PS2

Saksham Ahluwalia, Labib Chowdhury, Lisa Oh, Eric Yuan

2020-10-12

Abstract

Section 1: Introduction

Life satisfaction is recognized to be an important measure of an individuals well-being. There are a lot of factors that influence one's satisfaction of life. Factors like income [1], education [2], mental health [3], and physical health intuitively influence an individual's life satisfaction. Less intuitively, there is increasing evidence that ethnic minorities and immigrants typically have lower levels of life satisfaction when compared to non-visible minorities [4]. Additionally, we wanted to also look into other factors like how many hours an individual works in a week as an additional factor to income; by having the number of hours an individual works in a week, this gives us further insight on the responses where there is high income, but the respondent is overworked which leads to lower levels of life satisfaction.

In Section 2 we go over the design of the survey, as well as the data we chose and constructed. Section 3 provides the logistic regression model design. Section 4 provides the results of our model. Section 5 provides interpretation and analysis of our model and results.

Section 2: Data

Survey design

The following section defines the source of the data, the sampling procedure used in the study, some potential drawbacks of using this specific sampling procedure, how respondents were contacted and finally a brief discussion on how the Non-responses were handled. All of the following information is adapted from the user guide.

The data set used in this report was collected by GSS (General Social Survey) Canada from February 2nd to November 30th 2017 [12]. All respondents were interviewed via telephone. Households without telephones were excluded from the survey population. Telephone calls were made to randomly selected members within each household captured in the sample frame. During telephone calls if it was determined that a household did not have at least one person 15 years of age or older, the interview was terminated. The sampling technique employed by GSS to collect data for this survey was Stratified Random Sampling technique. Stratification was done based on the provinces. Each province had CMA[s] (Census Metropolitan Area) and non-CMA[s]. The following cities were considered their own strata: St. John's, Halifax, Saint John, Montreal, Quebec City, Toronto, Ottawa, Hamilton, Winnipeg, Regina, Saskatoon, Calgary, Edmonton and Vancouver. Three more stratas were created by grouping the CMA[s] in Quebec, Ontario and British Columbia that are not listed above. Additionally, non-CMA[s] of each province were considered their own strata. Therefore, in total 27 stratas were created. To summarize, 17 stratas were made up of CMA[s] and an additional 10 were made up of the non-CMA[s] of each province. After stratification, phone numbers that belonged to

the same address were grouped together. Each of these groups was then assigned to a stratum within the province they belonged to. Finally, a SRS (simple random sampling) without replacement of grouped phone numbers was performed in each stratum. Since the sampling followed SRS we assume that the observations are independent of each other. One of the drawbacks of a probability sampling like SRS is that it might be subject to non-participation from the randomly chosen participants. This can cause more time to be spent to clean and edit the data. For example, in this study non-responses for questions used for weighting were imputed based on other information from the questionnaire. Weights were also calibrated for responding telephones to represent non-responses. Some key-tasks of this survey were to aid the Canadian government with policy decisions. The survey was also designed to monitor changes in the living conditions and well-being of Canadian over time. Some disadvantages to the survey are that some of the survey questions are very long and difficult to understand. The survey is also very lengthy. This might cause some individuals to leave in between the survey. Some strengths of the survey are that if completed it provides lots of data points for analysis that can help the government with policy decisions.

Study population

The analysis was conducted based on the responses for the 2017 Canadian General Society Survey from non-institutionalized individuals that are 15 years of age or older that are living in the ten provinces of Canada. The target population of this survey is 30,530,800 (N=30520800); this figure was calculated under the assumption that the proportion of individuals that are 15 years of age or older is the same across all the provinces in the country in 2017 (p=0.8382249206) [5], and the population in the provinces in the country in 2017 was 30,622,177 [6]. Of the 20,602 responses that were obtained, our sample only consisted of 11,233 (n=11233) responses (54.5%) after removing responses from individuals who did not respond, skipped, or did not know how to answer a question.

