RegressionWeek9

Chris Kalra

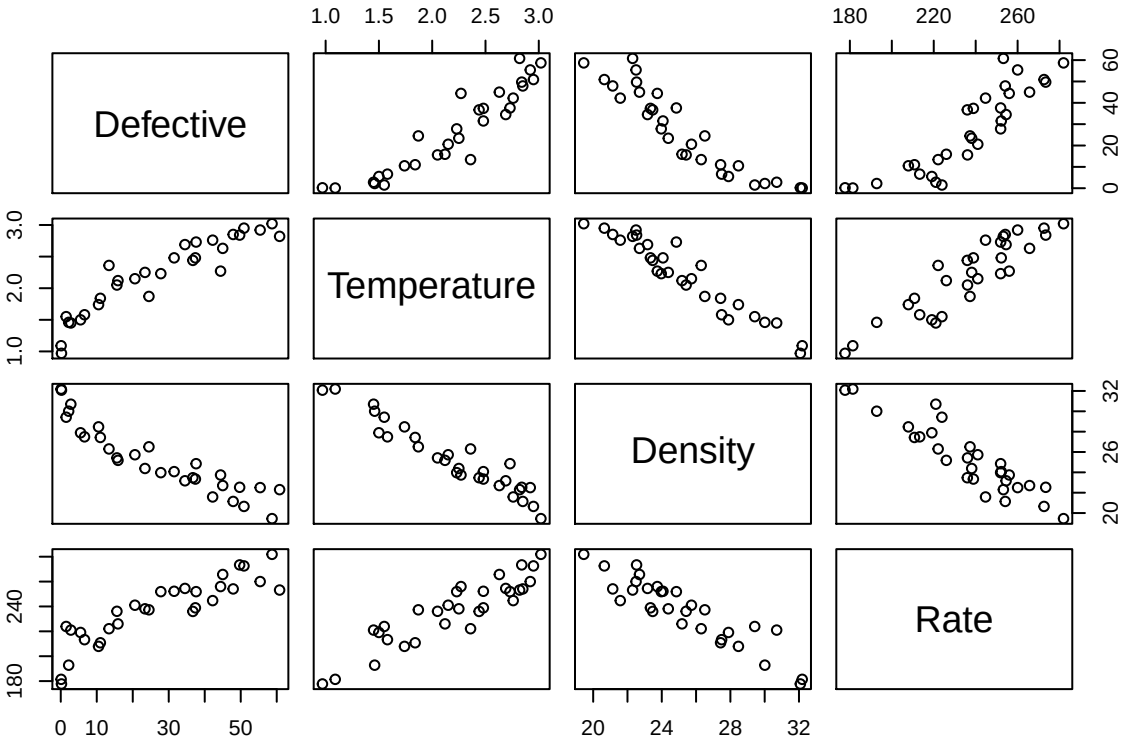
3/19/2019

Box Cox transformation

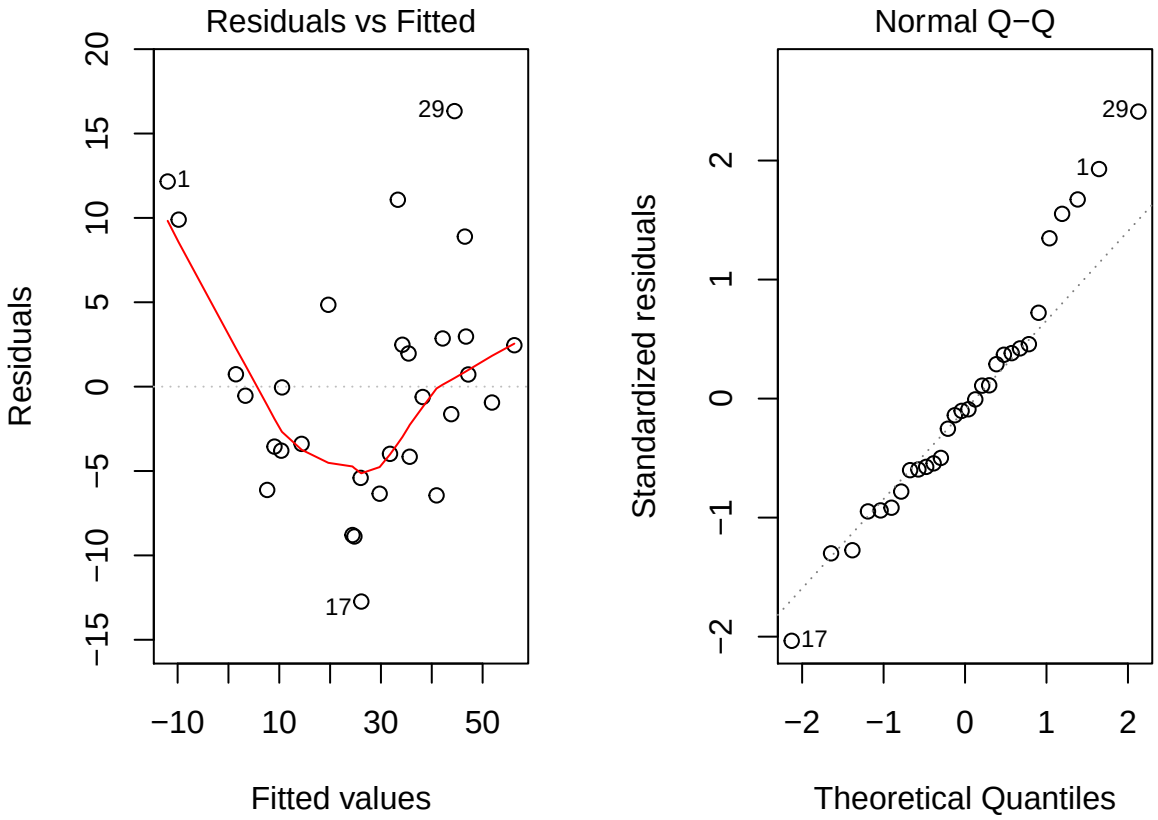
```
data_url = "https://www.stat.tamu.edu/~sheather/book/docs/datasets/defects.txt"
defects=read.table(data_url, header=T) # ; defects
library(MASS)
library(alr4)
```

```
## Loading required package: car
## Loading required package: carData
## Loading required package: effects
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
pairs(Defective ~ Temperature + Density + Rate, data=defects) # Correlation Matrix
```




```
lm1 <- lm(Defective ~ Temperature + Density + Rate, data=defects)
par(mfrow=c(1,2), mar=c(4.5, 4.5, 2, 2))
plot(lm1, c(1:2))
```




```
boxcox(lm1, lambda=seq(0.3,0.65,by=0.05))
