1 Introduction

The World Happiness Report is a representative survey that measures the level of global happiness. The first survey was carried in 2012, the next was in 2013, 2015, 2016 and so on annually. It ranked 155 countries by their happiness levels in beginning, but now, it ranks 156 countries. It seems like it is not a contributing survey for the world, but in fact, it is. The report was initially introduced at the United States at a celebration event of International Day of Happiness, and it continuously gains the interest of public organizations and governments of the world since it contributes a lot on informing the governments' policy making decisions.

The Happiness Report that this project refers uses the data from the Gallup World Poll, American Analytic and advisory company based in Washington, D.C. Gallup interviews approximately a thousand residents from each country. The target population is the entire civilian that is non-institutionalized, aged at least 15. Gallup asks the questions in the language that each interviewee prefers to minimize the chance of misunderstanding the interview question, and to produce statistically comparable results. Gallup interviews through phone call if the country of the respondents has telephone coverage of at least 80% and face-to-face interview if the country has less than 80% of telephone coverage.

The Happiness scores are based on the respondents' answers to the major life evaluation questions asked in the poll. The question, which is known as Cantril ladder, asks interviewees to think of a ladder with the most satisfying life for them being a 10 and the worst possible life being a 0 and to rate their own on that range. There are 8 categories that are considered in calculating the happiness scores: Dystopia, residuals, GDP per capita, social support, life expectancy, freedom to make life choices, generosity, perceptions of corruption.

Dystopia is an imaginary nation with the lowest happiness score. This index is established by setting up the lower threshold of the happiness score. It is set to have lowest generosity, lowest income level, lowest life expectancy, most corrupted, least freedom of making life choices and least life life expectancy, which makes it the most unpleasant country in the world. The value of it is fixed to 1.85 out of 10.

Residuals are also referred as unexplained components that affect the happiness score of each country and they differ for each country. The residuals reflect the six variables: GDP per capita, social support, life expectancy, freedom to make life choices, generosity, perceptions of corruption, either over- or under-evaluate the mean of 2014-2016 life evaluations. They are set to be approximately zero over the whole set of countries.

GDP per capita is also called as Gross Domestic Product per capita and it represents the personal opinion of each respondent about his or her country's

average income of each person in the country earns annually.

Social support represents the interviewees' personal ratings about their countries' level of family support that the citizens received from their own family.

Life expectancy is the expected number of years that people can live in that country.

Freedom to make life choices represents how freely the people in the country can make choices for their future.

Generosity represents degree of how much that the individuals of the countries donate and volunteer to help the people in danger(i.e. Population in poverty)

Perception of corruption represents level of public organizations' corruption in each country.

Simply, the Happiness score measures the index of the citizens opinion on what their life is rated out of their best possible life. However, each person has different threshold for his/her best possible life and has different weight for each variable. For example, Person A may think Healthy life expectancy is more important than social support to have better life, while person B thinks the other way around. Therefore, it is crucial to find out which variable the majority of people in each country takes as the most significant factor that makes their life better, and this project uses The World Happiness Report 2019 to explore which variable is the most effective factor to determine the country's Happiness Score.

2 Analysis

In the World Happiness Report 2019, we have 156 countries,one response variable and six explanatory variables, the data size is relatively large. There is some missing data in Social support and Generosity, and in order to make the data more complete and more understandable, Social support and Generosity are been converted to categorical variables.

As mentioned in introduction, Social support represents the interviewees' personal ratings about their countries' level of family support that the citizens received from their own family. The rating range is 0-2, assume any value among 0 and the mean of Social support rating means the interviewee received weak support, any value among the mean of Social support rating and 2 means the interviewee received strong support. Missing values will be treated as 0, no support received ever.

```
> Happiness$Social.support <- cut(Happiness$Social.support,c(0,mean(Happiness$Social.support),2),labels = c("0","1"))
> Happiness$Social.support
  [1] 1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1   
 [27] 1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1   
 [53] 1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    0    1    1    1    1   
 [79] 1    0    1    0    1    1    0    1    1    0    0    0    1    0    0    1    1    0    1    0    0    1    1    0    0    0   
[105] 0    1    0    1    0    1    0    0    1    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    1   
[131] 0    0    1    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    <NA> 0   
Levels: 0 1
> Happiness$Social.support[155] <- c("0")
> Happiness$Social.support
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
 [68] 1 1 1 1 1 1 0 1 1 1 1 1 0 1 0 1 1 0 1 1 0 0 0 1 0 0 1 1 0 1 0 0 1 1 0 0 0 0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0
[135] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Levels: 0 1
```