Measures of interest

Feelings of Life

The primary variable of interest of this study is how people felt about their life as a whole, rated from a scale from 0 to 10, where a 0 represents "Very dissatisfied" and a 1 represents "Very satisfied". Out of the 20,602 responses that we had, 20,331 (98.7%) responses did not answer "Valid skip", "Don't know", "Refusal" or "Not stated".

As seen in Figure 1, it is evident that there is a left skew in the life satisfaction scores. Thus we became more interested in predicting whether a person's life satisfaction score is above or below the mean value. Hence we constructed a new binary variable that represented feelings of life; if a response was below the 50th percentile they would have a feelings of life score of 0, and if a response was on or above the 50th percentile they would have a feelings of life score of 1. The value that represented the 50th percentile was 8.13, however since feelings of life scores are represented as integers, we rounded down to 8; moreover, all responses under 8 received a feelings of life of 0, otherwise 1.

```
library(tidyverse)
```

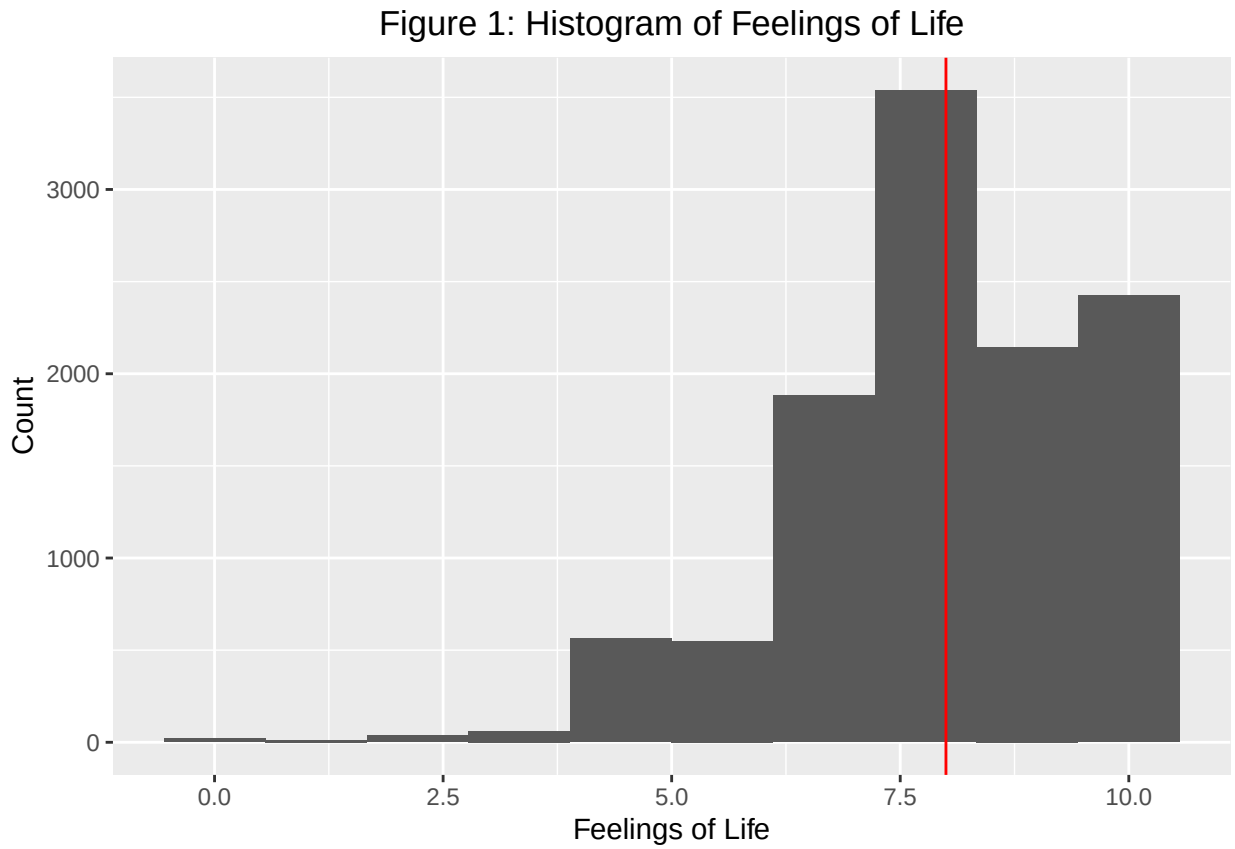
```
## -- Attaching packages ---------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.1     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# Import cleaned GSS data
setwd("~/Documents/sta304-ps2")
gss <- read_csv("gss.csv")
```

```
## Parsed with column specification:
## cols(
##   feelings_life = col_double(),
##   vis_minority = col_character(),
##   hours_worked = col_character(),
##   family_income = col_character(),
##   hh_type = col_character(),
##   education = col_character(),
##   self_rated_health = col_character(),
##   self_rated_mental_health = col_character(),
##   age = col_double(),
##   feelings_life_binary = col_double()
## )
```

```
ggplot(gss, aes(x=feelings_life)) +
  labs(title = "Figure 1: Histogram of Feelings of Life") +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab("Feelings of Life") +
  ylab("Count") +
  geom_histogram(bins = 10) +
  geom_vline(xintercept = round(mean(gss$feelings_life), 0), colour = "red")
```