summary(powerTransform(lm1)) # lambda{hat} = 0.4519, round to lambda{hat} = 0.5
```

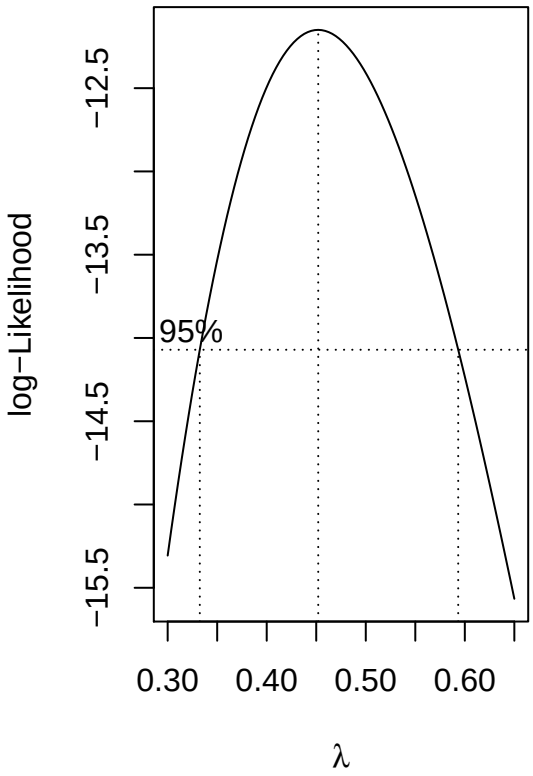
```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    0.4519         0.5       0.3253       0.5785
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                            LRT df      pval
## LR test, lambda = (0) 50.12714  1 1.441e-12
##
## Likelihood ratio test that no transformation is needed
##                            LRT df       pval
## LR test, lambda = (1) 30.03972  1 4.2329e-08
```

```
# str(powerTransform(lm1))   to analyze the structure in order to extract relevant values
powerTransform(lm1)$lambda  # lambda{hat} = 0.4519118
```

```
##        Y1
## 0.4519214
```

```
powerTransform(lm1)$roundlam    # Suggested rounding of lambda{hat} to 0.5
```

```
##  Y1
## 0.5
```

