An overview of Canadian retirement

Erica Zhou 1005687678

8/16/2021

Do Canadians plan for retirement?

Abstract

The Canadian Financial Capability Survey (CFCS) reveals how Canadian people understand their financial status and how they plan financially for the future. This project is based on the data 'cfcs-5159-E-2014_F1.csv', which was collected through CFCS. In this project, we look at some general financial information and focus on people's plans and expectations of retirement. We construct linear regression, frequentist, and Bayesian models to estimate parameters such as coefficients and means of important variables. We noticed that the wages or salaries are decreasing with the increase of age, and so we suggest saving for retirement earlier. Fortunately, we found that about 76% of the Canadians in the labor force have already planned for retirement financially. On the other side, we observed that most retired people feel satisfied with their life after retirement, of which the quality depends on advanced preparation, and it is mostly a good sign. Based on the information, we claimed that most Canadian people have some advanced financial plans for retirement and have a reasonable expectation of their life after retirement. On the other hand, it reflects that Canada has overall a good-functioning system of social welfare that balances the life quality of the retired people.

Introduction

The data 'cfcs-5159-E-2014_F1.csv' was collected from odesi (http://odesi2.scholarsportal.info/webview/). The size of the data is 6685, and that is exactly the number of samples that has taken the survey. We subset the data so that what is left are some variables related to our topic 'do Canadians plan for retirement'. According to Statistics Canada, between 1977 and 2013, the proportion of the overall employed Canadian population covered by Registered Pension Plans (RPPs) declined from 46% to 38%, mainly as a result of a drop in defined benefit plan coverage.[1] Plus, nowadays, there is generally a greater income gap between young employees and mid-aged or older employees, especially in industries such as social media. There is more and more demand for young labors, while the mid-aged or older group has less chance in the labor market. Therefore, it is important to save and plan for retirement earlier to avoid risks of unemployment and any changes in laws and rules. In this project, our goal is to overview the basic financial situation of Canadian people and estimate the proportion of people that have already had some plans for retirement to see if Canadian people attach importance to it. We will also explore if retired people feel satisfied with their life after retirement. Based on previous studies by Statistics Canada, we set the hypothesis to be that 78% of the labor force in the sample has financially planned for retirement ,and we will check if the population performs as the same. The goal of this project is to dig into our topic and to give people an overview of labor market and retirement with some advice.

[1].Government of Canada, Statistics Canada. "Insights on Canadian Society Financial Literacy and Retirement Planning." Financial literacy and retirement planning, April 28, 2016. https://www150.statcan.gc.ca/n1/pub/75-006-x/2016001/article/14360-eng.htm. (accessed August 20, 2021)

Data

```r
library(tidyverse)
```

```r
financial <- read.csv('/Users/ericazhou/Desktop/238/Project/cfcs-5159-E-2014/cfcs-5159-E-2014_F1.csv')
```

The data is collected from http://odesi2.scholarsportal.info/webview/. We explore the category 'Consumer Surveys', then we select the dataset 'Canadian Financial Capability Survey, 2014' and download directly to a local file. Among all the variables, we choose some of them that are related to our topic or can provide us with relevant information. The dataset 'financial' here is the one after sub-setting.

Important variables:

G2AGE: It shows the age group of all respondents. There are 6 years for each level. It starts at level 1, which represents 18-24 years old and ends at level 8, which represents 70 or more years old.

OCC6CURR: It shows the occupation group of respondents that are not retired. The occupation groups include 'Management Occupations', 'Business, Finance and Administrative Occupations', 'Sciences and Related Occupations, Health Occupations', 'Social Sci, Edu, Gvt, Religion, Art, Culture, Rec, Sport', 'Sales and Service Occupations', 'Trades, Transport, Primary Industry, Manufacturing, Utilities', and 'Not stated'. Plus, 'Valid skip' means that the respondent has retired and thus skipped this question.

