Better Household Living Conditions Improve Life Satisfaction

Rachel Oh, Yang Shang

19/10/2020

Abstract

This report describes the effects of household living conditions on life satisfaction. It is found that better household living conditions do improve life satisfaction to a certain degree. These results are important as it is fundamental that people know what makes them satisfied with their lives because no human would wish to have an unsatisfying life.

Introduction

In our studies, we established an interest about the relationship between people's life satisfaction and their living conditions. Resourced from Canadian GSS in 2017, our team is going to describe an individual's living conditions from the following aspects:

1. Household type: what type of residence is the respondent living in (house, apartment, etc.)?
2. Household size: how many residents (relatives and non-relatives), including the respondent, are living in the residence?
3. Ownership of the residence: is the respondent owning or renting the residence?
4. Cohabitation: Is the respondent living with a partner (married or not married)
5. Children in household: Is there any child (relatives and non-relatives) living in the residence? The detail of the above factors will be further described and explained in Section "Model".

Our prediction to the model is that: 1. people live in houses have higher life satisfaction; 2. the larger the household size, the higher the life satisfaction; 3. people owning the residence have higher life satisfaction; 4. people live with their partners have higher life satisfaction; 5. people live with children have higher life satisfaction.

Full code and data supporting this analysis is available at: https://github.com/hyunbani/household-living-conditions-and-life-satisfaction

Data

General Description of GSS

Statistics Canada's General Social Survey (GSS) is an annual survey with statistical information on the living conditions and well-being of people living in Canada. In this report, we are using the data collected in 2017. The main focus of that year's survey is "Family".

The Survey is divided into 86 minor sections, with around 200 questions in total, to describe the respondent's general identity, marriage, family relationships and identity, child care, social involvements, living conditions, career, and religion.

People of interest were those aged at 15 and older, living in all 10 provinces in Canada. However, it would be costly and impractical to survey each and every household in Canada, so random sampling was chosen as a method to determine the sample of 43,000 people, which reflects the population. In order to ensure that the sample is an accurate reflection of the population as a whole, the survey results from all sampled households were collected, and only one eligible person per household was interviewed.

Collection of GSS Data

The 2017 GSS was performed through telephone interviews. People who did not answer the telephone calls were excluded from the sample. However, there still exist some non responses for some of the questions being asked, and these are recorded as 'NA' in the dataset.

Since the answers are kept strictly confidential, the respondents must have provided the answers in a more honest manner, leading to an increase in accuracy of data information. However, as there are about 200 questions in total, even if some can be validly skipped according to respondent's status, it still took a lot of time to complete one servey. Besides, many people tend to refuse the phone calls. Thus, we believed that telephone interview might not be the most efficient collecting method.

About Our Studies

The team's interest of study is the relationship between life satisfaction and living conditions. We extract colume "feelings_life" to represent life satisfaction. We also extract columes "own_rent", "hh_type", "hh_size", "lives_with_partner", and "children_in_household" to describe household living conditions.

Our response variable "feelings_life" is a discrete numerical variable measured on a scale of intergers from 0 to 10, with 10 being the highest satisfaction level and 0 being the lowest. The team was previously interested in using data "self_rated_mental_health", which use categorical variables "poor", "fair", "good", "very good", "exellent" to describe the respondent's mental conditions. However, we experience difficulties creating and summerizing the model using this variable, even if we give it numerical expression. Then we switch to the current variable "feelings_life". We believe that it is an upper level data included the effect of respondent's mental health.

Figure 1 below is a bar graph showing the proportion of respondents' life satisfaction.

