

# 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 necessary 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, selfRated\_health and selfRated\_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 influence.

## Introduction

Our goal is to find a model to predict how many 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, selfRated\_health and selfRated\_mental\_health. Total children is the attribute we are interested in. Education and partner\_education are important because people who are enrolled 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 one 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 selfRated\_health and selfRated\_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.

## 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. The data are collected with the stratified random sampling 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 (partial or total) non-participation. This is handled by adjusting the weights to less for non-participation cases (Beaupré, 2020).

## Model

## 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](https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf)