

What determines the happiness of the Canadians? - a GSS 2017 Analysis

Names of your Group Members

Oct.19 2020

#Abstract

I used the 2017 GSS data. I am interested in the attribute called feelings_life in the survey data. I performed linear regression on the GSS data to find what variables affect one's feelings_life the most. And we find that wealth may not be that important to our happiness while health, family are more important in one's feelings of life.

#Introduction

The 2017 GSS data, conducted from February 2nd to November 30th 2017, is a sample survey with cross-sectional design. The target population includes all non-institutionalized persons 15 years of age and older, living in the 10 provinces of Canada.¹ I am interested in what contributes to the happiness of the Canadian people. Normally, we would think one's wealth or health may affect one's happiness. It is interesting to see what the data would tell us.

There is an attribute called feelings_life in the survey data. I see feelings of life as a person's happiness about his/her current life. Therefore I use feelings_life as my measure of a Canadian's happiness. I chose to perform a linear regression using feelings_life as the response variable and some other independent explanatory variables from the data.

#Data

The data I use is from GSS 2017. I downloaded the data from CHASS and cleaned them in R. The cleaned data has 81 variables. Except from the caseid, the identification variable, the cleaned data still has 80 variables.

I chose the GSS data because it contains enough demographic and personal data to perform analysis on one's happiness. Both the number of variables in the dataset and the number of respondents are large. Also, all the variables are easy to factor. We can easily apply regression models on these data.

However, many variables have too many NA values which may bring problems when we are doing the regression.

#Model

I chose to perform linear regression here since the response variable, feelings_life, is a scalar variable. Also, because there are more than one dependent variables, I used multiple regression instead of simple linear regression.

Since many variables have many NA values, I chose to first manually exclude those variables with too many NA values. Then I also excluded those variables that are too similar to another variable, and only keep one of them to avoid the collinearity problem. Thus, for the selected model, I chose age, total_children, sex, province, marital_status, education, own_rent, living_arrangement, selfRated_health, selfRated_mental_health, religion_has_affiliation, language_home, language_knowledge and income_respondent as the dependent variables. And feelings_life as the response variable. The mathematical notation is as follow:

$$feelingsLife = \beta_0 + \beta_1 age + \beta_2 totalChildren + \beta_3 sex + \beta_4 province + \beta_5 maritalStatus + \beta_6 education + \beta_7 ownRent + \beta_8 livingArrangement + \beta_9 selfRatedHealth + \beta_{10} selfRatedMentalHealth + \beta_{11} religionHasAffiliation + \beta_{12} languageHome + \beta_{13} languageKnowledge + \beta_{14} incomeRSespondent + \epsilon$$

where β is the coefficient of each variable, ϵ is the error term of the regression model, $feelingsLife$ is the true response where all the other variables are inputted from the sample.

I will use basic `lm` in R to perform this linear regression. I can not use `svyglm` since the population size N is unknown.

#Results

```
##
## Call:
## lm(formula = feelings_life ~ age + total_children + sex + province +
##      marital_status + education + own_rent + living_arrangement +
##      selfRatedHealth + selfRatedMentalHealth + religionHasAffiliation +
##      language_home + language_knowledge + income_respondent, data = data_gss)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3735 -0.7280  0.0870  0.9027  5.9857
##
## Coefficients:
##                                     Estimate
## (Intercept)                        6.8223697
## age                                0.0047756
## total_children                      0.0554699
## sexMale                           -0.1450347
## provinceBritish Columbia           0.0524975
## provinceManitoba                   0.0467648
## provinceNew Brunswick              0.2152939
## provinceNewfoundland and Labrador  0.1084957
## provinceNova Scotia                0.1424743
## provinceOntario                    0.0314903
## provincePrince Edward Island       0.1620071
## provinceQuebec                     -0.1140211
## provinceSaskatchewan               0.1670624
## marital_statusLiving common-law    0.1437891
## marital_statusMarried              0.2397163
## marital_statusSeparated            -0.2582725
## marital_statusSingle, never married 0.0691302
## marital_statusWidowed              0.0855406
## educationCollege, CEGEP or other non-university certificate or di... 0.0613975
## educationHigh school diploma or a high school equivalency certificate 0.1064444
## educationLess than high school diploma or its equivalent            0.3837654
## educationTrade certificate or diploma                               0.1129268
## educationUniversity certificate or diploma below the bachelor's level 0.0763078
## educationUniversity certificate, diploma or degree above the bach... -0.0098744
## own_rentOwned by you or a member of this household, even if it i... -0.2451715
## own_rentRented, even if no cash rent is paid                       -0.3244737
## living_arrangementLiving with one parent                          0.1326762
## living_arrangementLiving with two parents                        0.3168967
## living_arrangementNo spouse and non-single child(ren)           -0.1490905
## living_arrangementNo spouse and single child 25 years of age or older -0.0543665
## living_arrangementNo spouse and single child under 25 years of age -0.0674381
```

