STA 304 PS2: GSS Families Survey Analysis

Diana Azriel

October 19th, 2020

Health Perception Ratings and its Influencers in the GSS Survey 31: Families.

Diana Azriel

October 19th, 2020

Abstract

While many studies explore the health status of different populations, the mental aspect of this variable is often left out. Recent literature has shown that the way we view our own health is an important indicator of overall health, which is it is important to explore which variables influence people's mindsets. This report explores how factors such as age and income bracket influence one's perception of their health. To begin, self rated health results are divided into two categories: positive and negative. Then, a logistic regression model is run on the explanatory variables age and income bracket. The results of the model suggest that both age and income are significant indicators of one's perception of health. Specifically, the likelihood that one views their health positively decreases as age increases, and increases as one's income increases as well. Health is an important component of an individual's overall well-being and happiness in life, so it is important to examine which factors influence this perception in order to create appropriate social assistance programs to help the elderly and those with a lower income.

Introduction

When conducting extensive surveys on Canada's population, it may be difficult to capture an individual's health and well-being as that may encompass a wide variety of forms such as clinically diagnosed diseases, social function, or even an existence of a current injury. However, research has shown that perceived health is a good indicator of overall health status. Recent Stanford studies found that our perception of our own health has powerful effects on on our actual health, regardless of how accurate our perception may be. For this reason, I was interested in studying self rated health in the results of the General Social Survey on families. What factors influence the perception of our health? Does one's income bracket have an effect on the perception of one's well being, and ultimately one's true well-being?

The following sections will discuss the data and its strengths and weaknesses, as well as the model used to study the results. Finally, an exhibit and discussion of the results follows.

Data

The General Social Survey on Families was conducted during 2017 by Statistics Canada. The target population of the study was all non-institutionalized people living in Canada's 10 provinces aged 15 and older. This data was selected as it is the most recent study of Canadian families and contains many concepts of Canadian family life such as conjugal history, family origins, financial information, and mental and physical well-being.

Results were collected by telephone using Statistics Canada's Address Register. One strength of the methodology was that participants were contacted not only by landline, but also by cellular device. This

ensured that certain age demographics, like those who don't own a landline, were also represented. Another strength was that income information was obtained through tax data linkage, and thus was not subject to self-report bias. Successful linkage constituted for about 85% of participants. Each of the provinces as well as major metropolitan areas were divided into separate strata for sampling.

Some weaknesses of this data include that some questions have responses such as "NA" or "Valid Skip" which are not informative. Additionally, in some cases where the sex or age of the respondent was not inputted, values were manually selected based on an examination of the other response variables. Finally, reported dates were used to determine variables such as respondent's age at first marriage, but where a respondent was unable to provide the date, an estimate of their age at event occurring was provided. In some cases, these variables also required imputation, therefore subjecting the results to some reporting error.

Model

In the survey, self-rated-health had 7 answer options: "Don't know", "NA", "Poor", "Fair", "Good", "Very good", and "Excellent". I was interested in studying how other response factors can predict whether a person perceives their health as positive or negative First, "Don't know" and "NA" responses were removed because they make up less than 0.8% of responses and cannot be translated to positive or negative Then, "Poor" and "Fair" were dubbed negative, and "Good", "Very good", and "Excellent" were dubbed positive. Although "Fair" sounds neutral, it was dubbed as negative because many people whose life is impacted by long-term pre-existing conditions but is not completely debilitated would most likely answer "Fair" to indicate that they are functioning well, despite their health issues.

Next, I also removed all responses whose participant age was exactly 80. This is because there was clearly an input error - likely it was just that all participants 80 or older were recorded as exactly 80 years old, but since we do not know for sure, it is safest to remove them.

Then, using RStudio software, a logistic regression model was run on the explanatory variables age and family_income to model how likely a person is to view their health as negative or positive. The logistic regression model is appropriate here because the response variable is binary: {1,0}, or equivallently, {"positive", "negative"}. However, it is important to note that the conversion of the ordinal variable self_rated_health to a dichotomous one leads to a loss of some information.

