

A statistical study on how people's life satisfaction presents differently among five family income levels and the presence/absence status of children

Ruoning Guo Student No.1004772114
Yifan Zhu Student No.1006345849

Fanzhen Meng Student No.1002812824
Kaiou Guo Student No.1004151865

10/19/2020

Abstract

This paper studied people's life satisfaction score by analyzing the Canadian GSS from 2017. Our results showed that people with lower family income represent significantly different on life satisfaction compared with other income levels, whereas the presence of children does not. By creating a new variable that combines income levels with having/no children status, we discovered that: 1) when people don't have children, the lowest income groups were no differences in mean life satisfaction compared to the highest income group 2) when people having children, the lowest income groups were significantly different in mean satisfaction compare to other high income groups. The results may open a door for the discussion of increasingly family pressure in modern society.

Introduction

Life satisfaction at an individual level is extremely important to the well-being of society. In fact, it can be seen as a measurement for human welfare (Haybron, 2011). Empirical evidence suggests that individuals with higher life satisfaction are more productive (Bellet, et al). What's more, life satisfaction is also linked to better academic performance among students (Antaramian, 2017). Therefore, it is important to understand what factors could contribute to a person's life satisfaction in order to improve society as a whole.

This analysis will study how do different variables contribute to an individual's life satisfaction using the Canadian general social surveys (GSS) from 2017. The GSS recorded 20602 individual-level data regarding each respondent's gender, income, education, etc (Statistics Canada, 2020). In particular, there is a variable that asks the respondents about their "feelings about life as a whole", which can be seen as a subjective measurement of the respondents' level of life satisfaction (Statistics Canada,2020).

When conducting our analysis, we wanted to narrow down some of the factors contributing to overall life satisfaction. Hence two key factors-"children" and "income" have been discovered from the GSS survey. In the following study, we will be focusing on investigating the close relationship between these two variables and the Canadian's life satisfaction. This paper involves six parts: data, model choosing, results, discussion, weakness, and next step.

Data

Feelings_life	Children	Income
8	1	25-50k
10	1	75-100k
8	1	75-100k
10	1	100k+
8	0	50-75k
9	1	50-75k
4	1	below25k
10	1	below25k

Data understanding

To understand our data, we must do a simple introduction to the methodology first. So the GSS is an annual survey designed to study the living conditions and well-being of Canadians (Government of Canada, 2020). The target population of this survey is all Canadians; however, the sampling frame is Canadians older than 15 with functional phone number who lives in private households in all provinces excluding the 3 territories (Government of Canada, 2020). The sampling frame should be a good representation of the target population, therefore the GSS is expected to provide useful and reliable information about the living conditions of Canadians. The non-response questions have been named NA. This paper will randomly select 1000 observations from the sampled population of 20602 individuals, and conduct the study only using the 1000 newly sampled observations.

The table above is a preview of our data set. The variables in the dataset are:

- Children(two levels): the person has children(1) or not(0).
- Income(five levels): below25k, 25-50k, 50-75k, 75-100k, and 100k+, which indicates the annual family income.
- feeling_life: the person's "feelings of life as a whole" on a scale of 1-10, which can be seen as a level of life satisfaction.

Data preparation

After using the "gsscleaning" code provided by our instructor Rohan Alexander and Sam Caetano, we've made some additional selection and modification.

- Remove NA and other unused columns: To facilitate the analysis, all rows that give NA in feelings_life, total_children, and income_family have been taking out first.
- Select the first 1000 rows: Since the raw dataset was already in a random sequence, there is no need to randomize the sample again. So we've selected the first 1000 rows of respondents from the raw dataset.
- Shorten the income level's name and two income level "125,000 and more" and "\$100,000 to 124,999" are combined into one level "100k+". So there were 6 levels, now 5 levels.
- Create two new variables: 1) Children-by converting the number of children to a factor with 2 levels; 1 if the person has 1 or more than 1 children, and 0 if the number of children is 0, and 2) CE- a variable that combines children status and family income.

Replace the continuous variable-"total_children" to a binary variable might omit some of the variance bias since a family with 1 child or 5 children will be placed at the same level. However, it will contribute to a more intuitive analysis. More importantly, using a binary variable to combine with another categorical variable to make a deeper analysis would be more feasible.

Model

Model selection

Linear regression is known as a statistical model that can be used to predict a continuous outcome according to one or more continuous predictor variables. By contrast, the analysis of variance(ANOVA) model is used to predict a continuous

outcome according to one or more categorical predictor variables(P.Mean, 2010). If we used a linear regression model to apply and analyze the impact of life satisfaction on the two categorical variables, some responses will not fit linear regression assumptions and cause assumption violation.

In our study, a two-sample t.test procedure will firstly be used to investigate whether or not there is a difference in life satisfaction between children status(having children or not). Then the one-way analysis of variance will be used to predict life satisfaction(scale of 1-10) according to the categorical variable -Income with 5 levels. Specifically, it will show the difference between two or more means through significance tests and provides us a way to make multiple comparisons of several groups of means. This is crucially important to our study because to compare the mean life satisfaction among different income levels is one of the most important parts of our analysis. In addition, the three assumptions of one-way ANOVA are i)normality ii)homogeneity of variance iii)independence(STAT:5201, University of IOWA). They will be verified in the discussion section.