## living_arrangementOther living arrangement	0.1221476
## living_arrangementSpouse and non-single child(ren)	0.6526850
## living_arrangementSpouse and other	0.4246504
## living_arrangementSpouse and single child 25 years of age or older	0.0819071
## living_arrangementSpouse and single child under 25 years of age	0.2735570
## living_arrangementSpouse only	0.3621576
## self_rated_healthExcellent	0.9098238
## self_rated_healthFair	0.0901172
## self_rated_healthGood	0.4867632
## self_rated_healthPoor	-0.7279379
## self_rated_healthVery good	0.6959559
## self_rated_mental_healthExcellent	0.9117653
## self_rated_mental_healthFair	-1.0510524
## self_rated_mental_healthGood	-0.0432350
## self_rated_mental_healthPoor	-2.2486826
## self_rated_mental_healthVery good	0.4538622
## religion_has_affiliationHas religious affiliation	-0.0396088
## religion_has_affiliationNo religious affiliation	-0.1099705
## language_homeEnglish and French	-0.0748300
## language_homeEnglish and non-official language	0.0072994
## language_homeEnglish, French and non-official language	0.3823914
## language_homeFrench	0.0391257
## language_homeFrench and non-official language	-0.0586515
## language_homeMultiple non-official languages	-0.2515755
## language_homeNon-official languages	-0.0405059
## language_knowledgeDon't know	-2.1691619
## language_knowledgeEnglish only	-0.0068126
## language_knowledgeFrench only	0.0602461
## language_knowledgeNeither English nor French	-0.6702884
## income_respondent\$125,000 and more	0.0653333
## income_respondent\$25,000 to \$49,999	-0.0554597
## income_respondent\$50,000 to \$74,999	-0.0294838
## income_respondent\$75,000 to \$99,999	0.0376076
## income_respondentLess than \$25,000	-0.0869505
##	Std. Error
## (Intercept)	0.3884386
## age	0.0009187
## total_children	0.0084755
## sexMale	0.0209626
## provinceBritish Columbia	0.0440490
## provinceManitoba	0.0531653
## provinceNew Brunswick	0.0530169
## provinceNewfoundland and Labrador	0.0547438
## provinceNova Scotia	0.0507852
## provinceOntario	0.0388544
## provincePrince Edward Island	0.0633014
## provinceQuebec	0.0617112
## provinceSaskatchewan	0.0536683
## marital_statusLiving common-law	0.0853976
## marital_statusMarried	0.0802502
## marital_statusSeparated	0.0657174
## marital_statusSingle, never married	0.0443257
## marital_statusWidowed	0.0479065
## educationCollege, CEGEP or other non-university certificate or di...	0.0312020

