Analysis of Canadian Households' Income and Expenditure in 2017

Tong Su

8/19/2021

Abstract

This is the data analysis of Canadian Households' with their income and expenditure in 2017. We are going to investigate whether the household expenditure is reasonable with their income. The data come from the Statistics Canada survey which is conducted every two years. Some important variables include total income, total expenditure, annual expenditure on consumption, shelter, education, and health. We have used simple linear regression to estimate the relationship between total income and expenditure. We have also used bootstrap confidence interval to estimate the total expenditure on consumption, maximum likelihood estimation to approximate the distribution for education expenditure, and Bayesian credible interval to approximate the distribution for healthcare expenditure. Lastly, we have used hypothesis testing to verify the average expenditure in the shelter provided in statista. Based on our results, we get an appropriate linear regression model with a slope close to 1. It is not reasonable as people should not spend all that they earned. The government may choose to provide more subsidies to the household to reduce their expenditure. We reject our null hypothesis with a p-value close to 0. This result tells us that the average expenditure in the shelter provided in statista is not accurate.

Introduction

The data set we are going to use is the survey of Canadian household spending in 2017. The main problem we are investigating is whether the household expenditure is reasonable with their income in 2017. It is important as unreasonably high expenditure will lead to social instability and even social unrest.

From the global perspective, our analysis could be used to compare household expenditure with other countries. This will help the Canadian government understand household expenditure so that it could adjust the government policies accordingly. For example, if the expenditure on one category is higher than expected and occupies a high proportion of income, the government may want to put more subsidies in that area to reduce the burden of the household.

Of course, this analysis is also interesting for the general reader, especially for those people who consider investing but are not sure which area they should choose. We believe they will find the answer through our distribution and mean expenditure approximation. The businessmen could invest in a market with a high average expenditure.

We are going to approach this problem by analyzing the relationship between household total income and expenditure. We will use a simple linear regression model which is a model to predict the linear relation between two variables. For the other single variables such as annual expenditure on households' consumption, education, and health, different approaches including bootstrap confidence interval, maximum likelihood estimation, and Bayesian confidence interval will be used to determine the distribution. For the hypothesis testing, we will verify whether our sample data about the expenditure in shelter supports the official data of the average annual household expenditure in the shelter provided in statista. This analysis could also be used to correct the information provided by the current data.

Data

Data collection process

This data comes from the Statistics Canada survey accessed through ODESI (http://odesi2.scholarsportal.info/webview/). It is a comprehensive overview of Canadian households' income and expenditure. Since 2015, the survey is conducted every two years. The data are directly obtained from the household through two collection modes: a personal interview using a questionnaire on a laptop, and a daily expenditures diary over two weeks. This specific dataset is the survey conducted in 2017.

Data cleaning process

After downloading the data from ODESI, the data is cleaned by making the names of the data easier to use. Next, we select the variables that we are going to use for further analysis. Since some names of the columns are hard to understand, we choose to rename some of them for easier understanding. For the multiple-choice questions, the result is indicated as the discrete variables where the number is used to represent different choices. We choose to change them to categorical variables so that they could be used in graphical summaries. Finally, we want to get rid of duplicate rows to prevent repetitive responses. Since the number of observations did not change, we believe all responses are unique in this survey.

Data Description

Important variables

Before looking at the numerical and graphical summaries of different variables, it is important to understand all the variables in the survey first.

Table 1: Description of the important variables

Variables	Description
rp_age_grp	Age group of the reference person at the time of the interview
rp_sex	Sex of the reference person
rp_educ	Highest level of education attained by the reference person at the time of the interview
rp_tot_inc	Total income before taxes of the reference person
hh_tot_inc	Household total income before taxes
hh_tot_ex	Household total expenditure at the time of the interview
ex_educ	Expenditure spent on education
ex_consum	Expenditure spent on consumption
ex_health	Expenditure spent on health
ex_shelter	Expenditure spent on shelter

Numerical summaries

With all the variable introduced, we can then generate numerical summaries to help us better understand the data set. We will approach them through location and spread of different numerical variables.

Table 2: Numerical summaries of the location of the data

Variables	Min	Max	Mean	10% trimmed mean	First quartile	Median	Third quantile
Total income of the reference person	-9500	464500	50878	45242.7	22300	41150	67750
Household total income	200	571250	86699	78042	39100	71662	117450
Household total expenditure	720	633239	67150	59681	30272	54671	89225
Expenditure spent on education	0	104000	1337	289	0	0	376
Expenditure spent on consumption	720	459648	45160	40249	21970	36950	58650
Expenditure spent on health	0	43005	2248	1790	545	1505	3055
Expenditure spent on shelter	0	155598	17368	15276	8354	13686	5880

From the location of the data, we could observe that the minimum total income of the reference person goes down to negative. It means that the reference person does not earn money but instead losing it. Fortunately, the minimum household total income is 200. This indicates that every household has some income which is a good sign. For the expenditure session, we could observe that for the expenditure in education, both the first quantile and median are 0. This is not reasonable as the education cost could not be zero. We believe the reason is that over half of the household does not spend money on education. This situation makes us alert that if we want to conduct further analysis on the expenditure of a certain category, we may want to filter out those households that do not spend any money on that specific section.

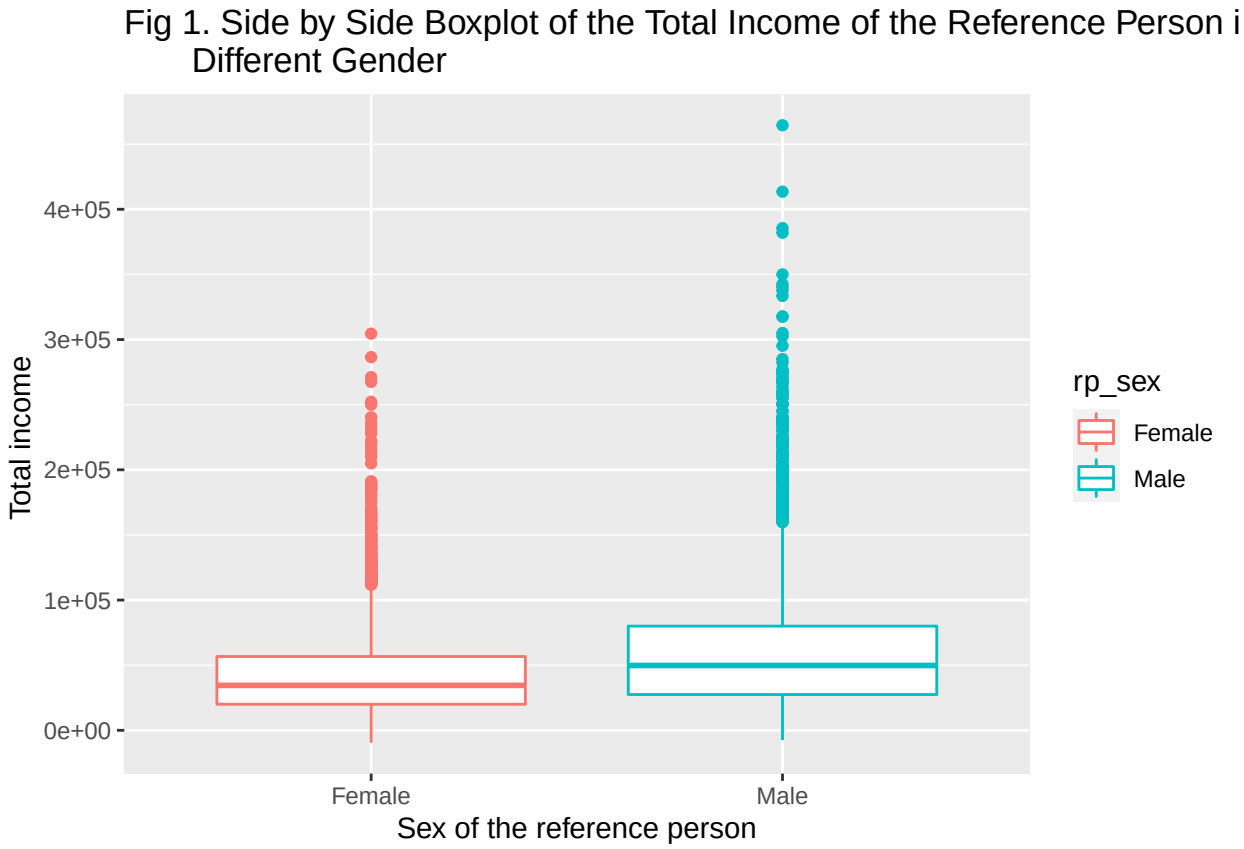
Table 3: Numerical summaries of the spread of the data

