

Analysis on key factors affecting life satisfaction

Woolim Kim, Yena Joo, Guemin Kim

Oct 19, 2020

Analysis on key factors affecting life satisfaction

Woolim Kim, Yena Joo, Guemin Kim

Oct 19, 2020

Abstract

Here is where you give a brief (one paragraph overview of your entire paper). This should include some background/introduction, some methodology, results and conclusions.

Introduction

Here is where you should give insight into the setting and introduce the goal of the analysis. Here you can introduce ideas and basic concepts regarding the study setting and the potential model. Again, this is the introduction, so you should be explaining the importance of the work that is ahead and hopefully build some suspense for the reader. You can also highlight what will be included in the subsequent sections.

Data

Introduce the data, explain why it was selected. Make sure to comment on important features and highlight any potential drawbacks to the data.

**** Guemin**

The data set we chose for this assignment contains responses of the General Social Survey conducted in 2017. The contents of the survey include some characteristics of diverse families in Canada, their socio-economic status, as well as other subjective information such as the respondent's life satisfaction and health conditions.

Since we want to identify some key factors affecting one's life satisfaction as mentioned previously, the focus of our analysis will be "Health and subjective well-being".

The reason for choosing the 2017 GSS is because it is the most recent¹ survey that includes the "Health and subjective well being" concept.

Since the original data set contains too many variables that are not necessary, we are going to clean the data set prior to analysis by removing them.

This is what our new data looks like:

```
## # A tibble: 6 x 9
##       ID   age sex  satisfaction_sc~ self_rated_heal~ self_rated_ment~
##   <dbl> <dbl> <chr>          <dbl>          <dbl>          <dbl>
## 1     1     1 52.7 Fema~           8             5             5
```

¹As it is stated in the documentation of the GSS, one of the primary objectives of the General Social Survey is to monitor the well being of Canadians over time. As a result, every survey conducted so far contain the responses related to the questions asking for the respondents' well being, and the 2017 GSS is the most recent survey with such responses.

```
## 2      2  51.1 Male           10           3           3
## 3      5   28  Male           8           3           3
## 4      9  63.8 Fema~        8           4           4
## 5     11  15.7 Male          10           5           5
## 6     12  40.3 Fema~         6           4           3
## # ... with 3 more variables: family_income <chr>,
## #   average_hours_worked <chr>, selfRatedTotalHealth <dbl>
```

Our data contains 8 variables: ID, age, sex, satisfaction_score, selfRatedHealth, selfRatedMentalHealth, family_income, and average_hours_worked.

Detailed descriptions of some variables:

- * satisfaction_score indicates the life satisfaction score on a scale of 0(very dissatisfied) to 10(very satisfied).
- * selfRatedHealth and selfRatedMentalHealth are the physical and mental health ratings, respectively, on a scale of 1(poor) to 5(Excellent) given by the respondent.
- * average_hours_worked indicates the average number of hours worked per week.

Since we want to observe how the life satisfaction score is related to potential factors such as health or financial conditions, the response variable of our analysis will be satisfaction_score and the predictors will be the potential factors: selfRatedHealth, selfRatedMentalHealth, family_income, and average_hours_worked.

Model

Introduce the selected model here. It is expected that you will use some mathematical notation here. If you do please ensure that all notation is explained. You may also want to discuss any special (hypothetical) cases of your model here, as well as any caveats.

```
# General Linear Regression Model
satisfaction_lm <- lm(satisfaction_score ~ as.factor(selfRatedHealth)
+ as.factor(selfRatedMentalHealth)
+ family_income
+ average_hours_worked, data = life_satisfaction_data)
summary(satisfaction_lm)
```

```
##
## Call:
## lm(formula = satisfaction_score ~ as.factor(selfRatedHealth) +
##   as.factor(selfRatedMentalHealth) + family_income + average_hours_worked,
##   data = life_satisfaction_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.1029 -0.7224  0.0663  0.8971  6.1204
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)      3.92502    0.29485  13.312
## as.factor(selfRatedHealth)2      0.51471    0.10379   4.959
## as.factor(selfRatedHealth)3      0.83475    0.09740   8.571
## as.factor(selfRatedHealth)4      0.98584    0.09793  10.067
## as.factor(selfRatedHealth)5      1.17522    0.09973  11.784
## as.factor(selfRatedMentalHealth)2      1.63274    0.11629  14.040
## as.factor(selfRatedMentalHealth)3      2.69951    0.10889  24.790
## as.factor(selfRatedMentalHealth)4      3.24499    0.10932  29.684
```

```

## as.factor(selfRatedMentalHealth)5      3.72145      0.11004      33.818
## family_income$125,000 and more          0.04865      0.03822       1.273
## family_income$25,000 to $49,999        -0.36600      0.04275      -8.561
## family_income$50,000 to $74,999        -0.14543      0.04231      -3.437
## family_income$75,000 to $99,999        -0.12047      0.04285      -2.811
## family_incomeLess than $25,000         -0.47951      0.05232     -9.166
## average_hours_worked0.1 to 29.9 hours   0.29608      0.27262       1.086
## average_hours_worked30.0 to 40.0 hours  0.23252      0.27176       0.856
## average_hours_worked40.1 to 50.0 hours  0.32056      0.27335       1.173
## average_hours_worked50.1 hours and more 0.39476      0.27448       1.438
##                                     Pr(>|t|)
## (Intercept)                          < 2e-16 ***
## as.factor(selfRatedHealth)2           7.17e-07 ***
## as.factor(selfRatedHealth)3           < 2e-16 ***
## as.factor(selfRatedHealth)4           < 2e-16 ***
## as.factor(selfRatedHealth)5           < 2e-16 ***
## as.factor(selfRatedMentalHealth)2     < 2e-16 ***
## as.factor(selfRatedMentalHealth)3     < 2e-16 ***
## as.factor(selfRatedMentalHealth)4     < 2e-16 ***
## as.factor(selfRatedMentalHealth)5     < 2e-16 ***
## family_income$125,000 and more         0.203047
## family_income$25,000 to $49,999       < 2e-16 ***
## family_income$50,000 to $74,999       0.000589 ***
## family_income$75,000 to $99,999       0.004939 **
## family_incomeLess than $25,000        < 2e-16 ***
## average_hours_worked0.1 to 29.9 hours  0.277474
## average_hours_worked30.0 to 40.0 hours 0.392231
## average_hours_worked40.1 to 50.0 hours 0.240935
## average_hours_worked50.1 hours and more 0.150400
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.298 on 12927 degrees of freedom
## (62 observations deleted due to missingness)
## Multiple R-squared:  0.2733, Adjusted R-squared:  0.2724
## F-statistic: 286 on 17 and 12927 DF, p-value: < 2.2e-16

# Linear Regression Model
# relationship between working hours and health condition
# (Response: selfRatedHealth, Predictor: averageHoursWorked)
health_lm <- lm(selfRatedHealth ~ averageHoursWorked, data = lifeSatisfactionData)
summary(health_lm)

##
## Call:
## lm(formula = selfRatedHealth ~ averageHoursWorked, data = lifeSatisfactionData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8085 -0.8085  0.1915  1.1915  1.2828
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                        2.9583      0.1933  15.306
## averageHoursWorked0.1 to 29.9 hours  0.7589      0.1943   3.905

```

```

## average_hours_worked30.0 to 40.0 hours    0.8502    0.1936    4.392
## average_hours_worked40.1 to 50.0 hours    0.7954    0.1948    4.084
## average_hours_worked50.1 hours and more    0.7932    0.1956    4.055
##                                     Pr(>|t|)
## (Intercept)                            < 2e-16 ***
## average_hours_worked0.1 to 29.9 hours    9.46e-05 ***
## average_hours_worked30.0 to 40.0 hours    1.13e-05 ***
## average_hours_worked40.1 to 50.0 hours    4.46e-05 ***
## average_hours_worked50.1 hours and more    5.04e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9469 on 13002 degrees of freedom
## Multiple R-squared:  0.00287,    Adjusted R-squared:  0.002563
## F-statistic: 9.357 on 4 and 13002 DF,  p-value: 1.507e-07

# Linear Regression Model (Predictor: physical and mental health ratings)
lm1 <- lm(satisfaction_score ~ as.factor(selfRatedHealth)
          + as.factor(selfRatedMentalHealth), data = life_satisfaction_data)
summary(lm1)

##
## Call:
## lm(formula = satisfaction_score ~ as.factor(selfRatedHealth) +
##     as.factor(selfRatedMentalHealth), data = life_satisfaction_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9825 -0.7522  0.0175  1.0175  6.0486
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        3.95144    0.12668   31.192 < 2e-16
## as.factor(selfRatedHealth)2         0.53787    0.10458    5.143 2.74e-07
## as.factor(selfRatedHealth)3         0.87842    0.09807    8.958 < 2e-16
## as.factor(selfRatedHealth)4         1.05311    0.09851   10.690 < 2e-16
## as.factor(selfRatedHealth)5         1.24952    0.10030   12.458 < 2e-16
## as.factor(selfRatedMentalHealth)2    1.65297    0.11731   14.091 < 2e-16
## as.factor(selfRatedMentalHealth)3    2.74768    0.10982   25.020 < 2e-16
## as.factor(selfRatedMentalHealth)4    3.30442    0.11022   29.979 < 2e-16
## as.factor(selfRatedMentalHealth)5    3.78157    0.11096   34.081 < 2e-16
##
## (Intercept)                        ***
## as.factor(selfRatedHealth)2         ***
## as.factor(selfRatedHealth)3         ***
## as.factor(selfRatedHealth)4         ***
## as.factor(selfRatedHealth)5         ***
## as.factor(selfRatedMentalHealth)2    ***
## as.factor(selfRatedMentalHealth)3    ***
## as.factor(selfRatedMentalHealth)4    ***
## as.factor(selfRatedMentalHealth)5    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.31 on 12936 degrees of freedom

```

```
## (62 observations deleted due to missingness)
## Multiple R-squared: 0.2595, Adjusted R-squared: 0.259
## F-statistic: 566.6 on 8 and 12936 DF, p-value: < 2.2e-16

# Linear Regression Model (Predictor: income of the family)
lm2 <- lm(satisfaction_score ~ family_income, data = life_satisfaction_data)
summary(lm2)

##
## Call:
## lm(formula = satisfaction_score ~ family_income, data = life_satisfaction_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.4145 -0.7925 -0.0706  0.9294  2.4080
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.31129    0.03692  225.100 < 2e-16 ***
## family_income$125,000 and more  0.10318    0.04409   2.340 0.019280 *
## family_income$25,000 to $49,999 -0.51883    0.04922 -10.540 < 2e-16 ***
## family_income$50,000 to $74,999 -0.24067    0.04879  -4.933 8.21e-07 ***
## family_income$75,000 to $99,999 -0.17738    0.04945  -3.587 0.000336 ***
## family_incomeLess than $25,000 -0.71931    0.05968 -12.053 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.499 on 12939 degrees of freedom
## (62 observations deleted due to missingness)
## Multiple R-squared: 0.03044, Adjusted R-squared: 0.03006
## F-statistic: 81.23 on 5 and 12939 DF, p-value: < 2.2e-16

# Linear Regression Model (Predictor: )
lm3 <- lm(satisfaction_score ~ average_hours_worked, data = life_satisfaction_data)
summary(lm3)

##
## Call:
## lm(formula = satisfaction_score ~ average_hours_worked, data = life_satisfaction_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2835 -1.0683 -0.1167  0.8833  2.9130
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)      7.0870    0.3170  22.354
## average_hours_worked0.1 to 29.9 hours  0.9813    0.3187   3.080
## average_hours_worked30.0 to 40.0 hours  1.0298    0.3175   3.244
## average_hours_worked40.1 to 50.0 hours  1.1269    0.3194   3.528
## average_hours_worked50.1 hours and more  1.1966    0.3207   3.731
##              Pr(>|t|)
## (Intercept)      < 2e-16 ***
## average_hours_worked0.1 to 29.9 hours  0.002078 **
## average_hours_worked30.0 to 40.0 hours  0.001183 **
```

```

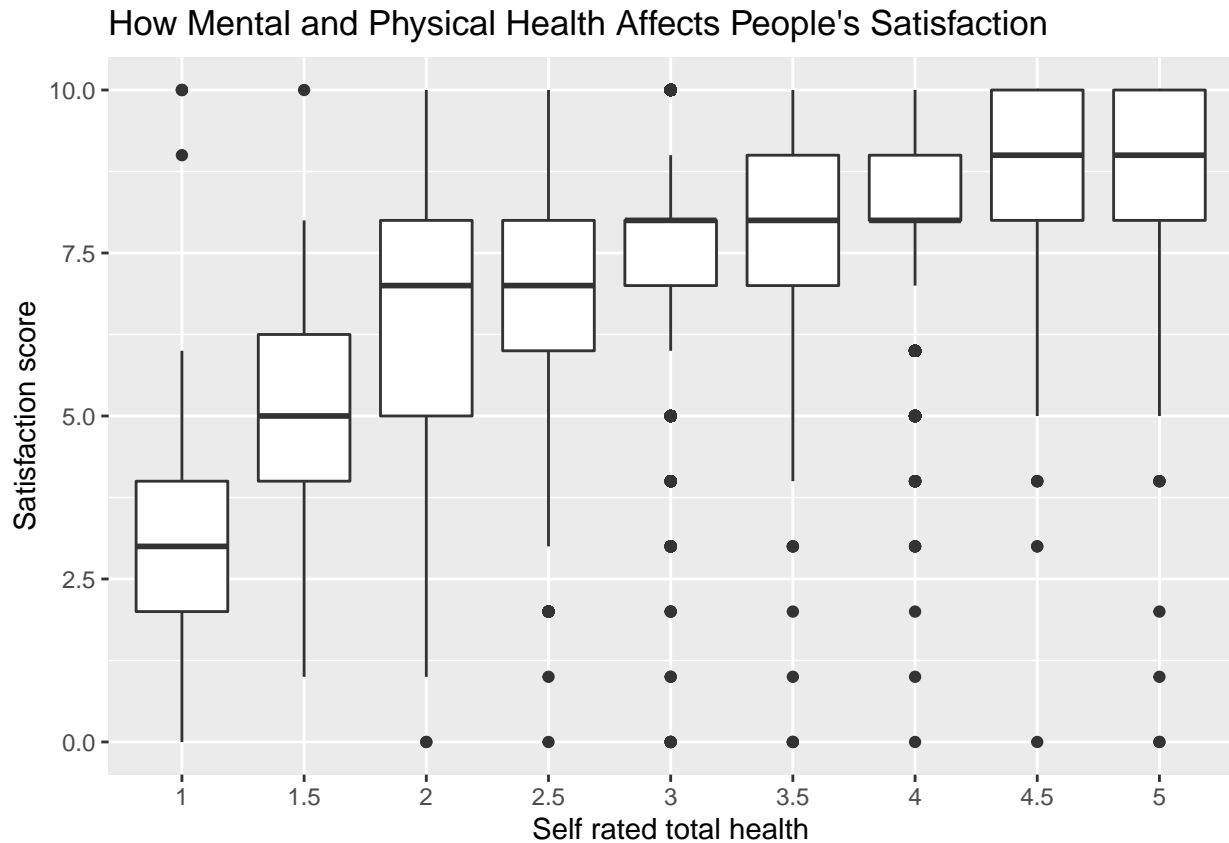
## average_hours_worked40.1 to 50.0 hours 0.000420 ***
## average_hours_worked50.1 hours and more 0.000192 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.52 on 12940 degrees of freedom
## (62 observations deleted due to missingness)
## Multiple R-squared:  0.002301, Adjusted R-squared:  0.001993
## F-statistic: 7.462 on 4 and 12940 DF, p-value: 5.333e-06
immigrant_data <- data %>% select(caseid, age, sex, feelings_life,
