The relationship between the intention of having children and religious belief for people in Canada

Jiafei Lyu (1005004727) Haichuan Xue (1004070346)
Wenzhuo Zeng (1005227050)

2020-10-18

Abstract

Nowadays, a significant decrease of fertility rate has shown in many countries. This report uses the 2017 Canadian General Social Survey (GSS) to propose a logistic regression model on fertility intentions using survey estimation. The model has shown that fertility intentions and extreme feelings toward religiosity have intimate connection with the control of other covariates. People who describe religion as "extremely important" tend to have higher fertility intentions compared to people for whom religion is "not important at all". This result can be caused by complicated reasons such as traditional religious family ideology, different perspectives in importance of family, parent's attitudes and the acceptance of non-marital sex relations.

Introduction

A study by Sarah R. Hayford and S. Philip Morgan in the US has shown that women who report that religion as "very important" have a higher fertility rate or higher intending fertility rate than women who consider religion as "somewhat important" or "not important". Due to this study in the US, we want to explore more about the relationship between religiosity and fertility intention in Canada using survey estimation from the data provided by Statistic Canada to see if the results are similar. A potential logistic model was built between religion, fertility rate and some other covariates to examine the connection. By controlling multiple covariates, the trend of fertility intention can be examined with respect to the increase of the importance of religion. This report will also discuss how the data was collected along with some weakness and strength of this approach method and the model constructed.

Data

A brief introduction to GSS

The target population for the 2017 Canadian General Social Survey is all persons aged fifteen and above in Canada, excluding residents of the Yukon, Northwest Territories, and Nunavut, and full-time residents in institutions. From statistics generated by Statistics Canada, such target population size comes to a total of 30,538,825. The ten provinces of interest were divided into strata based on their geographic location, which means that the population was divided into non-overlapping subgroups accordingly. The ten provinces formed a total of 27 strata.
The frame is a device from which a sample is drawn, which in this survey was a list of telephone numbers, where 86% of numbers were associated with one household, and the other 14% were not linked to any household. A simple random sample without replacement was performed in each stratum, which means that

every member in each stratum has an equal probability of being chosen and can only be chosen once.
All respondents were reached and interviewed by telephone. Several time periods were also set to be the best time to call interviewees. The selected interviewees who first refused to participate were re-contacted several times to explain the importance of the survey and were encouraged to participate. The overall response rate for this survey was 52.4%.

Strengths and weaknesses of GSS

By using a simple random sample without replacement, the results would be more accurate and unbiased. However, it is definitely expensive and time consuming to reach out to this many different people. The interview method, which is through telephone, is probably the easiest and the most efficient way to carry out the survey, and is probably the only way to get in contact with such a large sample spread widely in Canada. However, this can also cause problems: there is a high rate of interruption of the interview, respondents may have difficulties to provide personal information through the phone to a stranger. The repeated contact with a selected interviewee who first refused to do the survey helps to make every selected sample meaningful and helps to make the survey more complete, but its ethic issue should be questioned. The repeated contact may cause the interviewee's resistance in doing the survey and disturbance of their regular life.
Some missing information such as income and age where it can be found in Statistics Canada was imputed. Complete non-response households with no information available found in Statistics Canada were dropped.