Variables	Range	Interquantile range	Median Absolute Deviation	Standard deviation
Total income of the reference person	474000	45450	31358	40614
Household total income	571050	78350	54689	64049
Household total expenditure	632519	58953	41132	52107
Expenditure spent on education	104000	376	0	4612
Expenditure spent on consumption	458928	36680	25529	33048
Expenditure spent on health	43005	2510	1668	2649
Expenditure spent on shelter	155598	14158	9406	13366

From the spread of the data, we could observe that household total income has the largest interquartile range, median absolute deviation, and standard deviation. This means that the household total income has the largest spread with a lot of extreme values. The median absolute deviation for expenditure spent on education is 0. This means that the median of all absolute deviations in the data is 0. This may be due to same the fact that over half of the household does not spend money on education.

Graphical summaries

Apart from numerical summaries, it is also important to see graphical summaries to see trends of the numerical data with different categories.



This is the side-by-side boxplot of the total income of the reference person of different genders. From the graph, we could observe that for both sexes, the graphs are right-skewed. There are many outliers with high total income. However, the median total income of the male reference person is higher than that of the female reference person. The male reference person also has a larger interquartile range and range than the female reference. We could infer from this result that male reference people generally have a higher income than female reference people.

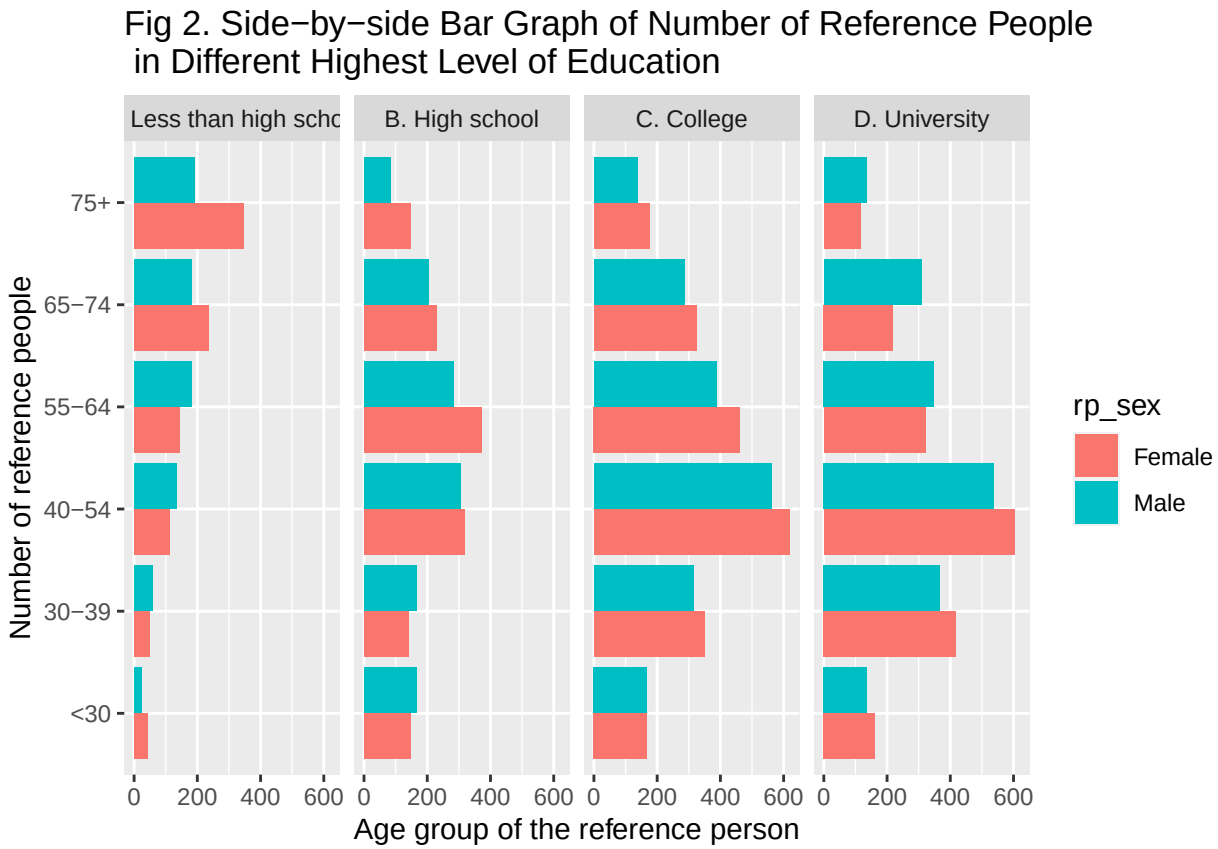


Fig 2 is a side-by-side bar graph of the number of reference people with different highest levels of education. From the graph, we could observe that most reference people have obtained a college or university diploma regardless of gender. In these two diplomas, we could further observe that more reference people are between 40 to 54 years old. For the reference people who are over 75 years old, most of them only obtained less than high school education. This is reasonable as younger people are more concerned about the education level in recent years compared with older people.