                                age_immigration, landed_immigrant, citizenship_status) %>%
  filter(landed_immigrant != "Don't know", citizenship_status != "Don't know")
table(immigrant_data$citizenship_status)

##
##           By birth By naturalization
##           83           2916
# Logit Regression Model
# lm4 <- lm(feelings_life ~ as.factor(landed_immigrant), data = immigrant_data, family = "binomial")
# summary(lm4)

life_satisfaction_data %>%
  ggplot(aes(x = as.factor(self Rated_total_health), y = (satisfaction_score))) + geom_boxplot() +
  ggtitle("How Mental and Physical Health Affects People's Satisfaction") +
  ylab("Satisfaction score") +
  xlab("Self rated total health")

## Warning: Removed 62 rows containing non-finite values (stat_boxplot).

```



Results

Here you will include all results. This includes descriptive statistics, graphs, figures, tables, and model results. Please ensure that everything is well formatted and in a report style. You must also provide an explanation of the results in this section. You can overflow to an Appendix if needed.

Please ensure that everything is well labelled. So if you have multiple histograms and plots, calling them Figure 1, 2, 3, etc. and referencing them as Figure 1, Figure 2, etc. in your report will be expected. The reader should not get lost in a sea of information. Make sure to have the results be clean, well formatted and digestible.

Discussion

Here you will discuss conclusions drawn from the results and comment on how it relates to the original goal of the study (which was specified in the Introduction).

Weaknesses

Here we discuss weaknesses of the study, data, analysis, etc. You can also discuss areas for improvement.

Next Steps

Here you discuss subsequent work to be done after this report. This can include next steps in terms of statistical analysis (perhaps there is a more efficient algorithm available, or perhaps there is a caveat in the data that would allow for some new technique). Future steps should also be specified in terms of the study setting (eg. including a follow-up survey on something, or a subsequent study that would complement the conclusions of your report).

References

1. GSS Data:
2. Data Cleaning Code: