Box-Cox-Transformation01

SP

16/10/2019

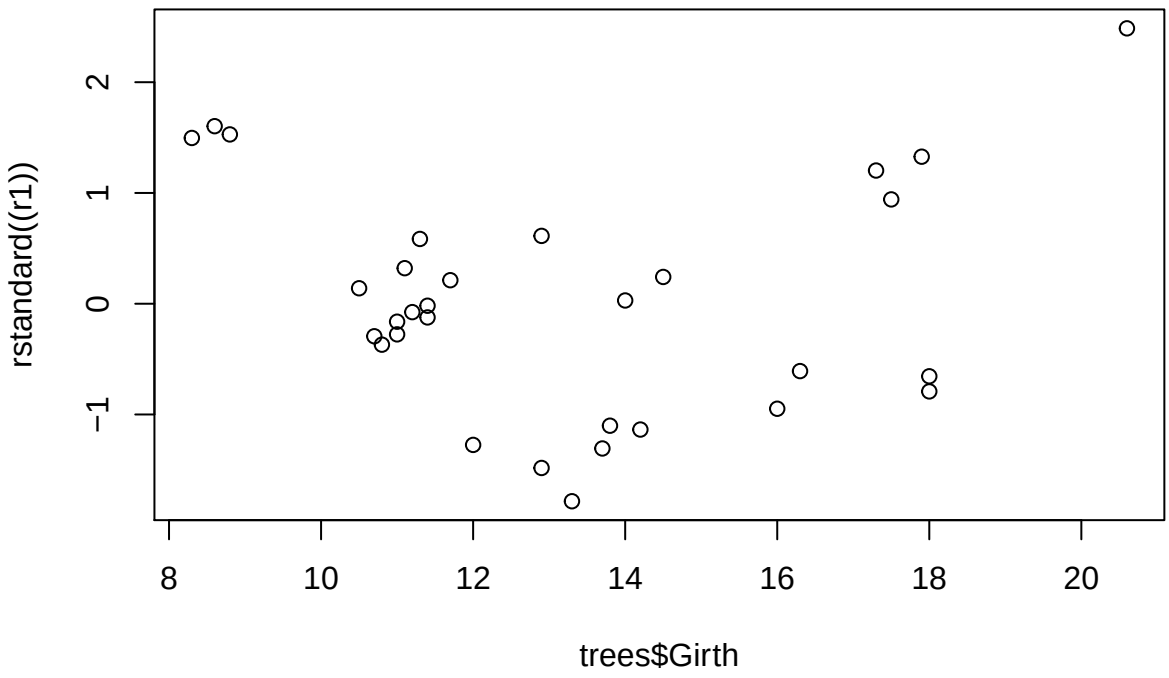
```
head(trees)
```

```
##   Girth Height Volume
## 1   8.3     70   10.3
## 2   8.6     65   10.3
## 3   8.8     63   10.2
## 4  10.5     72   16.4
## 5  10.7     81   18.8
## 6  10.8     83   19.7
```