## educationHigh school diploma or a high school equivalency certificate	0.0317711
## educationLess than high school diploma or its equivalent	0.0381495
## educationTrade certificate or diploma	0.0438978
## educationUniversity certificate or diploma below the bachelor's level	0.0566089
## educationUniversity certificate, diploma or degree above the bach...	0.0401846
## own_rentOwned by you or a member of this household, even if it i...	0.1898892
## own_rentRented, even if no cash rent is paid	0.1906013
## living_arrangementLiving with one parent	0.0697803
## living_arrangementLiving with two parents	0.0601678
## living_arrangementNo spouse and non-single child(ren)	0.3314175
## living_arrangementNo spouse and single child 25 years of age or older	0.0997535
## living_arrangementNo spouse and single child under 25 years of age	0.0572040
## living_arrangementOther living arrangement	0.0501760
## living_arrangementSpouse and non-single child(ren)	0.4373111
## living_arrangementSpouse and other	0.1234492
## living_arrangementSpouse and single child 25 years of age or older	0.1043497
## living_arrangementSpouse and single child under 25 years of age	0.0789449
## living_arrangementSpouse only	0.0764790
## self_rated_healthExcellent	0.2148929
## self_rated_healthFair	0.2154307
## self_rated_healthGood	0.2139675
## self_rated_healthPoor	0.2191854
## self_rated_healthVery good	0.2141527
## self_rated_mental_healthExcellent	0.2286211
## self_rated_mental_healthFair	0.2309593
## self_rated_mental_healthGood	0.2282276
## self_rated_mental_healthPoor	0.2413539
## self_rated_mental_healthVery good	0.2283528
## religion_has_affiliationHas religious affiliation	0.1174362
## religion_has_affiliationNo religious affiliation	0.1187418
## language_homeEnglish and French	0.0728371
## language_homeEnglish and non-official language	0.0428520
## language_homeEnglish, French and non-official language	0.2442469
## language_homeFrench	0.0571401
## language_homeFrench and non-official language	0.1650538
## language_homeMultiple non-official languages	0.2393893
## language_homeNon-official languages	0.0494959
## language_knowledgeDon't know	1.3626606
## language_knowledgeEnglish only	0.0389013
## language_knowledgeFrench only	0.0465791
## language_knowledgeNeither English nor French	0.2718504
## income_respondent\$125,000 and more	0.0671989
## income_respondent\$25,000 to \$49,999	0.0524688
## income_respondent\$50,000 to \$74,999	0.0532108
## income_respondent\$75,000 to \$99,999	0.0569679
## income_respondentLess than \$25,000	0.0536258
##	t value
## (Intercept)	17.564
## age	5.198
## total_children	6.545
## sexMale	-6.919
## provinceBritish Columbia	1.192
## provinceManitoba	0.880
## provinceNew Brunswick	4.061

## provinceNewfoundland and Labrador	1.982
## provinceNova Scotia	2.805
## provinceOntario	0.810
## provincePrince Edward Island	2.559
## provinceQuebec	-1.848
## provinceSaskatchewan	3.113
## marital_statusLiving common-law	1.684
## marital_statusMarried	2.987
## marital_statusSeparated	-3.930
## marital_statusSingle, never married	1.560
## marital_statusWidowed	1.786
## educationCollege, CEGEP or other non-university certificate or di...	1.968
## educationHigh school diploma or a high school equivalency certificate	3.350
## educationLess than high school diploma or its equivalent	10.060
## educationTrade certificate or diploma	2.572
## educationUniversity certificate or diploma below the bachelor's level	1.348
## educationUniversity certificate, diploma or degree above the bach...	-0.246
## own_rentOwned by you or a member of this household, even if it i...	-1.291
## own_rentRented, even if no cash rent is paid	-1.702
## living_arrangementLiving with one parent	1.901
## living_arrangementLiving with two parents	5.267
## living_arrangementNo spouse and non-single child(ren)	-0.450
## living_arrangementNo spouse and single child 25 years of age or older	-0.545
## living_arrangementNo spouse and single child under 25 years of age	-1.179
## living_arrangementOther living arrangement	2.434
## living_arrangementSpouse and non-single child(ren)	1.492
## living_arrangementSpouse and other	3.440
## living_arrangementSpouse and single child 25 years of age or older	0.785
## living_arrangementSpouse and single child under 25 years of age	3.465
## living_arrangementSpouse only	4.735
## self_rated_healthExcellent	4.234
## self_rated_healthFair	0.418
## self_rated_healthGood	2.275
## self_rated_healthPoor	-3.321
## self_rated_healthVery good	3.250
## self_rated_mental_healthExcellent	3.988
## self_rated_mental_healthFair	-4.551
## self_rated_mental_healthGood	-0.189
## self_rated_mental_healthPoor	-9.317
## self_rated_mental_healthVery good	1.988
## religion_has_affiliationHas religious affiliation	-0.337
## religion_has_affiliationNo religious affiliation	-0.926
## language_homeEnglish and French	-1.027
## language_homeEnglish and non-official language	0.170
## language_homeEnglish, French and non-official language	1.566
## language_homeFrench	0.685
## language_homeFrench and non-official language	-0.355
## language_homeMultiple non-official languages	-1.051
## language_homeNon-official languages	-0.818
## language_knowledgeDon't know	-1.592
## language_knowledgeEnglish only	-0.175
## language_knowledgeFrench only	1.293
## language_knowledgeNeither English nor French	-2.466
## income_respondent\$125,000 and more	0.972