However, ANOVA also has weaknesses. The results from the ANOVA indicated that several groups of means were not equal ($p\text{-value} < 0.05$), but it will not tell which means were different from which other means. Thus we will apply Tukey's HSD test to make further analysis. Tukey's test reports whether those means for every possible pair of levels are significantly different or not. Using Tukey's test, we will be able to effectively adjust the p-value of each comparison so that it corrects for multiple comparisons(Meng, 2019).

When conducting our model, we are concerned and wanted to investigate that: 1) is family income related to people's life satisfaction? 2) do people who have children possess a different life satisfaction score from those who do not have children? 3) whether or not there is a difference in life satisfaction among the 10 categories of people classified by the combination of their children status and family income?

Alternative model

Since the response variable, feelings_life is a categorical and ordered variable, it can take a number from 1 to 10, where higher numbers can be translated into higher life satisfaction. The ordinal logistic regression might be more suitable for modeling when the response is ordinal (Frost,2017). It could potentially provide a better estimation than the ANOVA model.

Furthermore, the income independent variable can be transformed into ordinal variables which will increase the interpretability of the estimated coefficient. We can construct a new variable denoted as X_1:

X_1=1, if family income is below \$25,000

X_1=2, if family income is between \$25,000 to 49,999

X_1=3, if family income is between \$50,000 to 74,999

X_1=4, if family income is between \$75,000 to 99,999

X_1=5, if family income is between \$100,000 to 124,999

X_1=6, if family income is \$125,000 and more.

This new variable can also replace the income variable used in the ANOVA model. Note if all we want to study is whether if family income has a statistically significant impact on life satisfaction, using or not would not make a large difference. However, the estimated coefficient on could provide real meanings under certain regression models.

We can then regress feelings_life on, children1, and the interaction term using ordinal logistic regression. A positive and significant coefficient would mean that the independent variable has a positive statistically significant impact on life satisfaction. For example, given the result from our ANOVA model, we would expect the coefficient to be positive. The coefficient on the interaction term can be used to study how does the impact of income on life satisfaction changes based on whether if the respondent has children or not.

Generally speaking, we would expect the ordinal logistic model to outperform linear models when the response variable is ordinal (Lee,2019). However, the ordinal logistic model also faces other limitations. Ordinal logistic regression has 4 key assumptions: ordinal response, variation in independent variables, no collinearity, and proportional odds (Lee,2019).

- The first two assumptions are clearly satisfied.

- The third assumption can be tested by studying the correlation between independent variables. However, given that we only have two independent variables, this should not be a large concern. Intuitively speaking, we would not expect having children or not to be significantly and directly correlated with income. Nevertheless, more rigorous methods should be used to test for multicollinearity (Lee,2019).
- The last assumption could be problematic because it requires that the relationship between each value of the response variable being the same (Lee,2019). Life satisfaction is a subjective term that cannot be exactly measured. One single value of the response by itself is quite meaningless, so intuitively, we would not know if this assumption holds or not. However, Brant's test can be used to test for proportional odds (Lee,2019).

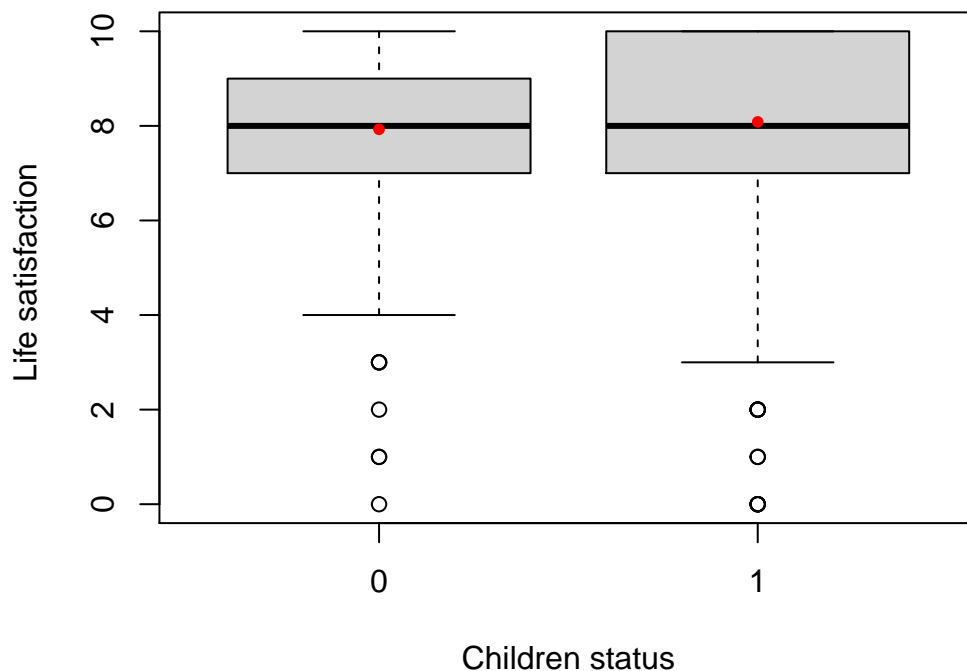
In summary, the ordinal logistic model is an alternative model that can be used in our study. It might outperform our current model if all assumptions are satisfied, however, it also faces its own limitations.

Results

The boxplots below show people's life satisfaction for people who have children and no children, for people classified by family income, and people classified by both children status and family income.

Figure 1: To compare life satisfaction between children and no-children status

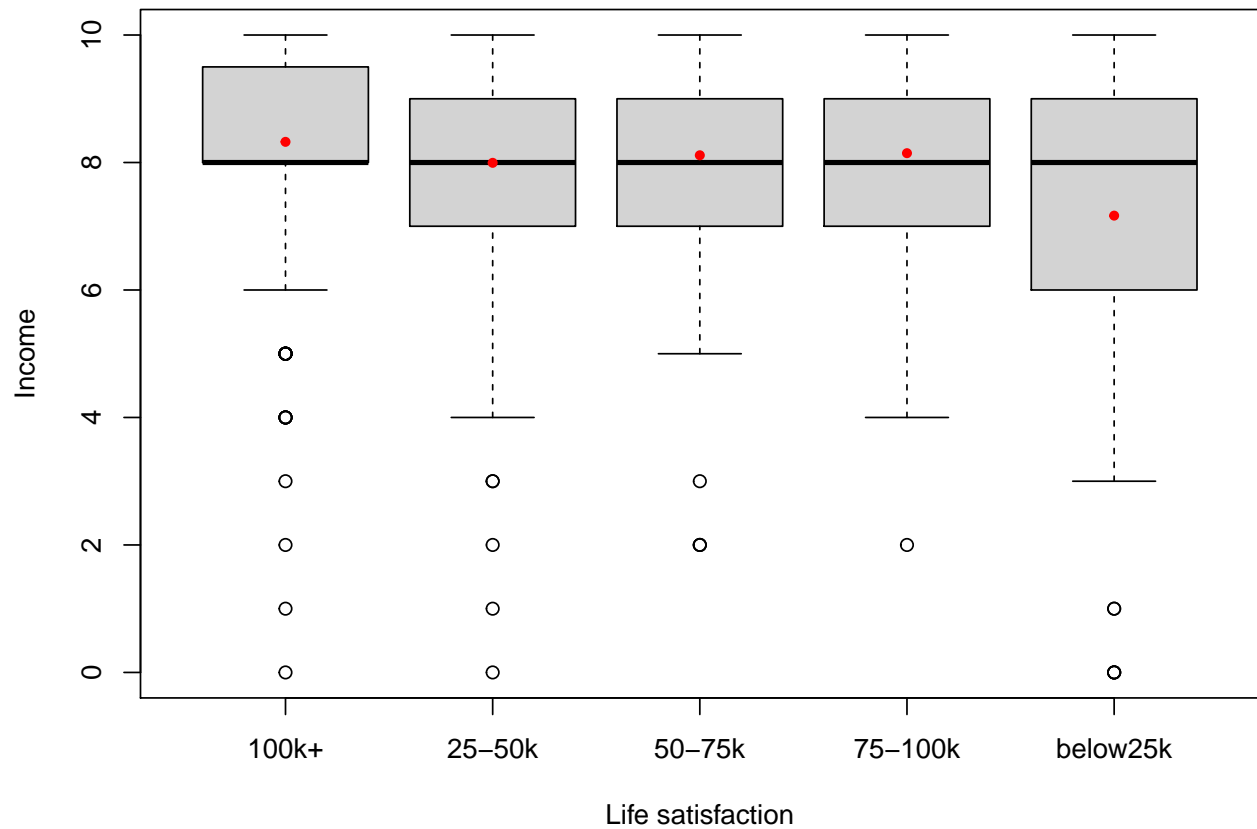
Figure 1: Life satisfaction by children status



- The distribution of life satisfaction among people who have children and no children seems similar in variance and mean (these red dots). In particular, people who have children look have a higher life satisfaction on average, but not much.