Figure1: Proportion of Respondents' Life Satisfaction
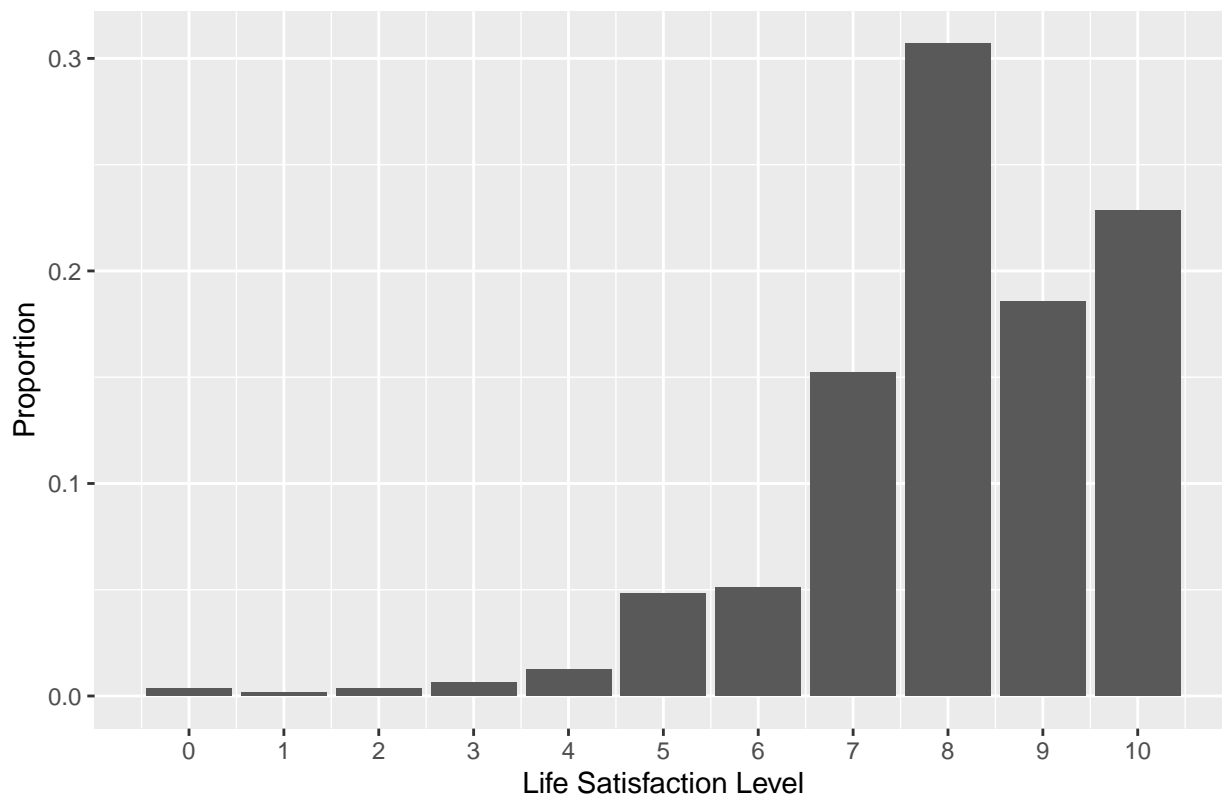
The graph in figure 1 shows that about 40% of the respondents rated their life satisfaction to 9 and 10, about 45% of the respondents rated 7 and 8 (with 30% in total rated 8), 10% rated 5 and 6, and less than 5% rated 4 and below. The median is 8 and mean life satisfaction is 8.101. We are pretty confident to say that most of the respondents have a pretty good attitudes toward life.

Variable "own_rent" illustrates if the respondents owned or rented their current residence. We believe that it is related to respondents' income level. However, factors like family history also affects respondents' ownership to a residence.

Variable "hh_type" shows the type of residence the respondents were living in, including "high rise apartment", "low rise apartment", "single detached house", and "other". The team has simplified the variable to two parameters, "single detached house" and "other" representing the other 3 residence types.

Variable "hh_size" is a categorical variable specified by integers from 1 to 6 showing the size of the household. "6" is specially representing a household of 6 or more people. In our model, we will treat this as a discrete numeric variable.

Variable "lives_with_partner" is a categorical variable showing if the respondent is living with a partner (married or not married). The variable includes some effect of the respondents' status of marriage. However, the variable doesn't measure the effect of parents in a household as some respondents probably lives together with their parents.

Variable "children_in_household" is a categorical variable showing how many children are living in the household. Non-relative children are also counted. The original data is classified by "no child", "one child", "two children", and "three or more children". In our model, we simplified it to just measure the effects of having child or not inside the household.

For the test set the team is using in the model, we have created some dummy variables for the data set. The default setting of a respondent is: "rent the residence", "lives in other types of residence", "lives without partner", "lives without child in household".

We have also modified names to the variables. The explanation of each explanatory variable is as follows:
1) 'Household Size' - the number of people each respondent lives with
2) 'Owned' - 1 if a respondent owns the house and 0 if he/she doesn't
3) 'Partner' - 1 if a respondent lives with a partner and 0 if he/she doesn't
4) 'No Child' - 1 if no child lives in the household with the respondent and 0 if he/she does
5) 'Single House' - 1 if respondent's house type is single detached house and 0 if it's not

Table 1 below shows the first ten rows of the data with the variables of interest.

Table 1: First 10 Rows of Raw Data

Case	Life Satisfaction	Household Size	Owned	Partner	No Child	Single House
1	8	1	1	0	1	0
2	10	2	1	0	1	1
3	8	2	1	0	1	1
4	10	2	1	0	1	0
5	8	2	0	1	1	0
6	9	2	1	0	1	1
7	4	1	0	0	1	0
8	10	1	0	0	1	0
9	8	1	1	0	1	0
10	5	6	1	0	1	1

Model

We are interested in explaining whether a person's life satisfaction improves when he/she has better household living conditions, based on household size, house ownership, whether he/she has a partner, no child, and lives in a single detached house. In order to do so, we are going to fit a multiple linear regression model.

