Statistics Project 2

David Rodden, Michael McIntosh, Khang Tran, Jeffrey Lazatin

April 23, 2017

```
sheet = read_excel("GDP.xlsx")
attach(sheet)
#Country, GDP, LEB, NLLEB, NLGDP
```

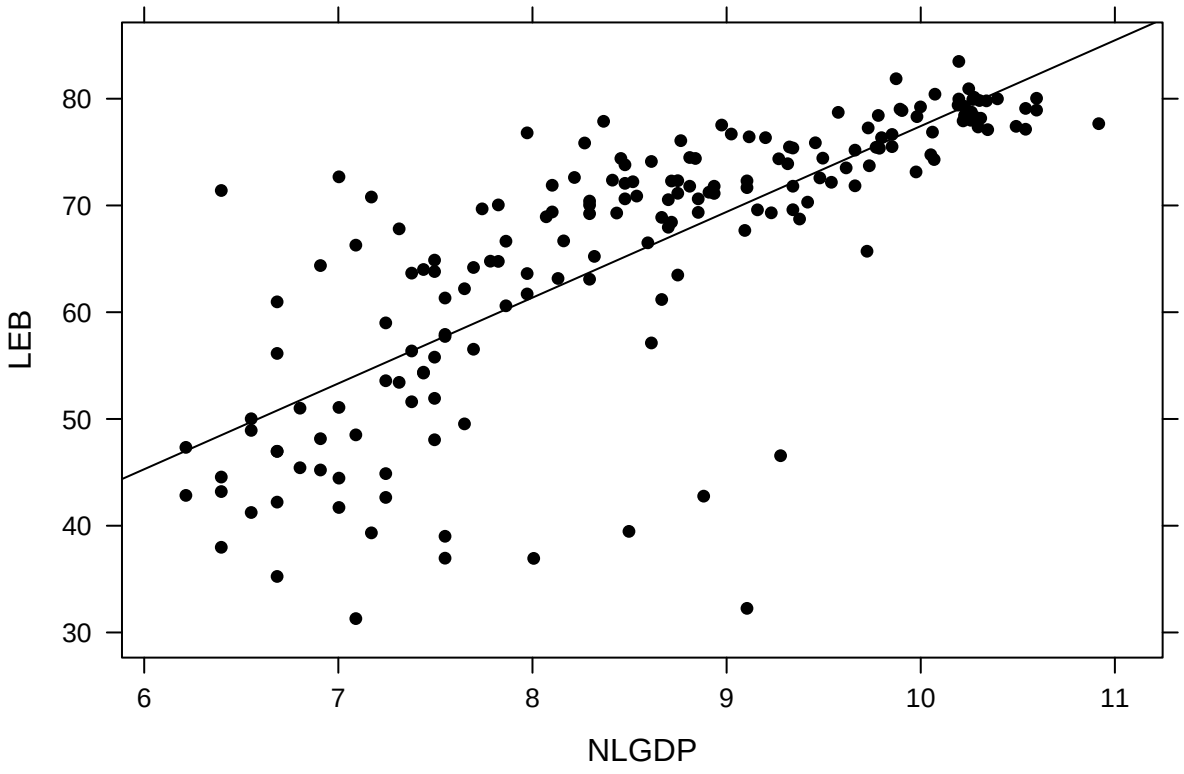
Introduction

We are investigating the relationship between Life Expectancy of a country based upon its GDP. The data we are using was collected in 2003 from the CIA Factbook; the data is across 180 countries. The investigation is looking to see if there is a positive correlation between life expectancy (LEB) and a country's GDP (NLGDP); using GDP as a predictor. In order to normalize the data, we use the natural log of the GDP. The data considers a country's life expectancy at birth and the GDP per capita (PPP). The data was collected from official reports that each nation compiles. We found the data from *Index Mundi*, who pulled from the CIA Factbook. The data is a sample of the world's countries, and is an observational study.

$H_0 : \rho = 0$ vs $H_a : \rho \neq 0$

Summary & Visualization

```
favs = favstats(LEB ~ NLGDP)
anova.b = anova(lm(LEB ~ NLGDP))
xyplot(LEB ~ NLGDP, type = c("p", "r"), pch=16, col="black")
```



The scatter plot shows that there are a few outliers which will influence the overall model. The outliers will impact the regression which we use to model and predict, based upon the data. There is a slight departure from linearity, a subtle curve in the data, but still increasing overall. The data does possesses changing variability, a fanning trend, wide to narrow from left to right. There appears to be a positive linear association between the two quantitative variables, LEB ~ NLGDP.

```
#five number summary
sum.sheet <- summary(sheet)
sum.sheet
```

```
##    Country               GDP             LEB            NLGDP
##  Length:180         Min.   :  500   Min.   :31.30   Min.   : 6.215
##  Class :character   1st Qu.: 1800   1st Qu.:57.87   1st Qu.: 7.496
##  Mode  :character   Median : 5650   Median :70.47   Median : 8.639
##                     Mean   :10051   Mean   :65.95   Mean   : 8.571
##                     3rd Qu.:15700   3rd Qu.:75.86   3rd Qu.: 9.661
##                     Max.   :55100   Max.   :83.49   Max.   :10.917
```

```
#standard deviation
leb.sd <- sd(sheet$LEB)
nlgdp.sd <- sd(sheet$NLGDP)
```

The sample size is 180 countries. The means for GDP, Life expectancy at birth, and Natural log GDP are Mean :10051 , Mean :65.95 , Mean : 8.571 , respectively.

Correlation Test

Regression

Teamwork