Data used in this study

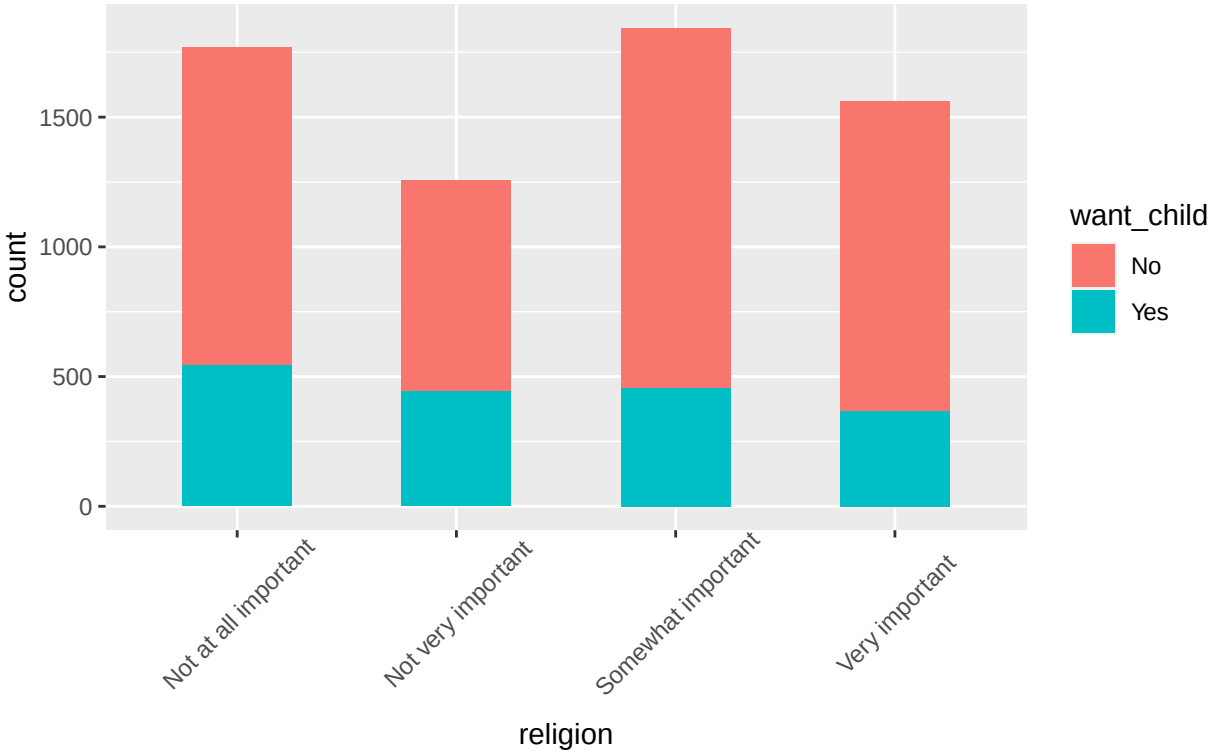
Our data is retrieved from the GSS 2017 dataset. After the cleaning and variable adjusting process, it contains 5141 observations with six variables. All the missing values and meaningless responses, such as unsure, skip, and do not know, are omitted since they do not contribute to our analysis.

The six variables contain the aspects of a subject's:

1. the intention of children in the future (Yes or No)
2. importance of religious belief (Not at all, Not very, Somewhat, or Very)
3. age (ranged from 15 to 59, with 21 as the mean)
4. sex (Male or Female)
5. self-rate of life quality (Bad/Average, Good/Superb)
6. whether the subject has a child before (Yes or No)

Figure 1

Histogram on Religion, with outcome of our interest



From Figure 1, we can see that in each level of the religious belief, the frequencies of having no intention of children in the future are generally close, and the frequencies of having the intention are various. However, Figure 1 does not have any other influencing factors included. We cannot give a proper justification of what we study based on this plot. Thus, the variables listed above from the 3rd to the 6th ones are considered as confounder variables. They are believed to have an influence on both the intention of children and the importance of religious belief.
Some potential drawbacks of the data are many of the variables may have answering biases from the responders, the sample size may be large enough to represent the target population, and factors of each variables are restricted and may not be representative.

Model

Methods

The GSS dataset provides information about how many children are wanted in the future from the survey participants. This information is what we are interested in and is converted into a binary variable, as the response variable Y. It indicates whether the outcome is to want children in the future. Thus, we propose a logistic regression model through R software to present the relationship between the response variable and the predictors. Due to the GSS was sampled based on different regions independently and aimed for the whole population in Canada as the target population, the infinite population correction and specifying the primary sampling unit is applied in the model through a survey design function in the survey package in R. The model is presented as the following:

$$log\left(\frac{P(Y_i|X_i)}{1-P(Y_i|X_i)}\right) = \beta_0 + \beta_1 \text{R. not. very} + \beta_2 R.somewhat + \beta_3 R.very + \beta_4 Age$$
$$+\beta_5 Sex.M + \beta_5 Sex.M + \beta_6 Rate.GS + \beta_7 Child.Y$$

Where:

- Y_i represents the result of whether planning to have children. If the ith subject wants children, then $Y_i = 1$. Otherwise, $Y_i = 0$.
- $P(Y_i|X_i)$ is the probability of planning to have a child for the ith subject, given the conditions of X_i.
- β's are the intercept and coefficients of each predictors to show the effectiveness

Model diagnose

We propose to use cross validation and the area under the curve (AUC) - receiver operating characteristics (ROC) curve – to perform the model diagnosis. In the cross validation, the data is divided into a training data to construct the model, containing 80% of the original data, and a testing data containing the rest data. The two statistic values of AUC are calculated for both training data and testing data. With a high value of AUC, it indicates that the model has a very good discriminate ability to distinguish the results between wanting a child and do not want a child.

