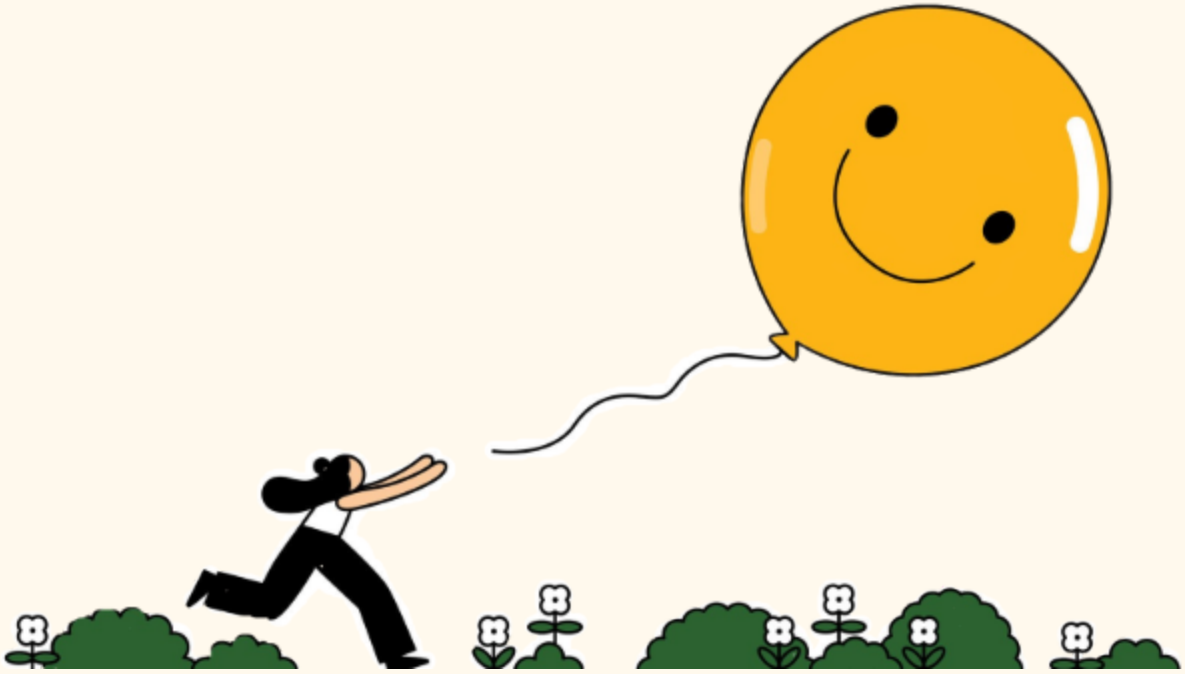
STATS 410
Fall 2020
Dr. Olga Korosteleva

Happiness Index
Regression Analysis

—

By Noah Gallagher



INTRODUCTION

For hundreds of years, researchers have been fascinated with the idea of happiness and have tried to come up with ways to measure happiness. World happiness is the sought after by many people but there are a number of factors that can get in the way of happiness. For this study, we will focus on a few factors that can contribute to a person or a country's happiness and see how these factors affect a total happiness score. A happiness score is a formula used to estimate the amount of happiness an individual or region has. This topic is very interesting as we as researchers may be able to determine what contributes to happiness through regression analysis. Overall, happiness is hard to generally calculate as it is a mental or emotional state, yet regression analysis allows us to discover new insights that can affect the lives of people and make a difference in the world.

BACKGROUND

The World Happiness Report is a survey of the state of global happiness based on several factors such as Economy, Family, Health, Freedom and Trust. The World Happiness 2019 Survey ranks 155 countries by their happiness score. This was released by the United Nations as an event celebrating International Day of Happiness on March 20th. In order to provide reference points, the top five happiest countries in the world as of 2019 are Finland, Denmark, Norway, Iceland and Netherlands. The bottom five happiest countries in the world as of 2019 are Rwanda, Tanzania, Afghanistan, Central African Republic and South Sudan.

DATA DESCRIPTION

Data was taken from a survey given to populations of people living in over 155 countries in the world. The happiness score is based on answers to the main life evaluation questions. Answers to these questions are based on a rating scale from zero to ten, with 0 representing the least it contributes to happiness and ten representing the most it contributes. The purpose of this study is to determine what factors contribute the most to a person or country's overall happiness. It is also interesting to analyze what factors have the most significant impact on the happiness score.

RESULTS

Looking at our outputs from both SAS and R, we can see that the data resembles a bell shaped curve and has a corresponding p-value that is greater than 0.05, thus the data is normally distributed. We are also given the parameter estimates with significant predictors of Economy, Family, Health and Freedom. By conducting a deviance test we are given a p-value of zero, which concludes that the model fits the data very well.

CONCLUSION

From this study, we can see that the most important factor that contributes to a person's happiness is Freedom. The more freedom a country has, the more likely we can expect to find happy citizens living within the country. The second most important factor in happiness is Family. The more family a person has will directly contribute to a person's overall happiness.

Happiness can be different for everyone and people can find happiness in many other factors different than the results from the regression analysis. As this study only considers six factors that affect happiness, it is important to consider the factors in this study but also to consider what else a person may find happiness within.

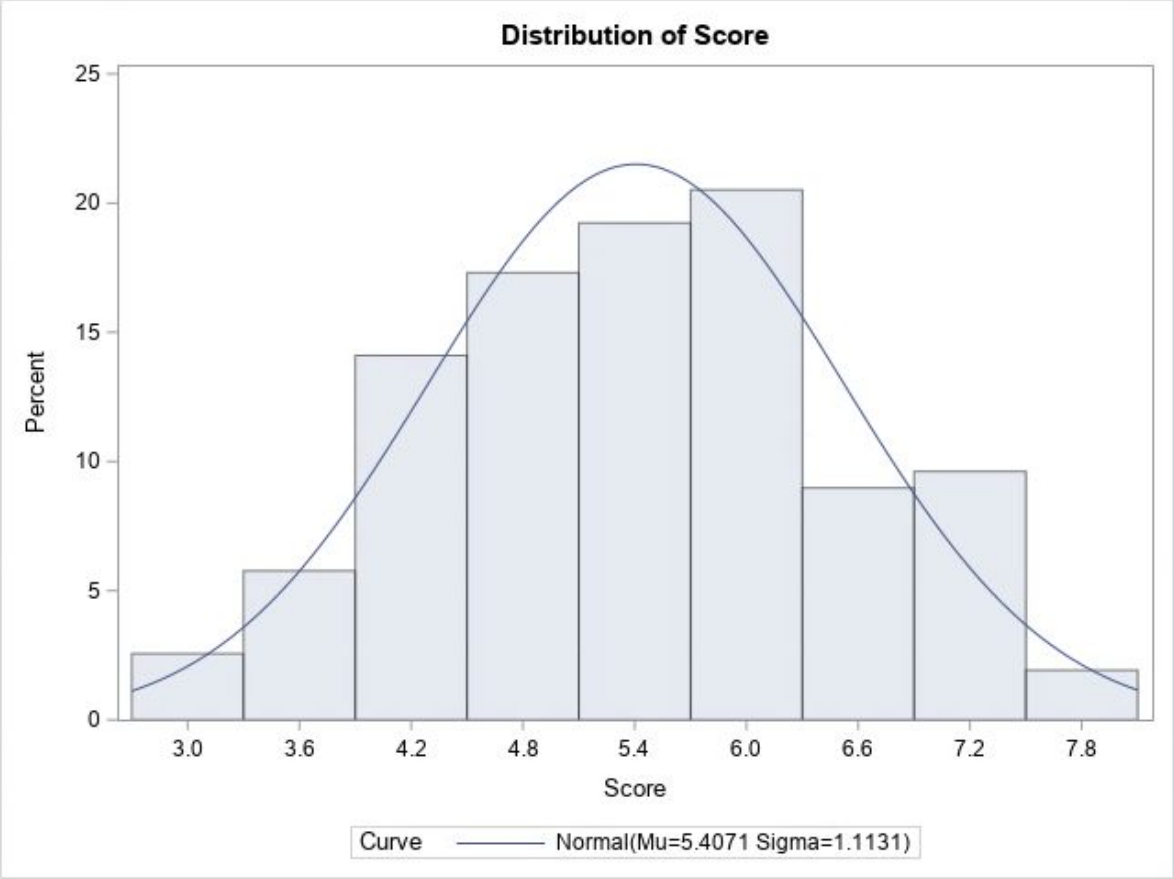
REFERENCE

Network, S. (2019, November 27). World Happiness Report. Retrieved November 30, 2020, from https://www.kaggle.com/unsdsn/world-happiness

APPENDIX A: SAS Code

```
/* Reading in Data and Renaming Variables */
proc import out=happy_data (rename= GDP_per_capita=economy
rename=Social_support=family rename=Healthy_life_expectancy=health
rename=Freedom_to_make_life_choices=freedom rename=Perceptions_of_corruption=trust)
datafile="C:\Users\017035809\Desktop\2019.csv"
dbms=csv replace;
run;

/* Normality Check */
proc univariate;
 var Score;
  histogram/normal;
run;

/*Fitting then general linear model*/
proc genmod;
 model Score = economy family health freedom trust generosity / dist=normal
link=identity;
run;

/* Intercept Only Model */
proc genmod;
 model Score = / dist=normal link=identity;
run;

/* Deviance Test */
data deviance_test;
 deviance=-2*(-237.5708-(-119.7648));
 p=1-probchi(deviance,6);
run;
proc print noobs;
run;
```