G_ASSETS: It represents the total assets owned by the respondents and/or their families. It includes group 'Less than 100k', 100k to under 200k', 200k to under 300k','300k to under 500k', and '500k or more'.

PINCQUIN: It shows the estimated total income before taxes and deductions of people that have income in 2013. The categories include 'Less than 13,001', '13,001 - 24,999', '25,000 - 39,999', '40,000 - 62,999', and '63,000 and over' corresponding to level 1 - 6.

RP_01: It represents if the respondents have financially planned for retirement.

RP_10: It represents how the retired respondents feel with their life quality after retirement comparing to their expectation. The categories include 'Much better than exp', 'Better than expected', 'As expected', 'Not as good as expct',' Much worse than exp' corresponding to level 1 - 6.

The dataset is mostly clean. What we need is to specify each occupation group with English instead of numbers.

```r
assets <- financial[financial$G_ASSETS != 9,]
summary(assets$G_ASSETS)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   3.000   3.091   5.000   5.000
```

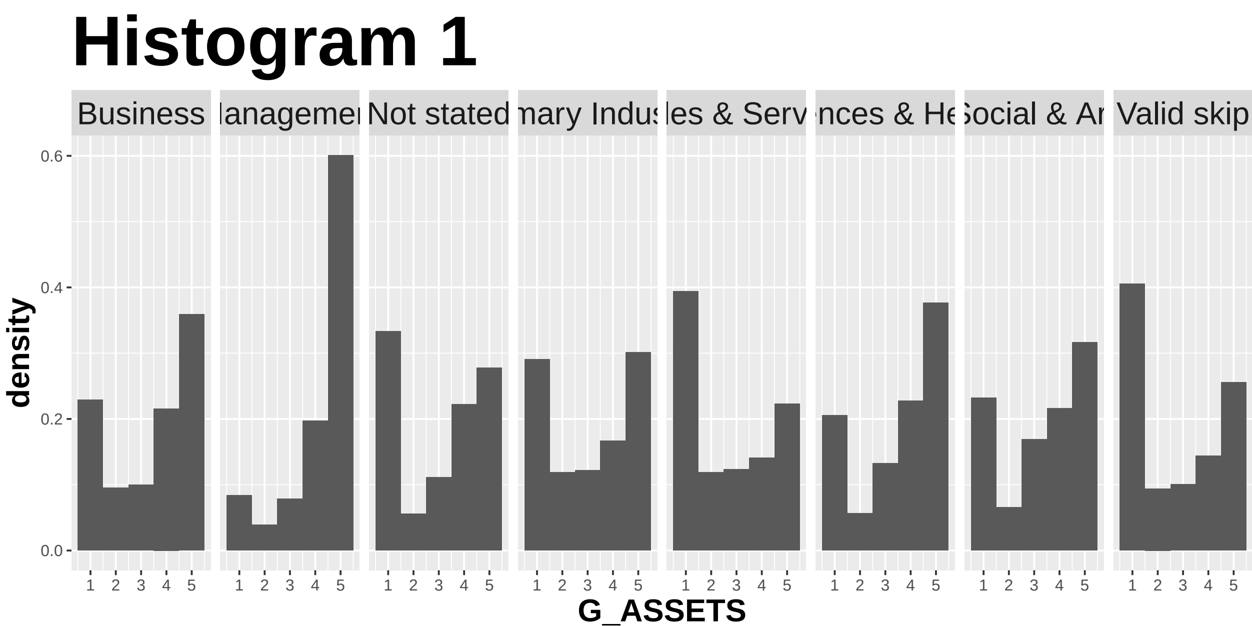
We summarized the basic numerical information of the total assets owned by the respondents and/or their families. When we look at the mean, which is 3.091, and it seems reasonable that the average total assets are around 200k to 300k. However, if we look at the 25% percentile, it turns out that 25% of the respondents have total assets less than 100k. Contrastively, the 75% percentile is 5. It means that at least 25% of the respondents own 500k or more assets. There is quite a gap between people's assets.

```r
summary(financial$PINCQUIN)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   3.124   4.000   5.000
```