TODO: ** FIGURE 1: SHOW HISTOGRAM OF FEELINGS OF LIFE SCORE, WITH 50% PERCENTILE LINE **

Visible Minority

The socioeconomic factors that we looked at in this study were whether or not the respondent was a visible minority, how many hours the respondent works a week, how much money do they make, and what level of education did the respondent obtain.

A respondent is determined to be not a visible minority if they answer "White", any other answers like "South Asian", "Chinese", "Black", "Filipino", "Latin American", "Arab", "Southeast Asian", "West Asian", "Korean", or "Japanese". Visible minority is represented as a 1, and not a visible minority is represented as a 2. Out of the 20,602 respondents, after filtering out "Valid skip", "Don't know", "Refusal", and "Not stated" entries, there were only 20,148 (97.8%) responses.

Other Covariates

Some other variables that we found to be interesting were state of physical health (Excellent, Very good, Good, Fair, Poor), state of mental health (Excellent, Very good, Good, Fair, Poor), highest level of education obtained (< high school diploma, trade certificate or diploma, College, CEGEP or other non-university certificate, University certificate or diploma below the bachelor's level, bachelor's degree, University certificate, diploma or degree above the bachelor's), hours worked in a week (0 hours, 0.1 to 29.9 hours, 30.0 to 40.0 hours, 40.1 to 50.0 hours, or 50.1 hours and more), family income (<$25,000, $25,000 to $49,999, $50,000 to $74,999, $75,000 to $99,999, $100,000 to $124,999, $125,000 or more), dwelling type (Single detached house, low-rise apartment (less than 5 stories), high rise apartment (5 or more stories), other), and their age in

years (ranges from 15 years of age to 80 years of age (80 years of age represents 80 years of age or older)) were retrieved directly from the survey responses.

Section 3: Model

Section 4: Results

Section 5: Discussion

References

Model

We will model the probability of a Canadian feeling satisfied about life as whole based on a set of given factors using Logistic Regression.

Let p = the probability of a Canadian feeling satisfied, where "feeling satisfied" is determined if an individual rates their feeling about life as a whole greater than 8. Then our GLM will be formulated as the following:

$$\log(\frac{p}{1-p}) = \beta_0 + \sum_{i=1}^{K} \beta_i x_i$$

Results

```
#install.packages("survey")
#install.packages("brms")
rm(list=ls())
library(janitor)
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(tidyverse)
library(survey)
```

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loading required package: survival