Multiple linear regression is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of it is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables.

As discussed in the data section, the dependent variable is Life Satisfaction and the explanatory variables are Household Size, Owned, Partner, No Child and Single House.

Formula and calculation of this model is as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \epsilon$$

where, for $i = 20331$ observations:
y_i = dependent variable
x_i = explanatory variables
β_0 = y-intercept (constant term)
$\beta_1...\beta_5$ = slope coefficients for each explanatory variable
ϵ = the model's error term (also known as the residuals)

Results

Table 2: Summary of the Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.41	0.05	135.47	0.00
Household Size	0.09	0.01	7.04	0.00
Owned	0.44	0.03	14.00	0.00
Partner	0.10	0.04	2.58	0.01
No Child	0.07	0.04	1.76	0.08
Single House	0.14	0.03	4.88	0.00

Figure 2: Relative Importance of Predictor Variables

% of R–Square
Total R–Square= 0.029
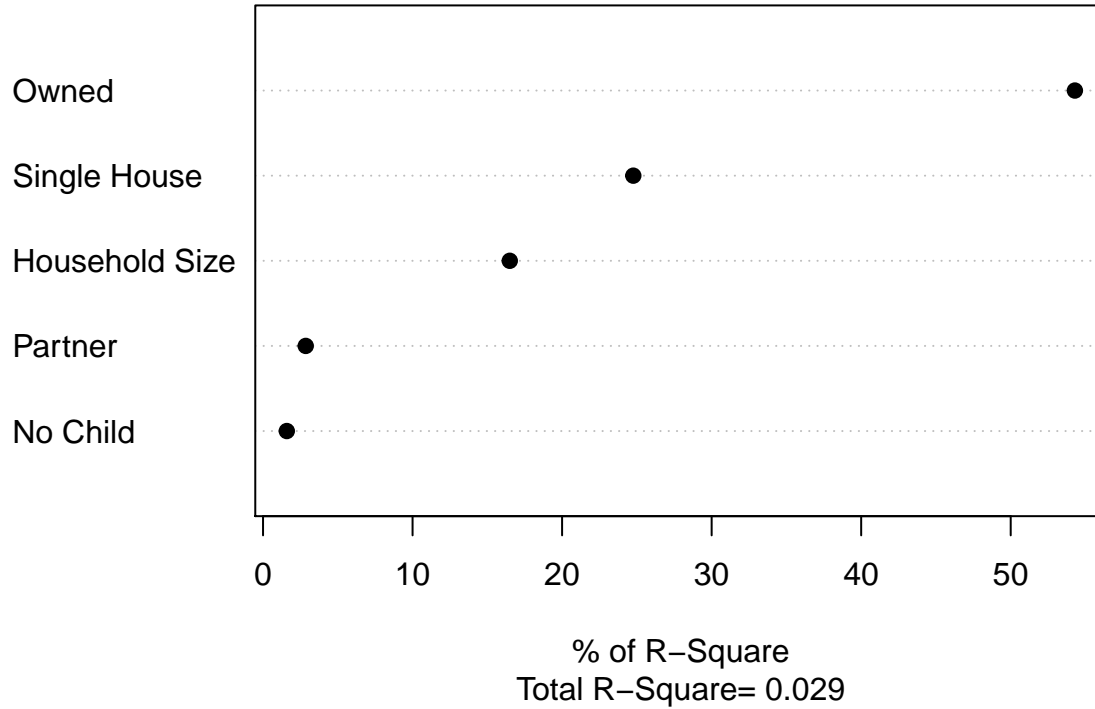
Discussion

From Table 2, We see that the p-values (represented as Pr(>\|t\|)) for all explanatory variables are less than 0.10. From this, we know that all these factors increase the life satisfaction level with a significance level of 0.10. It means that you will be more satisfied with your life if you have more people living in your household, own the house, live with a partner, live with no child, and live in a single detached house.

However, the significance for each explanatory variable is different on the response variable. If you look at Figure 2, you will see that whether you own the house or not has the biggest impact on your life satisfaction level. The variable with the least impact is having no child and this can also be seen through the graph.

References

- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

- JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2020). rmarkdown: Dynamic Documents for R. R package version 2.3. URL https://rmarkdown.rstudio.com.

- 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.

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

- Canadian General Social Surveys (2017). General social survey on Family (cycle 31). URL https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=335816

- Hadley Wickham and Evan Miller (2020). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.3.1. URL https://CRAN.R-project.org/package=haven

- Jacob Kaplan (2020). fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from Categorical Variables. R package version 1.6.2. URL https://CRAN.R-project.org/package=fastDummies

  Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30.

  Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963

  Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595

  H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.