Figure 2: To compare life satisfaction by family income

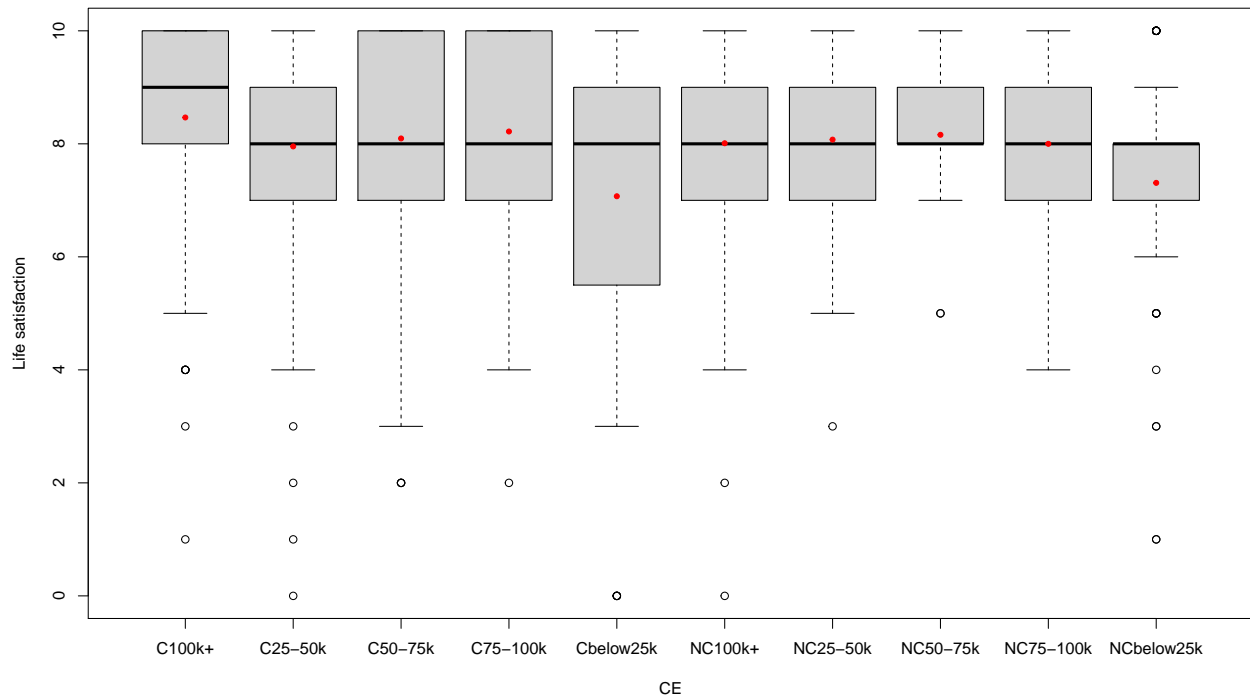
Figure 2: Life satisfaction by income



- From the plot above, we can observe that as the family income increases, life satisfaction increasing as well.

Figure 3: to compare life satisfaction among the 10 categories of the new factor-CE (people that classified by the combination of their children status and family income)

Figure 3: Life satisfaction by CE

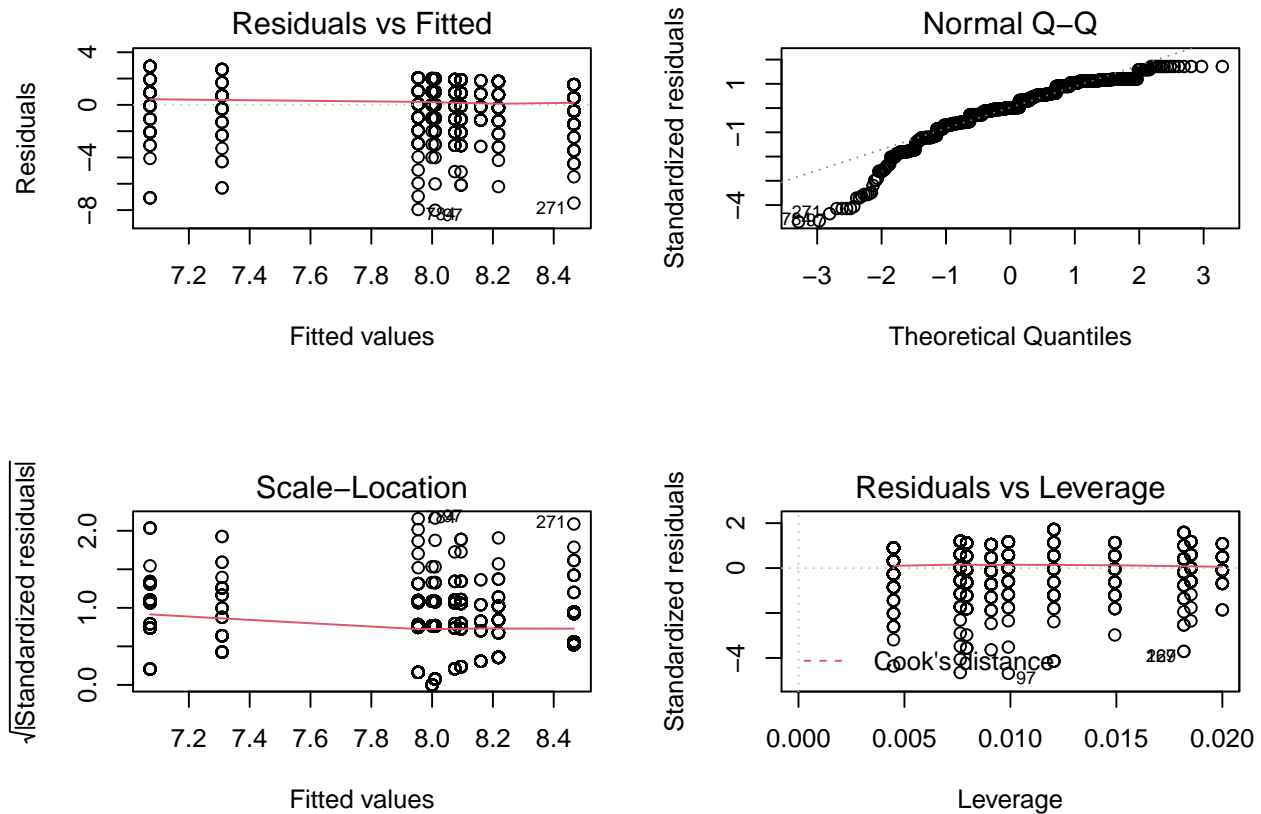


Recall: CE is a variable that combines children status and family income. (Note: C and NC indicates the children status- Having children or No children; family income level is following after. For example, NCbelow25k indicates people who don't have children and with annual family income below 25,000 dollars; C100+ indicate people who have children and with annual family income more than 100,000 dollars.)

- The boxplot shows that for all family income below25k, 25-50k, 50k-75k, people without children have a higher average life satisfaction than people who have children. On the contrary, for higher family income 75-100k and 100k+, people who have children have a higher average life satisfaction than those who don't have children.

Residual plots

Figure 4



- The Residuals versus Fitted values plot shows the fitted values (predicted values) on the x-axis and the residuals (errors) on the y-axis. It can be observed that there is a pattern that the variation of the residuals is not randomly distributed around 0. Hence the assumption of constant variance might not be held.
- Normal quantile plot: the majority of residual points line on the 45-degree reference line, which is a normal distribution. Hence the assumption of normality holds true.
- “Residuals vs Leverage” plot helps us to detect if there are any influential cases. Since they did not exceed the Cook’s distance, there are no influential points.

Discussion

Section 1: Investigate whether or not there is a difference in life satisfaction between children status(having children or not).

- Using F-test for variance, we got the p-value=0.04378 therefore we reject H_0 , that evidence both groups have different variances. Then we need to use Welch’s two-sample t-test to adjust the number of degrees of freedom when the variances are different.
- By using Walch’s two-sample t-test(Meng, 2019), we have $t = -1.384$, $df = 705$, p-value = 0.1668. Evidence shows(p=0.17) that people’s life satisfaction has no difference for people who have children or not.

Section 2: Investigate whether or not there is a difference in life satisfaction among people classified by family income.

- Applying a one-way analysis of variance. In one-way ANOVA, there are two possible hypotheses. The null hypothesis (H_0) is that there is no difference between 5 family income levels and equality between means. The alternative hypothesis (H_1) is that there is a difference between the means and levels(Meng, 2019).
- After that we have test statistics(df)=26.295(1,998) and P-value<0.0001, which reject H_0 . This is strong evidence that shows that life satisfaction means are very different from the 5 levels of family income.

- In addition, the Tukey's procedure((Meng, 2019)) also indicates the 5 levels are very different. We capture the people with an income level below 25k were significantly different($p\text{-value}<0.0001$) from all other income levels in mean life satisfaction.

Section 3: Investigate whether or not there is a difference in life satisfaction among the 10 categories of people classified by the combination of their children status and family income.

- By using ANOVA, there is strong evidence that shows that not all 10 categories of people have the same life satisfaction means($p\text{-value}<0.0001$).
- However, there are several notable points by observing Tukey's output and then comparing the p-values. Evidence shows that people(have children) with family income 0-25k, 25k-50k and 100k+ were no difference in mean life satisfaction compared to people who have no children($p\text{-value}=0.998, 0.999, 0.445$ respectively). In another word, it seems these three income group's life satisfaction level is not affected by children status. In addition, no-children people with income below 25k were no differences in mean life satisfaction compare to no-children people with income 75-100k and 100k+($p\text{-value}=0.25$ and 0.30).
- On the other hand, people who have children with income below 25k were significantly different in mean satisfaction compare to all other income groups($p\text{-value}<0.01$). This can also be observed in figure 3 boxplot, which shows the largest mean difference in life satisfaction is between C25k and C100k+(people who have children with income below 25k and people who have children with income above 100k).

Section 4: Assess whether the necessary assumptions of the model hold.

- Normality: From results-Figure 4 observation we know that the normality hold.
- Constant variance: the assumption of constant variance might not behold since there is a pattern that the variation of the residuals is not randomly distributed around 0.
- Independence: We assume that people include in this study were not related in any way. It is likely that with a sample this large that some people are related but no additional information was given to verify relatedness.

Perhaps a bit surprising after discussion, those who have children showed no statistically significant difference in the level of life satisfaction comparing to those who do not have children while controlling for their income. It could be the case that having children will make people happier, however, at the same time, it will also increase the emotional burden (Angeles, 2009). This could explain why having children did not have a statistically significant impact on life satisfaction. However, this could also be the result of not controlling for enough variables, which will be further addressed in the weakness section.