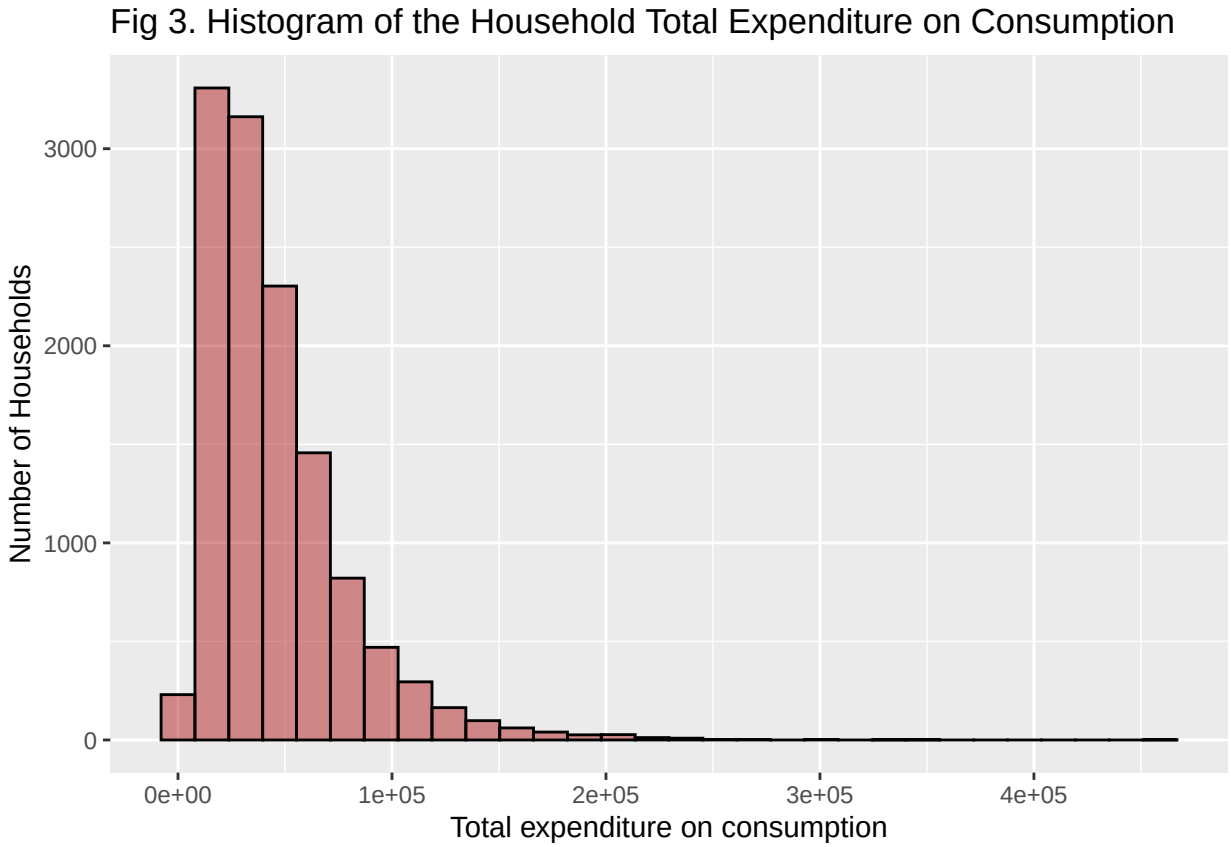


Fig 3 is a histogram of the household's total expenditure on consumption. From the graph, we could observe that this graph is right-skewed. This indicates that the mean value of the total expenditure on consumption is greater than its median. This result is also shown in our numerical summaries. The graph is unimodal with only one mode around 30000. The range of expenditure on consumption, from 0 to over 400000, is very large. Since there are some extreme values over 200000, the mean value of the expenditure on consumption is greatly impacted since it is not robust.

The total expenditure on consumption is the variable that we are going to perform the bootstrap on. Since the graph is right-skewed, it does not follow the normal distribution. In this case, it is hard for us to determine the true population mean from the sample. Therefore, a bootstrap method is suitable for us to approximate the mean value of total expenditure on consumption to normal distribution. To further improve its accuracy, a confidence interval is also used to determine a range of the mean total expenditure on consumption.

Methods

Income and Expenditure —- Simple Linear Regression

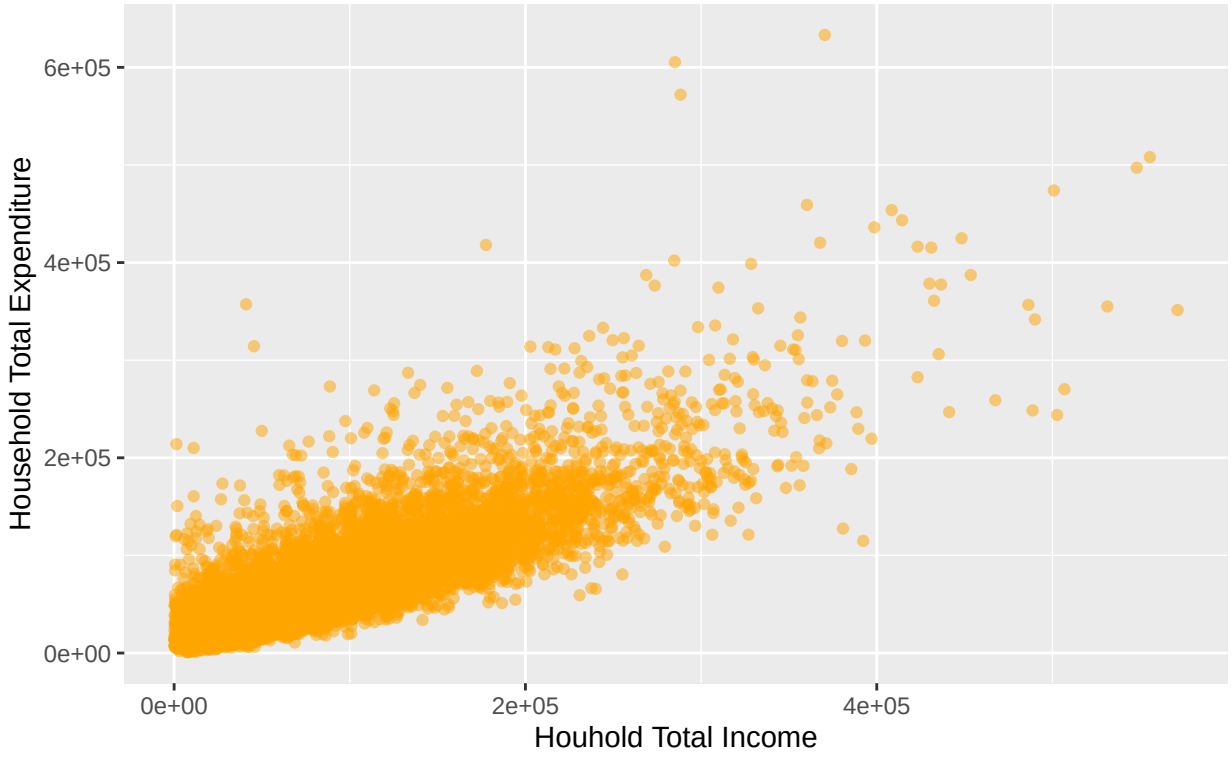
To estimate the relationship between household total income and household total expenditure, we are going to use a simple linear regression model. This is because both of the variables are numerical and continuous. We set the household total income as our independent variable and the household total expenditure as the dependent variable.

To use linear regression for these two variables, we have three assumptions in total.

1. There is a linear relationship between the household total income and the household total expenditure.
2. The error term for each pair of observations are independent of one another
3. Every error term has equal spread or constant variance.

