



Analysis of Life Expectancy Globally



Faaiz N., Mercedes D., Rameen U.



Dataset Information

- Kaggle.com - Life Expectancy (WHO)
- 2938 observations from 193 countries spanning from 2000-2015
- 20 predictor variables

Response Variable

Life Expectancy

Predictor Variables

Alcohol, hepatitis B, measles, polio, HIV/AIDS, diphtheria, BMI, schooling, population, GDP, total expenditure, percentage expenditure, thinness under 5, etc



Scientific Question

What factors are statistically the most significant when calculating life expectancy globally?



Variable Selection

```
Call:
lm(formula = lifeData$Life.expectancy ~ Alcohol + Measles + Hepatitis.B +
  thinness.5.9.years + Schooling + HIV.AIDS + percentage.expenditure +
  BMI + Polio + Total.expenditure + Diphtheria + GDP + Population +
  thinness.1.19.years + thinness.5.9.years + Income.composition.of.resources,
  data = lifeData)
```

Residuals:

Min	1Q	Median	3Q	Max
-16.8183	-2.5836	0.1287	2.6106	13.1406

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.594e+01	7.306e-01	62.880	< 2e-16	***
Alcohol	-1.776e-01	3.351e-02	-5.299	1.33e-07	***
Measles	1.267e-05	1.065e-05	1.190	0.234112	
Hepatitis.B	-7.920e-03	4.992e-03	-1.587	0.112801	
thinness.5.9.years	-7.875e-02	5.862e-02	-1.343	0.179346	
Schooling	1.049e+00	6.613e-02	15.857	< 2e-16	***
HIV.AIDS	-6.022e-01	1.762e-02	-34.177	< 2e-16	***
percentage.expenditure	4.773e-04	2.034e-04	2.346	0.019078	*
BMI	4.499e-02	6.748e-03	6.667	3.56e-11	***
Polio	1.552e-02	5.803e-03	2.675	0.007549	**
Total.expenditure	9.787e-02	4.592e-02	2.131	0.033229	*
Diphtheria	2.196e-02	6.658e-03	3.298	0.000993	***
GDP	1.095e-05	3.200e-05	0.342	0.732282	
Population	-6.305e-11	1.577e-09	-0.040	0.968122	
thinness.1.19.years	-5.600e-04	5.964e-02	-0.009	0.992509	
Income.composition.of.resources	1.236e+01	9.283e-01	13.316	< 2e-16	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.056 on 1633 degrees of freedom
Multiple R-squared: 0.7894, Adjusted R-squared: 0.7874
F-statistic: 408 on 15 and 1633 DF, p-value: < 2.2e-16

Reduced Linear Model will include variables:

Alcohol, Schooling, HIV.AIDS,
Percent Expenditure, BMI, Polio,
Total Expenditure, Diphtheria,
Income Composition of Resources



Comparing Reduced Model Data

```
Call:
lm(formula = lifeData$Life.expectancy ~ lifeData$Alcohol + lifeData$Schooling +
    lifeData$HIV.AIDS + lifeData$percentage.expenditure + lifeData$BMI +
    lifeData$Polio + lifeData$Total.expenditure + lifeData$Diphtheria +
    lifeData$Income.composition.of.resources, weights = wts)
```

weighted Residuals:

Min	1Q	Median	3Q	Max
-4.6637	-0.7757	-0.0223	0.7252	3.9673

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	4.459e+01	5.889e-01	75.723
lifeData\$Alcohol	-1.037e-01	2.849e-02	-3.638
lifeData\$Schooling	3.863e-01	7.013e-02	5.508
lifeData\$HIV.AIDS	-6.416e-01	2.465e-02	-26.031
lifeData\$percentage.expenditure	3.932e-04	5.638e-05	6.975
lifeData\$BMI	3.112e-02	5.414e-03	5.749
lifeData\$Polio	4.433e-03	4.988e-03	0.889
lifeData\$Total.expenditure	2.319e-01	4.087e-02	5.674
lifeData\$Diphtheria	1.416e-02	5.441e-03	2.602
lifeData\$Income.composition.of.resources	2.732e+01	1.298e+00	21.050

Pr(>|t|)

	Pr(> t)
(Intercept)	< 2e-16 ***
lifeData\$Alcohol	0.000283 ***
lifeData\$Schooling	4.22e-08 ***
lifeData\$HIV.AIDS	< 2e-16 ***
lifeData\$percentage.expenditure	4.42e-12 ***
lifeData\$BMI	1.07e-08 ***
lifeData\$Polio	0.374260
lifeData\$Total.expenditure	1.64e-08 ***
lifeData\$Diphtheria	0.009350 **
lifeData\$Income.composition.of.resources	< 2e-16 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.193 on 1639 degrees of freedom
 Multiple R-squared: 0.7852, Adjusted R-squared: 0.7841
 F-statistic: 665.8 on 9 and 1639 DF, p-value: < 2.2e-16

```
lm(formula = lifeData$Life.expectancy ~ Alcohol + Schooling +
    HIV.AIDS + percentage.expenditure + BMI + Polio + Total.expenditure +
    Diphtheria + Income.composition.of.resources, data = lifeData)
```

Residuals:

Min	1Q	Median	3Q	Max
-26.8836	-2.5661	-0.0456	2.5197	23.6831

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.492e+01	4.513e-01	99.532	< 2e-16 ***
Alcohol	-5.455e-02	2.606e-02	-2.093	0.0365 *
Schooling	1.063e+00	4.867e-02	21.834	< 2e-16 ***
HIV.AIDS	-6.620e-01	1.659e-02	-39.901	< 2e-16 ***
percentage.expenditure	4.422e-04	4.543e-05	9.734	< 2e-16 ***
BMI	5.119e-02	5.315e-03	9.631	< 2e-16 ***
Polio	2.926e-02	5.129e-03	5.705	1.30e-08 ***
Total.expenditure	3.433e-02	3.810e-02	0.901	0.3677
Diphtheria	3.120e-02	5.113e-03	6.101	1.21e-09 ***
Income.composition.of.resources	8.780e+00	6.718e-01	13.069	< 2e-16 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.27 on 2546 degrees of freedom
 (382 observations deleted due to missingness)
 Multiple R-squared: 0.7942, Adjusted R-squared: 0.7935
 F-statistic: 1092 on 9 and 2546 DF, p-value: < 2.2e-16

Original

$R^2 = .7841$

P value = 2.2e-16

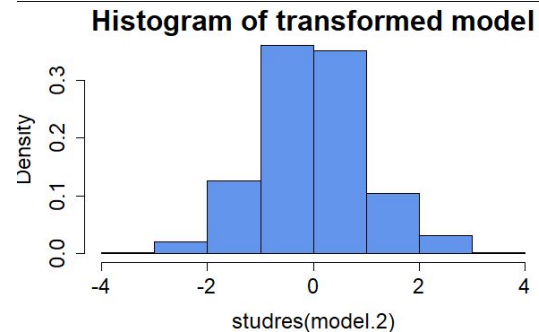
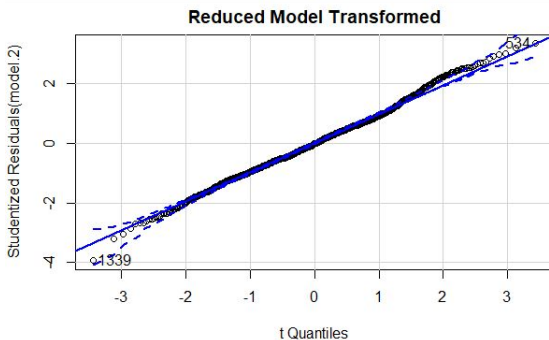
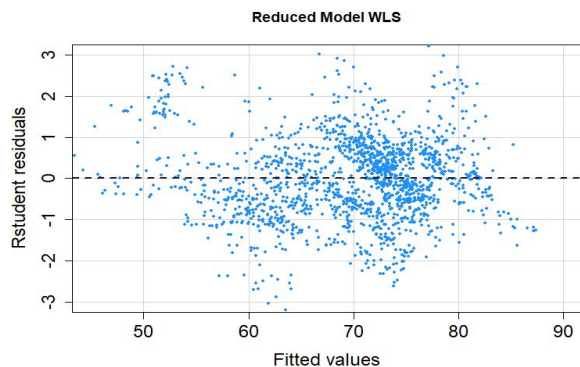
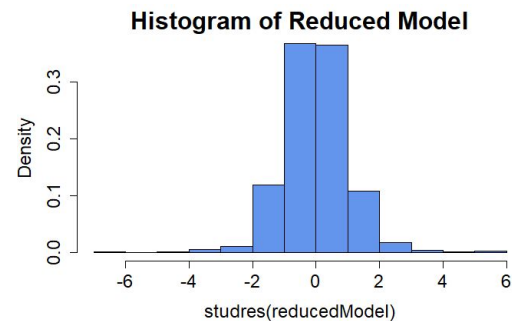
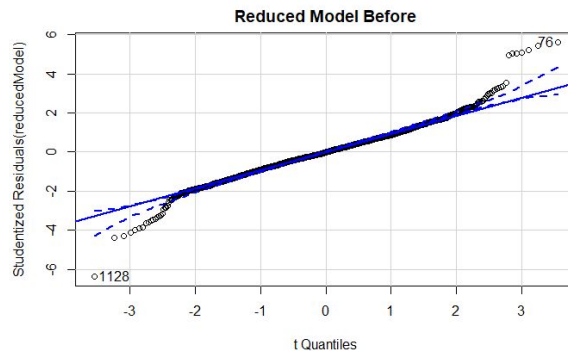
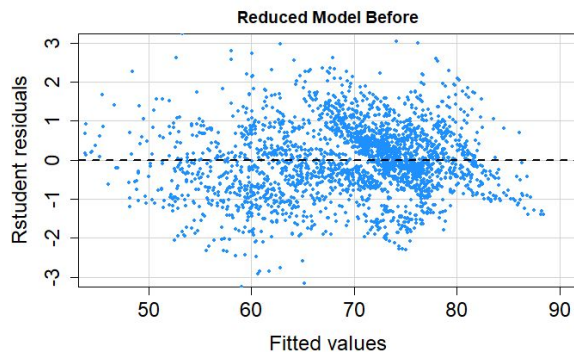
Reduced

$R^2 = .7935$

P value = 2.2e-16

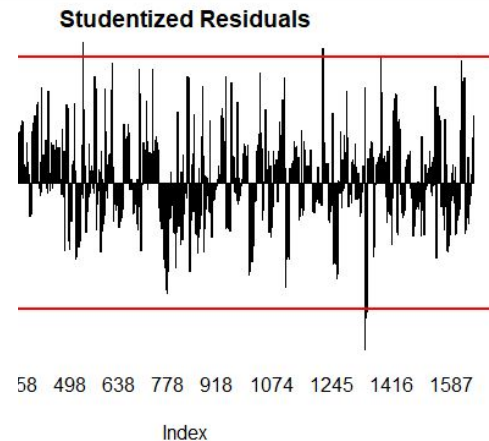
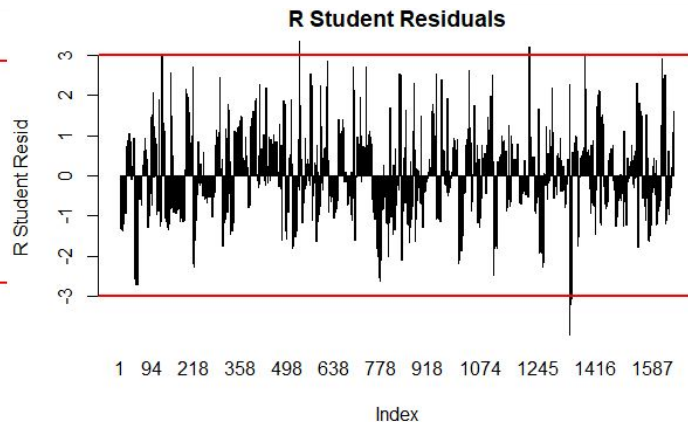
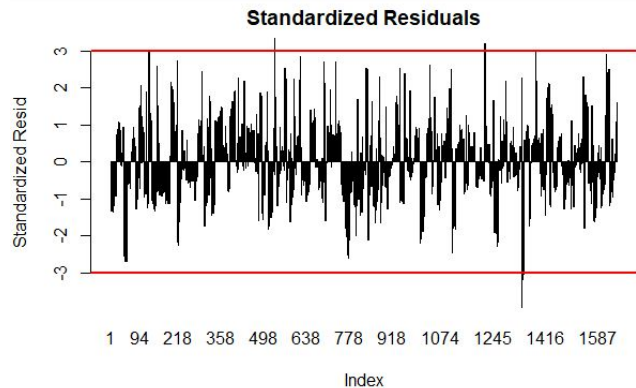


