-law                                    ***
## as.factor(marital_status)Married
## as.factor(marital_status)Separated
## as.factor(marital_status)Single, never married                                ***
## as.factor(marital_status)Widowed                                              .
## as.factor(education)College, CEGEP or other non-university certificate or di...  **
## as.factor(education)High school diploma or a high school equivalency certificate  ***
## as.factor(education)Less than high school diploma or its equivalent           ***
## as.factor(education)Trade certificate or diploma                              ***
## as.factor(education)University certificate or diploma below the bachelor's level  *
## as.factor(education)University certificate, diploma or degree above the bach...
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...  *
## as.factor(partner_education)High school diploma or a high school equivalency certi...  ***
## as.factor(partner_education)Less than high school diploma or its equivalent   ***
## as.factor(partner_education)Trade certificate or diploma                      ***
## as.factor(partner_education)University certificate or diploma below the bachelor's level
## as.factor(partner_education)University certificate, diploma or degree above the ba...
## as.factor(income_family)$125,000 and more                                     *
## as.factor(income_family)$25,000 to $49,999
## as.factor(income_family)$50,000 to $74,999                                    .
## as.factor(income_family)$75,000 to $99,999
## as.factor(income_family)Less than $25,000                                     *
## as.factor(self_rated_health)Excellent                                         ***
## as.factor(self_rated_health)Fair                                              ***
## as.factor(self_rated_health)Good                                              ***
## as.factor(self_rated_health)Poor                                              **
## as.factor(self_rated_health)Very good                                         ***
## as.factor(self_rated_mental_health)Excellent                                  **
## as.factor(self_rated_mental_health)Fair                                       *
## as.factor(self_rated_mental_health)Good                                       **
## as.factor(self_rated_mental_health)Poor                                       .
## as.factor(self_rated_mental_health)Very good                                  **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.263 on 12131 degrees of freedom
##   (8429 observations deleted due to missingness)
## Multiple R-squared:  0.1482, Adjusted R-squared:  0.1454
## F-statistic: 65.97 on 32 and 12131 DF,  p-value: < 2.2e-16
```

Results

Discussion

Weaknesses

Next Steps

In the next step, we may collect more data in other countries with different economic conditions and cultural backgrounds, like the country they immigrate from, the religious background to show the results more generally. Also, we can use the principal component analysis to narrow the variables of the model, which are strongly correlated with the interest. It is also worth considering to build a neural network model to make predictions based on our data, since the NN model is more robust to random cases.

References

1. NCHS Pressroom - 1997 Fact Sheet - Mothers Education and Birth Rate. (2009, November 17). Retrieved October 17, 2020, from https://www.cdc.gov/nchs/pressroom/97facts/edu2birt.htm
2. Pobric, A., & Robinson, G. M. (2015). Population ageing and low fertility: Recent demographic changes in Bosnia and Herzegovina. Journal of Population Research, 32(1), 23-43. doi:10.1007/s12546-014-9141-5
3. Martinez, G., Ph.D, Daniels, K., Ph.D, & Chandra, A., Ph.D. (2012, April 12). Fertility of Men and Women Aged 15–44 Years in the United States: National Survey of Family Growth, 2006–2010. Retrieved October 17, 2020, from https://www.cdc.gov/nchs/data/nhsr/nhsr051.pdf
4. Fertility rate, total (births per woman). (n.d.). Retrieved October 17, 2020, from https://data.worldbank.org/indicator/SP.DYN.TFRT.IN
5. 2017 General Social Survey (GSS): Families Cycle 31. (2017). Retrieved October 17, 2020, from Statistics Canada.
6. Beaupré, P. (2020). General Social Survey Cycle 31 : Families Public Use Microdata File Documentation and User's Guide. Ottawa, Canada: Authority of the Minister responsible for Statistics Canada. Retrieved October 17, 2020, from https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf