Life Expectancy Analysis

2022-11-10

# Team Members

#(Wednesday Morning Batch)

*Akash Mishra (U11134544)*

*Ojeswini Sesi Priya (U29976449)*

*Sonica Kolluri (U24764470)*

## *About Dataset*

*We will analyze data containing 1613 observations with 22 columns.The data-set related to life expectancy, health factors for 193 countries has been collected from the WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. The independent variables in our dataset are namely immunizations like polio, hepatatis other factors like HIV, BMI, alcohol consumption, country population and we will be doing regression analysis to see what the main factors among these variables effecting our dependant variable that is Life Expectancy*

# Preprocessing

*1) Loading the required Libraries and dataset*

rm(list = ls())  
library(rio)

## Warning: package 'rio' was built under R version 4.2.2

library(car)

## Warning: package 'car' was built under R version 4.2.2

## Loading required package: carData

library(corrplot)

## corrplot 0.92 loaded

Data = import("Life Expectancy Data.csv")  
colnames(Data) = tolower(make.names(colnames(Data)))  
attach(Data)

*2) Viewing the First few columns of Data and its Structure*

head(Data)

## country year status life.expectancy adult.mortality infant.deaths  
## 1 Afghanistan 2015 Developing 65.0 263 62  
## 2 Afghanistan 2014 Developing 59.9 271 64  
## 3 Afghanistan 2013 Developing 59.9 268 66  
## 4 Afghanistan 2012 Developing 59.5 272 69  
## 5 Afghanistan 2011 Developing 59.2 275 71  
## 6 Afghanistan 2010 Developing 58.8 279 74  
## alcohol percentage.expenditure hepatitis.b measles bmi under.five.deaths  
## 1 0.01 71.279624 65 1154 19.1 83  
## 2 0.01 73.523582 62 492 18.6 86  
## 3 0.01 73.219243 64 430 18.1 89  
## 4 0.01 78.184215 67 2787 17.6 93  
## 5 0.01 7.097109 68 3013 17.2 97  
## 6 0.01 79.679367 66 1989 16.7 102  
## polio total.expenditure diphtheria hiv.aids gdp population  
## 1 6 8.16 65 0.1 584.25921 33736494  
## 2 58 8.18 62 0.1 612.69651 327582  
## 3 62 8.13 64 0.1 631.74498 31731688  
## 4 67 8.52 67 0.1 669.95900 3696958  
## 5 68 7.87 68 0.1 63.53723 2978599  
## 6 66 9.20 66 0.1 553.32894 2883167  
## thinness..1.19.years thinness.5.9.years income.composition.of.resources  
## 1 17.2 17.3 0.479  
## 2 17.5 17.5 0.476  
## 3 17.7 17.7 0.470  
## 4 17.9 18.0 0.463  
## 5 18.2 18.2 0.454  
## 6 18.4 18.4 0.448  
## schooling  
## 1 10.1  
## 2 10.0  
## 3 9.9  
## 4 9.8  
## 5 9.5  
## 6 9.2

str(Data)

## 'data.frame': 1613 obs. of 22 variables:  
## $ country : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...  
## $ year : int 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 ...  
## $ status : chr "Developing" "Developing" "Developing" "Developing" ...  
## $ life.expectancy : num 65 59.9 59.9 59.5 59.2 58.8 58.6 58.1 57.5 57.3 ...  
## $ adult.mortality : int 263 271 268 272 275 279 281 287 295 295 ...  
## $ infant.deaths : int 62 64 66 69 71 74 77 80 82 84 ...  
## $ alcohol : num 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.03 0.02 0.03 ...  
## $ percentage.expenditure : num 71.3 73.5 73.2 78.2 7.1 ...  
## $ hepatitis.b : int 65 62 64 67 68 66 63 64 63 64 ...  
## $ measles : int 1154 492 430 2787 3013 1989 2861 1599 1141 1990 ...  
## $ bmi : num 19.1 18.6 18.1 17.6 17.2 16.7 16.2 15.7 15.2 14.7 ...  
## $ under.five.deaths : int 83 86 89 93 97 102 106 110 113 116 ...  
## $ polio : int 6 58 62 67 68 66 63 64 63 58 ...  
## $ total.expenditure : num 8.16 8.18 8.13 8.52 7.87 9.2 9.42 8.33 6.73 7.43 ...  
## $ diphtheria : int 65 62 64 67 68 66 63 64 63 58 ...  
## $ hiv.aids : num 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 ...  
## $ gdp : num 584.3 612.7 631.7 670 63.5 ...  
## $ population : num 33736494 327582 31731688 3696958 2978599 ...  
## $ thinness..1.19.years : num 17.2 17.5 17.7 17.9 18.2 18.4 18.6 18.8 19 19.2 ...  
## $ thinness.5.9.years : num 17.3 17.5 17.7 18 18.2 18.4 18.7 18.9 19.1 19.3 ...  
## $ income.composition.of.resources: num 0.479 0.476 0.47 0.463 0.454 0.448 0.434 0.433 0.415 0.405 ...  
## $ schooling : num 10.1 10 9.9 9.8 9.5 9.2 8.9 8.7 8.4 8.1 ...

*3) Creating a dataframe data1 to keep the original data safe*

data1=Data

## Data Analysis

*1) Checking the data for any missing values*

colSums(is.na(data1))

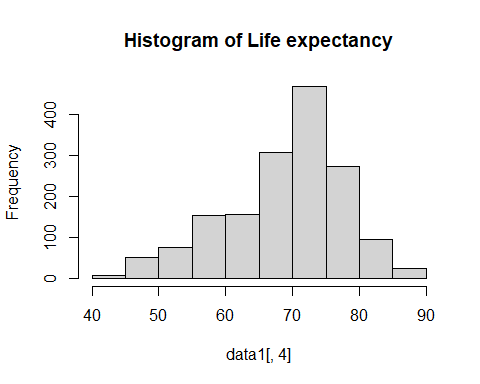
## country year   
## 0 0   
## status life.expectancy   
## 0 0   
## adult.mortality infant.deaths   
## 0 0   
## alcohol percentage.expenditure   
## 0 0   
## hepatitis.b measles   
## 0 0   
## bmi under.five.deaths   
## 0 0   
## polio total.expenditure   
## 0 0   
## diphtheria hiv.aids   
## 0 0   
## gdp population   
## 0 0   
## thinness..1.19.years thinness.5.9.years   
## 0 0   
## income.composition.of.resources schooling   
## 0 0

dim(data1)

## [1] 1613 22

*3) Checking distribution of dependent variable i.e. Life.expectancy-histogram*

hist(data1[,4], main="Histogram of Life expectancy")



## Regression Analysis

*1) Fitting a regression model with out the independent variables country, year and status*

model1<-lm(life.expectancy~.-country-year-status,data=data1)  
  
  
summary(model1)

##   
## Call:  
## lm(formula = life.expectancy ~ . - country - year - status, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.9097 -2.0683 -0.0256 2.2334 12.0424   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.316e+01 7.439e-01 71.456 < 2e-16 \*\*\*  
## adult.mortality -1.698e-02 9.538e-04 -17.801 < 2e-16 \*\*\*  
## infant.deaths 9.381e-02 1.072e-02 8.751 < 2e-16 \*\*\*  
## alcohol -6.025e-02 3.088e-02 -1.951 0.0512 .   
## percentage.expenditure 3.576e-04 1.819e-04 1.966 0.0494 \*   
## hepatitis.b -5.830e-03 4.476e-03 -1.302 0.1929   
## measles -9.746e-06 1.086e-05 -0.898 0.3694   
## bmi 3.457e-02 6.088e-03 5.679 1.61e-08 \*\*\*  
## under.five.deaths -7.023e-02 7.758e-03 -9.053 < 2e-16 \*\*\*  
## polio 6.964e-03 5.188e-03 1.342 0.1797   
## total.expenditure 8.172e-02 4.134e-02 1.977 0.0482 \*   
## diphtheria 1.391e-02 5.963e-03 2.332 0.0198 \*   
## hiv.aids -4.352e-01 1.797e-02 -24.224 < 2e-16 \*\*\*  
## gdp 1.654e-05 2.860e-05 0.579 0.5630   
## population -1.042e-09 1.764e-09 -0.591 0.5548   
## thinness..1.19.years -1.121e-02 5.332e-02 -0.210 0.8335   
## thinness.5.9.years -5.480e-02 5.261e-02 -1.042 0.2978   
## income.composition.of.resources 9.516e+00 8.407e-01 11.320 < 2e-16 \*\*\*  
## schooling 9.128e-01 6.092e-02 14.983 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.607 on 1594 degrees of freedom  
## Multiple R-squared: 0.8361, Adjusted R-squared: 0.8343   
## F-statistic: 451.8 on 18 and 1594 DF, p-value: < 2.2e-16

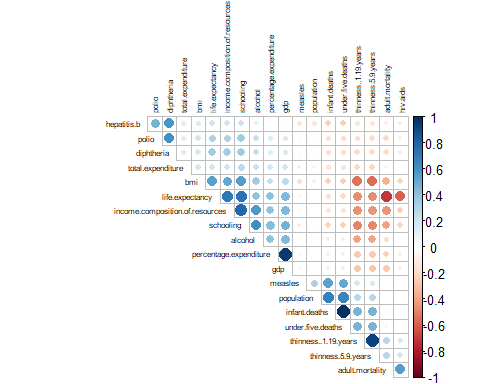
*Inference - The model fits quite good with the Multiple R-squared value 0.8361.We can also see that columns effect that dependent variable most due to their significant p value. Let see if we can improve the model a bit.*

*2) Checking for Mulitcollinearity among the independent variables*

cor(data1[,4:22])

## life.expectancy adult.mortality infant.deaths  
## life.expectancy 1.00000000 -0.703010377 -0.168992461  
## adult.mortality -0.70301038 1.000000000 0.043115640  
## infant.deaths -0.16899246 0.043115640 1.000000000  
## alcohol 0.40100831 -0.174230793 -0.107237163  
## percentage.expenditure 0.41142545 -0.237699795 -0.091920924  
## hepatitis.b 0.19647537 -0.104984154 -0.231087995  
## measles -0.06985580 -0.003377875 0.533045730  
## bmi 0.54665268 -0.351993234 -0.236953946  
## under.five.deaths -0.19221548 0.061091930 0.996903248  
## polio 0.32666000 -0.201876966 -0.155891074  
## total.expenditure 0.18245427 -0.089139690 -0.148508522  
## diphtheria 0.34019138 -0.192804089 -0.160673051  
## hiv.aids -0.59213787 0.551928327 0.006802911  
## gdp 0.44219659 -0.254231965 -0.098993867  
## population -0.02347217 -0.014040421 0.672879978  
## thinness..1.19.years -0.46373212 0.273978707 0.467372539  
## thinness.5.9.years -0.46358264 0.288766623 0.465848875  
## income.composition.of.resources 0.71931209 -0.440264037 -0.134814649  
## schooling 0.73001681 -0.418794263 -0.219258380  
## alcohol percentage.expenditure hepatitis.b  
## life.expectancy 0.40100831 0.41142545 0.19647537  
## adult.mortality -0.17423079 -0.23769980 -0.10498415  
## infant.deaths -0.10723716 -0.09192092 -0.23108799  
## alcohol 1.00000000 0.41901277 0.10582689  
## percentage.expenditure 0.41901277 1.00000000 0.01834742  
## hepatitis.b 0.10582689 0.01834742 1.00000000  
## measles -0.05189584 -0.06333256 -0.12624177  
## bmi 0.35933611 0.24441578 0.14644017  
## under.five.deaths -0.10201276 -0.09335014 -0.24002983  
## polio 0.23739208 0.13061372 0.46095121  
## total.expenditure 0.22520903 0.18521706 0.12073423  
## diphtheria 0.24047473 0.13685362 0.58750321  
## hiv.aids -0.02530000 -0.09584576 -0.09234514  
## gdp 0.44518656 0.95962346 0.04318323  
## population -0.03084519 -0.01725438 -0.13161981  
## thinness..1.19.years -0.40773709 -0.25481149 -0.13533678  
## thinness.5.9.years -0.39027954 -0.25546518 -0.13931662  
## income.composition.of.resources 0.56004016 0.40424126 0.18154714  
## schooling 0.61568587 0.42674012 0.21415408  
## measles bmi under.five.deaths  
## life.expectancy -0.069855800 0.5466527 -0.19221548  
## adult.mortality -0.003377875 -0.3519932 0.06109193  
## infant.deaths 0.533045730 -0.2369539 0.99690325  
## alcohol -0.051895838 0.3593361 -0.10201276  
## percentage.expenditure -0.063332565 0.2444158 -0.09335014  
## hepatitis.b -0.126241769 0.1464402 -0.24002983  
## measles 1.000000000 -0.1514704 0.51789958  
## bmi -0.151470425 1.0000000 -0.24486792  
## under.five.deaths 0.517899578 -0.2448679 1.00000000  
## polio -0.058913861 0.1938725 -0.17008801  
## total.expenditure -0.111601312 0.1842339 -0.14751236  
## diphtheria -0.059519540 0.1826469 -0.17721505  
## hiv.aids -0.003390587 -0.2126111 0.01853646  
## gdp -0.064658110 0.2655246 -0.10127106  
## population 0.321669440 -0.0790101 0.65982566  
## thinness..1.19.years 0.179877921 -0.5450997 0.46889685  
## thinness.5.9.years 0.174059191 -0.5522334 0.46638184  
## income.composition.of.resources -0.059052169 0.5131111 -0.14817030  
## schooling -0.118784707 0.5593334 -0.23107001  
## polio total.expenditure diphtheria  
## life.expectancy 0.32666000 0.18245427 0.34019138  
## adult.mortality -0.20187697 -0.08913969 -0.19280409  
## infant.deaths -0.15589107 -0.14850852 -0.16067305  
## alcohol 0.23739208 0.22520903 0.24047473  
## percentage.expenditure 0.13061372 0.18521706 0.13685362  
## hepatitis.b 0.46095121 0.12073423 0.58750321  
## measles -0.05891386 -0.11160131 -0.05951954  
## bmi 0.19387248 0.18423386 0.18264685  
## under.five.deaths -0.17008801 -0.14751236 -0.17721505  
## polio 1.00000000 0.12664175 0.60687991  
## total.expenditure 0.12664175 1.00000000 0.13660054  
## diphtheria 0.60687991 0.13660054 1.00000000  
## hiv.aids -0.10624966 0.04102153 -0.11576373  
## gdp 0.15906556 0.18109530 0.16049256  
## population -0.04673568 -0.07756652 -0.04109597  
## thinness..1.19.years -0.17003290 -0.20427715 -0.19327568  
## thinness.5.9.years -0.18062798 -0.21215235 -0.18701744  
## income.composition.of.resources 0.31444217 0.19203124 0.34258423  
## schooling 0.35488741 0.25911737 0.35450159  
## hiv.aids gdp population  
## life.expectancy -0.592137870 0.44219659 -0.023472167  
## adult.mortality 0.551928327 -0.25423197 -0.014040421  
## infant.deaths 0.006802911 -0.09899387 0.672879978  
## alcohol -0.025300003 0.44518656 -0.030845188  
## percentage.expenditure -0.095845762 0.95962346 -0.017254377  
## hepatitis.b -0.092345136 0.04318323 -0.131619813  
## measles -0.003390587 -0.06465811 0.321669440  
## bmi -0.212611082 0.26552464 -0.079010098  
## under.five.deaths 0.018536463 -0.10127106 0.659825659  
## polio -0.106249657 0.15906556 -0.046735683  
## total.expenditure 0.041021526 0.18109530 -0.077566524  
## diphtheria -0.115763734 0.16049256 -0.041095972  
## hiv.aids 1.000000000 -0.10840210 -0.027565387  
## gdp -0.108402097 1.00000000 -0.020551587  
## population -0.027565387 -0.02055159 1.000000000  
## thinness..1.19.years 0.175238851 -0.27658090 0.282574394  
## thinness.5.9.years 0.185906868 -0.27710190 0.277866141  
## income.composition.of.resources -0.246821588 0.44771535 -0.009010558  
## schooling -0.211755192 0.47124381 -0.042655794  
## thinness..1.19.years thinness.5.9.years  
## life.expectancy -0.4637321 -0.4635826  
## adult.mortality 0.2739787 0.2887666  
## infant.deaths 0.4673725 0.4658489  
## alcohol -0.4077371 -0.3902795  
## percentage.expenditure -0.2548115 -0.2554652  
## hepatitis.b -0.1353368 -0.1393166  
## measles 0.1798779 0.1740592  
## bmi -0.5450997 -0.5522334  
## under.five.deaths 0.4688969 0.4663818  
## polio -0.1700329 -0.1806280  
## total.expenditure -0.2042771 -0.2121524  
## diphtheria -0.1932757 -0.1870174  
## hiv.aids 0.1752389 0.1859069  
## gdp -0.2765809 -0.2771019  
## population 0.2825744 0.2778661  
## thinness..1.19.years 1.0000000 0.9271495  
## thinness.5.9.years 0.9271495 1.0000000  
## income.composition.of.resources -0.4591400 -0.4440031  
## schooling -0.4991614 -0.4804536  
## income.composition.of.resources schooling  
## life.expectancy 0.719312093 0.73001681  
## adult.mortality -0.440264037 -0.41879426  
## infant.deaths -0.134814649 -0.21925838  
## alcohol 0.560040163 0.61568587  
## percentage.expenditure 0.404241255 0.42674012  
## hepatitis.b 0.181547142 0.21415408  
## measles -0.059052169 -0.11878471  
## bmi 0.513111055 0.55933340  
## under.five.deaths -0.148170297 -0.23107001  
## polio 0.314442174 0.35488741  
## total.expenditure 0.192031237 0.25911737  
## diphtheria 0.342584228 0.35450159  
## hiv.aids -0.246821588 -0.21175519  
## gdp 0.447715353 0.47124381  
## population -0.009010558 -0.04265579  
## thinness..1.19.years -0.459140017 -0.49916145  
## thinness.5.9.years -0.444003087 -0.48045357  
## income.composition.of.resources 1.000000000 0.78502994  
## schooling 0.785029942 1.00000000

library(corrplot)  
corrplot(cor(data1[,4:22]),type = "upper", tl.pos = "td",  
 method = "circle", tl.cex = 0.5, tl.col = 'black',  
 order = "hclust", diag = FALSE)



*Inference -* *From the correlation plot we can see what all variables are correlated with each other in our data set*

vif(model1)

## adult.mortality infant.deaths   
## 1.795522 212.286384   
## alcohol percentage.expenditure   
## 1.935541 12.942997   
## hepatitis.b measles   
## 1.647306 1.515607   
## bmi under.five.deaths   
## 1.793389 202.006031   
## polio total.expenditure   
## 1.706833 1.123909   
## diphtheria hiv.aids   
## 2.084074 1.485347   
## gdp population   
## 13.610946 1.950755   
## thinness..1.19.years thinness.5.9.years   
## 7.535955 7.511196   
## income.composition.of.resources schooling   
## 2.969561 3.550960

sqrt(vif(model1))>3

## adult.mortality infant.deaths   
## FALSE TRUE   
## alcohol percentage.expenditure   
## FALSE TRUE   
## hepatitis.b measles   
## FALSE FALSE   
## bmi under.five.deaths   
## FALSE TRUE   
## polio total.expenditure   
## FALSE FALSE   
## diphtheria hiv.aids   
## FALSE FALSE   
## gdp population   
## TRUE FALSE   
## thinness..1.19.years thinness.5.9.years   
## FALSE FALSE   
## income.composition.of.resources schooling   
## FALSE FALSE

*Inference :*

*Multicollinearity can be detected using a statistic called the variance inflation factor (VIF). For any predictor variable, the square root of the VIF indicates the degree to which the confidence interval for that variable’s regression parameter is expanded relative to a model with uncorrelated predictors.* *Sq. root vif >3 indicates a multicollinearity problem. We see that multicollinearity is there in our data.Infant deaths, under 5 deaths, polio, thinness 1 to 19 yrs and thinness 5 to 9 yrs are highly correlated.*

*3) Improving model performance*

#We will be excluding columns which seems to be highly colinear

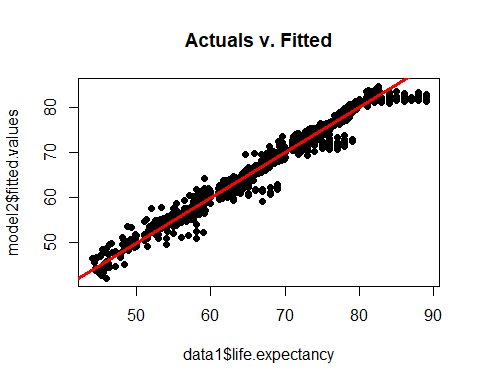
model2<-lm(life.expectancy~.-infant.deaths-under.five.deaths-polio-thinness..1.19.years-thinness.5.9.years,data=data1)  
summary(model2)

##   
## Call:  
## lm(formula = life.expectancy ~ . - infant.deaths - under.five.deaths -   
## polio - thinness..1.19.years - thinness.5.9.years, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.5431 -0.8029 -0.1684 0.3762 7.8777   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.958e+02 3.439e+01 -11.509 < 2e-16 \*\*\*  
## countryAlbania 1.558e+01 7.251e-01 21.489 < 2e-16 \*\*\*  
## countryAlgeria 1.377e+01 7.895e-01 17.445 < 2e-16 \*\*\*  
## countryAngola -7.483e+00 7.748e-01 -9.658 < 2e-16 \*\*\*  
## countryArgentina 1.452e+01 9.689e-01 14.987 < 2e-16 \*\*\*  
## countryArmenia 1.394e+01 7.077e-01 19.700 < 2e-16 \*\*\*  
## countryAustralia 2.038e+01 1.223e+00 16.663 < 2e-16 \*\*\*  
## countryAustria 2.148e+01 9.730e-01 22.080 < 2e-16 \*\*\*  
## countryAzerbaijan 1.154e+01 7.092e-01 16.275 < 2e-16 \*\*\*  
## countryBangladesh 1.096e+01 6.700e-01 16.364 < 2e-16 \*\*\*  
## countryBelarus 1.030e+01 9.090e-01 11.327 < 2e-16 \*\*\*  
## countryBelgium 2.020e+01 1.015e+00 19.910 < 2e-16 \*\*\*  
## countryBelize 9.856e+00 7.425e-01 13.273 < 2e-16 \*\*\*  
## countryBenin -4.168e-01 6.358e-01 -0.656 0.512201   
## countryBhutan 7.445e+00 6.549e-01 11.369 < 2e-16 \*\*\*  
## countryBosnia and Herzegovina 1.627e+01 7.893e-01 20.612 < 2e-16 \*\*\*  
## countryBotswana 1.475e+00 7.482e-01 1.971 0.048860 \*   
## countryBrazil 1.351e+01 8.210e-01 16.461 < 2e-16 \*\*\*  
## countryBulgaria 1.333e+01 8.396e-01 15.873 < 2e-16 \*\*\*  
## countryBurkina Faso -4.354e-01 7.413e-01 -0.587 0.557082   
## countryBurundi -1.635e+00 6.748e-01 -2.423 0.015517 \*   
## countryCabo Verde 1.306e+01 7.254e-01 18.006 < 2e-16 \*\*\*  
## countryCambodia 6.916e+00 7.246e-01 9.544 < 2e-16 \*\*\*  
## countryCameroon -2.309e+00 6.994e-01 -3.302 0.000983 \*\*\*  
## countryCanada 2.170e+01 9.930e-01 21.851 < 2e-16 \*\*\*  
## countryCentral African Republic -5.418e+00 8.392e-01 -6.456 1.45e-10 \*\*\*  
## countryChad -5.052e+00 7.941e-01 -6.362 2.66e-10 \*\*\*  
## countryChile 1.911e+01 9.401e-01 20.327 < 2e-16 \*\*\*  
## countryChina 1.534e+01 8.149e-01 18.822 < 2e-16 \*\*\*  
## countryColombia 1.387e+01 7.197e-01 19.276 < 2e-16 \*\*\*  
## countryComoros 2.884e+00 6.620e-01 4.357 1.41e-05 \*\*\*  
## countryCosta Rica 1.900e+01 7.521e-01 25.257 < 2e-16 \*\*\*  
## countryCroatia 1.632e+01 9.666e-01 16.889 < 2e-16 \*\*\*  
## countryCyprus 1.968e+01 8.570e-01 22.968 < 2e-16 \*\*\*  
## countryDominican Republic 1.329e+01 7.515e-01 17.685 < 2e-16 \*\*\*  
## countryEcuador 1.501e+01 7.469e-01 20.093 < 2e-16 \*\*\*  
## countryEl Salvador 1.214e+01 7.234e-01 16.782 < 2e-16 \*\*\*  
## countryEquatorial Guinea -9.101e-01 1.742e+00 -0.523 0.601347   
## countryEritrea 4.419e+00 7.380e-01 5.987 2.68e-09 \*\*\*  
## countryEstonia 1.404e+01 9.811e-01 14.308 < 2e-16 \*\*\*  
## countryEthiopia 4.399e+00 7.338e-01 5.995 2.56e-09 \*\*\*  
## countryFiji 8.374e+00 8.015e-01 10.448 < 2e-16 \*\*\*  
## countryFrance 2.223e+01 9.943e-01 22.356 < 2e-16 \*\*\*  
## countryGabon 5.419e+00 8.001e-01 6.772 1.83e-11 \*\*\*  
## countryGeorgia 1.411e+01 7.398e-01 19.072 < 2e-16 \*\*\*  
## countryGermany 2.084e+01 1.027e+00 20.287 < 2e-16 \*\*\*  
## countryGhana 3.034e+00 6.471e-01 4.688 3.01e-06 \*\*\*  
## countryGreece 2.092e+01 9.458e-01 22.125 < 2e-16 \*\*\*  
## countryGuatemala 1.376e+01 7.129e-01 19.308 < 2e-16 \*\*\*  
## countryGuinea -7.345e-01 7.347e-01 -1.000 0.317630   
## countryGuinea-Bissau -4.209e-01 8.118e-01 -0.519 0.604185   
## countryGuyana 7.118e+00 6.977e-01 10.202 < 2e-16 \*\*\*  
## countryHaiti 3.396e+00 1.262e+00 2.690 0.007218 \*\*   
## countryHonduras 1.412e+01 6.647e-01 21.247 < 2e-16 \*\*\*  
## countryIndia 7.078e+00 8.774e-01 8.067 1.49e-15 \*\*\*  
## countryIndonesia 8.072e+00 6.981e-01 11.562 < 2e-16 \*\*\*  
## countryIraq 1.092e+01 7.041e-01 15.513 < 2e-16 \*\*\*  
## countryIreland 2.129e+01 1.240e+00 17.163 < 2e-16 \*\*\*  
## countryIsrael 2.038e+01 9.170e-01 22.219 < 2e-16 \*\*\*  
## countryItaly 2.177e+01 9.509e-01 22.896 < 2e-16 \*\*\*  
## countryJamaica 1.489e+01 7.547e-01 19.733 < 2e-16 \*\*\*  
## countryJordan 1.294e+01 7.634e-01 16.949 < 2e-16 \*\*\*  
## countryKazakhstan 6.569e+00 8.139e-01 8.071 1.44e-15 \*\*\*  
## countryKenya 1.292e+00 6.603e-01 1.957 0.050513 .   
## countryKiribati 6.003e+00 6.979e-01 8.601 < 2e-16 \*\*\*  
## countryLatvia 1.285e+01 9.930e-01 12.939 < 2e-16 \*\*\*  
## countryLebanon 1.474e+01 8.578e-01 17.183 < 2e-16 \*\*\*  
## countryLesotho -3.743e+00 7.536e-01 -4.967 7.58e-07 \*\*\*  
## countryLiberia 2.096e+00 7.676e-01 2.731 0.006387 \*\*   
## countryLithuania 1.268e+01 9.730e-01 13.036 < 2e-16 \*\*\*  
## countryLuxembourg 2.110e+01 9.407e-01 22.435 < 2e-16 \*\*\*  
## countryMadagascar 4.586e+00 6.411e-01 7.154 1.32e-12 \*\*\*  
## countryMalawi -3.692e+00 6.970e-01 -5.297 1.36e-07 \*\*\*  
## countryMalaysia 1.369e+01 7.403e-01 18.493 < 2e-16 \*\*\*  
## countryMaldives 1.580e+01 6.988e-01 22.613 < 2e-16 \*\*\*  
## countryMali -2.143e+00 6.571e-01 -3.262 0.001131 \*\*   
## countryMalta 2.069e+01 8.785e-01 23.549 < 2e-16 \*\*\*  
## countryMauritania 4.796e+00 6.942e-01 6.909 7.26e-12 \*\*\*  
## countryMauritius 1.259e+01 7.673e-01 16.411 < 2e-16 \*\*\*  
## countryMexico 1.620e+01 7.396e-01 21.899 < 2e-16 \*\*\*  
## countryMongolia 6.070e+00 7.220e-01 8.407 < 2e-16 \*\*\*  
## countryMontenegro 1.430e+01 8.875e-01 16.109 < 2e-16 \*\*\*  
## countryMorocco 1.300e+01 6.563e-01 19.809 < 2e-16 \*\*\*  
## countryMozambique -1.173e+00 6.470e-01 -1.813 0.070107 .   
## countryMyanmar 5.742e+00 6.668e-01 8.611 < 2e-16 \*\*\*  
## countryNamibia 5.917e+00 8.972e-01 6.595 5.91e-11 \*\*\*  
## countryNepal 7.832e+00 6.728e-01 11.640 < 2e-16 \*\*\*  
## countryNetherlands 1.869e+01 1.351e+00 13.841 < 2e-16 \*\*\*  
## countryNicaragua 1.438e+01 6.727e-01 21.379 < 2e-16 \*\*\*  
## countryNiger 3.427e+00 8.647e-01 3.963 7.76e-05 \*\*\*  
## countryNigeria -4.104e+00 7.460e-01 -5.501 4.45e-08 \*\*\*  
## countryPakistan 6.345e+00 6.713e-01 9.453 < 2e-16 \*\*\*  
## countryPanama 1.691e+01 7.674e-01 22.031 < 2e-16 \*\*\*  
## countryPapua New Guinea 3.757e+00 6.136e-01 6.123 1.17e-09 \*\*\*  
## countryParaguay 1.324e+01 7.472e-01 17.724 < 2e-16 \*\*\*  
## countryPeru 1.384e+01 7.854e-01 17.625 < 2e-16 \*\*\*  
## countryPhilippines 8.420e+00 6.958e-01 12.101 < 2e-16 \*\*\*  
## countryPoland 1.548e+01 9.078e-01 17.049 < 2e-16 \*\*\*  
## countryPortugal 1.988e+01 9.631e-01 20.644 < 2e-16 \*\*\*  
## countryRomania 1.435e+01 8.330e-01 17.222 < 2e-16 \*\*\*  
## countryRussian Federation 7.973e+00 8.566e-01 9.308 < 2e-16 \*\*\*  
## countryRwanda 3.146e+00 6.646e-01 4.734 2.41e-06 \*\*\*  
## countrySamoa 1.412e+01 7.378e-01 19.138 < 2e-16 \*\*\*  
## countrySao Tome and Principe 6.718e+00 6.803e-01 9.875 < 2e-16 \*\*\*  
## countrySenegal 5.310e+00 6.745e-01 7.872 6.74e-15 \*\*\*  
## countrySerbia 1.422e+01 8.518e-01 16.689 < 2e-16 \*\*\*  
## countrySeychelles 1.271e+01 7.597e-01 16.736 < 2e-16 \*\*\*  
## countrySierra Leone -1.010e+01 7.341e-01 -13.757 < 2e-16 \*\*\*  
## countrySolomon Islands 9.266e+00 6.173e-01 15.010 < 2e-16 \*\*\*  
## countrySouth Africa 3.978e+00 7.926e-01 5.019 5.82e-07 \*\*\*  
## countrySpain 2.171e+01 9.767e-01 22.226 < 2e-16 \*\*\*  
## countrySri Lanka 1.326e+01 7.992e-01 16.596 < 2e-16 \*\*\*  
## countrySuriname 1.132e+01 7.786e-01 14.540 < 2e-16 \*\*\*  
## countrySwaziland 2.695e+00 8.254e-01 3.265 0.001121 \*\*   
## countrySweden 2.046e+01 1.190e+00 17.195 < 2e-16 \*\*\*  
## countrySyrian Arab Republic 1.492e+01 7.889e-01 18.917 < 2e-16 \*\*\*  
## countryTajikistan 7.508e+00 6.729e-01 11.157 < 2e-16 \*\*\*  
## countryTimor-Leste 6.492e+00 8.385e-01 7.742 1.81e-14 \*\*\*  
## countryTogo -1.089e+00 7.931e-01 -1.373 0.170041   
## countryTonga 1.234e+01 8.045e-01 15.342 < 2e-16 \*\*\*  
## countryTrinidad and Tobago 1.195e+01 7.591e-01 15.741 < 2e-16 \*\*\*  
## countryTunisia 1.406e+01 7.985e-01 17.610 < 2e-16 \*\*\*  
## countryTurkey 1.409e+01 7.359e-01 19.141 < 2e-16 \*\*\*  
## countryTurkmenistan 5.764e+00 6.972e-01 8.267 3.04e-16 \*\*\*  
## countryUganda 2.755e-01 6.979e-01 0.395 0.693062   
## countryUkraine 1.033e+01 8.367e-01 12.348 < 2e-16 \*\*\*  
## countryUruguay 1.584e+01 8.720e-01 18.162 < 2e-16 \*\*\*  
## countryUzbekistan 8.405e+00 7.063e-01 11.901 < 2e-16 \*\*\*  
## countryVanuatu 1.261e+01 6.526e-01 19.330 < 2e-16 \*\*\*  
## countryZambia 4.267e-01 7.412e-01 0.576 0.564894   
## countryZimbabwe -1.497e+00 7.264e-01 -2.061 0.039519 \*   
## year 2.248e-01 1.738e-02 12.936 < 2e-16 \*\*\*  
## statusDeveloping NA NA NA NA   
## adult.mortality -5.701e-04 5.447e-04 -1.047 0.295452   
## alcohol -7.149e-02 3.033e-02 -2.357 0.018543 \*   
## percentage.expenditure -7.430e-05 1.232e-04 -0.603 0.546505   
## hepatitis.b 3.030e-03 2.407e-03 1.259 0.208317   
## measles -6.374e-06 6.418e-06 -0.993 0.320836   
## bmi -1.988e-03 3.429e-03 -0.580 0.562137   
## total.expenditure -2.442e-02 2.660e-02 -0.918 0.358848   
## diphtheria 1.257e-03 2.844e-03 0.442 0.658657   
## hiv.aids -3.010e-01 1.571e-02 -19.164 < 2e-16 \*\*\*  
## gdp 1.348e-05 1.859e-05 0.725 0.468451   
## population -4.582e-11 8.525e-10 -0.054 0.957143   
## income.composition.of.resources 1.096e+00 6.160e-01 1.780 0.075310 .   
## schooling 2.805e-01 7.907e-02 3.548 0.000400 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.669 on 1468 degrees of freedom  
## Multiple R-squared: 0.9677, Adjusted R-squared: 0.9645   
## F-statistic: 305.3 on 144 and 1468 DF, p-value: < 2.2e-16