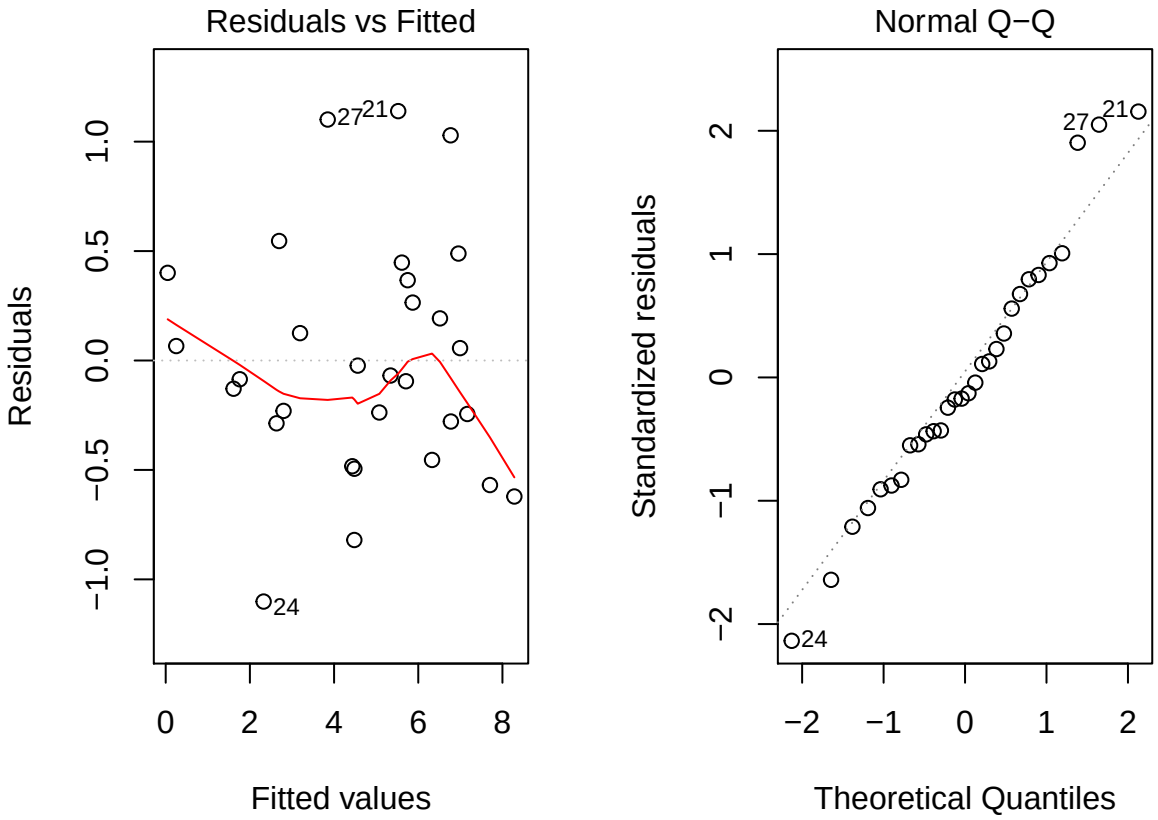



```
lm2 <- lm(sqrt(Defective) ~ Temperature + Density + Rate,data=defects)
# Since We are using lambda{hat} = 0.5, we use the square root of Defective

summary(lm2)
```

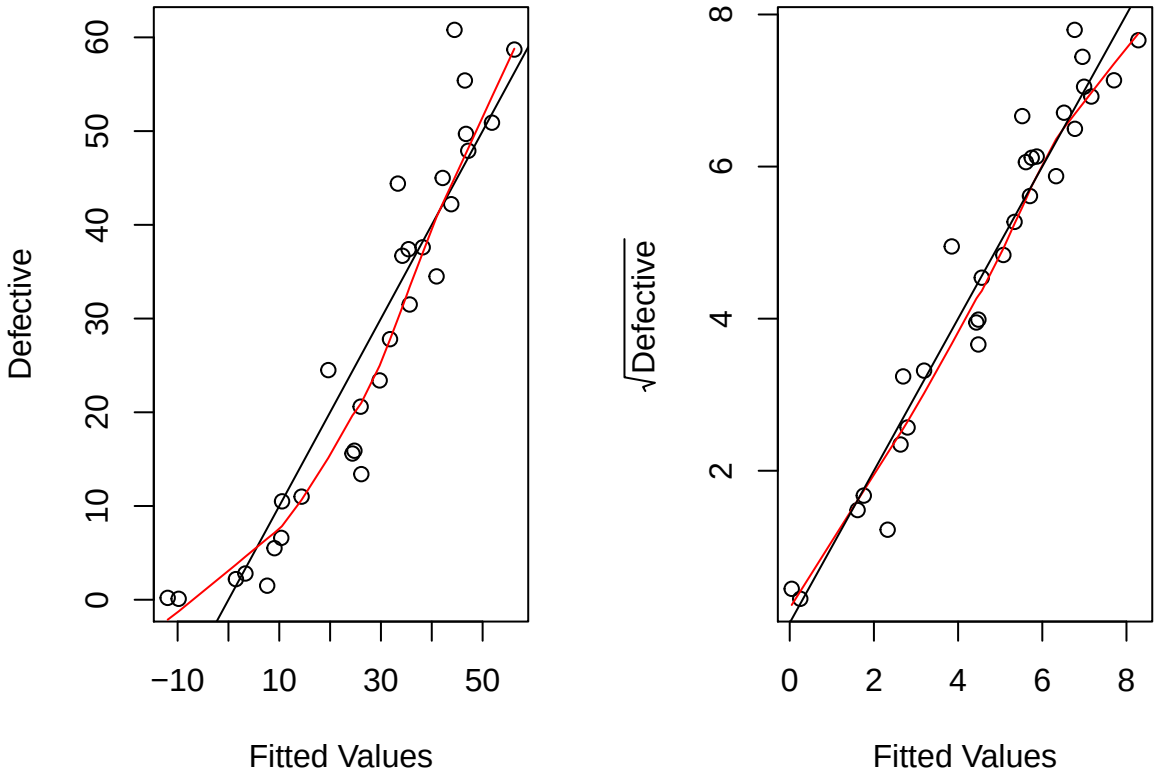
```
##
## Call:
## lm(formula = sqrt(Defective) ~ Temperature + Density + Rate,
##     data = defects)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10147 -0.28502 -0.07716  0.34139  1.13951
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.59297    5.26401   1.062   0.2978
## Temperature  1.56516    0.66226   2.363   0.0259 *
## Density     -0.29166    0.11954  -2.440   0.0218 *
## Rate         0.01290    0.01043   1.237   0.2273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5677 on 26 degrees of freedom
## Multiple R-squared:  0.943,  Adjusted R-squared:  0.9365
## F-statistic: 143.5 on 3 and 26 DF,  p-value: 2.713e-16
```

```
par(mfrow=c(1,2), mar=c(4.5, 4.5, 2, 2))
plot(lm2, c(1:2))
```




```
par(mfrow=c(1,2), mar=c(4.5, 4.5, 2, 2))
plot(predict(lm1), defects$Defective,xlab = "Fitted Values", ylab = "Defective")
abline(0,1) ; lines(lowess(predict(lm1), defects$Defective), col='red')

plot(predict(lm2), sqrt(defects$Defective),xlab = "Fitted Values", ylab =
      expression(sqrt(Defective)))
lines(lowess(predict(lm2), sqrt(defects$Defective)), col='red') ; abline(0,1)
```