We believe this model is appropriate for these two variables. Based on our common sense, an increase in income will trigger an increase in expenditure. This is also the reason why we set the household total income as the independent variable but not vice versa. Furthermore, from Fig 4.1, we could see that there is a linear relationship between household total income and expenditure. In this case, the first assumption of the linear relationship is satisfied.

Fig 4.1 The Scatter Plot of the Total Expenditure against the Houhold Total Income



We will use a linear regression model to predict the result.

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \epsilon_i \sim N(0, \sigma^2)$$

The linear regression model will predict the linear relationship between the independent variable x_i and the dependent variable Y_i. This means that the relationship between these two variables could be represented by a straight line. A one=unit increase in x_i is going to cause a corresponding change of the dependent variable Y_i. This change is also indicated by β_1 which is the slope of the straight line.

In this particular model,

$i = 1, ..., n$ where n is the number of households in this dataset

x_i is the i^{th} selected household's total income

Y_i is the i^{th} selected household's total expenditure

β_1 is the slope of the model. It represents the average change in the household's total expenditure for a 1-unit change in the household's total income

β_0 is the intercept of the model. It represents the average change in the household's total expenditure when the household's total income is 0

ϵ_i is the i^{th} error term

We are interested in estimating β_1.

Total Expenditure on Consumption —- Bootstrap Confidence Interval

In order to estimate the true mean value of the total expenditure on consumption, we are going to use a bootstrap confidence interval. From Fig 3. in the "Data" section, the sample total expenditure on consumption does not follow a normal distribution. We also do not know the true variance of the total expenditure on consumption. In this case, we will use the bootstrap method to get an estimation of the T distribution by using the bootstrapped samples to compute many studentized means:

$$t^* = \frac{\overline{x}_n^* - \overline{x}_n}{s_n^*/\sqrt{n}}$$

In this particular model:

n is the number of observations in this sample dataset

$\overline{x}_n$ is the sample mean of the total expenditure on consumption

s_n^* is the bootstrapped sample standard deviations of the total expenditure on consumption

$\overline{x}_n^*$ is the bootstrapped sample mean of the total expenditure on consumption

After generating many bootstrap dataset with studentized means, we will estimate c_l^* and c_u^* using the $\alpha/2$ and $(1-\alpha)/2$ percentiles of the bootstrapped studentized means. c_l and c_u are the critical values for the confidence interval and α is the level of significance.

In this particular model:

α is set to 5%

c_l^* is the 2.5th percentile of the studentized means of the total expenditure on consumption

c_u^* is the 97.5th percentile of the studentized means of the total expenditure on consumption

The confidence interval is then

$$\left(\overline{x}_n - c_u^* \cdot \frac{s_n}{\sqrt{n}}, \overline{x}_n - c_l^* \cdot \frac{s_n}{\sqrt{n}}\right)$$

arised from

$$P\left(c_l^* < \frac{\overline{X}_n^* - \mu^*}{S_n^*/\sqrt{n}} < c_u^*\right) \approx 1 - \alpha$$

where s_n is the sample standard deviations of the total expenditure on consumption

Shelter Expenditure —- Hypothesis Test

In order to verify the true mean value of the total expenditure on shelter, we are going to use hypothesis testing for this variable. From the average annual household expenditure in Canada in 2017 in statista, we have obtained the average expenditure in shelter is 18637. We want to test whether this is true.

Here are our hypotheses:

$$H_0 : \mu = 18637$$
$$H_1 : \mu \neq 18637$$

where μ is the the average expenditure in shelter.

Since the true variance remains unknown, in this case we are going to use t-test

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim T_{n-1}$$

In this particular model:

n is the number of observations in this sample data set

$\overline{x}$ is the sample mean of the total expenditure on shelter

s_n is the sample standard deviations of the total expenditure on shelter

We need to determine whether to reject the null hypothesis or not based on the p-value calculation

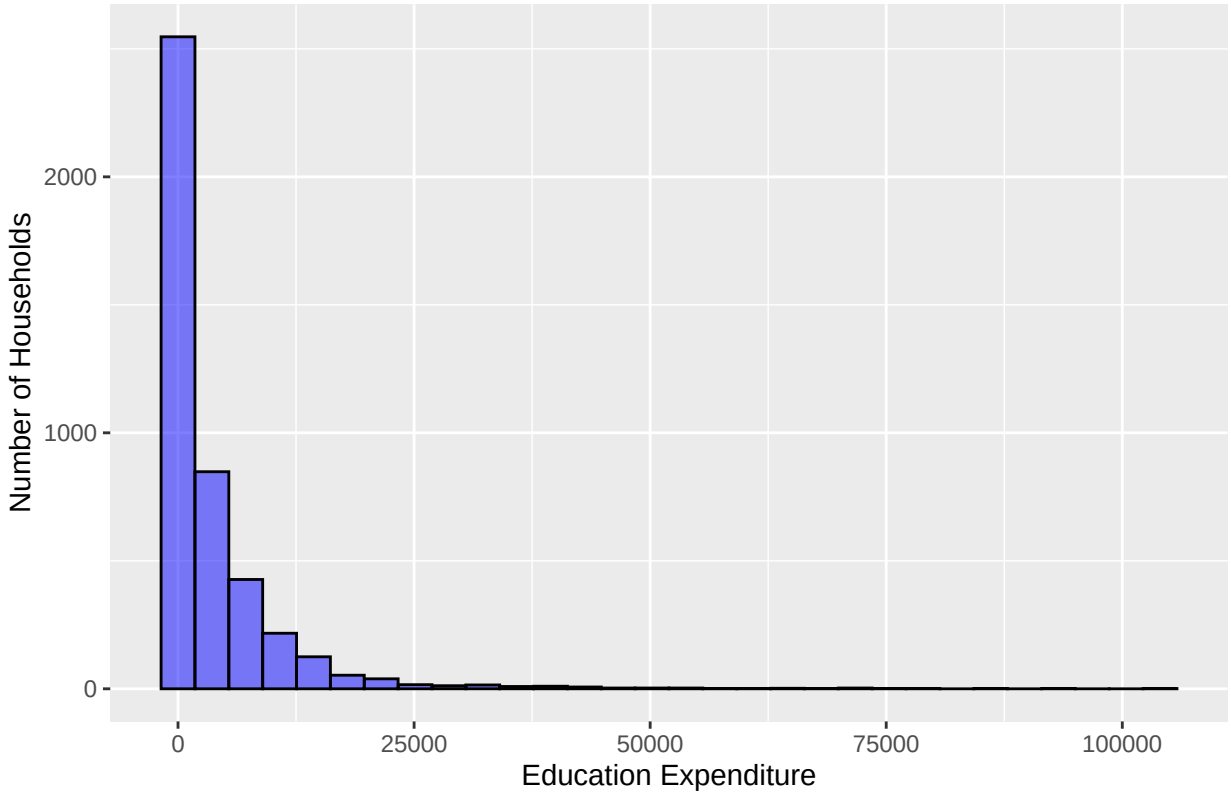
$$P(|T| \geq |t|)$$

Education Expenditure —- Maximum Likelihood Estimation

The variable we are going to analyze is the education expenditure in the household. We are going to use exponential distribution and fit this model using a frequentist approach.

This model is appropriate for this variable as the expenditure in education is a numerical continuous variable. Furthermore, from the graph shown below (Fig 5.1), we could observe that the graph follows a similar trend as the graph of an exponential distribution.

Fig 5.1 Distribution of Education Expenditure



the exponential model we will fit is:

$$X_i \sim Exp(\lambda)$$

For this specific dataset,

$i = 1, ..., n$ where n is the number of households in this dataset

X_i is the expenditure spent on education for the i_{th} household

$1/\lambda$ is the mean value of the expenditure spent on education in 2017

Based on our calculation (the derivation is shown in Appendix), the maximum likelihood estimator of λ is $1/\bar{x}$. It is calculated by using 1 over the total expenditure on education divided by the total number of households. It is also the reciprocal of the mean expenditure on education of one household for this specific data set. It is an unbiased estimator for λ.

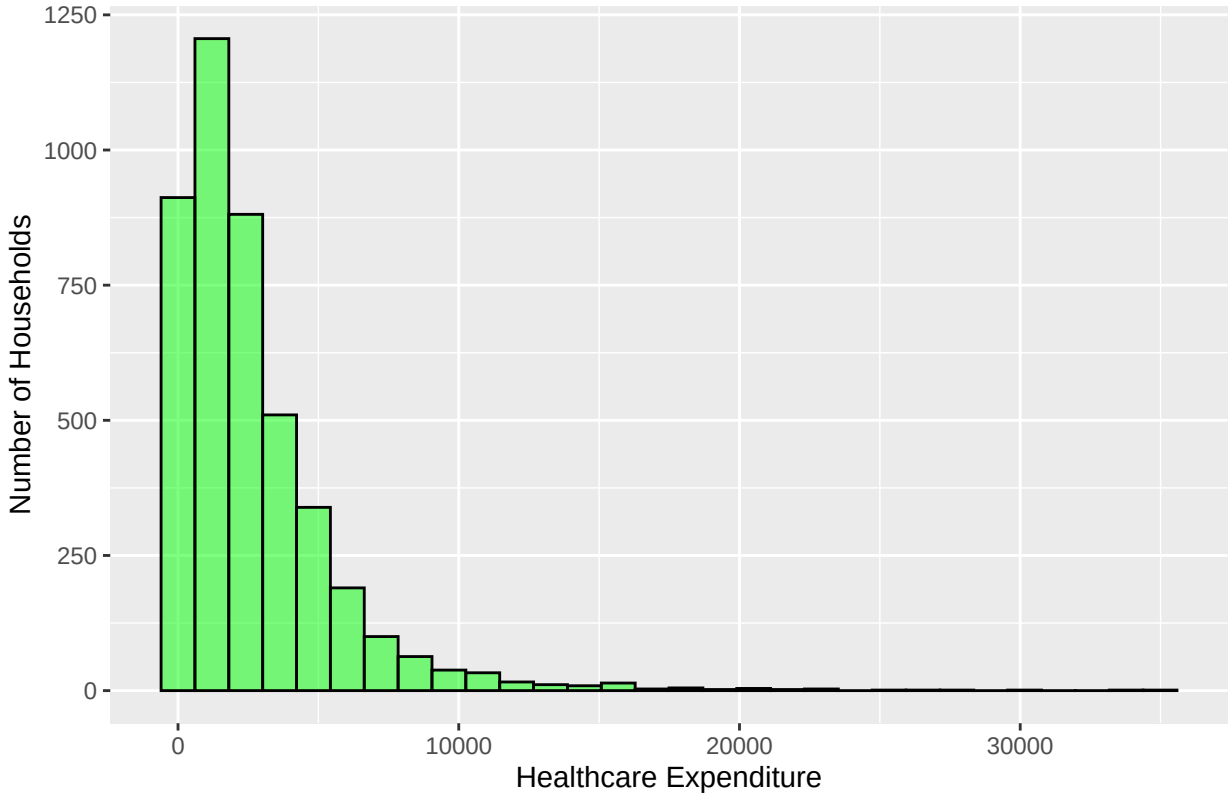
Healthcare Expenditure —- Bayesian Credible Interval

For the healthcare expenditure, we are going to use the exponential model using a Bayesian framework with credible interval.

For $(1-\alpha)100\%$ credible interval, we choose $\alpha = 0.05$ as it is one of the common value of α

This model is appropriate for this variable as the expenditure in healthcare is a numerical continuous variable. Furthermore, from the graph shown below (Fig 6), we could observe that the graph follows a similar trend as the graph of exponential distribution.

Fig 6. Distribution of Healthcare Expenditure



the exponential model we will fit is:

$$Y_i \sim Exp(\theta)$$

For this specific dataset,

$i = 1, ..., n$ where n is the number of households in this dataset

Y_i is the expenditure spent on healthcare for the i_{th} household

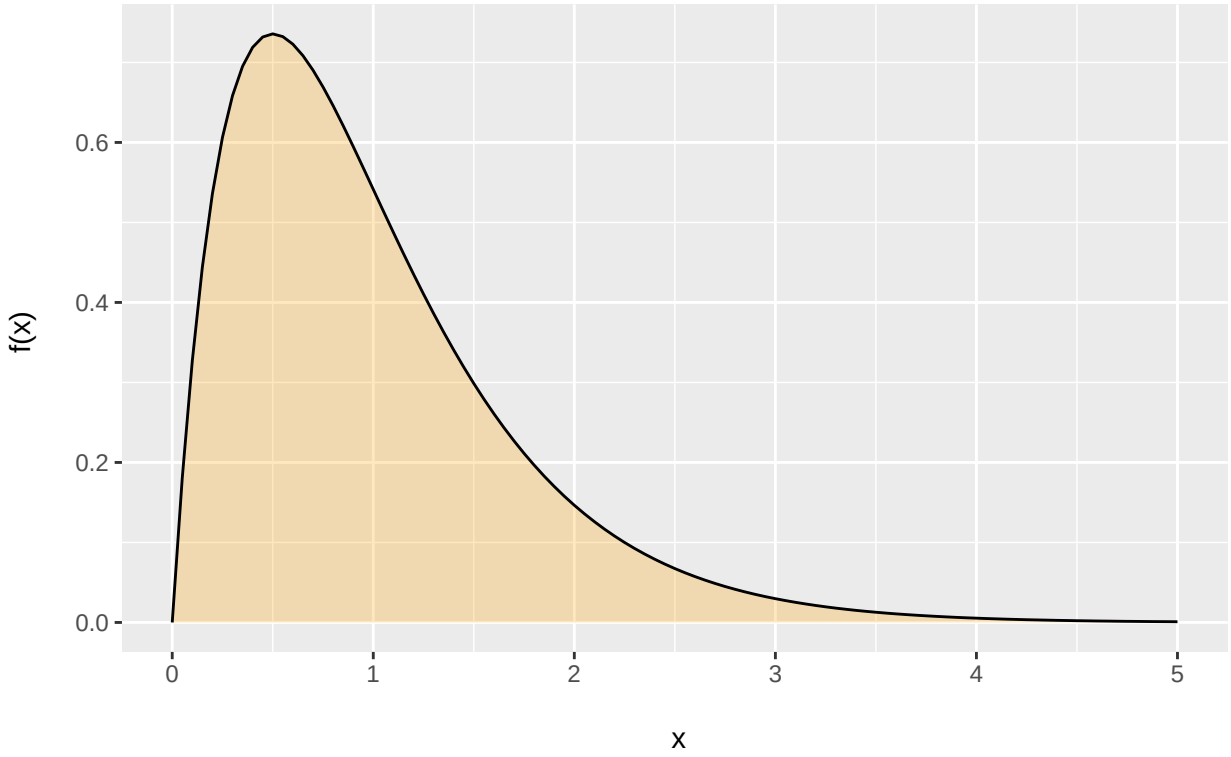
$1/\lambda$ is the mean value of the expenditure spent on healthcare in 2017