##
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
data <- read_csv("gss.csv")
```

```
## Parsed with column specification:
## cols(
##   feelings_life = col_double(),
##   vis_minority = col_character(),
##   hours_worked = col_character(),
##   family_income = col_character(),
##   hh_type = col_character(),
##   education = col_character(),
##   self_rated_health = col_character(),
##   self_rated_mental_health = col_character(),
##   age = col_double(),
##   feelings_life_binary = col_double()
## )
```

```
# glm model
N = 30530800
n = length(data$feelings_life_binary)
fpc.srs = rep(N, n)

satisfied.design <- svydesign(id=~1, data=data, fpc=fpc.srs)

satisfied.glm <- svyglm(feelings_life_binary ~ age + as.factor(vis_minority) + as.factor(hours_worked) +
                          as.factor(family_income) + as.factor(self_rated_health) + as.factor(self_rated_
                          as.factor(education),
                        design=satisfied.design, family="binomial")
```

```
satisfied.glm %>%
  broom::tidy() %>%
  knitr::kable()
```

term	estimate	std.error	statistic	
(Intercept)	1.4105758	0.5258994	2.6822162	0
age	0.0086502	0.0016917	5.1132311	0
as.factor(vis_minority)Visible minority	-0.1857261	0.0694006	-2.6761458	0
as.factor(hours_worked)0.1 to 29.9 hours	0.4564724	0.5034842	0.9066272	0
as.factor(hours_worked)30.0 to 40.0 hours	0.4000247	0.5014473	0.7977402	0
as.factor(hours_worked)40.1 to 50.0 hours	0.5616753	0.5056691	1.1107564	0
as.factor(hours_worked)50.1 hours and more	0.6593767	0.5094448	1.2943046	0
as.factor(hh_type)Low-rise apartment (less than 5 stories)	-0.0299018	0.1167173	-0.2561903	0
as.factor(hh_type)Other	0.0242468	0.1126168	0.2153033	0
as.factor(hh_type)Single detached house	0.3419775	0.1052559	3.2490096	0
as.factor(family_income)$125,000 and more	0.0913551	0.0842856	1.0838757	0
as.factor(family_income)$25,000 to $49,999	-0.5562833	0.0904600	-6.1494956	0
as.factor(family_income)$50,000 to $74,999	-0.2531424	0.0889939	-2.8444934	0
as.factor(family_income)$75,000 to $99,999	-0.2136649	0.0906834	-2.3561624	0
as.factor(family_income)Less than $25,000	-0.7216504	0.1090438	-6.6179843	0
as.factor(self_rated_health)Fair	-1.0882304	0.1128428	-9.6437757	0
as.factor(self_rated_health)Good	-0.5570374	0.0806275	-6.9087734	0
as.factor(self_rated_health)Poor	-1.3484761	0.2022180	-6.6684267	0
as.factor(self_rated_health)Very good	-0.2757175	0.0769053	-3.5851584	0
as.factor(self_rated_mental_health)Fair	-2.7082551	0.1170944	-23.1288163	0
as.factor(self_rated_mental_health)Good	-1.5715485	0.0758266	-20.7255661	0
as.factor(self_rated_mental_health)Poor	-3.4512074	0.2701833	-12.7735787	0
as.factor(self_rated_mental_health)Very good	-0.6709117	0.0738969	-9.0790213	0
as.factor(education)College, CEGEP or other non-university certificate or di...	0.0926732	0.0659374	1.4054739	0
as.factor(education)High school diploma or a high school equivalency certificate	0.1463421	0.0682101	2.1454621	0
as.factor(education)Less than high school diploma or its equivalent	0.4918342	0.0994396	4.9460603	0
as.factor(education)Trade certificate or diploma	0.1841633	0.0956106	1.9261814	0
as.factor(education)University certificate or diploma below the bachelor's level	0.2722627	0.1374129	1.9813480	0