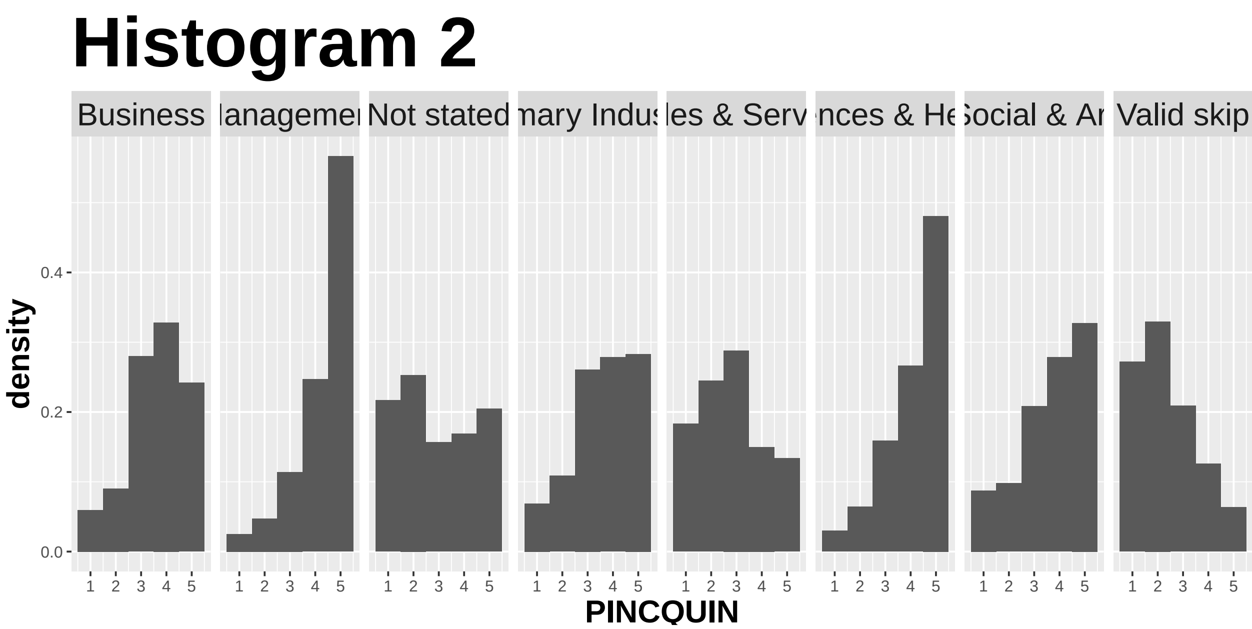
We also summarized the numerical information of the estimated total income made by the respondents in 2013. The mean is 3.124, which means that the average total annual income is about 25,000 - 39,999. Most people have an annual income of about 13,001 - 62,999, and few are out of these boundaries.

```r
assets %>% ggplot(aes(x = G_ASSETS, y = ..density..)) +
  geom_histogram(binwidth = 1) +
  facet_grid(cols = vars(OCC6CURR), scales = 'free_y') +
  ggtitle('Histogram 1') +
  theme(strip.text.x = element_text(size = 18), axis.title=element_text(size=18,face="bold"),
        plot.title = element_text(size = 40, face = "bold"))
```



On histogram 1, we see how people's total assets differ by their occupations. We notice that respondents that work in management mostly have a lot of assets. For other occupation groups, most people have little or a lot of assets. Surprisingly, the heights do not concentrate in the middle of the graphs. It reflects the gap between people's assets.

```r
financial %>% ggplot(aes(x = PINCQUIN, y = ..density..)) +
  geom_histogram(binwidth = 1) +
  facet_grid(cols = vars(OCC6CURR), scales = 'free_y') +
  ggtitle('Histogram 2') +
  theme(strip.text.x = element_text(size = 18), axis.title=element_text(size=18,face="bold"),
        plot.title = element_text(size = 40, face = "bold"))
```



