Statistical Analysis on Factors Influencing Life Expectancy

Group 1
Will Bodeau, Priscilla Dixon, Chloe Guo, Ivy Le, Bryan Mui

TABLE OF CONTENTS

01

Intro

RESEARCH QUESTION:

How do various socioeconomic factors affect a country's average life expectancy?

Motivation and research question

- **Context:**

❑ Addressed gaps in past research by formulating a multiple linear regression model

❑ **Identifying key predictors contributing to life expectancy**

❑ Providing targeted recommendations for improving population health

- **Source:**

❑ The data set was collected under WHO and United Nations website

- **Information about dataset**

❑ Year: 2000 - 2015 for 193 countries

❑ 20 predicting variables

❑ Narrowed down to 5 variables for our study

02

Methods

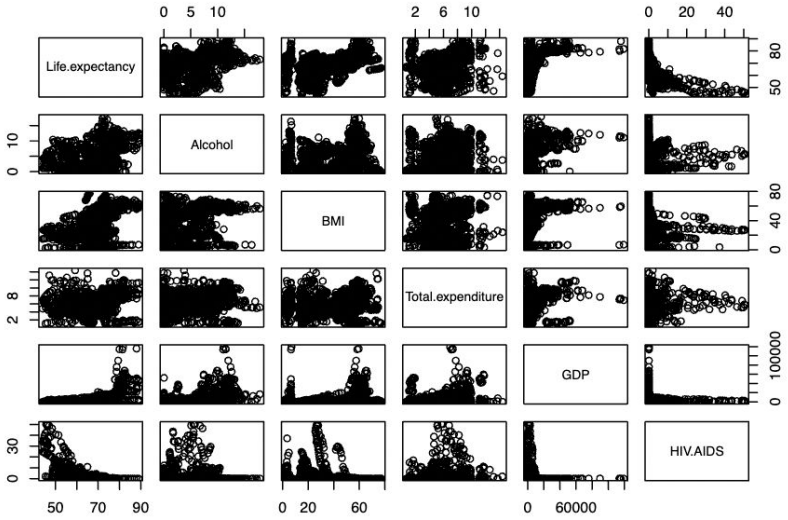
Variable Selection

NAME		DESCRIPTION	Transformed?
Alcohol		Alcohol consumption per capita (liters)	✗
BMI		Average Body Mass Index	✗
Expenditure		Government expense in health as a % of total money spent	✗
GDP		Monetary value of goods/services	Log ✓
HIV.AIDS		HIV/AIDS deaths per 1000 births	^(-½) ✓

Transformation Selection (Full Model)

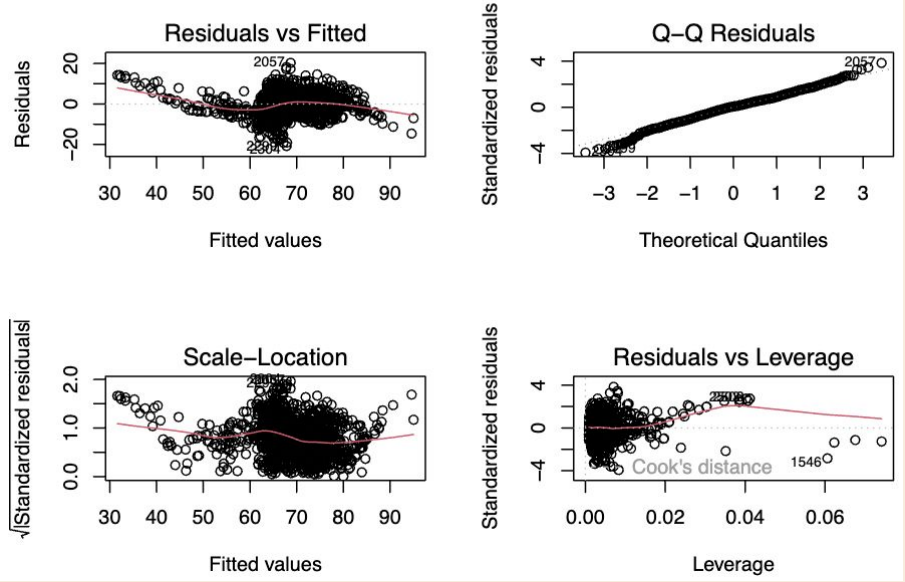
Full model, no transformations

```
life <- life[, c("Life.expectancy", "Alcohol", "BMI", "Total.expenditure", "GDP", "HIV.AIDS")]
life_cleaned_og <- life
pairs(life)
```