On average, people with different family income levels showed statistically significant different levels of life satisfaction, while controlling for the presence/absence of children. What's more, a notable finding is that: for people who don't have children, the lower income groups were no differences in mean life satisfaction compared to the high income group. While for people who have children, the lowest income groups were significantly different in mean satisfaction compare to other high income groups. This result is expected, previous empirical research has suggested that richer people are often "happier" people (Ortiz-Ospina, et al,2013). Individuals with higher incomes are not only able to purchase a greater variety and amount of goods and services, but also better able to raise children financially.

Although the no-children family with low income were no differences in mean life satisfaction, with the birth of children, higher family income seems to play an important role to break the barrier such that leading to a higher level of life satisfaction.

Weaknesses

The work might have some drawbacks as follows. First of all, we have employed various statistical methods to verify the necessary assumptions for our models. It appears that the variation of the residual is not randomly distributed around 0. Therefore, the results may need to be validated again in the future since the assumption of the model might not be fully held.

Secondly, the goal of this study is to study variables that contribute to life satisfaction, however, we only used the Canadian GSS data from 2017. So further study can be conducted using the Canadian GSS survey from other years. If we observe similar results from other datasets, then it makes our conclusion more reliable. Moreover, we do not know if our result

from studying the Canadian population would translate well to other parts of the world or not. Income and the presence of children might impact life satisfaction differently in other countries. For example, intuitively, we might expect the presence/absence of children to have a greater impact on life satisfaction in countries with low birth rates, such as Japan.

Thirdly, our research might have omitted variable bias problems. Intuitively speaking, apart from income and having children or not, there are many other variables such as education, sex, religion, etc. that could potentially have a significant impact on life satisfaction. By omitting these variables we have induced bias into our estimated coefficients, it might also affect the significance of the variables we included. This is because the predictor variables are no longer independently distributed with the error term, leading to invalid inferences.

Next Steps

The results may need to be validated again since the homogeneity of variance assumption of the model might not be held. In the next step, we have a few considerations:

- Find a transformation to deal with the unconstant variance. For example, applying the variance-stabilizing transformation to the response variable(y) or Box-Cox transformations to achieve constant variance(STAT:5201, University of IOWA).
- Applying multinomial regression and comparing the results with the current results.
- More rigorous statistical methods should be used when selecting relevant independent variables. The omitted variable bias problem was discussed earlier, however, including too many independent variables will lead to a multi-collinearity problem and will over-complicate the problem. Multi-collinearity will increase the standard error of our estimates, so there is a tradeoff between omitted variable bias and multi-collinearity. We can potentially look at the Akaike information criterion or the Bayesian information criterion to come up with a better set of relevant independent variables for this study.

References

- Angeles, L. (2009). Children and Life Satisfaction. *Journal of Happiness Studies*, 11(4), 523-538. doi:10.1007/s10902-009-9168-z
- Antaramian, S. (2017). The importance of very high life satisfaction for students' academic success. Retrieved October 19, 2020, from <https://www.tandfonline.com/doi/abs/10.1080/2331186X.2017.1307622>
- Bellet, C., Neve, J. D., & Ward, G. (2019). Does Employee Happiness Have an Impact on Productivity? *SSRN Electronic Journal*. doi:10.2139/ssrn.3470734
- Frost, J., Tamanani, R., Hedvat, J., Bumadian, I., Manfredo-Thomas, A., Padua, A., . . . Siddiqui, A. (2017, June 13). Regression Tutorial with Analysis Examples. Retrieved October 19, 2020, from <https://statisticsbyjim.com/regression/regression-tutorial-analysis-examples/>
- Government of Canada, S. (2017, February 27). The General Social Survey: An Overview. Retrieved October 19, 2020, from <https://www150.statcan.gc.ca/n1/pub/89f0115x/89f0115x2013001-eng.html>
- Government of Canada, S. (2020, October 17). Statistics Canada: Canada's national statistical agency. Retrieved October 19, 2020, from <https://www.statcan.gc.ca/eng/start>
- Haybron, D. M. (2001). 3. Life satisfaction. *Happiness*, 31-41. doi:10.1093/actrade/9780199590605.003.0003
- Lee, E. (2019, May 29). Ordinal Logistic Regression on World Happiness Report. Retrieved October 19, 2020, from <https://medium.com/evangelinelee/ordinal-logistic-regression-on-world-happiness-report-221372709095>
- Ortiz-Ospina, E., & Roser, M. (2013, May 14). Happiness and Life Satisfaction. Retrieved October 19, 2020, from <https://ourworldindata.org/happiness-and-life-satisfaction>
- One-way ANOVA Part 1. (n.d.). Retrieved October 19, 2020, from <https://vault.hanover.edu/~altermattw/courses/220/spss/oneway/one-way-stats1.html>

Meng, F. (2019). STA303 notes. Toronto, Ontario: University of Toronto.