Results

	Abbr.	Exponential Estimate	Std. Error	t value	P value
Intercept		6.58	0.24	7.75	0.00
religion Not very important	R.not.very	1.25	0.14	1.62	0.11
religion Somewhat important	R.somewhat	1.2	0.13	1.39	0.16
religion Very important	R.very	2.17	0.16	4.93	0.00
age		0.83	0.01	-24.98	0.00
sex Male	Sex.M	2.17	0.1	7.62	0.00
feelings life Good/Superb	Rate.GS	2.02	0.21	3.3	0.00
ever have child Yes	Child.Y	0.11	0.12	-17.77	0.00

Table 1. Summary of Model

Figure 2: AUC−ROC curve using training data

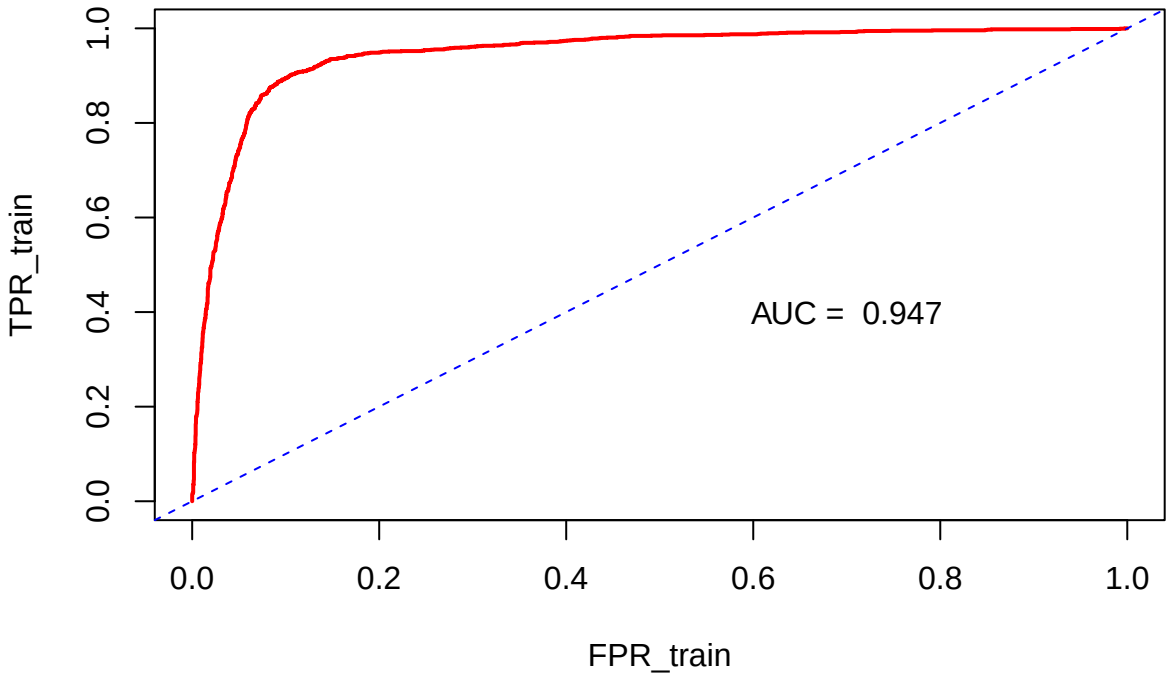
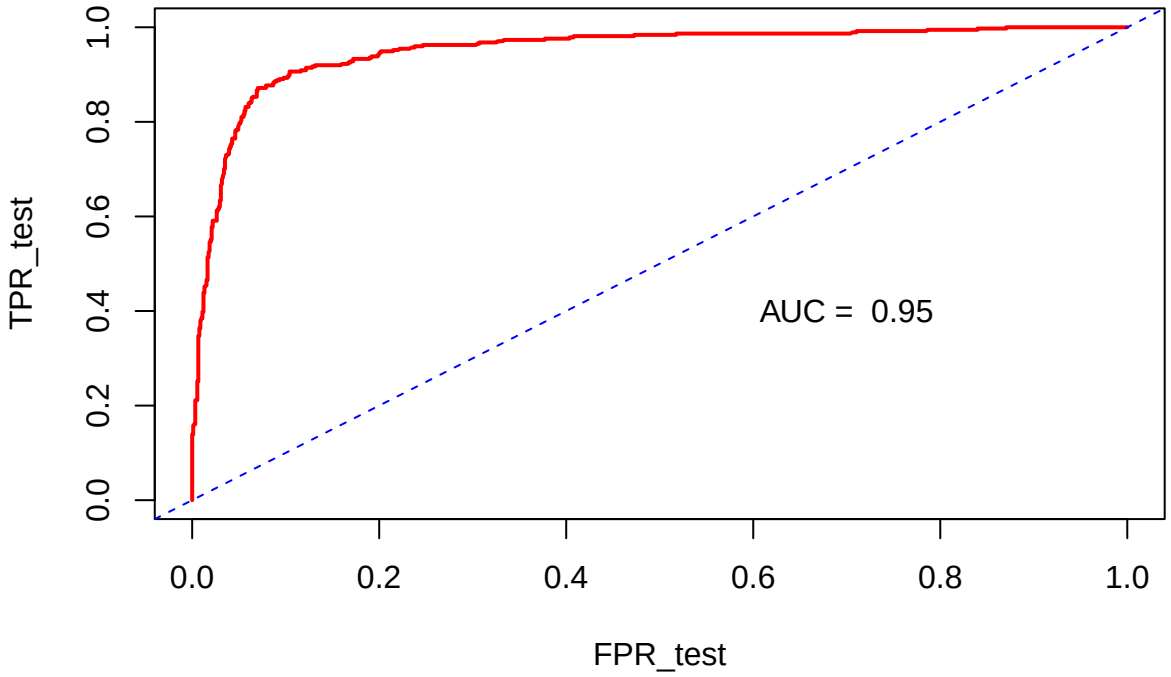


