Statistics Project 2

David Rodden, Michael McIntosh, Khang Tran, Jeffrey Lazatin

April 23, 2017

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

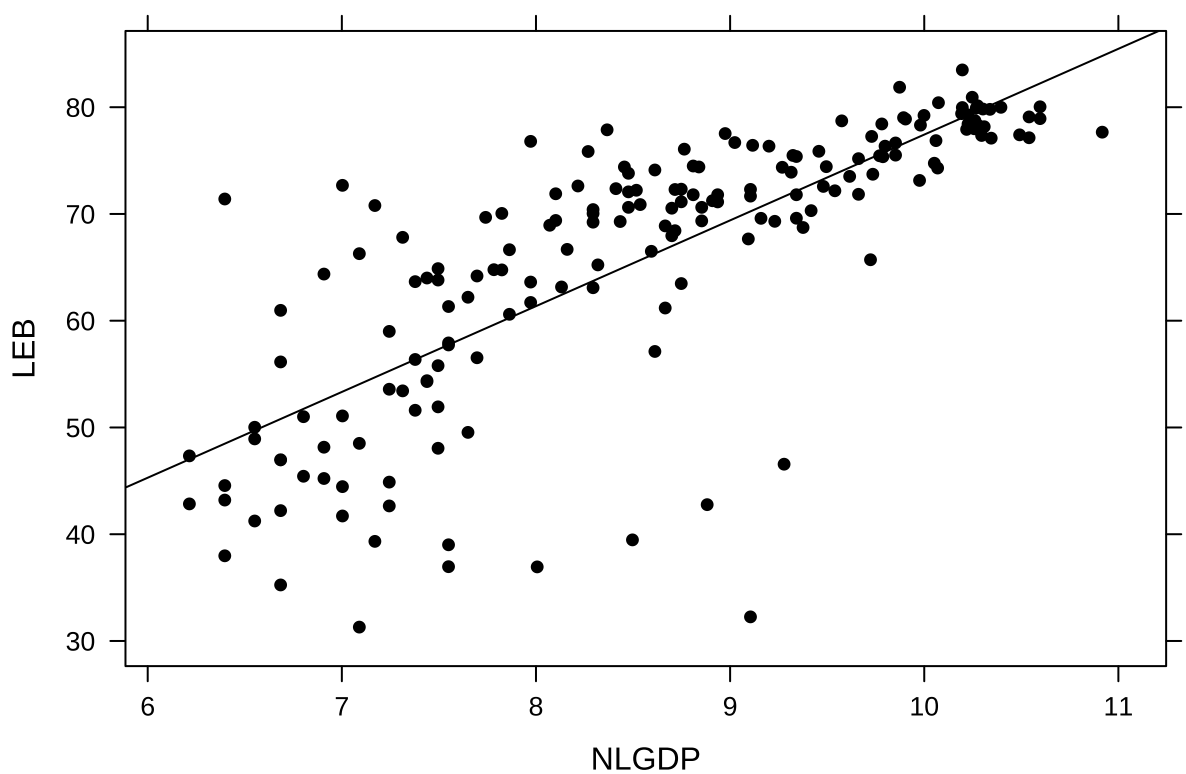
Introduction

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

```r
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
gdp.sd = sd(sheet$GDP); gdp.sd
```

```
## [1] 10757.43
```

```
leb.sd = sd(sheet$LEB); leb.sd
```

```
## [1] 12.75888
```

```
nlgdp.sd = sd(sheet$NLGDP); nlgdp.sd
```

```
## [1] 1.223437
```

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

The standard deviation for GDP, Life expectancy, at birth and Natural log of GDP are 1.0757429×10^4, 12.7588816, 1.2234365 respectively.

Correlation Test

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

```
cor = cor.test(NLGDP, LEB)
cor.p.value = cor$p.value
```

The p-value $\approx 1.1930935 \times 10^{-36}$. As the p-value is very small, we reject the null hypothesis in favor of the alternative hypothesis that there is a non-zero correlation between Life expectancy at birth and the natural log of GDP.