On histogram 2, we see how people's annual income differs by their occupations. We notice that people that work in management and sciences & health mostly have a high income. For the category 'Valid skip', there are mostly retired people but still have sources of income. Thus, it is reasonable that only a few people in this category have high incomes. Most people in this category are covered by pensions such as Canada Pension Plan (CPP) and Old Age Security (OAS) pension.(https://www.canada.ca/en/financial-consumer-agency/services/retirement-planning/sources-retirement-income.html.).

Reference

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Method & Result

Linear regression

$$Y_i = \beta + \alpha x_i + \epsilon_i \epsilon_i \sim N(0, \sigma^2)$$

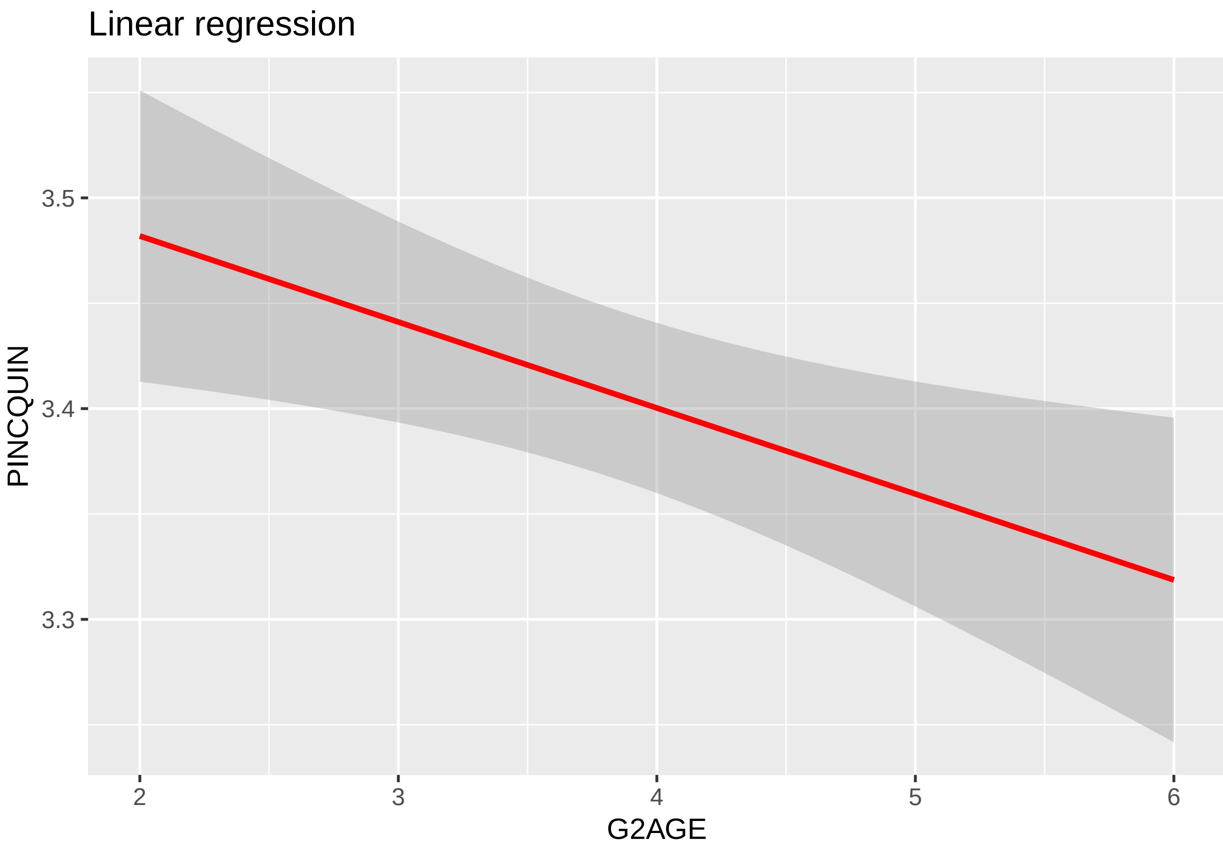
In the linear regression model, x_i is the age group of the respondents, and Y_i is their estimated total income in 2013. We want to estimate α, which is the slope of the linear function. In other words, we are interested in at what rate the income is increasing/decreasing with respect to the age.

```
labor_force <- financial[!financial$G2AGE %in% c(1,7,8),]
linear_model <- lm(PINCQUIN~G2AGE, data = labor_force)
linear_model
```

```
##
## Call:
## lm(formula = PINCQUIN ~ G2AGE, data = labor_force)
##
## Coefficients:
## (Intercept)        G2AGE
##      3.5635      -0.0408
```