Since we know Gamma and Exponential distribution are common conjugate priors, we are going to choose gamma distribution as the prior for θ which is

$$\theta \sim Gamma(2, 2)$$

with probability density function is shown in Fig 7.

Fig 7. Probability Density Function of Gamma Distribution With Rate and Shape = 2



Results

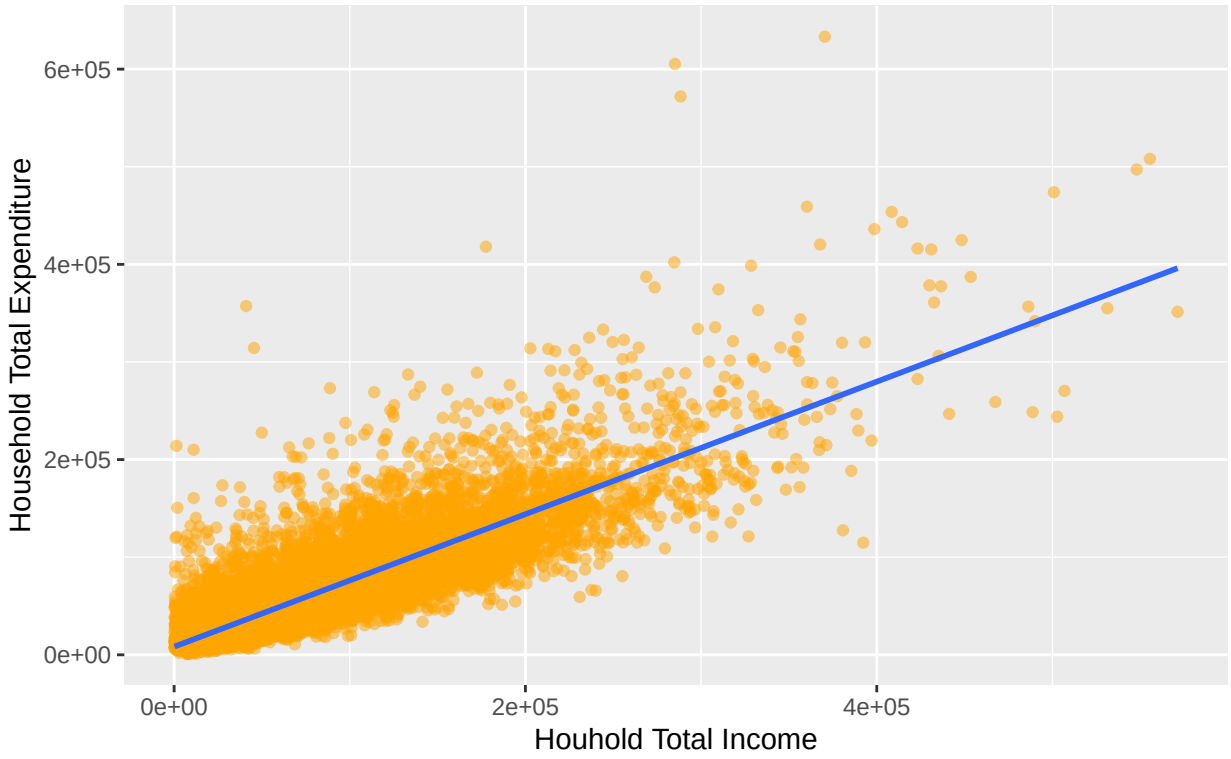
Income and Expenditure —- Simple Linear Regression

Based on our calculation, we get that the fitted value of household total expenditure $\hat{y}$ when the household total income is equal to x is:

$$\hat{y} = 17823.353968 + 1.025704x$$

By drawing this line as shown in Fig 4.2, we could observe that the graph generally follows this line trend.

Fig 4.2 The Scatter Plot of the Total expenditure against the Houhold total income with the Best Fit Line

