Life Expectancy

Jeremy Miranda, Jakin Kolars, Chris Faires

General Information

Research Questions:

1. What factors contribute most to increasing life expectancy?
2. Is life expectancy affected more by physical traits or environmental conditions?

Data Set Description:

The data comes from the The Global Health Observatory (GHO) data repository under the World Health Organization (WHO) that keeps track of the health status as well as many other related factors for all countries. From the original data set we took the variables country (193 total countries), year (years in data set range from 2000 to 2015), alcohol consumption (per capita consumption of alcohol in liters) life expectancy, percentage expenditure (expenditure on health as a percentage of GDP, per capita), BMI (average body mass index of entire country population), and population (population of the country). Since each country is in the data set 15 times we took out 14 of the years to focus on just 1 year, which will be 2013.

Independent Variables

Alcohol consumption: per capita on alcohol percentage in liters (mean = 4.166763 and standard deviation = 4.218792

Health Expenditure: this is the expenditure on health as a percentage of GDP, per capita (mean = 980.3671 and standard deviation = 2650.263)

BMI: average body mass index of the entire country population (mean = 40.08273 and standard deviation = 20.48858)

Population: population of the entire country (mean = 12513081 and standard deviation = 28071960)

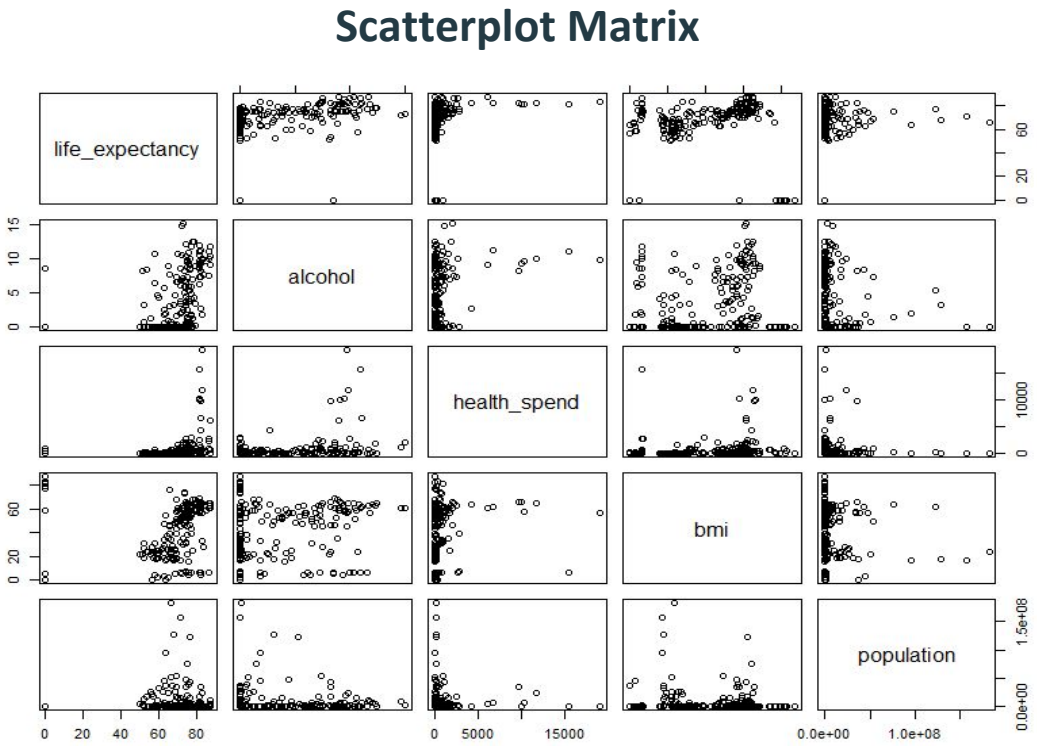
All of the independent variables that we used are quantitative.

Graph and Table Information

Correlation Matrix Table

	Life Expectancy	Alcohol	Health Spend	BMI	Population
Life Expectancy	1	0.5261328	0.38193291	0.51019294	-0.07678363
Alcohol	0.52613281	1	0.35745785	0.33909982	-0.13450972
Health Spend	0.38193291	0.3574579	1	0.18148669	-0.05056412
BMI	0.51019294	0.3390998	0.18148669	1	-0.08107548
Population	-0.07678363	-0.1345097	-0.05056412	-0.08107548	1

Graph and Table Information (continued)



Full Model Summary

Coefficients:

	Estimate	P-value
Intercept	6.141e+01	< 2e-16
Alcohol	6.805e-01	1.15e-05***
Health Spend	6.400e-04	0.00536
BMI	1.518e-01	6.50e-07***
Population	2.279e-09	0.90984

Confidence Intervals for slopes:

	Confidence Interval
Alcohol	3.851870e-01 and 9.759056e-01
Health Spend	1.928648e-04 and 1.087138e-03
BMI	9.433987e-02 and 2.092610e-01
Population	-3.744297e-08 and 4.200001e-08

Full Model Summary

Full Model:

	Adjusted R^2	F Test Statistic	P-value
Full Model	0.4182	25.8	7.319e-16***

Final Model

Description:

- The final model has alcohol, health expenditure, and BMI as the independent variables

Method:

- Using both forward selection and backward elimination, both methods came up with the same final model of having the three independent variables above

Final Model Summary

Final Model:

	Estimate	P-value
Intercept	6.145e+01	<2e-16
Alcohol	6.787e-01	1.01e-05
Health Spend	6.400e-04	.00519
BMI	1.57e-01	5.93e-07

	Adjusted R^2	F Test Statistic	P-value
Full Model	.4224	34.64	<2.2e-16

Check of Assumptions

Independence: Assumption is met because the data came from WHO, the data collected by should be representative of the populations

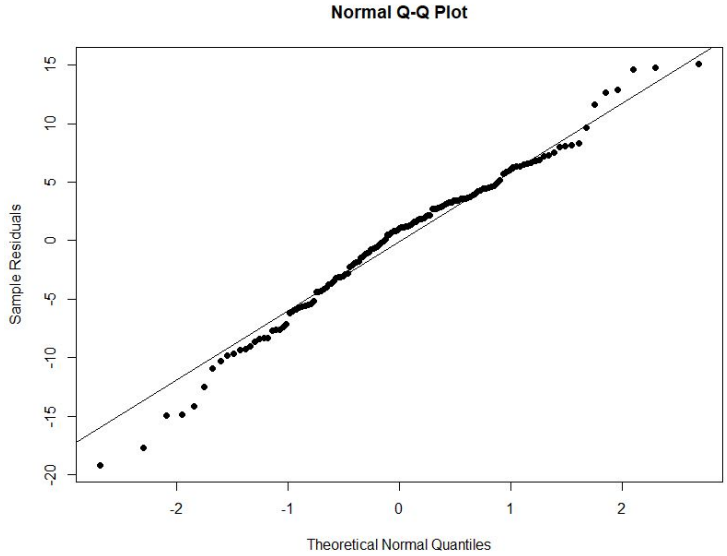
Normality: Assumption met, the points in the normal qq plot do follow a straight line, the Shapiro-Wilks test has a somewhat large p-value (0.08851), and the histogram of the residuals does represent a normal distribution

Linearity: Assumption met, no patterns appear in the residual plot

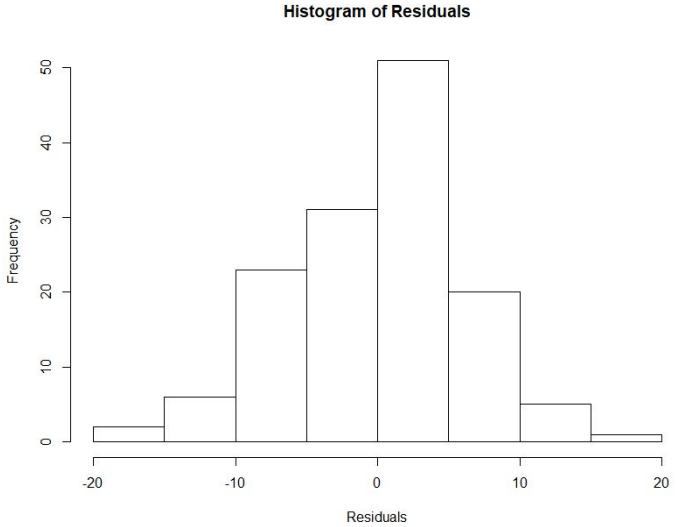
Constant Variance: Assumption met, the Levene test produced a somewhat large p-value, the standardized residual plot seems to be distributed evenly above and below the 0 line (0.06343)

Normality Assumption Graphs

Normality:

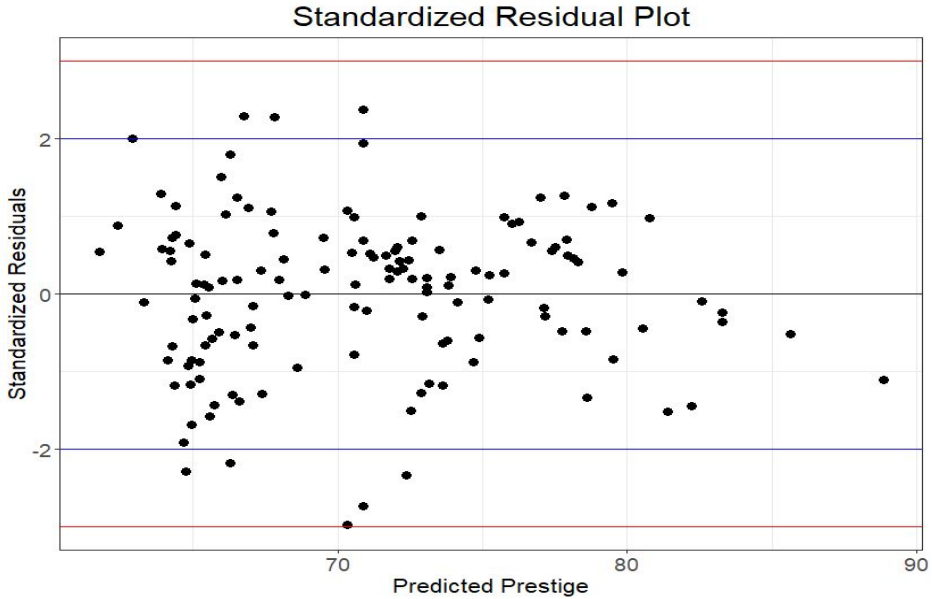


Normality:



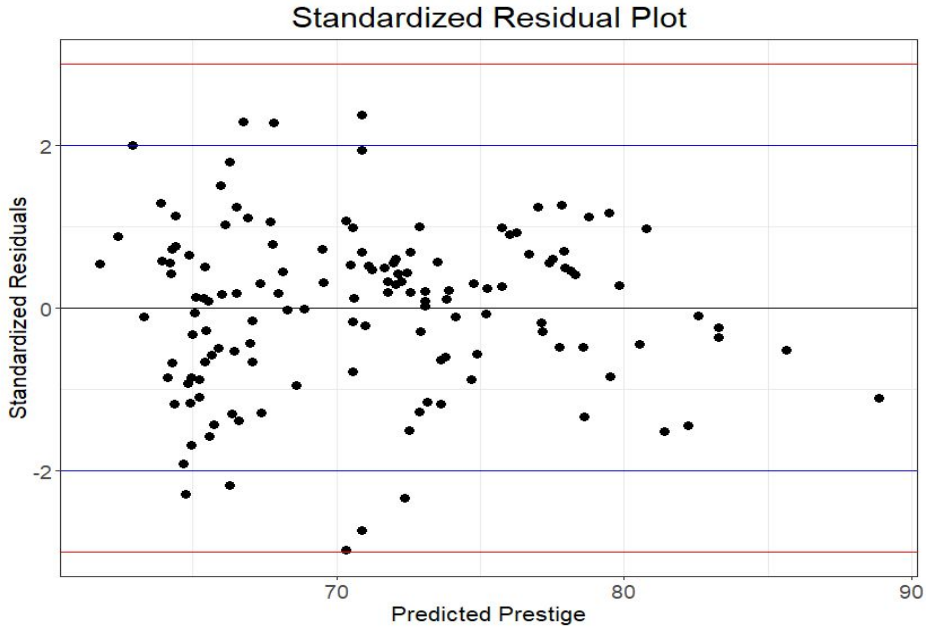
Linearity Assumption Graph

Linearity:



Constant Variance Assumption Graph

Constant Variance:



Residuals, leverage, and influential points

Residuals

- There are 8 total points that are somewhat unusual (no very unusual points)
  - 3 points are greater than 2
  - 5 points are less than -2

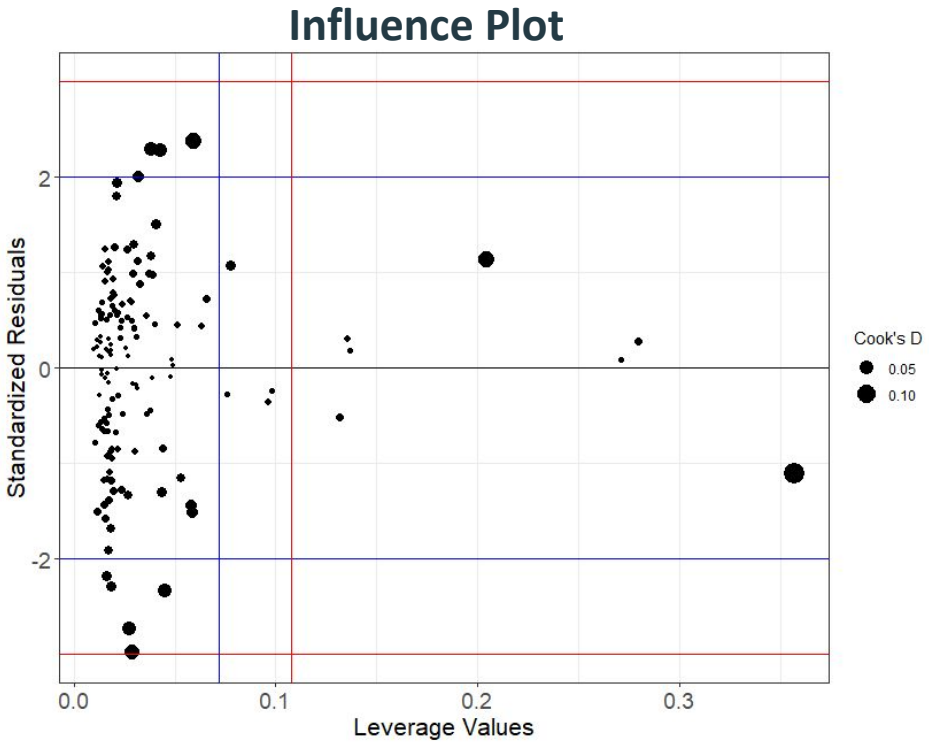
Leverage

- There are 10 total points above the low and high cutoff
  - 4 points are above the low cutoff (.0719)
  - 7 points are above the high cutoff (.1079)

Cook’s D

- There are no points considered to be influential (no points greater than .5 or 1)

Residuals, leverage, and influential points (continued)



Project Summary

1. **What factors contribute most to increasing life expectancy?**

   Based off the final model, the variables that were left in the model where alcohol, health expenditure, and BMI. Therefore we can conclude that alcohol, health expenditure, and BMI contribute more to life expectancy then the population of the country.

2. **Is life expectancy affected more by physical traits or environmental conditions?**

   We can conclude that physical traits play more of a role in life expectancy than environmental conditions. Traits that affect you physically such as alcohol consumption and BMI contribute more to life expectancy than traits you can’t control such as such as population and health expenditure.