The coefficient α is -0.0408. It represents that when the age increases, the income decreases.

```r
labor_force %>% ggplot(aes(x = G2AGE, y = PINCQUIN)) +
  stat_smooth(formula = y ~ x, method = 'lm', col = 'red') +
  ggtitle('Linear regression')
```



Since the income decreases as the age increases, there is less chance for mid-aged or old employees to earn as much as when they were young. While there old people still need consumption. Some elderly people may also want to leave some assets to the next generation. Therefore, it is essential to save money earlier and to understand the concept of retirement.

Frequentist model

We want to figure out how many of our respondents have financially prepared for retirement. Here, we only consider ones who have answered the question with 'yes' or 'no'. We let the random variable Y_i be that the ith respondent has financially planned for retirement. Since it is a 'yes' or 'no' question, Y_i follows a Bernoulli distribution, with parameter p_i, which represents the probability of answering 'yes' by each respondent. $0 \leq p_i \leq 1$.

$$Y_i \sim Bern(\ p_i)$$

MLE

```r
# only consider people who have or have not plan for retirement financially
n <- sum(labor_force$RP_01 == 1) + sum(labor_force$RP_01 == 2)
# the probability that one has financially palnned for retirement based on data (only used for sampling
x_bar <- sum(labor_force$RP_01 == 1)/n

cat('The MLE of parameter p is',  x_bar)
```

```
## The MLE of parameter p is 0.7640999
```

The MLE is the estimator that is most likely to output the data we see. The sample mean is an unbiased and consistent estimator of the parameter p_i of the Bernoulli distribution, and is the MLE of p_i. It turns out that 76.4% of the Canadian labor force have advanced financial plans for retirement. Comparing to the 81% [2] in 2009, it has decreased 4.6%. It turns out that people's quality of life improves, consumption improves. Few people are worried about retirement.

[2].Statistics Canada, Government of Canada. "Insights on Canadian Society Financial Literacy and Retirement Planning." Financial literacy and retirement planning, April 28, 2016. https://www150.statcan.gc.ca/n1/pub/75-006-x/2016001/article/14360-eng.htm. (accessed August 20, 2021)

95% CI

The value of p_0 is assumed by the existing studies on Statistics Canada mentioned in the previous section.

```r
set.seed(238)

# same sampling as the MLE section (just getting the sample)
plan_sample <- rbinom(n, 1, x_bar)
# sample mean <- our estimator
x_bar1 <- mean(plan_sample)

p_0 <- 0.78
s_2 <- p_0 * (1 - p_0)

cl <- x_bar1 - qnorm(0.975) * sqrt(s_2/n)
cu <- x_bar1 + qnorm(0.975) * sqrt(s_2/n)

cat('The 95% CI is', '[', cl, ',', cu, '].')
```

```
## The 95% CI is [ 0.7487528 , 0.7748115 ].
```

Based on the calculation, we are 95% confident that the true mean of the population is between 0.749 and 0.775. We are pretty confident that as low as 74.9% and as high as 77.5% of the labor force have financially planned for retirement in Canada in 2014.