*Inference - After removing the columns that were highly collinear and adding the country column in the analysis we can see that the model fit increased. With the Multiple R-squared value 0.9677 we expect 96 percent of the variation in life expectancy is explained by the variation in the independent variables like country, alcohol consumption and others. We can also see that the models p value which is quite significant hence we can reject the null hypothesis which states that there is no linear relation between the independent and dependent variable ie Life expectancy.*

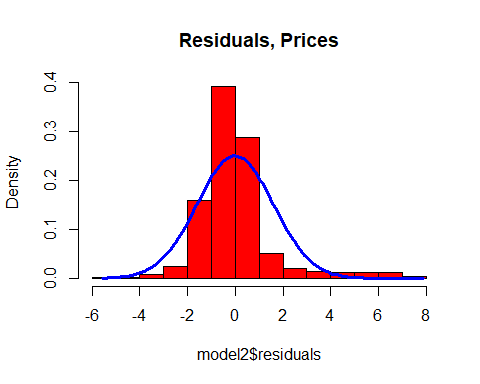
*4) Determine whether our model meets the LINE assumptions of regression.*

#Linearity Assumption  
  
plot(data1$life.expectancy,model2$fitted.values,pch=19,main="Actuals v. Fitted")  
abline(0,1,col="red",lwd=3)



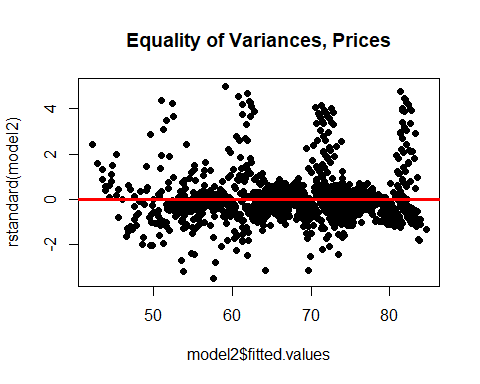
*Inference - From the plot we can see that there might be a linear relation although there are few outliers*

#Normality assumption  
  
hist(model2$residuals,col="red",main="Residuals, Prices",probability=TRUE)  
curve(dnorm(x,mean(model2$residuals), sd(model2$residuals)),from=min(model2$residuals),to=max(model2$residuals),lwd=3,col="blue",add=TRUE)



*Inference - From the histogram we can see that our model confirms to normality assumption*

#Equality of Variances  
  
plot(model2$fitted.values,rstandard(model2), pch=19,main="Equality of Variances, Prices")  
abline(0,0,lwd=3,col="red")



*Inference - The equality of variances assumption cannot be confirmed as the points are quite scattered away. Hence we cannot confirm the assumption*