Box Cox: transforming multiple predictor variables simultaneously

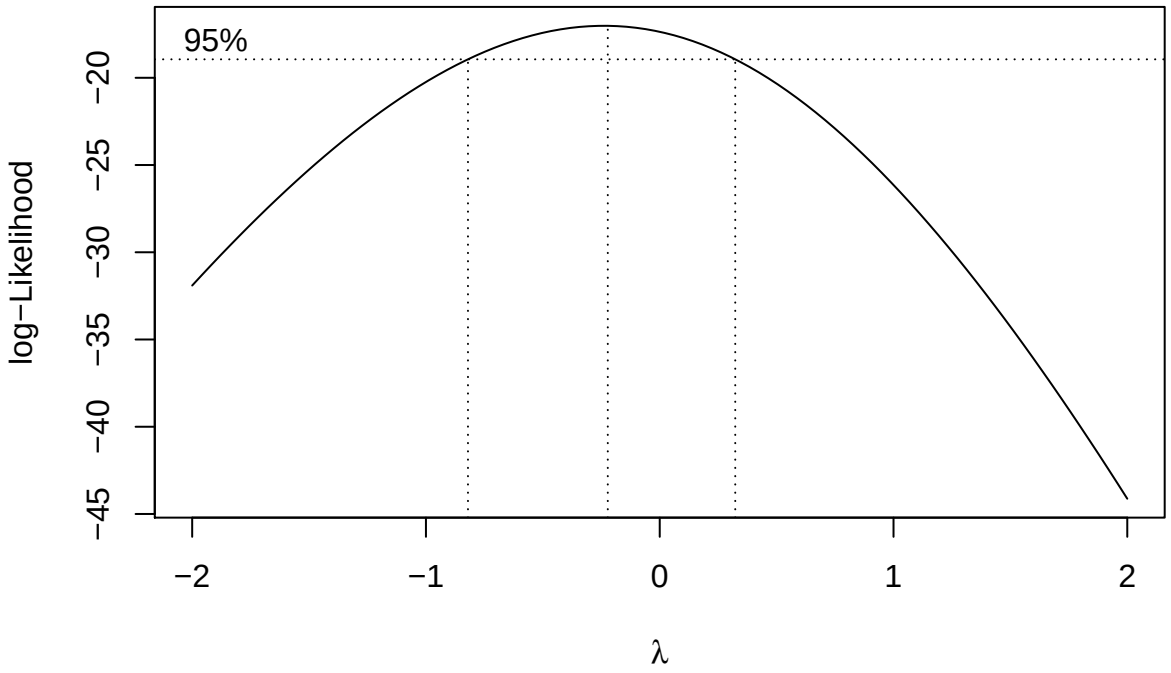
```
Highway$sigs1 <- with(Highway, (sigs * len + 1)/len)
# Transforming "sigs" in order to ensure all values are positive

summary(powerTransform(cbind(len, adt, trks, shld, sigs1) ~ 1,Highway))
```

```
## bcPower Transformations to Multinormality
##       Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## len      0.1437           0      -0.2732       0.5607
## adt      0.0509           0      -0.1854       0.2872
## trks    -0.7028           0      -1.9134       0.5078
## shld     1.3456           1       0.6341       2.0570
## sigs1   -0.2408           0      -0.5341       0.0525
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                     LRT df      pval
## LR test, lambda = (0 0 0 0 0) 23.32447  5 0.0002926
##
## Likelihood ratio test that no transformations are needed
##                                     LRT df       pval
## LR test, lambda = (1 1 1 1 1) 132.8574  5 < 2.22e-16
```

```
# Notice that 4/5 variables have rounded lambda{hat} values = 0,
# so they should be transformed via the "log()" function, whereas "shld"
# need not be transformed, as it has a rounded lambda{hat} value of 1
```

```
lm3 <- lm(rate ~ log(len) +log(adt) + log(trks) +slim + shld + log(sigs1), data=Highway)
boxcox(lm3)
```




```
summary(powerTransform(lm3))
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1   -0.2384           0       -0.805       0.3282
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                               LRT df    pval
## LR test, lambda = (0) 0.6884455  1 0.40669
##
## Likelihood ratio test that no transformation is needed
##                             LRT df       pval
## LR test, lambda = (1) 18.2451  1 1.9422e-05
```

```
# Notice that 0 is the rounded estimate, so the log of rate should be taken

lm4 <- lm(log(rate) ~ log(len) + log(adt) + log(trks) +slim + shld + log(sigs1), data=Highway)
summary(lm4)
```

```
##
## Call:
## lm(formula = log(rate) ~ log(len) + log(adt) + log(trks) + slim +
##     shld + log(sigs1), data = Highway)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38998 -0.16631 -0.02273  0.18706  0.62695
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.44433    0.69627   6.383  3.6e-07 ***
## log(len)    -0.26028    0.09684  -2.688   0.0113 *
## log(adt)    -0.05375    0.05792  -0.928   0.3603
## log(trks)   -0.33622    0.22777  -1.476   0.1497
## slim        -0.02598    0.01366  -1.901   0.0663 .
## shld        -0.02082    0.02514  -0.828   0.4137
## log(sigs1)   0.08000    0.05618   1.424   0.1641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2759 on 32 degrees of freedom
## Multiple R-squared:  0.7008, Adjusted R-squared:  0.6447
## F-statistic: 12.49 on 6 and 32 DF,  p-value: 3.252e-07
```