Hypothesis test

Our null hypothesis is that p = 0.78, where p is the probability that a Canadian has financially planned for retirement.

H_0: p = 0.78

Our alternative hypothesis is that p $\neq$ 0.78.

H_A: p $\neq$ 0.78

```r
#sample proportion is x_bar1, hypothesized value is p_0, sample variance is s_2,and n is found
#test statistic:
Z <- (x_bar1 - p_0)/sqrt(s_2/n)
# 0.05 significance level
alpha <- 0.05
#prop.test(n*x_bar, n, p=0.5, correct=FALSE)
p_value <- 2 * pnorm(Z)
cat('The p-value is', p_value)
```

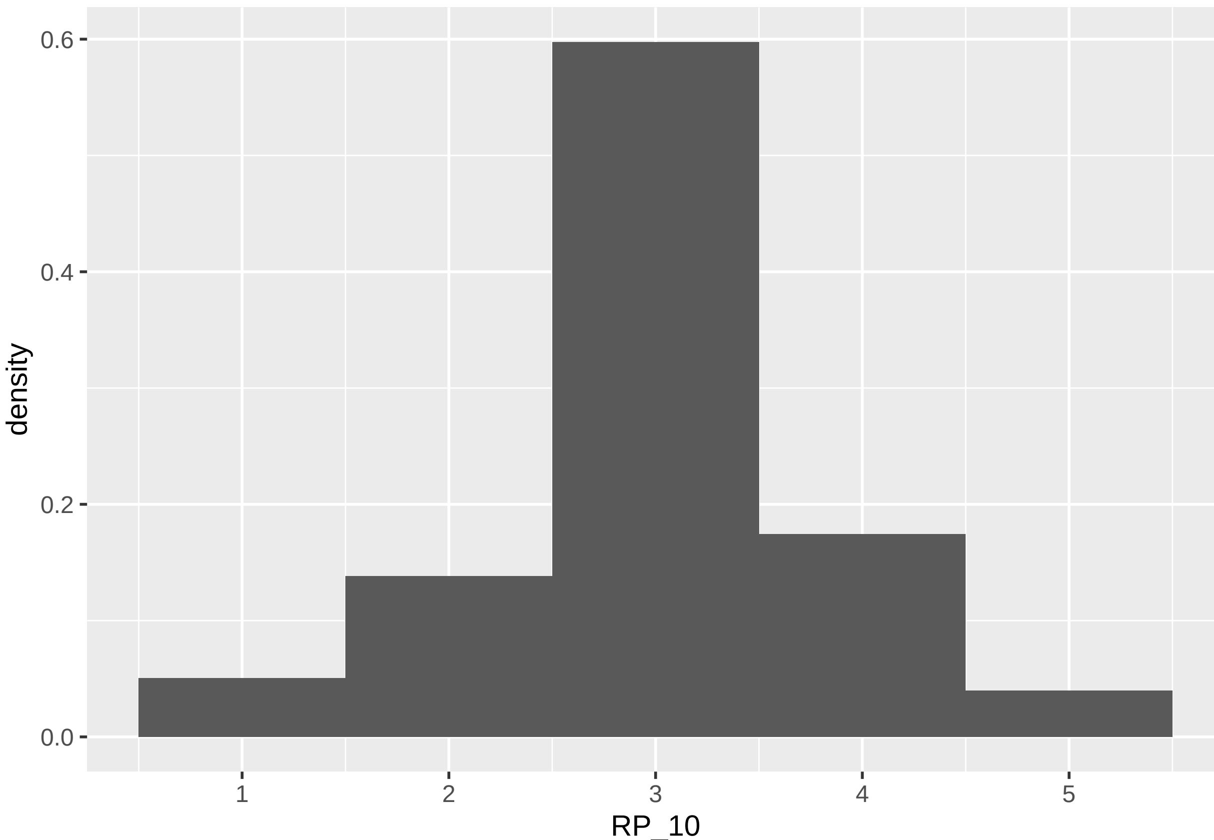
```
## The p-value is 0.006135413
```

The p-value is small, thus, we conclude that there is strong evidence against H_0. In previous sections, we know the sample mean is around 76.4%. If H_0 is true, then there is only a 0.6% chance that we will observe the 1.6% deviation from the average 78%. Therefore, there is strong evidence that the actual proportion is different from 78%.

Bayesian model

```
retired <- financial[!financial$RP_10 %in% c(6,7,8,9),]

retired %>% ggplot(aes(x = RP_10, y = ..density..)) + geom_histogram(binwidth = 1)
```



Now we turn to the retired group of people. We want to see how the life quality has met the expectation or not for retired people. We noticed that 60% of the retired people think their life is just as expected. The histogram looks like a normal distribution. Therefore, we assume that our data follows a normal distribution. Since we don't have prior studies or knowledge of μ, we assume that μ follows a standard normal distribution.

Because we know neither μ nor σ, we assume σ to be 1. Then the Bayesian estimator will also follow a normal distribution.

Bayesian estimation

$$X_1, ..., X_n \sim N(\mu, \sigma_0^2)\mu \sim N(0,1)\hat{\mu}_{\text{bayes}} \sim (\frac{\mu_0/\tau_0^2 + n\bar{x}/\sigma_0^2}{1/\tau_0^2 + n/\sigma_0^2}, \frac{1}{1/\tau_0^2 + n/\sigma_0^2})$$

```
living_sd <- mean(retired$RP_10)