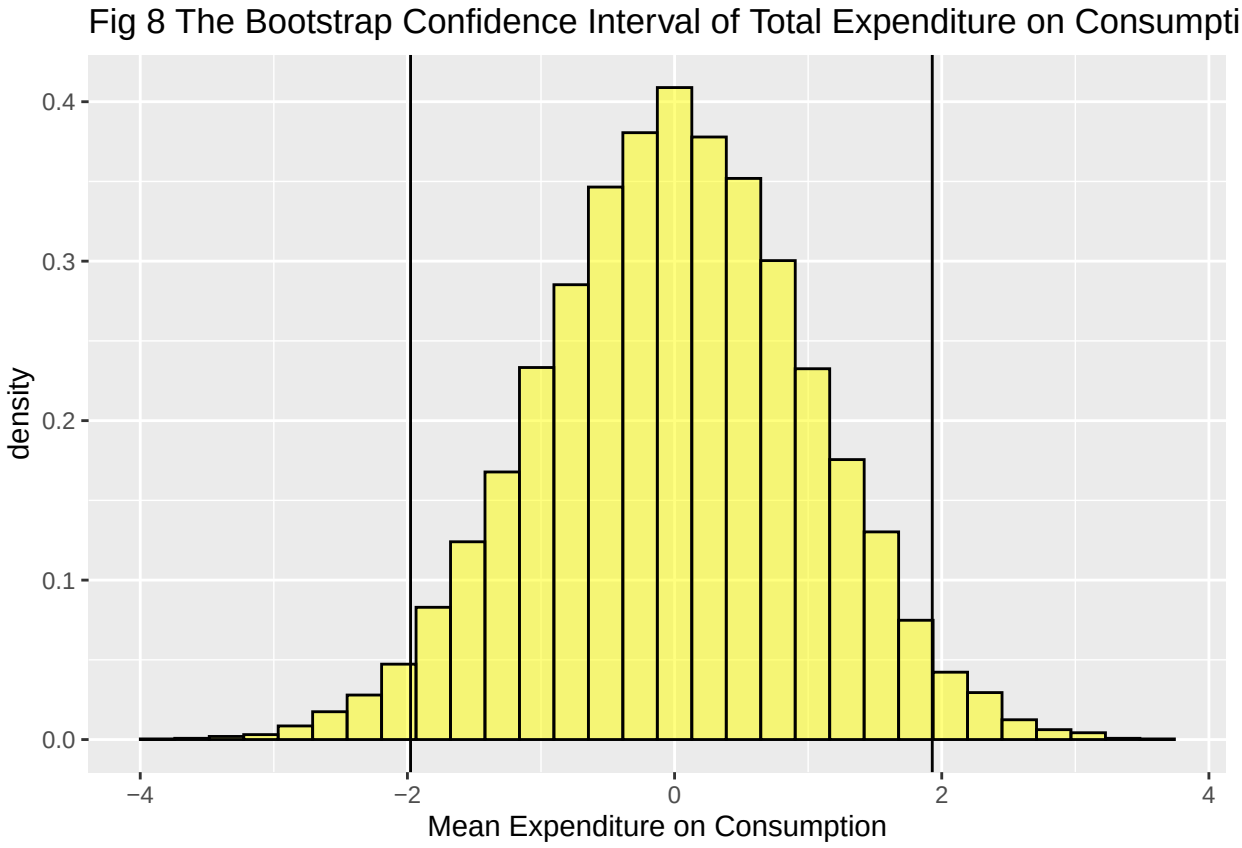


The estimate of the $\hat{\beta}_0$ is $\hat{\beta}_0 = 17823.353968$. This means that when the household total income is equal to 0, the mean household total expenditure is 17823.353968. This result is reasonable as even there is no income, the households still need to spend money for the basic needs of the family.

The estimate of $\hat{\beta}_1$ is $\hat{\beta}_1 = 1.025704$. This means that each additional dollar increase of the household total income is associated with an increase of 1.025704 in the household total expenditure. Although this value is positive and close to 1, this result seems unreasonable as the value is greater than 1. It indicates that one dollar income will trigger a more than more dollar increase in expenditure. In reality, people tend to spend less than proportionate income as some of the income is saved for emergency.

There are also some limitations to this result. From Fig 4.2 with the best-fitted line, we could observe that the error term does not have an equal spread around the best-fitted line. Therefore the third assumption of the linear regression may not hold. This may affect the accuracy of our result and a more appropriate model might be used.

Total Expenditure on Consumption —- Bootstrap Confidence Interval



Based on our calculation, we get:

$c_l^* = -1.976294$ It means that the 2.5th percentile of the studentized means of the total expenditure on consumption is -1.976294

$c_u^* = 1.928968$ It means that the 97.5th percentile of the studentized means of the total expenditure on consumption is 1.928968

$\overline{x}_n = 45159.55$ It means that the sample mean of the total expenditure on consumption is 45159.55

$s_n = 33047.6$ is the sample standard deviations of the total expenditure on consumption is 33047.6

Substitute these values to our formula,

$$\left(\overline{x}_n - c_u^* \cdot \frac{s_n}{\sqrt{n}}, \overline{x}_n - c_l^* \cdot \frac{s_n}{\sqrt{n}}\right)$$

we have obtained that the 95% confidence interval for the true mean value of the total expenditure on consumption is (44589.19, 45743.9).

This implies that we are 95% confident that the true mean value of the total expenditure on consumption μ is as low as 44589.19 and as high as 45743.9. This value seems reasonable as this confidence interval has captured the sample mean of the total expenditure on consumption. Since our sample size is very big, the sample mean will be close to the true mean value of the total expenditure on consumption.

Shelter Expenditure —- Hypothesis Test

Based on our calculation, we get

$n = 12492$ It means that the number of observations in this sample data set is 12492

$\overline{x} = 17368.13$ It means that the sample mean of the total expenditure on shelter is 17368.13

$s_n = 13366.44$ It means that is the sample standard deviations of the total expenditure on shelter is 13366.44.

Substituting these value to our formula, we get p-value

$$P(|T| \geq 10.61007) = 3.457936 \times 10^{-26}$$

Since the p-value is extremely small, we have strong evidence against the null hypothesis. This value means that if the the average expenditure in shelter is 18637, there is 3.457936×10^{-24} % chance that a random sample of 12492 will have an average expenditure in shelter that deviates by more than 1268.87.

Education Expenditure —- Maximum Likelihood Estimation

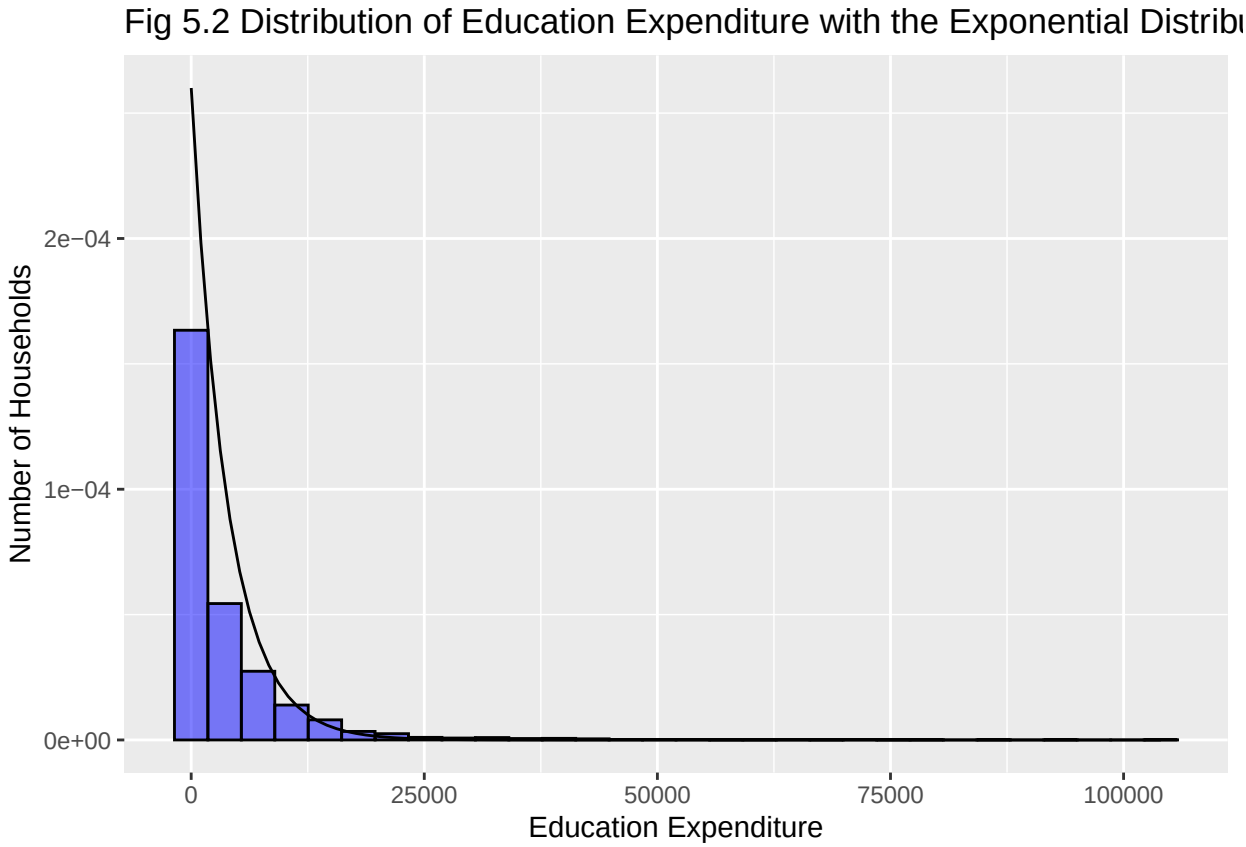
Based on our calculation, the value of $1/\bar{x}$ is 0.0002603494. This is calculated by using 1 divided by the mean expenditure in education. This means that if the education expenditure follows an exponential distribution, then the most likely value of λ would be 0.0002603494.