Mean, P. (n.d.). Regression and Anova. Retrieved October 19, 2020, from <http://www.pmean.com/10/Interesting2010.html>

Pitman, E. J. (1979). Some basic theory for statistical inference. London, UK: Chapman and Hall.

STAT:5201 Applied Statistics II. (n.d.). Retrieved October 19, 2020, from <http://homepage.divms.uiowa.edu/~rdecook/stat5201.html>

Technology, A. (n.d.). Data Centre. Retrieved October 19, 2020, from <http://dc.chass.utoronto.ca/myaccess.html>

Software used in producing the report: Rstudio

Packages used in producing the report:

readr: Read Rectangular Text Data

dplyr: A Grammar of Data Manipulation

FrF2: Fractional Factorial Designs with 2-Level Factors

kableExtra: Construct Complex Table with 'kable' and Pipe Syntax

ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics

Retrieved October 19, 2020, from https://cran.r-project.org/web/packages/available_packages_by_name.html

Appendix

#Section 1:

```
children_mean <- tapply(feelings_life, Children, mean)
children_mean
```

```
##           0           1
## 7.926606 8.084695
```

```
var.test(feelings_life~Children)
```

```
##
## F test to compare two variances
##
## data:  feelings_life by Children
## F = 0.82207, num df = 326, denom df = 672, p-value = 0.04378
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6837250 0.9945108
## sample estimates:
## ratio of variances
##           0.8220664
```

```
t.test(feelings_life~Children, var.equal=FALSE) #Welch Two Sample t-test
```

```
##
## Welch Two Sample t-test
##
## data:  feelings_life by Children
## t = -1.384, df = 705.66, p-value = 0.1668
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.38235831 0.06617853
```

```
## sample estimates:
## mean in group 0 mean in group 1
##      7.926606      8.084695

#Section 2:

income_mean <- tapply(feelings_life,income,mean)

income_mean

##      100k+    25-50k    50-75k    75-100k below25k
## 8.324074 7.994949 8.113636 8.146341 7.166667

anova(lm(feelings_life~income))

## Analysis of Variance Table
##
## Response: feelings_life
##           Df Sum Sq Mean Sq F value    Pr(>F)
## income      4  134.56   33.641   11.395 4.785e-09 ***
## Residuals 995 2937.35    2.952
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(aov(feelings_life~as.factor(income)))

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov.default(formula = feelings_life ~ as.factor(income))
##
## $`as.factor(income)`
##              diff              lwr              upr              p adj
## 25-50k-100k+    -0.3291246 -0.7526719  0.09442274 0.2107969
## 50-75k-100k+    -0.2104377 -0.6501088  0.22923342 0.6863234
## 75-100k-100k+   -0.1777326 -0.6277065  0.27224123 0.8171870
## below25k-100k+ -1.1574074 -1.6346956 -0.68011918 0.0000000
## 50-75k-25-50k    0.1186869 -0.3677412  0.60511497 0.9634367
## 75-100k-25-50k   0.1513920 -0.3443681  0.64715205 0.9198737
## below25k-25-50k -0.8282828 -1.3489609 -0.30760476 0.0001478
## 75-100k-50-75k   0.0327051 -0.4768991  0.54230928 0.9997883
## below25k-50-75k -0.9469697 -1.4808461 -0.41309329 0.0000143
## below25k-75-100k -0.9796748 -1.5220674 -0.43728215 0.0000092

#Section 3:

CE_mean <- tapply(feelings_life,CE,mean)

CE_mean

##      C100k+    C25-50k    C50-75k    C75-100k    Cbelow25k    NC100k+    NC25-50k
## 8.466368 7.954198 8.095238 8.218182 7.072289 8.009901 8.074627
##      NC50-75k    NC75-100k    NCbelow25k
## 8.160000 8.000000 7.309091

tapply(feelings_life,CE,var)

##      C100k+    C25-50k    C50-75k    C75-100k    Cbelow25k    NC100k+    NC25-50k
```

```
##      2.150891    3.628655    3.142857    2.282235    5.848369    2.989901    2.524649
##      NC50-75k    NC75-100k    NCbelow25k
##      1.280000    1.471698    4.476768
```

```
t.test(feelings_life~Children, var.equal=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: feelings_life by Children
## t = -1.384, df = 705.66, p-value = 0.1668
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.38235831  0.06617853
## sample estimates:
## mean in group 0 mean in group 1
##      7.926606      8.084695
```

```
lm(feelings_life~CE)
```

```
##
## Call:
## lm.default(formula = feelings_life ~ CE)
##
## Coefficients:
## (Intercept)      CEC25-50k      CEC50-75k      CEC75-100k      CECbelow25k
##      8.4664      -0.5122      -0.3711      -0.2482      -1.3941
##      CENC100k+      CENC25-50k      CENC50-75k      CENC75-100k      CENCbelow25k
##      -0.4565      -0.3917      -0.3064      -0.4664      -1.1573
```

```
anova(lm(feelings_life~CE))
```

```
## Analysis of Variance Table
##
## Response: feelings_life
##           Df Sum Sq Mean Sq F value    Pr(>F)
## CE           9  153.42   17.047    5.7825 7.296e-08 ***
## Residuals  990  2918.49    2.948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(aov(feelings_life~CE))
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov.default(formula = feelings_life ~ CE)
##
## $CE
##              diff              lwr              upr              p adj
## C25-50k-C100k+ -0.51216924 -1.11149798  0.087159499  0.1710079
## C50-75k-C100k+ -0.37112962 -0.97790306  0.235643820  0.6418550
## C75-100k-C100k+ -0.24818589 -0.88252998  0.386158189  0.9656309
## Cbelow25k-C100k+ -1.39407856 -2.09411522 -0.694041892  0.0000000
## NC100k+-C100k+ -0.45646672 -1.10946335  0.196529904  0.4459249
## NC25-50k-C100k+ -0.39174085 -1.15024981  0.366768116  0.8293723
## NC50-75k-C100k+ -0.30636771 -1.15828110  0.545545671  0.9804042
```