Transformation



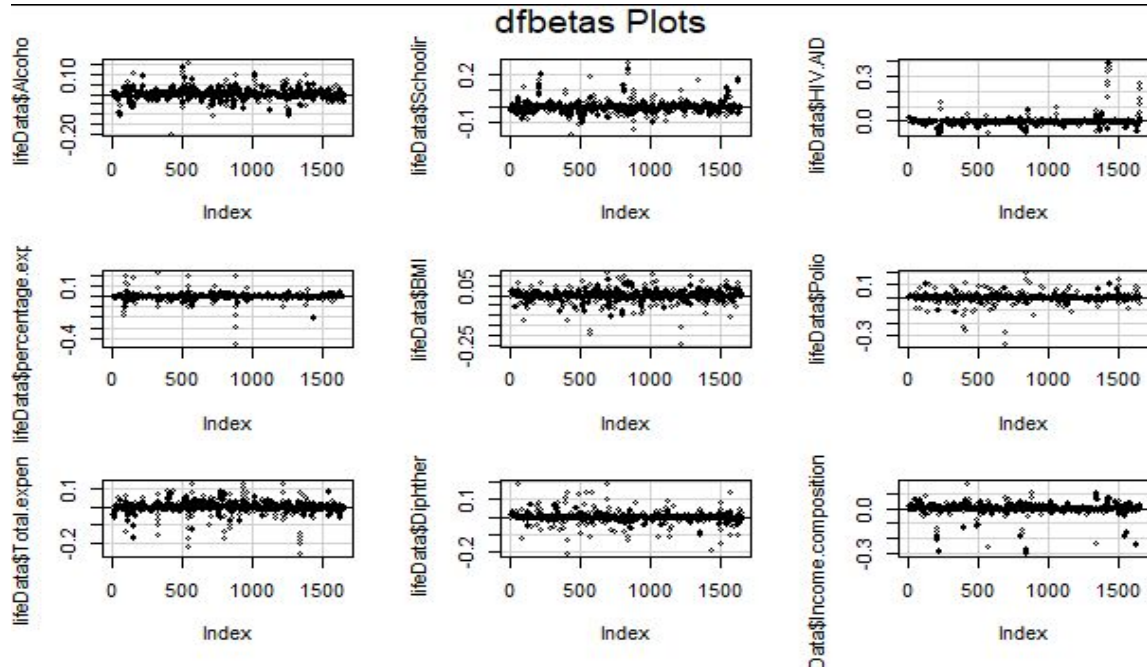


Residual Analysis for Reduced Model



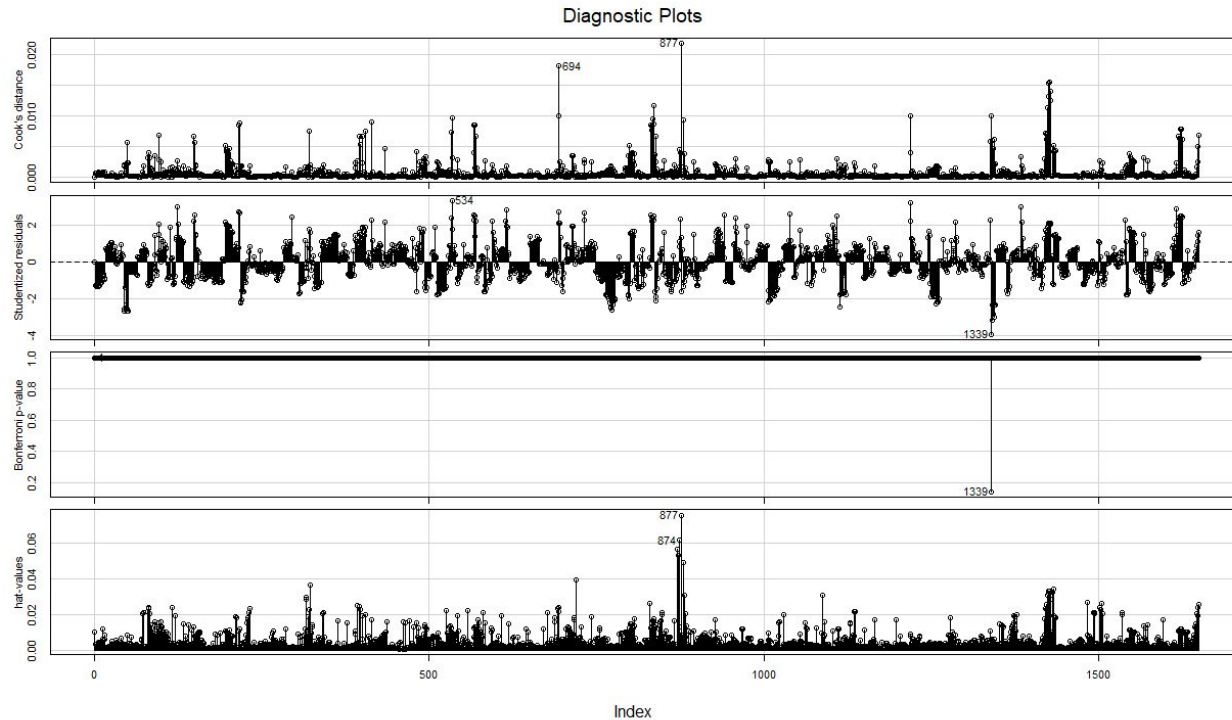