Diagnostics of full model

```
m_full <- lm(Life.expectancy ~ Alcohol + BMI + Total.expenditure + GDP + HIV.AIDS, data = life)
par(mfrow=c(2,2))
plot(m_full)
```



Transformation Selection (Predictors)

Power transformations for predictors

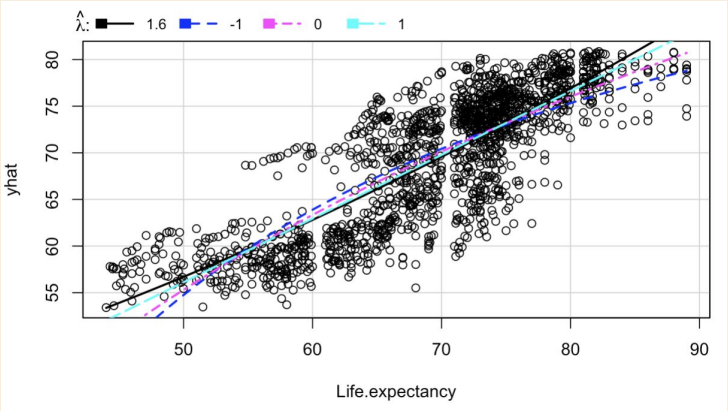
```
pwrTransform <- powerTransform(cbind(life$GDP, life$HIV.AIDS,
                               life$Total.expenditure, life$Alcohol, life$BMI) ~ 1)
summary(pwrTransform)

## bcPower Transformations to Multinormality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    0.0740          0.07       0.0524       0.0957
## Y2   -0.6241         -0.62      -0.6703      -0.5779
## Y3    0.8388          0.84       0.7505       0.9271
## Y4    0.4050          0.40       0.3773       0.4327
## Y5    1.0676          1.00       0.9926       1.1425
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                     LRT df       pval
## LR test, lambda = (0 0 0 0 0) 3485.373  5 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                                     LRT df       pval
## LR test, lambda = (1 1 1 1 1) 15934.79  5 < 2.22e-16
```

Y1 : GDP
Y2: HIV.AIDS
Y3: Total.expenditure
Y4: Alcohol
Y5: BMI

Conclusion:
Transform log(GDP) and HIV.AIDS^(½)

Transformation Selection (Y Variable)

Transformation of Y variable:

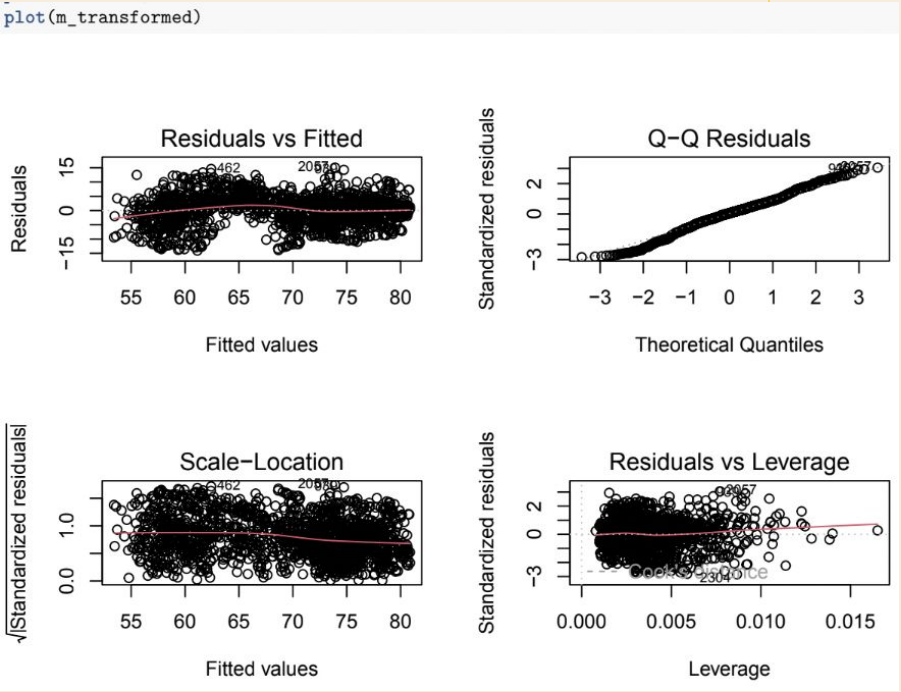


lambda <dbl>	RSS <dbl>
1.599614	27418.88
-1.000000	31136.65
0.000000	28821.26
1.000000	27612.35

Conclusion: No transform on Y

Diagnostics of transformed model:

```
plot(m_transformed)
```



Model Selection

```
library(leaps)
m_life_back <- step(m_transformed, direction = "backward")
```

```
## Start:  AIC=5287.78
## Life.expectancy ~ Alcohol + BMI + Total.expenditure + logGDP +
##     transformedHIV
##
##                      Df Sum of Sq   RSS    AIC
## <none>                             40429 5287.8
## - Total.expenditure   1       301 40730 5298.0
## - Alcohol             1      1420 41850 5342.7
## - BMI                 1      1766 42195 5356.3
## - logGDP              1      2742 43171 5394.0
## - transformedHIV      1     33156 73586 6273.4
```

VIFs < 5:

Alcohol	BMI	Total.expenditure	logGDP	transformedHIV
1.361146	1.450318	1.069556	1.484039	1.360665

03

Result Analysis

Summary (R-squared)

```
summary(m_transformed)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Alcohol + BMI + Total.expenditure +
##     logGDP + transformedHIV, data = life)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0395  -2.7862   0.2937   3.0959  15.0827
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       47.258475   0.597961  79.033  < 2e-16 ***
## Alcohol            0.268829   0.035382   7.598 5.03e-14 ***
## BMI                0.063114   0.007449   8.472  < 2e-16 ***
## Total.expenditure  0.192117   0.054959   3.496 0.000485 ***
## logGDP             0.896935   0.084974  10.555  < 2e-16 ***
## transformedHIV     4.682749   0.127569  36.708  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.961 on 1643 degrees of freedom
## Multiple R-squared:  0.683,  Adjusted R-squared:  0.682
## F-statistic: 707.9 on 5 and 1643 DF,  p-value: < 2.2e-16
```

- R-squared: 0.683
- P-values:

```
## Alcohol              Pr(>F)
## BMI                < 2.2e-16 ***
## Total.expenditure  < 2.2e-16 ***
## logGDP              0.007039 **
## transformedHIV     < 2.2e-16 ***
## transformedHIV     < 2.2e-16 ***
```

Summary (Interpretation)

- Our **intercept** coefficient was **47.258475**, therefore if all the predictors have a value of **0**, then on average life expectancy for that country would be **47.258475** years

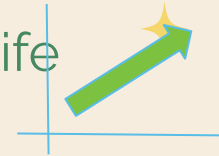
- Our **Alcohol** coefficient was **0.268829**, therefore if a country increases their per capita transformation of alcohol consumption by 1 liter, then on average expectancy goes up by **0.268829** years

- Our **BMI** coefficient was **0.063114**, meaning if the country's average BMI score increases by 1 point, then the life expectancy would go up by **0.063114** years

Summary (Interpretation)

- Our **Total expenditure** coefficient was **0.192117,** therefore if a country increased their expenditure ratio of health vs total expenditure by 1% life expectancy would increase by about **0.192117** years

- Our **logGDP** coefficient was **0.896935**, meaning that GDP and life expectancy are **positively correlated**



- Out **transformedHIV** coefficient was 4.682749, since we took the ^(-½) **HIV**, and **increase in HIV deaths would imply a decrease in life expectancy**

Increasing Life Expectancy

Healthcare funding

Countries with a greater percent of GDP spent on healthcare have longer life expectancies

More Developed

Countries that can afford to consume more alcohol/people have higher BMI seem to live longer

Thanks!