The model is of the form:

$log(\frac{\hat{p}}{1-\hat{p}}) = \hat{\beta}_0 + \hat{\beta}_1 x_{age} + \hat{\beta}_2 x_{\$25{,}000-\$49{,}999} + \hat{\beta}_3 x_{\$50{,}000-\$74{,}999} + \hat{\beta}_4 x_{\$75{,}000-\$99{,}999} + \hat{\beta}_5 x_{\$100{,}000-\$124{,}999} + \hat{\beta}_6 x_{\$125{,}000\ and\ more}$

Where:

$$\hat{\beta}_0 = 2.392$$
$$\hat{\beta}_1 = -0.02669$$
$$\hat{\beta}_2 = 0.5618$$
$$\hat{\beta}_3 = 0.9628$$
$$\hat{\beta}_4 = 1.237$$
$$\hat{\beta}_5 = 1.411$$
$$\hat{\beta}_6 = 1.679$$

Below is a quick summary of the model results:

Table 1: Fitting generalized (binomial/logit) linear model: health_rating ~ age + as.factor(income_family)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.392	0.08701	27.49	3.613e-163
age	-0.02669	0.00134	-19.91	2.442e-87
as.factor(income_family)$25,000 to $49,999	0.5618	0.06206	9.052	1.535e-19
as.factor(income_family)$50,000 to $74,999	0.9628	0.06871	14.01	2.161e-44
as.factor(income_family)$75,000 to $99,999	1.237	0.07928	15.6	1.513e-54
as.factor(income_family)$100,000 to $ 124,999	1.411	0.09322	15.14	1.841e-51
as.factor(income_family)$125,000 and more	1.679	0.07756	21.65	1.009e-102

In table 1 in the rightmost column, we see a significance test (also called a p-value) for each of the betas. Since they are all significantly smaller than 0.05, we can conclude that there is a significant relationship between the explanatory variables and the outcome, namely age and family income, and health rating.

As expected β_1, which is the slope for age, is negative. It is logical that with an increase of age, the probability that a person feels positively about their health decreases, thus the log odds of that probability decrease as well. β_2 - β_6 are all positive and follow an increasing pattern as we move up the income bracket categories. This indicates that the higher your income, the larger the impact it has on the probability of having a positive outlook on one's health.

Results

Figure 1 shows respondent counts by income bracket, which is further broken down into health outlook within each bracket. Firstly, we can observe that the income bracket with the largest frequency is $125,000 and more of family income. Next, it can be observed that as we move down along the income bracket, a larger proportion of respondents consider their health negative. It is apparent that income bracket is indicative of how likely one is to perceive their health as negative or positive.