## NC75-100k-C100k+	-0.46636771	-1.29210543	0.359370008	0.7410375
## NCbelow25k-C100k+	-1.15727680	-1.97694894	-0.337604666	0.0003591
## C50-75k-C25-50k	0.14103962	-0.53831603	0.820395274	0.9997068
## C75-100k-C25-50k	0.26398334	-0.44010650	0.968073193	0.9740318
## Cbelow25k-C25-50k	-0.88190932	-1.64571651	-0.118102123	0.0098952
## NC100k+-C25-50k	0.05570252	-0.66523763	0.776642661	0.9999999
## NC25-50k-C25-50k	0.12042839	-0.69730380	0.938160589	0.9999839
## NC50-75k-C25-50k	0.20580153	-0.69924351	1.110846559	0.9993835
## NC75-100k-C25-50k	0.04580153	-0.83464886	0.926251912	1.0000000
## NCbelow25k-C25-50k	-0.64510756	-1.51987182	0.229656696	0.3649553
## C75-100k-C50-75k	0.12294372	-0.58749388	0.833381322	0.9999360
## Cbelow25k-C50-75k	-1.02294894	-1.79261152	-0.253286352	0.0011296
## NC100k+-C50-75k	-0.08533711	-0.81247792	0.641803706	0.9999978
## NC25-50k-C50-75k	-0.02061123	-0.84381534	0.802592884	1.0000000
## NC50-75k-C50-75k	0.06476190	-0.84523017	0.974753981	1.0000000
## NC75-100k-C50-75k	-0.09523810	-0.98077293	0.790296743	0.9999990
## NCbelow25k-C50-75k	-0.78614719	-1.66602876	0.093734385	0.1268243
## Cbelow25k-C75-100k	-1.14589266	-1.93747267	-0.354312654	0.0002143
## NC100k+-C75-100k	-0.20828083	-0.95858222	0.542020559	0.9970256
## NC25-50k-C75-100k	-0.14355495	-0.98728680	0.700176892	0.9999445
## NC50-75k-C75-100k	-0.05818182	-0.98678507	0.870421429	1.0000000
## NC75-100k-C75-100k	-0.21818182	-1.12283112	0.686467485	0.9990108
## NCbelow25k-C75-100k	-0.90909091	-1.80820714	-0.009974681	0.0449727
## NC100k+-Cbelow25k	0.93761183	0.13100717	1.744216499	0.0090479
## NC25-50k-Cbelow25k	1.00233771	0.10816639	1.896509024	0.0144419
## NC50-75k-Cbelow25k	1.08771084	0.11305044	2.062371249	0.0152549
## NC75-100k-Cbelow25k	0.92771084	-0.02415536	1.879577050	0.0633170
## NCbelow25k-Cbelow25k	0.23680175	-0.70980741	1.183410915	0.9986776
## NC25-50k-NC100k+	0.06472588	-0.79311770	0.922569453	1.0000000
## NC50-75k-NC100k+	0.15009901	-0.79134465	1.091542672	0.9999681
## NC75-100k-NC100k+	-0.00990099	-0.92772589	0.907923906	1.0000000
## NCbelow25k-NC100k+	-0.70081008	-1.61318181	0.211561647	0.3056762
## NC50-75k-NC25-50k	0.08537313	-0.93209796	1.102844226	0.9999999
## NC75-100k-NC25-50k	-0.07462687	-1.07028433	0.921030602	1.0000000
## NCbelow25k-NC25-50k	-0.76553596	-1.75616880	0.225096883	0.2971599
## NC75-100k-NC50-75k	-0.16000000	-1.22852906	0.908529063	0.9999814
## NCbelow25k-NC50-75k	-0.85090909	-1.91475776	0.212939577	0.2504116
## NCbelow25k-NC75-100k	-0.69090909	-1.73391454	0.352096354	0.5275291

```

#Section 4:
par(mfrow=c(2,2))
lm_CE <- lm(feelings_life~CE, newdata1)
plot(lm_CE)

```