Regression

```
mod = lm(LEB ~ NLGDP)
y.hat = makeFun(mod); y.hat
```

```
## function (NLGDP, ..., transformation = function (x) 
## x) 
## return(transformation(predict(model, newdata = data.frame(NLGDP = NLGDP), 
##     ...)))
## <environment: 0x000000000f1c2000>
## attr(,"coefficients")
## (Intercept)       NLGDP 
##   -2.920003    8.035419
```

```r
mod.prection = predict(mod)
mod.resid = resid(mod); mod.resid
```

```
##            1            2            3            4            5
##   -4.176979138  0.323020862 -3.922009299 -5.282009299 22.917990701
##            6            7            8            9           10
## -10.502009299  0.299325350 -0.790674650 -8.480674650 -3.823655383
##           11           12           13           14           15
##   -8.583655383 -3.823655383  5.346344617 10.176344617 -15.543655383
##           16           17           18           19           20
##   -0.730091463 -6.310091463 -4.436707387 -7.366707387 11.783292613
##           21           22           23           24           25
##   -8.892564647 19.327435353 -2.272564647 -11.642564647 -5.541737547
##           26           27           28           29           30
## -22.751737547 12.228262453 -15.364914267 16.095085733 -12.640402898
##           31           32           33           34           35
##    3.709597102 -1.710402898 -10.410402898 -2.414789551 11.965210449
##           36           37           38           39           40
##    7.296616369 -4.753383631  0.006616369  7.149472112 -2.550527888
##           41           42           43           44           45
##   -2.470527888 -9.259819712 -1.519819712  7.570180288 -5.379819712
##           46           47           48           49           50
##    6.500180288  3.585727494 -20.784272506  0.175727494 -0.014272506
##           51           52           53           54           55
## -18.734272506 -9.008485063  3.651514937  5.267707105 -2.392292895
##           56           57           58           59           60
##   10.400518553  5.158534205  4.810512361 10.100512361  6.385357485
##           61           62           63           64           65
##    0.335357485  2.477895384  0.567895384 15.657895384 -24.474517799
##           66           67           68           69           70
##    7.006888121  9.709624940  7.209624940  0.739743863  4.026816849
##           71           72           73           74           75
##    9.520289793 12.327275320  6.313836117  5.503836117  6.683836117
##           76           77           78           79           80
##   -0.636163883  1.305420620 13.572709275  7.697400036  4.440790304
##           81           82           83           84           85
##    9.377978927  6.878805956  5.428805956  8.618805956 -25.886878663
##           86           87           88           89           90
##    6.700784113  5.201661699  0.362369876 -9.165073148  7.834926852
##           91           92           93           94           95
##    2.168167136 -5.521832864  0.975753952  3.555753952  5.172934080
##           96           97           98           99          100
##    1.312934080  3.763704524  4.943704524 -3.896295476  8.567159872
##          101          102          103          104          105
##    6.619061041  3.929061041  6.292708140  2.397088601  1.127088601
##          106          107          108          109          110
## -25.679276208  2.570561545  2.246270997  2.916270997  8.335183658
##          111          112          113          114          115
##    7.098292157 -2.492546057 -37.982328212  2.057671788  1.427671788
##          116          117          118          119          120
##    6.098881720 -1.086781007  5.341814527 -1.938066550  2.812840754
##          121          122          123          124          125
## -25.072609747  1.982479381  3.480412356  3.238188833 -0.341811167
##          126          127          128          129          130
```

```
##   -2.541811167  -3.688922053  -2.442389793   2.797431624  -0.678725412
##            131           132           133           134           135
##    1.049523481  -1.565037829   4.700995543  -0.827026300   0.466474116
##            136           137           138           139           140
##   -2.873525884  -9.499697013   2.002330192  -1.585357896  -0.135691652
##            141           142           143           144           145
##    2.752995783  -0.362274305   0.537943539   0.403490745  -0.736509255
##            146           147           148           149           150
##    5.456080120   2.442086279   2.231327616  -4.099799226   1.042913379
##            151           152           153           154           155
##    1.805470355  -3.101289361  -1.050263341  -3.688650301   2.397373245
##            156           157           158           159           160
##    0.419643530   4.479604545   0.959604545  -1.268308686   0.033251643
##            161           162           163           164           165
##   -0.826748357  -1.115809580   1.510440266  -1.524591420  -0.792736525
##            166           167           168           169           170
##    0.341267295   0.485658635  -1.382002044  -2.458894952  -0.033005724
##            171           172           173           174           175
##   -1.746754549  -0.354271919  -3.106113514  -0.619228474  -3.971784709
##            176           177           178           179           180
##   -2.683834137  -4.633834137  -3.298400633  -2.188400633  -7.141906485
```

Using the regression model, $\hat{Y} = b_0 + b_1 x$.

Teamwork