If we relate this value to our data set, this means that the expected education expenditure would be $1/0.0002603494 = 3840.992$ This value is considered reasonable as by checking the fact sheet from Pan-Canadian Education Indicators Program in 2009, from 1997 to 2007, the average education expenditure per household increases from 2000 to 3000. Following this trend, we could predict that a possible education expenditure in 2017 would be around 4000 and 3840.992 is very near this value.

Therefore, after calculation, now our model looks like:

$$X \sim Exp(\lambda = 0.0002603494)$$

By drawing the probability density function as shown in in Fig 5.2, we could observe that the graph generally follows this exponential curve trend.



There is also some limitations in this result. Since our observation is not the rate of time, exponential distribution may not be a good distribution despite the fitness. This may affect the accuracy of our result and a more appropriate model might be used instead.

Healthcare Expenditure —- Bayesian Credible Interval

Based on our calculation, the posterior distribution is

$$\theta \sim Gamma\left(n+2, \sum_{i=1}^{n} x_i + 2\right)$$

where n is the number of households in this sample

$\sum_{i=1}^{n} x_i$ is the total healthcare expenditure in 2017.

Based on our calculation, the 95^th credible interval for the value of θ is (0.000404333, 0.0004193448). It means that there is a 95% probability that θ is between 0.000404333 and 0.0004193448.

For this specific variable, this value means that there is a 95% probability that the true mean total healthcare expenditure in 2017 is between 2384.673 and 2473.209. This value is not reasonable as from the average annual household expenditure in Canada in 2017 in statista, we have obtained the average expenditure in healthcare is 2579. However, this value is not included in our credible interval.

There are some limitations to this result. For the exponential model we used to estimate the healthcare expenditure, we do not have much information about the θ. We choose our prior just based on the common conjugate priors of the exponential model. In this case, the prior we used for analysis may not be appropriate for the analysis of the Bayesian credible interval. More information needs to be investigated and a more suitable prior may be used.

Conclusions

Interpretation of the Results

From the previous sessions, we have used simple linear regression to estimate the relationship between income and expenditure. It shows that a one-dollar increase in income will trigger more than more dollar increase in expenditure. Based on this result, we could infer that the household expenditure in Canada might be too high. The government may choose to provide more subsidies to the household to reduce their expenditure. We have also used bootstrap confidence interval to estimate the total expenditure on consumption, maximum likelihood estimation to approximate the distribution for education expenditure, and Bayesian credible interval to approximate the distribution for healthcare expenditure. Based on our results, consumption has the highest mean expenditure. The government may want to implement policies to reduce its cost. From the investors' perspective, they may want to invest in a related area due to the high expenditure. Last but not least, when we use a hypothesis test to verify the average expenditure in the shelter provided in statista, an extremely small p-value tells us that we have strong evidence against the null hypothesis. This result tells us that the average expenditure in the shelter provided in statista is not accurate.

Limitations

There are also some limitations in our analysis. For the simple linear regression model we used to estimate the relationship between household total income and household total expenditure, the error term does not have an equal spread around the best-fitted line. Therefore, the third assumption for linear regression may not hold. This may affect the accuracy of our linear relationship result. For the maximum likelihood estimation we used to estimate the education expenditure, our observation is not the rate of time. In this case, exponential distribution may not be a good distribution despite the fitness. A more appropriate model might be used in this case. For the Bayesian credible interval we used to estimate the healthcare expenditure, we do not have much information about the θ. Our prior choice is just based on the common conjugate priors. In this case, the prior chosen may not be appropriate for the analysis of the Bayesian credible interval. More information needs to be investigated and a more suitable prior may be used.

Next Step

The next step of our analysis is to include more expenditure analysis such as transportation and recreation. However, we think that due to COVID-19, the expenditure in these areas would be lower than the normal time. In this case, the analysis of the expenditure in these areas will not help adjust the current policy in these areas. In later years of the analysis, we may include these areas when the country's expenditure preference resumes to normal.

Furthermore, due to the limitation of the data set, we are not able to find the more recent data later than 2017. If the recent year data can be found, we are willing to analyze the more recent data which will be more relevent as well.

Bibliography

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

2. Sam Firke (2021). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.1.0. https://CRAN.R-project.org/package=janitor

3. Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2021). skimr: Compact and Flexible Summaries of Data. R package version 2.1.3. https://CRAN.R-project.org/package=skimr

4. Tierney N (2017). "visdat: Visualising Whole Data Frames." *JOSS*, *2*(16), 355. doi: 10.21105/joss.00355 (URL: https://doi.org/10.21105/joss.00355), <URL: http://dx.doi.org/10.21105/joss.00355>.

5. Statistics Canada. (2017). Survey of Household Spending, 2017 [Canada]: Interview File. Data Liberation Initiative. Retrieved from http://odesi2.scholarsportal.info.myaccess.library.utoronto.ca/webview/index.jsp?object=http%3A%2F%2F142.150.190.128%3A80%2Fobj%2FfStudy%2Fshs-62M0004-E-2017-int&v=2&mode=download

Appendix

The derivation of the maximum likelihood estimation of λ

For the variable the education expenditure in the household, we are going to assume that it follows exponential distribution. This model is appropriate for this variable as the expenditure in education is a numerical continuous variable. Furthermore, from the graph shown (Fig 5.), we could observe that the graph follows a similar trend as the graph of exponential distribution.

$$X_i \sim Exponential(\lambda)$$

and

$$f(x) = \lambda e^{-\lambda x}$$

Our parameter of interest is λ, this parameter is investigated as this is the only unknown variable in the exponential model.

$1/\lambda$ represents the mean value of the expenditure spent on education in 2017

The likelihood function is

$$\begin{aligned} L(\lambda) &= f(x_1) \cdots f(x_n) \\ &= \lambda e^{-\lambda x_1} \cdot \ldots \cdot \lambda e^{-\lambda x_n} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \end{aligned}$$

The loglikelihood function is

$$l(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

Next, we will differentiate the log likelihood function with respect to λ, by setting $l'(\lambda) = 0$, we will obtained the critical value of λ

$$l'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

Lastly, we will do the second derivative test to verify that the critical value we obtained is the maximum value.

$$\begin{aligned} l''(\lambda) &= -\frac{n}{\lambda^2} \\ substitute\ \lambda &= \frac{1}{\bar{x}}, \\ \ell''(\lambda) &= -\frac{n}{\left(\frac{1}{\bar{x}}\right)^2} < 0 \end{aligned}$$

Therefore, we have proven that $\hat{\lambda}_{MLE} = \frac{1}{\bar{x}}$