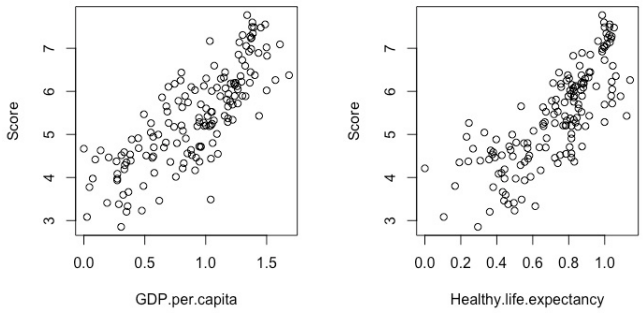
Similarly, Generosity represents the degree of how much that the individuals of the countries donate and volunteer to help the people in danger. The rating range is also 0-2, assume any value among 0 and the mean of Generosity rating means the interviewee rarely donate or volunteer, any value among the mean of Generosity rating and 2 means the interviewee is very active in donating and volunteering. Missing values will be treated as 0, never donating or volunteering.

```
> Happiness$Generosity <- cut(Happiness$Generosity,c(0,mean(Happiness$Generosity),2),labels = c("0","1"))
> Happiness$Generosity
  [1] 0    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    0    1    0    1    1    0    0    1    1   
 [27] 0    0    1    0    0    0    0    1    0    0    1    0    1    0    1    0    0    0    1    1    0    0    1    0    0    1   
 [53] 0    0    0    0    1    0    1    0    0    0    0    1    0    0    1    0    0    0    0    0    0    1    0    1    0    1   
 [79] 0    1    0    <NA> 1    1    1    1    1    0    0    0    0    1    0    0    1    1    0    1    0    1    0    0    0    0   
[105] 1    0    0    0    1    0    0    1    0    1    0    0    1    1    0    1    1    0    1    0    0    0    1    0    1    1   
[131] 1    1    1    1    0    1    0    1    0    1    1    1    0    0    0    0    1    0    1    1    0    1    1    0    1    1   
Levels: 0 1
> Happiness$Generosity[82] <- c("0")
> Happiness$Generosity
  [1] 0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 0 1 1 0 0 1 1 0 0 1 0 0 0 0 1 0 0 1 0 1 0 1 0 0 0 1 1 0 0 1 0 0 1 0 0 0 0 1 0 1 0 0 0 0 1 0 0 1
 [68] 0 0 0 0 0 0 1 0 1 0 1 0 1 0 0 1 1 1 1 1 0 0 0 0 1 0 0 1 1 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0 1 0 1 0 0 1 1 0 1 1 0 1 0 0 0 1 0 1 1 1 1 1 1
[135] 0 1 0 1 0 1 1 1 0 0 0 0 1 0 1 1 0 1 1 0 1 1
Levels: 0 1
```

Now the data is much better now. To get a general idea of the relationships between the response variable Score and each of the explanatory variables. We use scatter plot to visualize the relationship between two continuous variables. When one variable is categorical and the other variable is continuous, box-plot is an appropriate way to visualize the relationship.

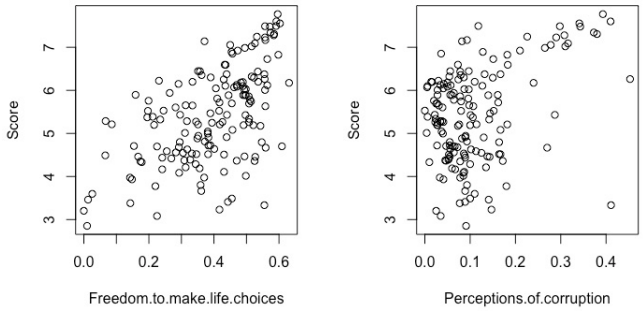
Scatterplots:

```
> plot(Score~GDP.per.capita, data = Happiness)
> plot(Score~Healthy.life.expectancy, data = Happiness)
> plot(Score~Freedom.to.make.life.choices, data = Happiness)
> plot(Score~Perceptions.of.corruption, data = Happiness)
```



There is a positive linear relationship between Happiness Score and GDP.per.capita. That is as GDP.per.capita increases, the Happiness Score increases.

There is a positive linear relationship between Happiness Score and Healthy.life,expectancy. As Healthy.life,expectancy increases, the Happiness Score increases.

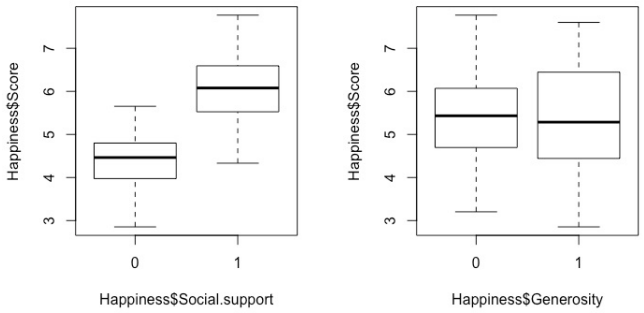


There is a positive linear relationship between Happiness Score and Freedom.to.make.life.choices. As Freedom.to.make.life.choices increases, the Happiness Score increases.

There is a positive non-linear relationship between Happiness Score and Perceptions.of.corruption. As Perceptions.of.corruption increases, the Happiness Score increases.

Boxplots:

```
> Happiness$Social.support <- as.factor(Happiness$Social.support)
> boxplot(Happiness$Score~Happiness$Generosity)
> Happiness$Generosity <- as.factor(Happiness$Generosity)
> boxplot(Happiness$Score~Happiness$Social.support)
```



In the box-plot of Score and Social support, the IQR for "0" (received weak or no support) is smaller than "1" (received strong support) levels, the median, maximum and minimum Score of "0" group are all lower than "1" group, the variation is similar for two groups.

In the box-plot of Score and Generosity, the IQR for"0" group (rarely donate or volunteer) is smaller than "1" group (often donate or volunteer) , the median, maximum and minimum Score of "0" group are all higher than "1" group, the variation is similar for two groups.

Now the general idea of how the response variable is related with the explanatory variables are clear, variable selection will be performed to find the best model of each model size and compare $AdjR^2$, C_p and AIC for each model of each size to get the best model among all the model size. Therefore, some non-significant variables will be ruled out from this step. Then according to the change in R^2 for the last variable added to the best model, find the variable that produces the largest R^2 increase when it is the last variable added to the model.

Best model/ Variable selection has three methods: Exhaustive search, Forward selection and Backward selection. Since the data size is too large here, Exhaustive search cannot work in this case. As the fact that when all variables in full model are significant, the Backward selection won't start.

The full model is:

$$\begin{aligned}\text{Score} = \beta_0 + \beta_1 \times \text{GDP.per.capita} + \beta_2 \times \text{Healthy.life.expectancy} \\ + \beta_3 \times \text{Freedom.to.make.life.choices} + \beta_4 \times \text{Perceptions.of.corruption} \\ + \beta_5 \times \text{Social.support} + \beta_6 \times \text{Generosity}\end{aligned}$$

Full model:

```
> #Backward
> full_model <- lm(Score~GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support+Generosit
y,data = Happiness)
> summary(full_model)

Call:
lm(formula = Score ~ GDP.per.capita + Healthy.life.expectancy +
    Freedom.to.make.life.choices + Perceptions.of.corruption +
    Social.support + Generosity, data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7203 -0.3062  0.0528  0.3709  1.1355

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                    2.77539    0.17416  15.936  < 2e-16 ***
GDP.per.capita                 0.91003    0.20494   4.441 1.74e-05 ***
Healthy.life.expectancy        0.83804    0.34779   2.410   0.0172 *
Freedom.to.make.life.choices   1.63773    0.36494   4.488 1.43e-05 ***
Perceptions.of.corruption      1.16171    0.54277   2.140   0.0340 *
Social.support1                0.67074    0.13632   4.920 2.26e-06 ***
Generosity1                    0.05125    0.09328   0.549   0.5836
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5318 on 149 degrees of freedom
Multiple R-squared:  0.7806,    Adjusted R-squared:  0.7717
F-statistic: 88.34 on 6 and 149 DF,  p-value: < 2.2e-16
```

The p-value of Generosity is $0.5836 > 0.05$, so Generosity is not significant. Not all the variables are significant, Backward selection can be applied here.

Backward Selection:

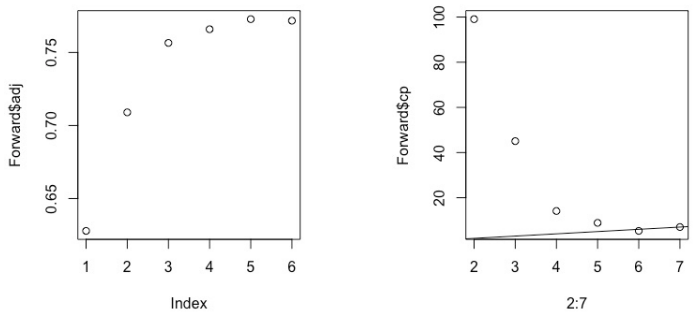
```
> Back <- regsubsets(Score~GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support+Generos
ity, data=Happiness, method="backward")
> Backward <- summary(Back)
> Backward$which
  (Intercept) GDP.per.capita Healthy.life.expectancy Freedom.to.make.life.choices Perceptions.of.corruption Social.support1 Generosity1
1        TRUE           TRUE                  FALSE                         FALSE                      FALSE           FALSE       FALSE
2        TRUE           TRUE                  FALSE                          TRUE                      FALSE           FALSE       FALSE
3        TRUE           TRUE                  FALSE                          TRUE                      FALSE            TRUE       FALSE
4        TRUE           TRUE                   TRUE                          TRUE                      FALSE            TRUE       FALSE
5        TRUE           TRUE                   TRUE                          TRUE                       TRUE            TRUE       FALSE
6        TRUE           TRUE                   TRUE                          TRUE                       TRUE            TRUE        TRUE
> Backward$adj
[1] 0.6278490 0.7089994 0.7565378 0.7658966 0.7728019 0.7717394
> Backward$cp
[1] 99.078186 45.053820 14.122879  8.865176  5.301825  7.000000
```

Similarly, perform Forward selection.

Forward Selection:

```
> #Forward
> For <- regsubsets(Score ~GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support+Generos
ity, data=Happiness, method="forward")
> Forward <- summary(For)
> Forward$which
  (Intercept) GDP.per.capita Healthy.life.expectancy Freedom.to.make.life.choices Perceptions.of.corruption Social.support1 Generosity1
1        TRUE           TRUE                  FALSE                         FALSE                      FALSE           FALSE       FALSE
2        TRUE           TRUE                  FALSE                          TRUE                      FALSE           FALSE       FALSE
3        TRUE           TRUE                  FALSE                          TRUE                      FALSE            TRUE       FALSE
4        TRUE           TRUE                   TRUE                          TRUE                      FALSE            TRUE       FALSE
5        TRUE           TRUE                   TRUE                          TRUE                       TRUE            TRUE       FALSE
6        TRUE           TRUE                   TRUE                          TRUE                       TRUE            TRUE        TRUE
> Forward$adjr2
[1] 0.6278490 0.7089994 0.7565378 0.7658966 0.7728019 0.7717394
> plot(Forward$adj)
> abline(a=0,b=1)
```

```
> Forward$cp
[1] 99.078186 45.053820 14.122879  8.865176  5.301825  7.000000
> plot(2:7,Forward$cp)
> abline(a=0,b=1)
```




```
> #AIC
> model_5 <- lm(Score ~GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support, data=Happi
ness)
> AIC(model_5)
[1] 252.8449
>
> model_6 <- lm(Score ~GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support+Generosity,
 data=Happiness)
> AIC(model_6)
[1] 254.5292
```

In Backward selection and Forward selection, the value of $AdjR^2$ and C_p for the best models of each model size are the same.$model5$ has the biggest $adjR^2$=0.7728019 and smallest C_p=5.301825, p=6,AIC=252.8449. While $model_6$ has $adjR^2$=0.7717394 which is similar to $model5$'s $adjR^2$, its C_p=7 is the second smallest and p=7,AIC=254.5292. Using 2-fold cross validation can test the prediction power of a model, smaller prediction error indicates "better fit" of the model.

2-fold cross validation:

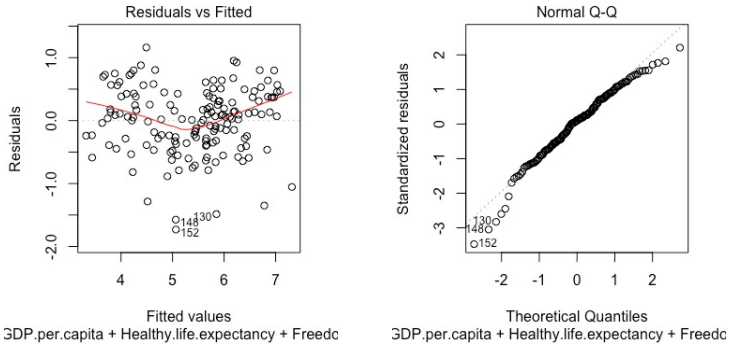
```
> #2-fold cross validation
> train <- 1:as.integer(dim(Happiness)[1]/2)
>
> reg5_1 <- lm(Score ~ GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support, data=Happi
ness[train,])
> error5_1 <- sum((Happiness$Score[-train] - predict(reg5_1, Happiness[-train,]))^2)
> reg5_2 <- lm(Score ~ GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support, data=Happi
ness[-train,])
> error5_2 <- sum((Happiness$Score[train] - predict(reg5_2, Happiness[train,]))^2)
> error5 <- (error5_1 + error5_2)/dim(Happiness)[1]
> error5
[1] 1.316205
>
> reg6_1 <- lm(Score ~ GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support+Generosity,
 data=Happiness[train,])
> error6_1 <- sum((Happiness$Score[-train] - predict(reg6_1, Happiness[-train,]))^2)
> reg6_2 <- lm(Score ~ GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support+Generosity,
 data=Happiness[-train,])
> error6_2 <- sum((Happiness$Score[train] - predict(reg6_2, Happiness[train,]))^2)
> error6 <- (error6_1 + error6_2)/dim(Happiness)[1]
> error6
[1] 1.31413
```

Since R^2 will increase as the model get more complicated, a better model should have a larger $adjR^2$, smaller C_p which should be close to p(the number of

betas) a smaller AIC and a smaller prediction error. The difference between the prediction error of the two model is 1.316205-1.31413 = 0.002075 which is really small. $Model5$ has the biggest $adjR^2$ and its C_p is way smaller than $model6$ so $model5$ is the best model.

```
> model_5 <- lm(Score ~GDP.per.capita+Healthy.life.expectancy+Freedom.to.make.life.choices+Perceptions.of.corruption+Social.support, data=Happi
ness)
> summary(model_5)

Call:
lm(formula = Score ~ GDP.per.capita + Healthy.life.expectancy +
    Freedom.to.make.life.choices + Perceptions.of.corruption +
    Social.support, data = Happiness)

Residuals:
     Min       1Q   Median       3Q      Max
-1.73163 -0.32373  0.05885  0.37932  1.16133

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                     2.7908     0.1715  16.274  < 2e-16 ***
GDP.per.capita                  0.8984     0.2034   4.418 1.90e-05 ***
Healthy.life.expectancy         0.8332     0.3469   2.402   0.0175 *
Freedom.to.make.life.choices    1.6793     0.3562   4.714 5.49e-06 ***
Perceptions.of.corruption       1.2378     0.5236   2.364   0.0194 *
Social.support1                 0.6684     0.1359   4.917 2.28e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5306 on 150 degrees of freedom
Multiple R-squared:  0.7801,    Adjusted R-squared:  0.7728
F-statistic: 106.4 on 5 and 150 DF,  p-value: < 2.2e-16

> plot(model_5)
```

Model5: Score = 2.7908+0.8984× GDP.per.capita+0.8332 × Healthy.life.expectancy +1.6793 × Freedom.to.make.life.choices+1.2378 × Perceptions.of.corruption +0.6684 × Social.support



The Residual vs fitted graph looks reasonably random, however the fitted trend suggests the the mean of the residuals is greater than 0. This could suggest a non random distribution since in general the actual data appears to be greater than the fitted data. This could potentially be explained as a consequence of excluding the generosity explanatory variable, or as a result of making the social support variable categorical. The Normal Q-Q looks reasonable except the ends taper off in a non linear fashion. The linearity of the points between -1.5 and 2 suggests that those data points are normally distributed, but outside

of that range there are outliers / extremes. These extreme data points could explained by the countries with NA values in Social Support which is missing a key influencer in predicting the Happiness score. This could contribute to the extreme standardized residual in the otherwise linear Normal Q-Q plot.

change in R-squared:

When an independent variable is the last one be fitted into the model, the change in R-squared represents how the goodness-of-fit is improved by the last variable. It illustrates the percentage of the variance the last variable explains while the other variables cannot explain. And the variable with the largest change in R-squared should be the most important variable. Removing one explanatory variable from model 5 and refit the model calculate the difference between the original model 5 R-square and the R-square of the refitted model. Repeat this step until all the variables have been removed once.

```
> OGmodel <- lm(Score~GDP.per.capita + Social.support + Healthy.life.expectancy + Freedom.to.make.life.choices  + Perceptions.of.corruption, da
ta = Happiness) # all explanatory variables
> OGrSquared <- summary(OGmodel)$r.squared
> OGrSquared
[1] 0.7801308
>
> modelGDP <- lm(Score~  Social.support + Healthy.life.expectancy + Freedom.to.make.life.choices + Perceptions.of.corruption, data = Happiness)
 # model without GDP per capita
> GDPrSquared <- summary(modelGDP)$r.squared
> OGrSquared-GDPrSquared
[1] 0.0286061
>
>
> modelSocial <- lm(Score~  GDP.per.capita + Healthy.life.expectancy + Freedom.to.make.life.choices + Perceptions.of.corruption, data = Happine
ss) # model without Social support
> SocialrSquared <- summary(modelSocial)$r.squared
> OGrSquared-SocialrSquared
[1] 0.03543642
>
> modelLifeExp <- lm(Score~ GDP.per.capita + Social.support + Freedom.to.make.life.choices  + Perceptions.of.corruption, data = Happiness) #mod
el without life expectancy
> LifeExpRsquared <- summary(modelLifeExp)$r.squared
> OGrSquared-LifeExpRsquared
[1] 0.008457447
>
> modelFreedom <- lm(Score~ GDP.per.capita + Social.support + Healthy.life.expectancy   + Perceptions.of.corruption, data = Happiness) #model w
ithout Freedom to make life choices variable
> FreedomRsquared <- summary(modelFreedom)$r.squared
> OGrSquared-FreedomRsquared
[1] 0.03257847
>
> modelCorruption <- lm(Score~ GDP.per.capita + Social.support + Healthy.life.expectancy + Freedom.to.make.life.choices , data = Happiness)
> CorruptionRsquared <- summary(modelCorruption)$r.squared
> OGrSquared-CorruptionRsquared
[1] 0.008192863
```

Obviously, the model without Social support has the biggest difference 0.03543642. Therefore Social support is the most important variable in this study.

3 Conclusion

Our research question was, "which variable in the 2019 World Happiness Report has the most impact in estimating a country's Happiness Score". In order to answer this research question we compared Adjusted R^2 values, Mallow's Cp values, and AIC (Akaike information criterion) of different models to find which variable would have the greatest impact on maximizing the idealized values of these criterion. Using Backwards, forwards, and two-Fold selection we where able to find that the most accurate 5 variable model was a model excluding generosity. Then we used an approach of fitting $k-1$ explanatory variables to the responding variable of happiness score, comparing the R^2 value to the original model R^2 value with k explanatory variable (which in our context $k = 5$ because we exclude the generosity variable as it was shown to be least important). Using this approach, we found that excluding the variable social support resulted in the largest change in R^2 from the original model. We can then suggest using the results from the Backwards, forwards, and two-Fold selection, and R^2 rankings that social support has the largest impact on happiness score.

However, certain assumptions and limitations should be noted. One assumption we made was to consider linear regression without interaction terms. We made this assumption because the description of the 2019 World happiness report stated, "If you add all these factors up, you get the happiness score", in which the "factors" refer to the explanatory variables. Another assumption we made was in treating multicollinearity. Multicollinearity, or the linearly dependency of explanatory variables was considered to be potentially a problem, so we converted two of the variables, social support and generosity, to categorical variables. We did this under the assumption that if there was multicollinearity between the variables, it would lessened by making social support and generosity non continuous variables. Besides assumptions we also had limitations in our analysis. One limitation was having to deal with NA values in social support and generosity, which was another reason for making these variables categorical. Another limitation was only working with 2019 data. As a result, we are limited in generalizing our findings to any other year. However despite these assumptions and limitations, we were able to conclude that social support is the most important variable for estimating the Happiness score in 2019. These finding could potentially be useful in raising happiness of nations if countries focused on improving social support of their citizens. However, there is much more work to be done, such as analyzing the data sets of previous years, and seeing if social support is the most important variable in previous years as well.

Bibliography

FAQ, World Happiness Report, worldhappiness.report/faq/.

Home, World Happiness Report, worldhappiness.report/.

Kaminer, Alex. Generosity = Happiness, Gross National Happiness USA(GNHUSA), 14 Aug. 2019, gnhusa.org/serious-about-happiness/generosity-happiness/.

McFadden, Robert D. "Louis Harris, Pollster at Forefront of American Trends, Dies at 95." The New York Times, The New York Times, 19 Dec. 2016, www.nytimes.com/2016/12/19/us/louis-harris-pollster-at-forefront-of-american-trends-dies-at-95.html.

Network, Sustainable Development Solutions. "World Happiness Report." Kaggle, 27 Nov. 2019, www.kaggle.com/unsdsn/world-happiness.