## income_respondent\$25,000 to \$49,999	-1.057
## income_respondent\$50,000 to \$74,999	-0.554
## income_respondent\$75,000 to \$99,999	0.660
## income_respondentLess than \$25,000	-1.621
##	Pr(> t)
## (Intercept)	< 2e-16
## age	2.03e-07
## total_children	6.11e-11
## sexMale	4.70e-12
## provinceBritish Columbia	0.233355
## provinceManitoba	0.379082
## provinceNew Brunswick	4.91e-05
## provinceNewfoundland and Labrador	0.047507
## provinceNova Scotia	0.005030
## provinceOntario	0.417680
## provincePrince Edward Island	0.010496
## provinceQuebec	0.064667
## provinceSaskatchewan	0.001855
## marital_statusLiving common-law	0.092244
## marital_statusMarried	0.002820
## marital_statusSeparated	8.52e-05
## marital_statusSingle, never married	0.118872
## marital_statusWidowed	0.074184
## educationCollege, CEGEP or other non-university certificate or di...	0.049112
## educationHigh school diploma or a high school equivalency certificate	0.000809
## educationLess than high school diploma or its equivalent	< 2e-16
## educationTrade certificate or diploma	0.010104
## educationUniversity certificate or diploma below the bachelor's level	0.177680
## educationUniversity certificate, diploma or degree above the bach...	0.805896
## own_rentOwned by you or a member of this household, even if it i...	0.196674
## own_rentRented, even if no cash rent is paid	0.088702
## living_arrangementLiving with one parent	0.057272
## living_arrangementLiving with two parents	1.40e-07
## living_arrangementNo spouse and non-single child(ren)	0.652818
## living_arrangementNo spouse and single child 25 years of age or older	0.585754
## living_arrangementNo spouse and single child under 25 years of age	0.238450
## living_arrangementOther living arrangement	0.014926
## living_arrangementSpouse and non-single child(ren)	0.135585
## living_arrangementSpouse and other	0.000583
## living_arrangementSpouse and single child 25 years of age or older	0.432505
## living_arrangementSpouse and single child under 25 years of age	0.000531
## living_arrangementSpouse only	2.20e-06
## self_rated_healthExcellent	2.31e-05
## self_rated_healthFair	0.675724
## self_rated_healthGood	0.022920
## self_rated_healthPoor	0.000898
## self_rated_healthVery good	0.001157
## self_rated_mental_healthExcellent	6.68e-05
## self_rated_mental_healthFair	5.38e-06
## self_rated_mental_healthGood	0.849752
## self_rated_mental_healthPoor	< 2e-16
## self_rated_mental_healthVery good	0.046876
## religion_has_affiliationHas religious affiliation	0.735910
## religion_has_affiliationNo religious affiliation	0.354389