n_retired <- length(retired)
```

```
mu_0 <- 0
tau_sq_0 <- 1
sigma_sq_0 <- 1
# plug in the expression
numerator <- mu_0/tau_sq_0 + n*living_sd / sigma_sq_0
denominator <- 1/tau_sq_0 + n/sigma_sq_0
est_b <- numerator / denominator
cat('The Bayesian estimate of the mean is', est_b)
```

```
## The Bayesian estimate of the mean is 3.013997
```

Under our assumptions, the estimate seems to be reasonable according to the graph, while some more prior knowledge is required for a more accurate estimation. According to the graph and estimate, we confirmed that the life quality has mostly met their expectation of life after retirement on average. Yet, since the estimate is greater than 3, it reflects that more people tend to answer 4 or 5 rather than 1 or 2. It means that more people think their life doesn't at least meet the expectation. Therefore, for these people, more detailed plans before retirement should have been helpful.

The posterior distribution is

$$\mu|X_1, ..., X_n, \sigma_0^2 \sim N(\ 3.014,\ 0.0026)$$

By plugging in the values, we got the posterior distribution of μ. It has an expectation of about 3.014 and variance of about 0.0026 under our simple assumption.

95% Credible interval

```
living_sd <- mean(retired$RP_10 )
n_retired <- length(retired)

cl_bayes <- qnorm(0.025, 3.014, sqrt(0.0026))
cu_bayes <- qnorm(0.975, 3.014, sqrt(0.0026))

cat('The 95% credible interval for mu is [' ,cl_bayes, ',', cu_bayes, ']')
```

```
## The 95% credible interval for mu is [ 2.914061 , 3.113939 ]
```

The 95% credible interval tells us that the true μ can be contained in this interval with a 95% probability. In other words, there is a 95% chance for the true value of μ to be any value between 2.914 and 3.114. In practice, there may be more people to answer 'better' or more people to answer 'worse' if we cover every Canadian in the survey, but most people are currently satisfied with their life after retirement.

Conclusion

In this project, we have used a linear regression model to estimate the relationship between age and income. We have applied MLE, confidence interval and hypothesis test to estimate the probability that a person in the labor force has financially planned for retirement based on a frequentist model. We have also used the Bayesian estimation and credible interval to estimate the people's average degree of satisfaction with their life after retirement. According to our research, in the Canadian labor force in 2014, the income decreases as the age grows. 76.4% of the labor force have financially planned for retirement. It turns out that most people in the labor force plan for retirement in advance but the proportion is lower than the one in 2009. The economic development and diversification of consumption have improved people's consumption, and thus fewer young people are worried about retirement. However, it is still necessary to understand the financial situation and concepts of retirement to avoid any problem when changes such as the decreasing of coverage of employment insurance are made to social welfare. With full preparation, people can make a more appropriate expectation of the rest of life. In 2014, about 60% of the retired people have met their expectation after retirement, but 20% of people feel upset. There may be different reasons like the cut in pension or lack of preparations.

Because the survey didn't cover the question if these retired people had financial plans before retirement, we didn't explore the exact relationship between the degree of planning and the life quality after retirement. In future studies, it will be an important concept if we want to dig deeper in this topic. While according to Wei-Ting Yen, the professor at The Ohio State University, Department of Political Science, people have higher and higher retirement insecurity mostly because of pension investment risk. She also claimed that the risk has reduced people's expectation of post-retirement income, which has caused a pessimistic perceptions about retirement. [3] Thus, we believe that a better plan would reduce part of this anxiety and increase the quality of retirement. We advice people to always have some plans for retirement.

[3]Wei-Ting, Yen. "Pension Plans and Retirement Insecurity." Ageing International 43, no. 4 (2018): 438–63. https://doi.org/10.1007/s12126-018-9326-x. (accessed August 20, 2021)

Bibliography

Canada, Financial Consumer Agency of. "Sources of Retirement Income." Canada.ca. / Government du Canada, May 11, 2021. https://www.canada.ca/en/financial-consumer-agency/services/retirement-planning/sources-retirement-income.html. (accessed August 20, 2021)

Government of Canada, Statistics Canada. "Insights on Canadian Society Financial Literacy and Retirement Planning." Financial literacy and retirement planning, April 28, 2016. https://www150.statcan.gc.ca/n1/pub/75-006-x/2016001/article/14360-eng.htm. (accessed August 20, 2021)

Yen,Wei-Ting. "Pension Plans and Retirement Insecurity." Ageing International 43, no. 4 (2018): 438–63. https://doi.org/10.1007/s12126-018-9326-x. (accessed August 20, 2021)

Appendix

$$X_1, X_2, ..., X_n \sim Bern(\theta) \Rightarrow P_\theta(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}$$

$$Likelihood\,function:$$

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n|\theta)$$

$$= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{x_1+x_2+...+x_n}(1-\theta)^{n-x_1-x_2-...-x_n}$$

$$= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$Log-\ likelihood\,function:$$

$$l(\theta) = logL(\theta) = log[\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}]$$

$$= (\sum x_i)log\theta + (n - \sum x_i)log(1-\theta)$$

$$Take\,the\,first\,derivative:$$

$$Let\, l'(\theta) = \frac{dl}{d\theta} = \frac{\sum x_i}{\theta} - \frac{n - \sum(x_i)}{1-\theta} = 0$$

$$\Rightarrow \frac{\sum x_i}{\theta} = \frac{n - \sum x_i}{1-\theta} = n\theta - \theta \sum x_i$$

$$\Rightarrow \hat{\theta}_{ML} = \frac{\sum x_i}{n}$$

$$2^{nd}\, derivative\, test:$$

$$l''(\theta) = \frac{-\sum x_i}{\theta^2} - \frac{n - \sum x_i}{(1-\theta)^2} < 0 \Rightarrow is\, a\, maximum$$

$Thus,\ the\ MLE\ of\ the\ parameter\ of\ the\ Bernoulli\ distribution\ is\ the\ sample\ mean.$