DfBetas Plots





Influential Points



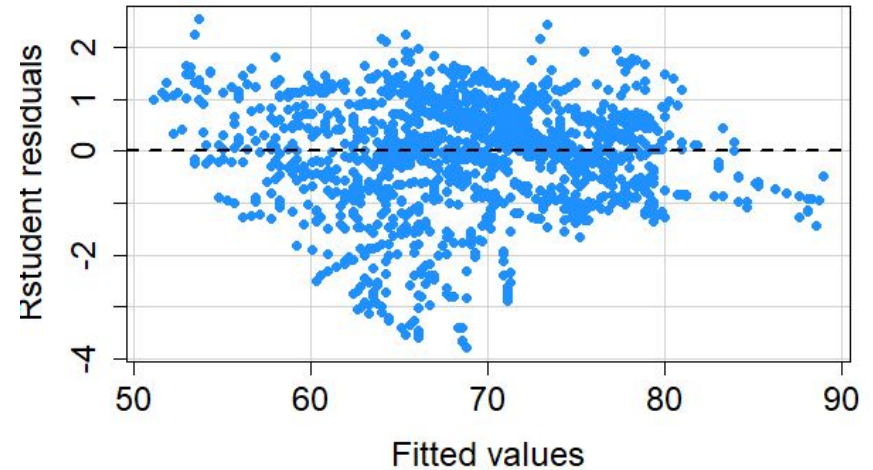


What is the impact of schooling on the life expectancy of people?