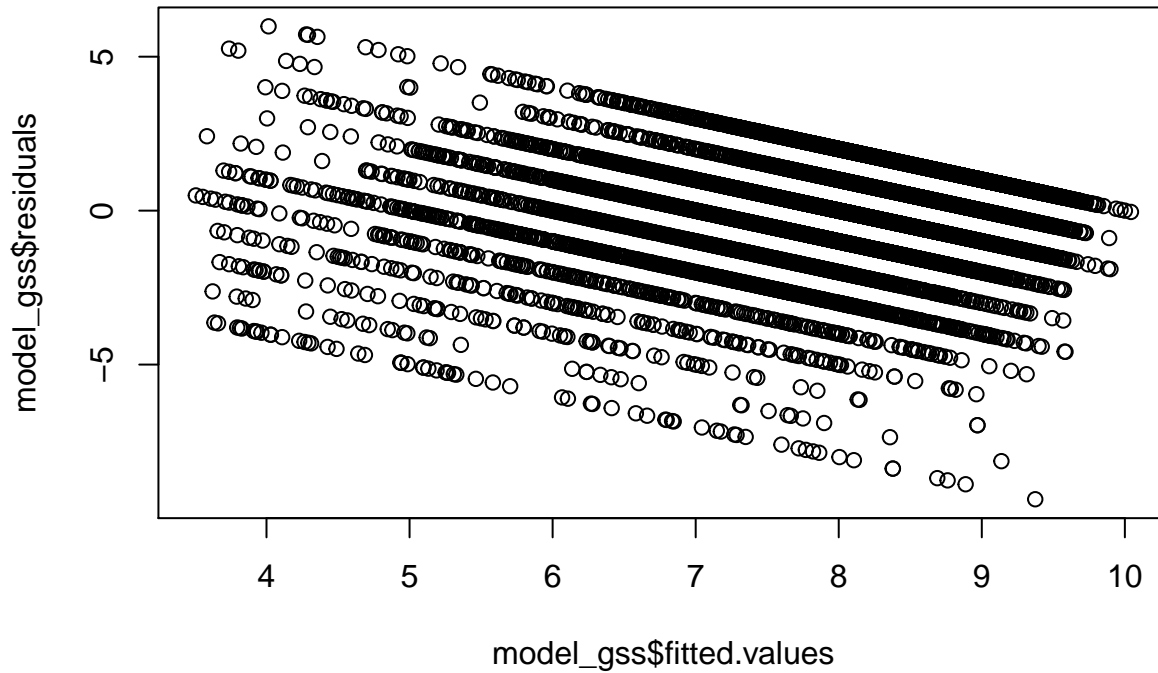
## language_homeEnglish and French	0.304263
## language_homeEnglish and non-official language	0.864745
## language_homeEnglish, French and non-official language	0.117460
## language_homeFrench	0.493520
## language_homeFrench and non-official language	0.722333
## language_homeMultiple non-official languages	0.293315
## language_homeNon-official languages	0.413157
## language_knowledgeDon't know	0.111433
## language_knowledgeEnglish only	0.860983
## language_knowledgeFrench only	0.195883
## language_knowledgeNeither English nor French	0.013685
## income_respondent\$125,000 and more	0.330945
## income_respondent\$25,000 to \$49,999	0.290523
## income_respondent\$50,000 to \$74,999	0.579521
## income_respondent\$75,000 to \$99,999	0.509163
## income_respondentLess than \$25,000	0.104942
##	
## (Intercept)	***
## age	***
## total_children	***
## sexMale	***
## provinceBritish Columbia	
## provinceManitoba	
## provinceNew Brunswick	***
## provinceNewfoundland and Labrador	*
## provinceNova Scotia	**
## provinceOntario	
## provincePrince Edward Island	*
## provinceQuebec	.
## provinceSaskatchewan	**
## marital_statusLiving common-law	.
## marital_statusMarried	**
## marital_statusSeparated	***
## marital_statusSingle, never married	
## marital_statusWidowed	.
## educationCollege, CEGEP or other non-university certificate or di...	*
## educationHigh school diploma or a high school equivalency certificate	***
## educationLess than high school diploma or its equivalent	***
## educationTrade certificate or diploma	*
## educationUniversity certificate or diploma below the bachelor's level	
## educationUniversity certificate, diploma or degree above the bach...	
## own_rentOwned by you or a member of this household, even if it i...	
## own_rentRented, even if no cash rent is paid	.
## living_arrangementLiving with one parent	.