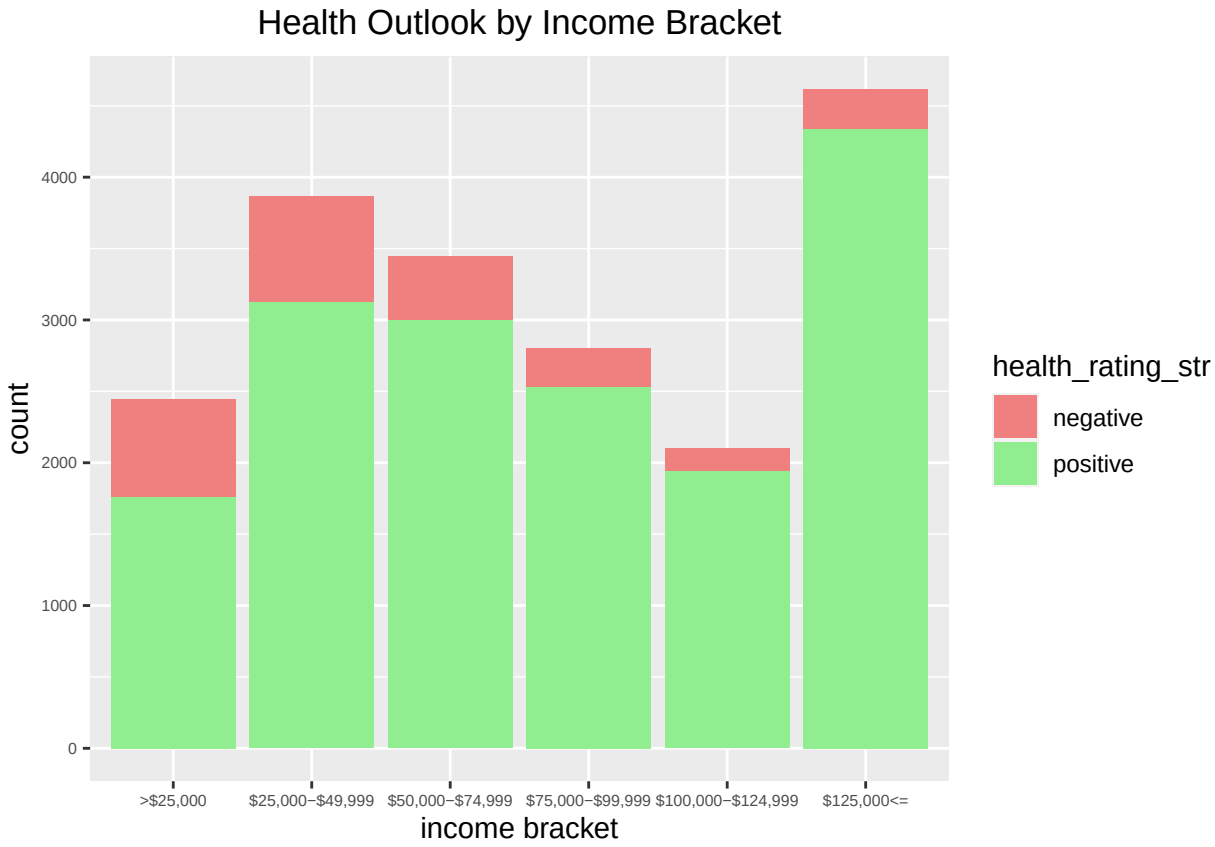

Figure 1: Health Outlook count by Income Bracket

In figure 2, we can see the density of respondents' self rated health outlook broken down by age. In smaller ages, namely 15 through around 50, most respondents felt their health was excellent. Each subsequent category was represented lower down the line. That is, the next most common response was "very good", then "good", then "fair", and finally "poor". Around age 50, all density lines meet, meaning that all responses were roughly equally common. From there on, the density lines follow an inverse pattern as to what was seen before. Now, the lowest quality of health response is most common, and so on. It is interesting to note that responses "Poor" and "Fair" follow a very similar pattern, while the other 3 responses follow their own trend. This fact further justifies the categorization of the former two as an overall negative outlook on one's health, while the latter three as positive.