APPENDIX B: SAS Output



Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.05780215	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.06192554	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.42309674	Pr > A-Sq	>0.250

deviance	pvalue
235.612	0

Log Likelihood		-119.7648
Full Log Likelihood		-119.7648

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.7952	0.2063	1.3909	2.1995	75.74	<.0001
economy	1	0.7754	0.2133	0.3574	1.1934	13.22	0.0003
family	1	1.1242	0.2315	0.6704	1.5780	23.58	<.0001
health	1	1.0781	0.3269	0.4373	1.7189	10.87	0.0010
freedom	1	1.4548	0.3668	0.7359	2.1738	15.73	<.0001
trust	1	0.9723	0.5301	-0.0666	2.0112	3.36	0.0666
Generosity	1	0.4898	0.4864	-0.4636	1.4432	1.01	0.3140
Scale	1	0.5214	0.0295	0.4666	0.5826		

APPENDIX C: R Code

```
#reading in data
happiness.data<-read.csv(file="/Users/noahgallagher1/Documents/FALL 2020/STAT
410/Project Data/2019.csv", header=TRUE, sep=",")

#plotting histogram with fitted normal dist
library(rcompanion)
plotNormalHistogram(happiness.data$Score)

#normality check
shapiro.test(happiness.data$Score)

#Renaming the variables
Economy <- happiness.data$GDP.per.capita
Family <- happiness.data$Social.support
Health <- happiness.data$Healthy.life.expectancy
Freedom <- happiness.data$Freedom.to.make.life.choices
Trust <- happiness.data$Perceptions.of.corruption
Generosity <- happiness.data$Generosity

#fitting general linear model
summary(fitted.model<- glm(Score~ Economy + Family + Health + Freedom + Trust +
Generosity,data = happiness.data, family=gaussian(link=identity)))

#outputting estimated sigma
sigma(fitted.model)

#checking model fit
null.model<- glm(Score ~ 1, data = happiness.data,family = gaussian(link=identity))
print(deviance<- -2*(logLik(null.model)-logLik(fitted.model)))
print(p.value<- pchisq(deviance, df=7, lower.tail=FALSE))
```

APPENDIX D: R Output

```
Call:
glm(formula = Score ~ Economy + Family + Health + Freedom + Trust +
    Generosity, family = gaussian(link = identity), data = happiness.data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.75304  -0.35306   0.05703   0.36695   1.19059

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.7952     0.2111   8.505 1.77e-14 ***
Economy       0.7754     0.2182   3.553 0.000510 ***
Family        1.1242     0.2369   4.745 4.83e-06 ***
Health        1.0781     0.3345   3.223 0.001560 **
Freedom       1.4548     0.3753   3.876 0.000159 ***
Trust         0.9723     0.5424   1.793 0.075053 .
Generosity    0.4898     0.4977   0.984 0.326709
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.2846424)

    Null deviance: 192.051  on 155  degrees of freedom
Residual deviance:  42.412  on 149  degrees of freedom
AIC: 255.53

Number of Fisher Scoring iterations: 2
```

```
> #outputting estimated sigma
> sigma(fitted.model)
[1] 0.5335189
>
> #checking model fit
> null.model<- glm(Score ~ 1, data = happiness.data,family = gaussian(link=identity))
> print(deviance<- -2*(logLik(null.model)-logLik(fitted.model)))
'log Lik.' 235.6121 (df=2)
> print(p.value<- pchisq(deviance, df=7, lower.tail=FALSE))
'log Lik.' 3.18462e-47 (df=2)
>
> #Prediction
> print(predict(fitted.model, data.frame(Economy=1.433, Family=1.457, Health=0.874, Freedom=0.454, Trust=0.128, Generosity=0.280)))
       1
6.408657
```

APPENDIX E: Predicted Results

1) SAS Code and Output

```
/* Prediction */
data prediction;
input economy family health freedom trust generosity;
cards;
1.433 1.457 0.874 0.454 0.128 0.280
;
```

```
/* Combining Data Sets */
data happy_data;
 set happy_data prediction;
run;
```

```
/* Computing Prediction */
proc genmod;
 model Score = economy family health freedom trust generosity / dist=normal link=identity;
 output out=outdata p=pscore;
run;
```

```
/* Printing Results */
proc print data=outdata(firstobs=157);
var pscore;
run;
```

Obs	pscore
157	6.40866

2) R Code and Output

```
> print(predict(fitted.model, data.frame(Economy=1.433, Family=1.457, Health=0.874, Freedom=0.454, Trust=0.128, Generosity=0.280)))
       1
6.408657
```