## living_arrangementLiving with two parents	***
## living_arrangementNo spouse and non-single child(ren)	
## living_arrangementNo spouse and single child 25 years of age or older	
## living_arrangementNo spouse and single child under 25 years of age	
## living_arrangementOther living arrangement	*
## living_arrangementSpouse and non-single child(ren)	
## living_arrangementSpouse and other	***
## living_arrangementSpouse and single child 25 years of age or older	
## living_arrangementSpouse and single child under 25 years of age	***
## living_arrangementSpouse only	***

```

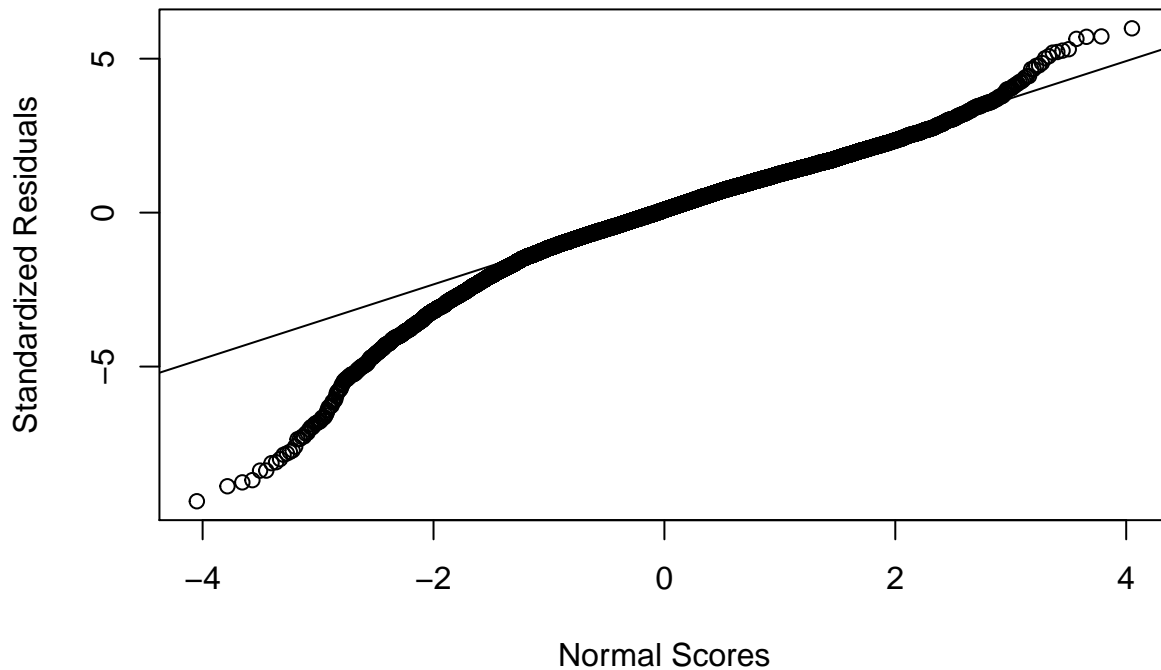
## selfRated_healthExcellent ***
## selfRated_healthFair
## selfRated_healthGood *
## selfRated_healthPoor ***
## selfRated_healthVery good **
## selfRated_mental_healthExcellent ***
## selfRated_mental_healthFair ***
## selfRated_mental_healthGood
## selfRated_mental_healthPoor ***
## selfRated_mental_healthVery good *
## religion_has_affiliationHas religious affiliation
## religion_has_affiliationNo religious affiliation
## language_homeEnglish and French
## language_homeEnglish and non-official language
## language_homeEnglish, French and non-official language
## language_homeFrench
## language_homeFrench and non-official language
## language_homeMultiple non-official languages
## language_homeNon-official languages
## language_knowledgeDon't know
## language_knowledgeEnglish only
## language_knowledgeFrench only
## language_knowledgeNeither English nor French *
## income_respondent$125,000 and more
## income_respondent$25,000 to $49,999
## income_respondent$50,000 to $74,999
## income_respondent$75,000 to $99,999
## income_respondentLess than $25,000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.361 on 19461 degrees of freedom
## (1076 observations deleted due to missingness)
## Multiple R-squared:  0.3091, Adjusted R-squared:  0.3068
## F-statistic: 136 on 64 and 19461 DF, p-value: < 2.2e-16