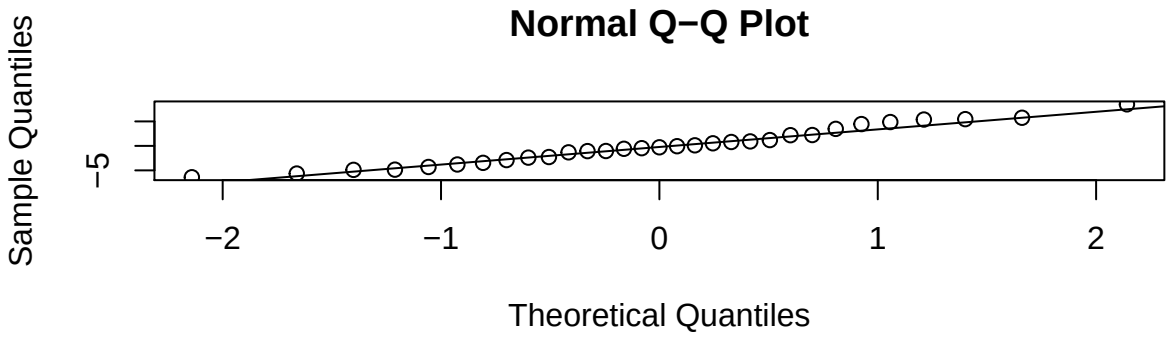
```
r1 = lm(Volume~Height+Girth, trees)
summary(r1)
```

```
##
## Call:
## lm(formula = Volume ~ Height + Girth, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4065 -2.6493 -0.2876  2.2003  8.4847
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -57.9877     8.6382  -6.713 2.75e-07 ***
## Height        0.3393     0.1302   2.607   0.0145 *
## Girth         4.7082     0.2643  17.816  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.882 on 28 degrees of freedom
## Multiple R-squared:  0.948,  Adjusted R-squared:  0.9442
## F-statistic:   255 on 2 and 28 DF,  p-value: < 2.2e-16
```

```
plot(trees$Girth, rstandard((r1)))
```



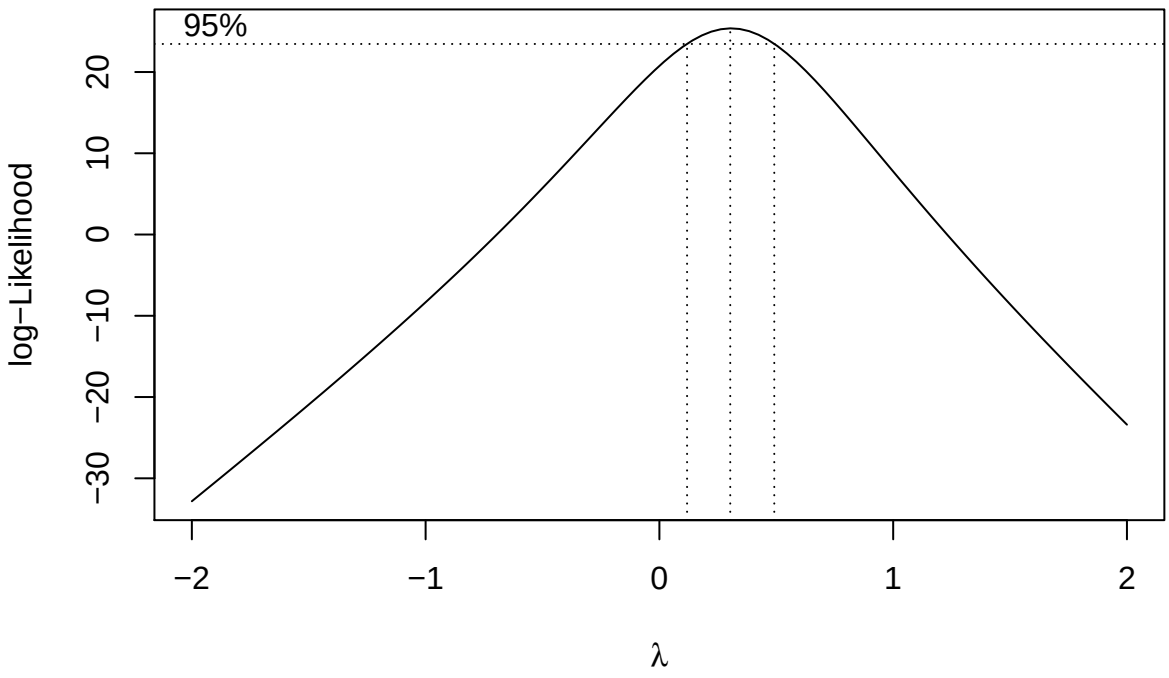

```
par(mfrow=c(2,1))
hist(r1$residuals)
qqnorm(r1$residuals)
qqline(r1$residuals)
```




```
library(moments)
skewness(r1$residuals)
```

```
## [1] 0.3102985
```

```
library(MASS)
b=boxcox(Volume~Height+Girth, data=trees)
```




```
#b
```

```
lmda = b$x
lik = b$y
bc=cbind(lmda,lik)
#bc
```