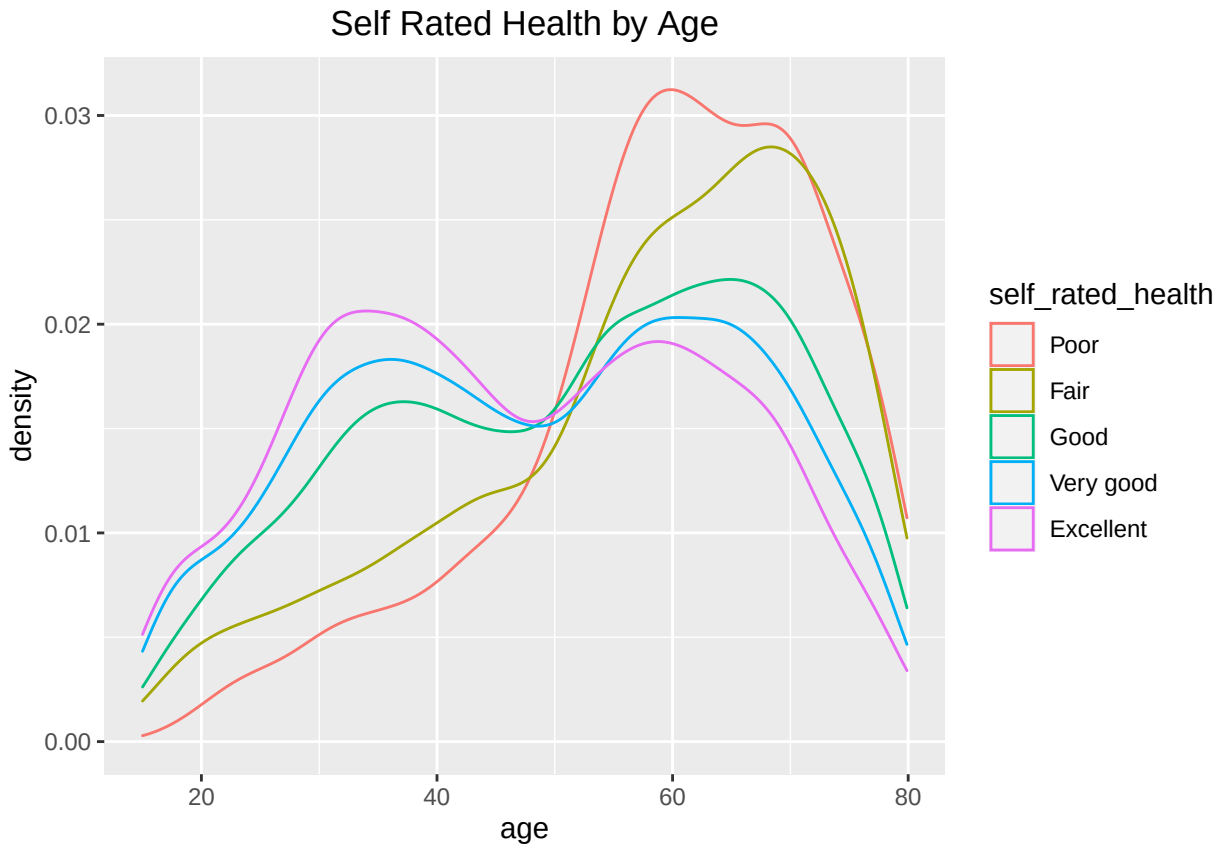

Figure 2: Self Rated Health Density by Age

Figure 3 depicts a similar breakdown of density of health rating responses by age, but here we examine the health rating as it was categorized in our model - positive or negative. It is evident that participants under 50 more commonly view their health as positive, while the opposite is true for respondents over 50.