```


1. Residual plot



2. Normal Q-Q



From the output of the summary, we can see the following results:

1. From the F-test, we have $p\text{-value} < 2.2e-16 < 0.05$, which implies we have strong evidence to reject the null hypothesis and concluded that there must be at least one variable with coefficient 0. This means at least one variable is truly contributed to feelings_life.
2. From individual t-test for each variable, if the p-value is less than 0.05, we have strong evidence to reject

the null hypothesis and concluded that this variable is significant and should be included in the model to predict one's happiness. If the p-value is larger than 0.05, we would accept the null hypothesis and conclude that this variable may not be significant and should not be included in the model of predicting one's happiness. Thus, we have age, total_children, sexMale, province, marital_status, education, living_arrangement, selfRated_health, selfRated_mental_health and language_knowledge should be significant variables with p-value > 0.05.

3. From the coefficient of each variable, if it is positive, then there is a positive relationship between the variable and feelings_life, vice versa.
4. Note that for some variables, only certain levels of the variable are significant. We need to include those variables, but keep in mind that only those levels are significant. For other levels, we may not have strong evidence.

However, as we can see from the plot 1 and 2, there may exist problems for this model:

1. Plot 1 is a residual plot where there are certain patterns exist
2. Plot 2 is a normal q-q plot where the qqline is not very straight.

The above plots may imply the data may not satisfy the assumptions of the linear regression.

#Discussion

From the results, we have age, total_children, sexMale, province, marital_status, education, living_arrangement, selfRated_health, selfRated_mental_health and language_knowledge to be the significant variables while own_rent, religion_has_affiliation and income_respondent are less important.

In the introduction, I made an assumption that wealth and health should contribute the most to one's happiness. However, from the result, we know that income_respondent are less important. The income of the respondent is not very important for his/her happiness. Thus, wealth may not play a key role in one's happiness. Similarly, own_rent, which is whether the respondent own the place living or not, is less important. Own_rent is part of wealth, and the above also implies that wealth may not be the key to one's happiness.

religion_has_affiliation is whether one has religion or not. The less importance of religion_has_affiliation shows that anyone could have happiness regardless of religion.

For those important variables, I found all of them reasonable. For example, both health and mental health should be important for one's feelings of life. A bad health or mental health can bring pain.

There are some significant variables worth mentioning. The sexMale has negative coefficient which implies that man may be less happy than woman in Canada. This maybe because the social pressure is more heavy on male than female. Furthermore, we could see that language_knowledgeNeither English nor French is significant and with the coefficient negative. We can learn that people in Canada with no knowledge of English or French are not happy. Canada's official languages are English and French. People who do not know those languages, but live in Canada may experience difficulties in their lives, and result in unhappiness.

#Weaknesses

As we can see from the plots in the result, the data may not satisfy the assumptions of the linear regression. We need to find a way to improve the model to meet the assumptions.

Also, the model contain many non-significant variables. We need to find a way to clean the variables for a better model. We could perform model selection, and drop non-significant variables from the linear regression. We could then use AIC/BIC to select the best model for predicting one's happiness.

Lastly, those variables in the data may not be enough to cover all that influence one's happiness. Many other variables could be considered.

#Next Steps

Based on the weakness, we first need to make adjustments or improvements accordingly.

Then, we could do a follow-up survey after we suggesting people what we think is important for a Canadian's happiness to see whether there is truly an improvement in people's happiness.

Also, we need to keep adding more useful variables to the study to better understand happiness.

#References

Bibliography:

1. Statistics Canada. "General Social Survey, Cycle 31: Families, Public Use Microdata File Documentation and User's Guide" 2020.