Regression

Chinki

April 23, 2017

# Installing package   
# install.packages("alr4") If required   
# library(alr4) If reqired

#Reading the data from package   
UN11=alr4::UN11

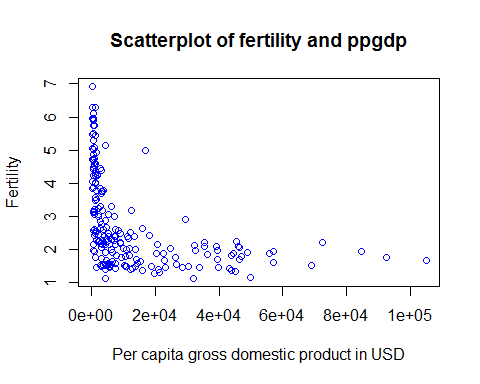
Un11 is the data of National health, welfare, and education statistics for 210 places, mostly UN members, but also other areas like Hong Kong that are not independent countries.

#Code to get discription of dataset  
help("UN11")

## starting httpd help server ...

## done

#Scatterplot of fertility and ppgdp  
plot(UN11$fertility~UN11$ppgdp,xlab="Per capita gross domestic product in USD",ylab = "Fertility",main="Scatterplot of fertility and ppgdp",col="blue")

 From the scatterplot, It is easy to conclude that fertility and ppgdp is not linear.

# Correlation test   
cor.test(UN11$ppgdp,UN11$fertility)

##   
## Pearson's product-moment correlation  
##   
## data: UN11$ppgdp and UN11$fertility  
## t = -6.877, df = 197, p-value = 7.903e-11  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.5456842 -0.3205140  
## sample estimates:  
## cor   
## -0.4399891

From the boxcox, We can get transformation. We will use log transformation.

#Tranforming log transformation  
y=log(UN11$fertility)  
x=log(UN11$ppgdp)

result1=lm(y~x)  
summary(result1)

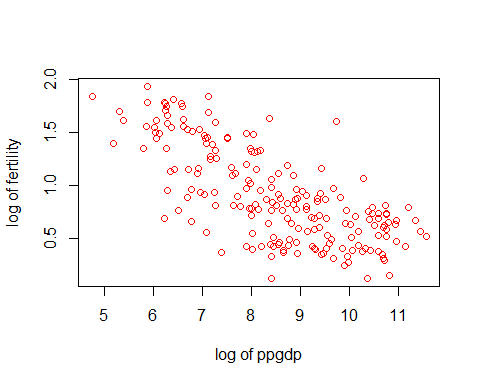
##   
## Call:  
## lm(formula = y ~ x)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.79828 -0.21639 0.02669 0.23424 0.95596   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.66551 0.12057 22.11 <2e-16 \*\*\*  
## x -0.20715 0.01401 -14.79 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3071 on 197 degrees of freedom  
## Multiple R-squared: 0.526, Adjusted R-squared: 0.5236   
## F-statistic: 218.6 on 1 and 197 DF, p-value: < 2.2e-16

y=2.66551+(-0.20715)\*x

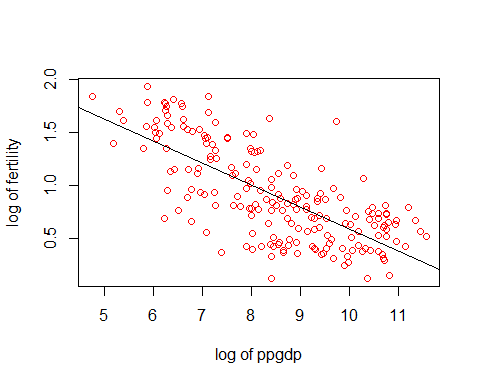
Next we will look for dignosis of the model. 1).. Normality test

r=residuals(result1)  
rs=rstandard(result1)  
f=fitted.values(result1)

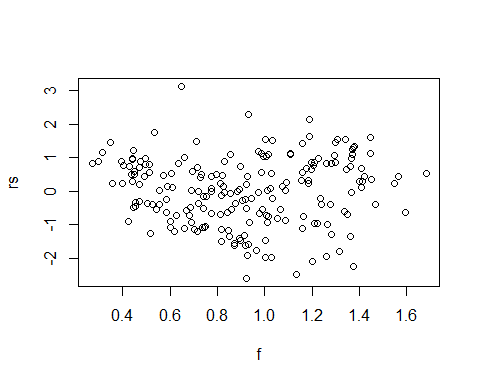
plot(y~x,xlab="log of ppgdp",ylab="log of fertility",col="red")



plot(y~x,xlab="log of ppgdp",ylab="log of fertility",col="red")  
abline(lm(y~x))



#plot of residuals and fitted value   
plot(rs~f)



#Shapiro test for normality   
shapiro.test(rs)

##   
## Shapiro-Wilk normality test  
##   
## data: rs  
## W = 0.9905, p-value = 0.2145

As p value is greater than 0.05.So conclude that formality followed by residuals as well as y.

#Brown forsytha test for equal variance  
result1$group=ifelse(x<8,c("1"),c("2"))  
result1$group

## [1] "1" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "1" "2" "2" "2"  
## [18] "2" "1" "2" "1" "1" "2" "2" "2" "2" "2" "1" "1" "1" "1" "2" "2" "2"  
## [35] "1" "1" "2" "2" "2" "1" "1" "2" "2" "1" "2" "2" "2" "2" "1" "2" "1"  
## [52] "2" "2" "1" "2" "1" "2" "2" "1" "2" "1" "2" "2" "2" "2" "2" "1" "1"  
## [69] "2" "1" "2" "2" "2" "1" "1" "1" "2" "1" "1" "2" "2" "2" "1" "1" "2"  
## [86] "1" "2" "2" "2" "2" "2" "2" "2" "1" "1" "2" "1" "1" "2" "2" "1" "1"  
## [103] "2" "2" "2" "2" "1" "1" "2" "2" "1" "2" "2" "1" "2" "2" "1" "1" "1"  
## [120] "2" "1" "1" "1" "2" "2" "1" "2" "2" "2" "2" "1" "1" "1" "1" "2" "2"  
## [137] "1" "2" "1" "2" "1" "1" "2" "1" "2" "2" "2" "2" "2" "2" "2" "1" "2"  
## [154] "2" "1" "2" "1" "2" "2" "1" "2" "2" "2" "1" "1" "2" "2" "1" "2" "1"  
## [171] "2" "2" "2" "2" "1" "1" "1" "2" "2" "1" "2" "2" "2" "2" "2" "2" "1"  
## [188] "2" "2" "2" "2" "2" "1" "1" "2" "1" "1" "1" "1"

# install.packages("lawstat") if required   
library(lawstat)

levene.test(r,result1$group)

##   
## modified robust Brown-Forsythe Levene-type test based on the  
## absolute deviations from the median  
##   
## data: r  
## Test Statistic = 0.40048, p-value = 0.5276

As p value is greter than 0.05. Equal varinace assumption is satisfied.

#constant variance test  
 # install.packages("lmtest")  
library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

##   
## Attaching package: 'lmtest'

## The following object is masked from 'package:VGAM':  
##   
## lrtest

library(car)

##   
## Attaching package: 'car'

## The following object is masked from 'package:lawstat':  
##   
## levene.test

## The following object is masked from 'package:VGAM':  
##   
## logit

ncvTest(result1)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 2.696278 Df = 1 p = 0.1005828

As p value is grester than 0.05. The constant variance assumption is satisfied.

so all the assumptions are good and satisfied. This is perfect example for regression.Transformation is working good.