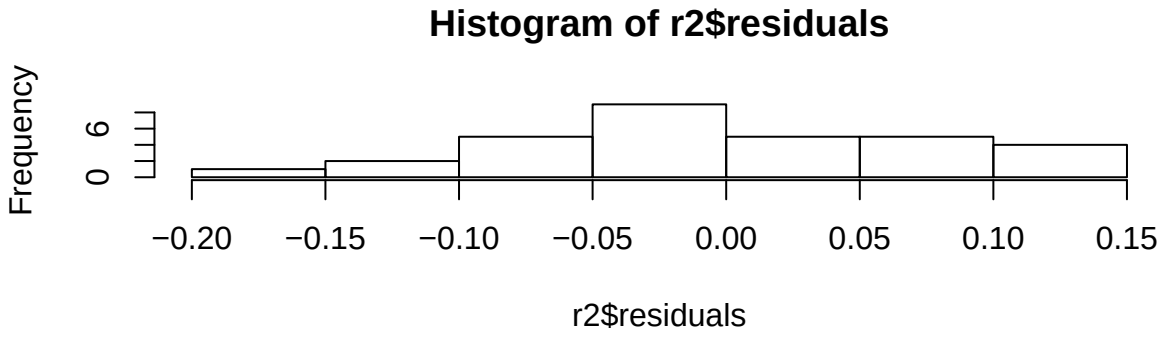
```
#bc[order(lik),]
```

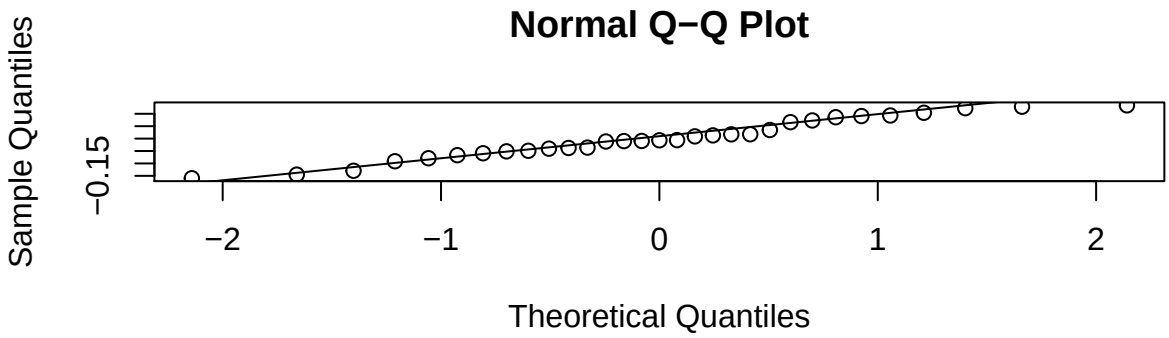
```
# from the above box-cox plot, the lamba is around 0.3 ( i.e 1/3)
r2 = lm(Volume^(1/3)~Height+Girth, trees)
summary(r2)
```

```
##
## Call:
## lm(formula = Volume^(1/3) ~ Height + Girth, data = trees)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.159602 -0.050200 -0.006827  0.069649  0.133981
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.085388   0.184315  -0.463    0.647
## Height       0.014472   0.002777   5.211 1.56e-05 ***
```

```
## Girth        0.151516   0.005639  26.871  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08283 on 28 degrees of freedom
## Multiple R-squared:  0.9777, Adjusted R-squared:  0.9761
## F-statistic: 612.5 on 2 and 28 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,1))
hist(r2$residuals)
qqnorm(r2$residuals)
qqline(r2$residuals)
```






```
skewness(r2$residuals)
```

```
## [1] -0.06991731
```

```r
plot(trees$Girth, rstandard((r2)))
```