Affect of Schooling on Life Expectancy

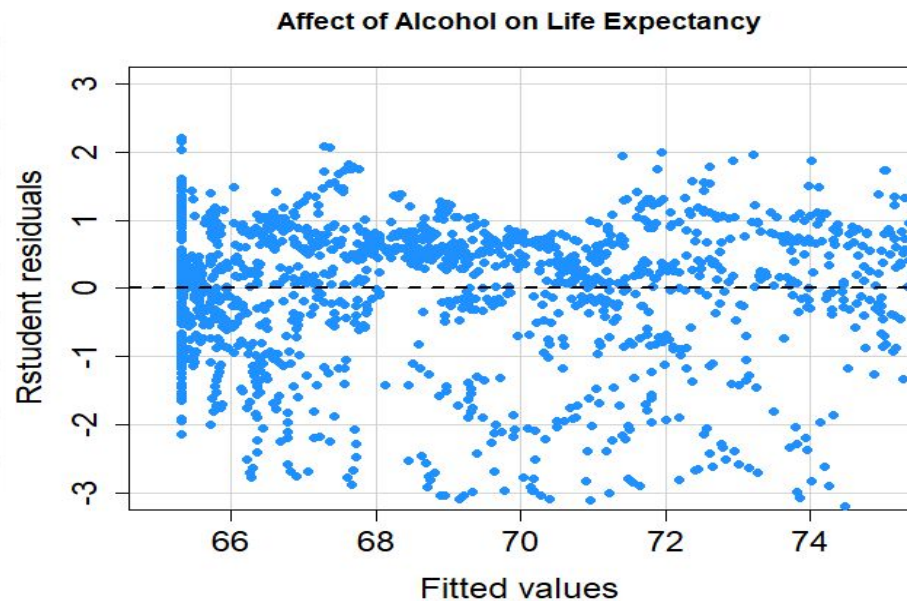
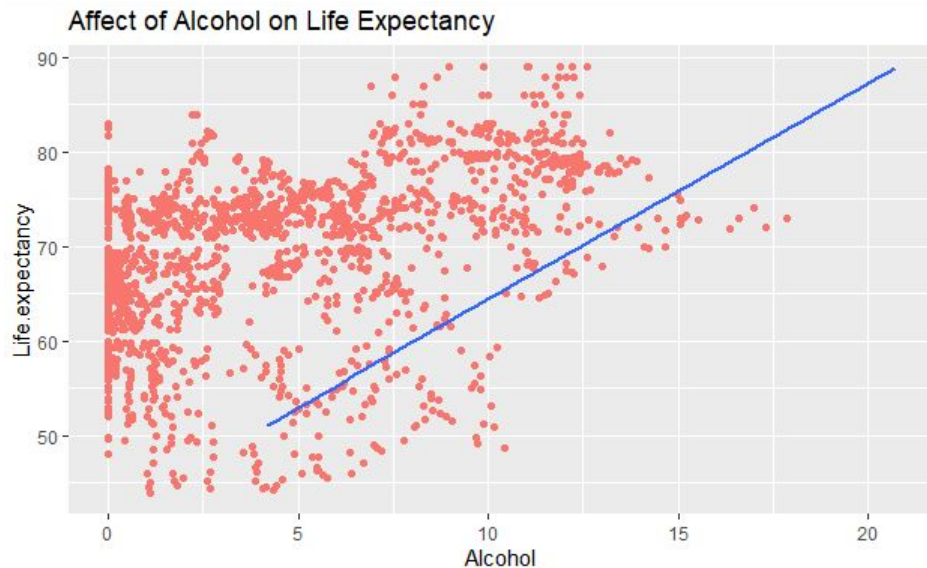


Affect of Schooling on Life Expectancy





How does alcohol consumption affect life expectancy?





Future Direction

- ❑ Analyze the difference in life expectancy between developing and developed countries
- ❑ Examine more recent years to see how life expectancy has changed over time
- ❑ Determine if additional predictor variables are more influential to the response variable



Thanks!

Any questions ?