Figure 2: AUC−ROC curve using testing data



Discussion

Discussion of model & model diagnose

From Table 1, we can see that the odds of planning to have children for a 15-year-old woman, who thinks religion is not important to her at all, does not feel good or superb about her life, and does not have a child before, is 6.58. As the age increases by one unit, the odds will decrease when controlling the other conditions. For the same woman mentioned above, if her religious belief becomes very important to her, the odds of planning to have children will be 117% higher. If the religion is not very important or somewhat important to her, the odds will be 25% and 20% higher, respectfully. However, we can only conclude that taking religions as a very important matter or as an absolutely not important matter is significantly associated with the response variable since both of their p-values are less than 0.05. If the p-value of an estimation of beta is greater than 0.05, it means that the interpretation of this estimation is not significant or statistically supported. To sum, we can see that when religious belief is extremely important, compared to not important at all, to a person, the person is more likely to plan to have children with the control of other covariates.
From Figure 2, the AUC, equaling to 0.947, is calculated based on the training data for the model. It indicates that the model can distinguish between planning to have children and not planning to have children for a subject in the training data at 94.7% of the times. In other words, the model is well built. From Figure 3, instead of using the training data, the testing data is used for calculating the AUC value. The AUC, equaling to 0.95, indicates that at 95% of the times, the model can distinguish between planning to have children and not planning to have children for a subject. Thus, the prediction ability of the model is highly qualified.

Weakness of model

The aspects covered by the model are limited. There might be other confounding variables which are more decisive but left out. Also, the sex variable here is limited only to males and females. It does not consider the diversity of sex identities.

Discussion of conclusions

The GSS was to gather data of general Canadians and to provide information on specific social issues of current or emerging interest. Results from Sarah R. Hayford and S. Philip Morgan's study was examined using data from the GSS. Our built model agrees with the result that if religion is highly important to someone, their fertility intention will be very strong. However, one other interesting finding in this study that was not expected is that, for a person whose religion is not important at all to themselves, their fertility intention is also very strong, although not as strong as people who value their religion heavily. People who value their religion has not very important or somewhat important did not show a significant relationship between religiosity and fertility intention. The reasons for this result are yet to be discovered more deeply, But it could be that the people who believe their religion is very important to them are more traditional and want a traditional family. It also could be that people who have a traditional family value religion more heavily.

Next step

Increasing the sample size and collecting more data on relevant aspects are highly recommended to build a better model. Besides, the construction of the model is based on the GSS 2017 dataset only. It is recommended to use a different dataset from another year (from recent years) and study on the similar topic. Seeking the pattern among recent years can also be done for further investigation.

Appendix

GitHub link (pick one that works)

https://github.com/HaichuanXue/The-relationship-between-the-intention-of-having-children-and-religious-belief-for-people-in-Canada.git

gh repo clone HaichuanXue/The-relationship-between-the-intention-of-having-children-and-religious-belief-for-people-in-Canada

References

1.Computing in the Humanities and Social Sciences.(n.d). Retrieved from http://www.chass.utoronto.ca/

2.Hayford, S. R., & Morgan, S. P. (2008). Religiosity and Fertility in the United States: The Role of Fertility Intentions. Social forces; a scientific medium of social study and interpretation, 86(3), 1163–1188. https://doi.org/10.1353/sof.0.0000

3.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

4.Statistics Canada. (n.d.). Retrieved from https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1710000501

5.T. Lumley (2020) "survey: analysis of complex survey samples". R package version 4.0.

6.Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

7.Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77. DOI: 10.1186/1471-2105-12-77 http://www.biomedcentral.com/1471-2105/12/77/