```
glm_step_bic <- step(satisfied.glm, k=log(n), trace=0)
glm_step_bic
```

```
## Independent Sampling design
## svydesign(id = ~1, data = data, fpc = fpc.srs)
##
## Call:  svyglm(formula = feelings_life_binary ~ age + as.factor(vis_minority) +
##     as.factor(hh_type) + as.factor(family_income) + as.factor(self_rated_health) +
##     as.factor(self_rated_mental_health), design = satisfied.design,
##     family = "binomial")
##
## Coefficients:
##                                                (Intercept)
##                                                   1.897663
##                                                        age
##                                                   0.009071
##                         as.factor(vis_minority)Visible minority
```

```
##                                                -0.231841
## as.factor(hh_type)Low-rise apartment (less than 5 stories)
##                                                 0.002945
##                                     as.factor(hh_type)Other
##                                                 0.071711
##                       as.factor(hh_type)Single detached house
##                                                 0.407965
##                    as.factor(family_income)$125,000 and more
##                                                 0.079630
##                  as.factor(family_income)$25,000 to $49,999
##                                                -0.511838
##                  as.factor(family_income)$50,000 to $74,999
##                                                -0.231350
##                  as.factor(family_income)$75,000 to $99,999
##                                                -0.206377
##                   as.factor(family_income)Less than $25,000
##                                                -0.659741
##                            as.factor(self_rated_health)Fair
##                                                -1.037924
##                            as.factor(self_rated_health)Good
##                                                -0.533986
##                            as.factor(self_rated_health)Poor
##                                                -1.313539
##                       as.factor(self_rated_health)Very good
##                                                -0.268632
##                     as.factor(self_rated_mental_health)Fair
##                                                -2.702249
##                     as.factor(self_rated_mental_health)Good
##                                                -1.573656
##                     as.factor(self_rated_mental_health)Poor
##                                                -3.411204
##                 as.factor(self_rated_mental_health)Very good
##                                                -0.682286
##
## Degrees of Freedom: 11232 Total (i.e. Null);  11214 Residual
## Null Deviance:       13290
## Residual Deviance: 11030     AIC: 11070
```

```
glm_step_bic %>%
  broom::tidy() %>%
  knitr::kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	1.8976629	0.1513209	12.5406544	0.0000000
age	0.0090709	0.0016683	5.4371338	0.0000001
as.factor(vis_minority)Visible minority	-0.2318406	0.0684268	-3.3881536	0.0007061
as.factor(hh_type)Low-rise apartment (less than 5 stories)	0.0029447	0.1151999	0.0255615	0.9796075
as.factor(hh_type)Other	0.0717112	0.1109879	0.6461170	0.5182168
as.factor(hh_type)Single detached house	0.4079653	0.1032408	3.9515893	0.0000781
as.factor(family_income)$125,000 and more	0.0796304	0.0839051	0.9490530	0.3426141
as.factor(family_income)$25,000 to $49,999	-0.5118378	0.0899014	-5.6933236	0.0000000
as.factor(family_income)$50,000 to $74,999	-0.2313499	0.0887323	-2.6072796	0.0091386
as.factor(family_income)$75,000 to $99,999	-0.2063770	0.0905459	-2.2792520	0.0226708

term	estimate	std.error	statistic	p.value
as.factor(family_income)Less than $25,000	-0.6597414	0.1074126	-6.1421235	0.0000000
as.factor(self_rated_health)Fair	-1.0379238	0.1123828	-9.2356133	0.0000000
as.factor(self_rated_health)Good	-0.5339862	0.0804200	-6.6399652	0.0000000
as.factor(self_rated_health)Poor	-1.3135389	0.2001738	-6.5619913	0.0000000
as.factor(self_rated_health)Very good	-0.2686322	0.0769387	-3.4915080	0.0004822
as.factor(self_rated_mental_health)Fair	-2.7022487	0.1174399	-23.0096325	0.0000000
as.factor(self_rated_mental_health)Good	-1.5736560	0.0757369	-20.7779413	0.0000000
as.factor(self_rated_mental_health)Poor	-3.4112038	0.2695301	-12.6561167	0.0000000
as.factor(self_rated_mental_health)Very good	-0.6822855	0.0738080	-9.2440578	0.0000000