2017 Canadian Marriage Investigation

xinyi zhang

2020/10/19

2017 Canadian Marriage Investigation

Xinyi Zhang 2020 10 17

Abstract

Based on the analyzed result of the 2017 General Social Survey published by Statistics Canada, I investigated how combined variables, such as age, education, and aboriginal identity, affect people's experience (whether they have experience of marriage). Since we are investigating the relationship between the variable ever_married and the other predictors, a logistic regression model would be implemented to

Introduction

In this report, I will mainly investigate on experience of marriage of Canadians based on the 2017 General Social Survey. Our hypothesis is the probability of ever married is affected by age, education and whether being aboriginal or not. To research this hypothesis, I will use logistic regression model considering the ever_marriage variable is a identical variable, and I will include various predictors such as age, education level and whether grandparents still living since I am curious to see if these variables have effect on the response variable. In the following section, I will be setting up an appropriate model to testify my assumption and use the model and statistics to find any significant evidence.

Data

I am using the 2017 General Social Survey data set, which includes various aspects of Canadians 15 years of age and older, living in Canada's ten provinces. This data is released from statistics Canada which further guarantees its legitimacy and reliability. A governmental institution will have more legal and valid information(eg.phone number) to conduct the research compared to the other companies.

The dataset has several advantages. First of all, it has a different primary focus each year, and 2017 focuses on family. Considering the topic that I am interested in, the 2017 dataset would be an appropriate and up-to-date option. Also, the data is completed and through considering its response rate and the number of cases collected. The response rate is 20 percent higher than the previous years' survey and ended up with around 30 million cases collected. It is outstanding that the survey designers recognized the shift of ways that people get connected to. They have changed the sampling frame, which is the device they use to connect to the target population they want to gain information from, to better adapt to the shift of trend. A higher response rate and more sample cases enable us to achieve a more unbiased insight into the population and question.

From the perspective of a user of this data, there exist minor drawbacks. Under the variable ever_married, which is a yes or no question, it is unnecessary to include the other options. I found some people responded with "Don't Know," which might be a mistake considering the extremely low percentage of people choosing

this option. But they can avoid this problem by not giving out the "Don't know" option since it also take extra steps for the data users to verify and handle the minor cases.

In terms of the questionnaire, the most outstanding feature is that it is a profound questionnaire. There are both general questions and questions that ask for more depth and details. Also, there is no short answer question, which makes the data easier to quantify and visualize.

Their target population for the 2017 GSS included all persons 15 years of age and older in Canada, excluding residents of the Yukon, Northwest Territories, and Nunavut and full-time residents of institutions. Since they are governmental institutions, they have legal information on Canadian residents' phone numbers, home phone numbers, addresses, etc. It is feasible for them to reach out to the target population. The frame, which is the device they use to reach out to the target population, lists a phone number, and address register. They also changed the frame this year to better capture the target population since many Canadians today don't have home telephones anymore. Therefore, cell phone numbers would make that portion of the sample reachable. They have used a stratified sampling method to collect data from the ten provinces. They first divided the population into homogeneous groups by province geographically. Homogeneous means every sample under strata (a province in the case) has no significant difference. Since every province has a different population, it can reduce bias and variability of results by applying the stratified sampling method. In terms of the non-response problem, Statistics Canada strived to reach the randomly selected member. They will try several times and explain the survey's importance even when the member refuses to participate. Also, people who are not convenient at the moment of the call will receive a phone call at another adjusted time. All the effort the interviewers paid would reduce the ono-response problem at some level.

Model

I decided to apply the logistic regression model since the dependent variable that I would like to investigate is binary. Binary refers to variables with only Yes or No answers(Yes=1, No=0). Thus it would not be appropriate for me to use a linear model since points are not scattered evenly around a straight line. The logistic equation that contains my response variable and prediction variables is:

$$\ln Odds = \beta_0 + \beta_1 x_{Age} + \beta_2 x_{education} + \beta_3 I_{aboriginal}$$

Results

Table 1: Modeling odds of ever married

	Estimate	Std. Error	z value	Pr(>|z|)
Baseline	0.049	1.078	0.000000e+00	1.000
age	1.084	1.001	3.052988e+25	1.000
educationCollege, CEGEP or other non-university certificate or di...	1.034	1.063	1.734000e+00	1.790
educationHigh school diploma or a high school equivalency certificate	0.791	1.063	2.200000e-02	1.000
educationLess than high school diploma or its equivalent	0.537	1.077	0.000000e+00	1.000
educationTrade certificate or diploma	0.865	1.086	1.720000e-01	1.082
educationUniversity certificate or diploma below the bachelor's level	1.061	1.131	1.613000e+00	1.883
educationUniversity certificate, diploma or degree above the bach...	1.048	1.088	1.744000e+00	1.783
aboriginalYes	0.868	1.093	2.040000e-01	1.119

For analysing the results, we are going to look at the table above showing the odds of ever married. First of all, with every increasing unit of average age, the odds of having ever married will increase by 8.4%. In terms of education level contributing to the odds of ever married, people with college diploma or any non-university certificates, university certificate no matter above or below bachelor's level have a higher odds compared

to people with bachelor's level. Being about 3.1%, 6.1% and 4.8% higher than people with bechelor's level respectively. On the other hand, those who got high school level or lower diplama and those who got trade certificate will have a lower odds of having ever married. They are 79.1%, 53.7% and 86.5% odds of those people who got bachelor's level. Finally, the aboriginal people will have a lower odds of having ever married compared with non-aboriginal people, being 86.8% of the later.

Discussion

Based on the result shown above, we observed that higher age than average will lead higher chance of getting married, which corresponds the social phenomenon. On the other hand, what surprised me was the lower education will cause smaller possibilities of having ever married in general. On the contrast, people with above bechelor's level have a very decent chance of getting married. Last but not least, the aboriginal people have slightly lower chance of ever married than those who are not aboriginal.

Weaknesses

To see the impact of independent variables on marriage experiences, I have run the logistic model to see the relationship. However, I have not testified whether outliers or the other affecting factors do not drive the independent variables. For example, education might not be the most original factor affecting the result of marriage; there might be another variable that affects education. Also, although the sample size of data is significant, there is an amount of N/A under many variables, making it complicated to run an ANOVA. ANOVA can explain how well the regression might explain the variation we observe in the response, and it requires the data to be clean.

Next Steps

To make the predictor more accurate and crucial, it would be better to pick more than three more promising predictors rather than follow my instinct only to choose 3. I can then compare the AIC and BIC of all the promising independent variables to pick out the best predictors. In such a way, I can fit a more accurate model and predict more accurate results on ever married variable.

References

2017 General Social Survey Cycle 31 : Families Public Use Microdata File Documentation and User's Guide. (n.d.). Retrieved October 19, 2020, from https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/index.htm

2017 General Social Survey: Families Cycle 31: Public Use Microdata File (n.d.). Retrieved October 19, 2020, from https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/index.htm

GSS.cleaning, Rohan Alexander and Sam Caetano. Retrieved October 19, 2020, from https://q.utoronto.ca/courses/184060/files/9422740?module_item_id=1867317