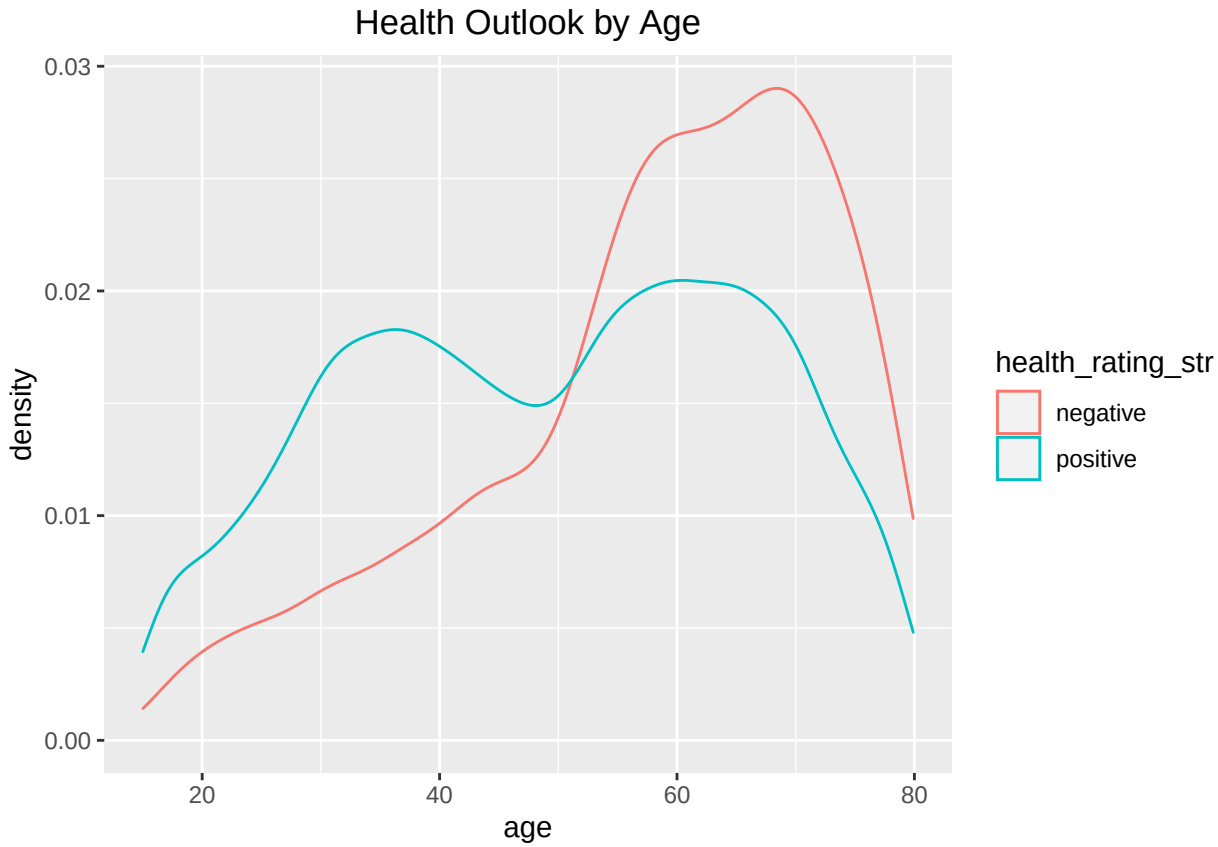

Figure 3: Health Outlook Density by Age

Discussion

Through examining the validity of a logistic regression model that used age and income bracket to determine the (log) likelihood of a person viewing their health as negative or positive, we determine that age and income bracket are significant indicators of one's perception of health. Figure 1 further showed that income bracket, alone, can determine the likelihood of a person viewing their health positively or negatively by demonstrating that a greater proportion of respondents viewed their health negatively in lower income brackets. Finally, figures 2 and 3 demonstrated that age is a predictor of health outlook.

It is not surprising that as we age, we are more likely to view our health as negative. After all, physical deterioration is a natural component of aging. However, societal changes can be implemented in order to prevent those with a lower income of feeling negatively about their health. Canada already has universal healthcare, so we must ask ourselves what further steps can be taken. It is important to note that while general healthcare and doctor visits are fully covered by provincial health plans, most drug plans are not. Therefore, individuals with lower income may not have private insurance plans to cover their prescription costs and may not be able to afford the outright cost. This means that while they may receive a diagnosis, they may be unable to afford treatment, leading them to perceive their health negatively. In addition, lower income jobs often have fewer benefits, such as paid sick days. In order to minimize the impact of income on one's perception of health, there ought to be social programs that equalize accessibility to medical treatment and provide paid sick leave.

Weaknesses

The analysis is subject to weaknesses of categorizing self rated health into two dichotomies: positive or negative. It is likely that some participants would have chosen a different answer had they known that an ambiguous response such as "fair" would be categorized as negative. It is also possible that results would differ if "fair" was categorized as positive. In addition, the removal of participants categorized as exactly 80 years old poses some limitations to studying the effect of aging on health perception. For future surveys, I recommend that age categorization above 80 remains consistent with recording of all other ages. In addition, there may be other variables that affect both income level and health perception, such as hours worked, that should be examined in order to rule out the possibility of common cause.

Next Steps

This analysis should be refreshed when GSS releases survey results on Families in Canada in 2022. In addition, other surveys such as the 2016 GSS: Canadian at Work and Home survey contain information on income, age, and health rating. It would be beneficial to run a similar analysis and compare the results.

References

Alexander, R. and Caetano, S. (2020). "gss_cleaning,.R". Retrieved from: https://q.utoronto.ca/courses/184060

Daróczi, G. (2014). pander: An R Pandoc Writer. R package version 0.5.1. Retrieved from: http://cran.r-project.org/package=pander

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

Martinovich, M. (2017) How Your Perception of Health May Extend your Life. Standford Graduate School of Business. Retrieved from: https://www.gsb.stanford.edu/insights/how-your-perception-health-may-extend-your-life

Satistics Canada. (2018). Canada at a Glance 2018. Retrieved from https://www150.statcan.gc.ca/n1/pub/12-581-x/12-581-x2018000-eng.htm

Statistics Canada. (2020). General Social Survey Cycle 31: Families. 45250001 Issue no. 2019001. pages 3-11.