#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#

#Author: Vivek Gupta, Date 4/1/17

#Purpose: To Demonstrate box cox and keep it records for reference.

#Great tool for feature engineering or running statistical analysis on data

#careful on running this for CI about mean and sigma, retransforms dont work mathematically

y1 <- sort(c(19.7,21.6,21.9,23.5,24.2,24.4,24.9,25.1,

26.4,26.9,27.6,27.7,27.9,28.4,29.8,30.7,

31.1,31.1,31.7,31.8,32.6,34.0,34.8,34.9,

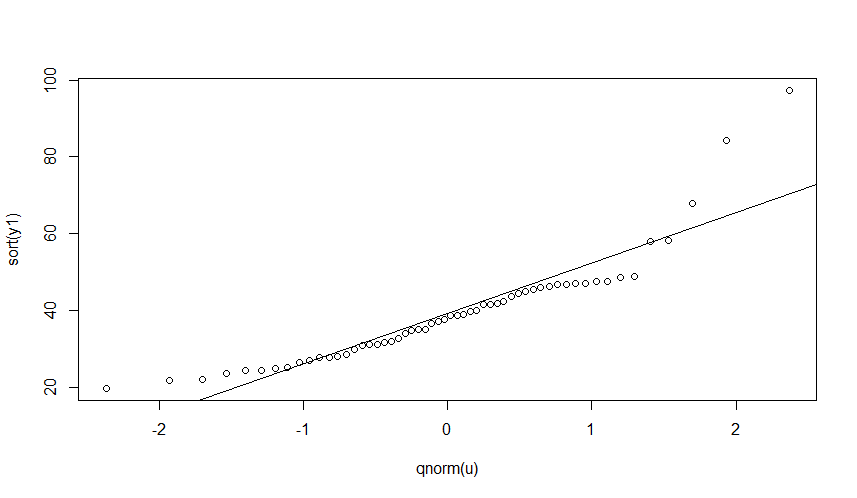
35.1,36.6,37.0,37.7,38.7,38.7,39.0,39.6,

40.0,41.4,41.4,41.8,42.2,43.5,44.5,45.0,

45.5,45.9,46.3,46.7,46.7,47.0,47.0,47.4,

47.6,48.6,48.8,57.9,58.3,67.9,84.2,97.3))

Reference dist plot:

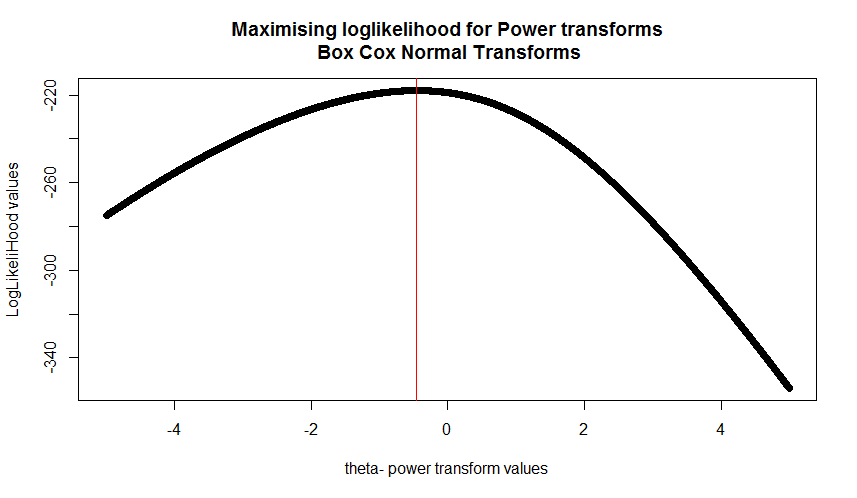


Shapiro-Wilk normality test

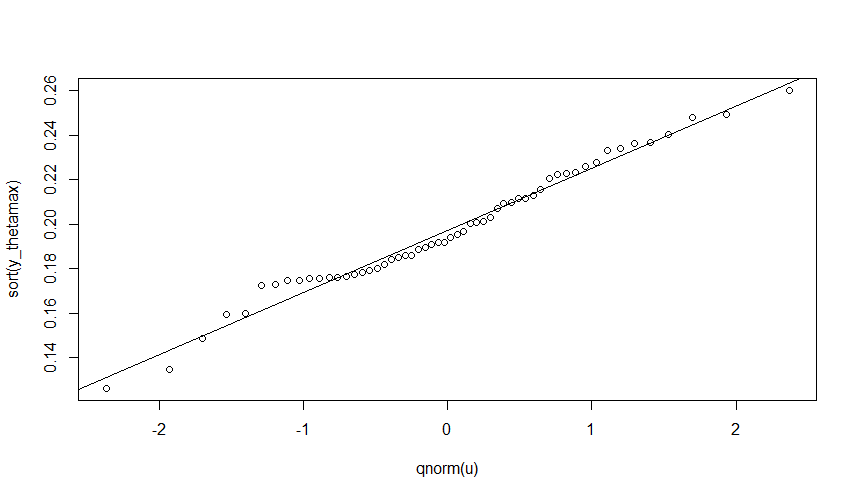
data: y1

W = 0.85279, p-value = 6.915e-06

Get the transform:



Reference dist plot again after transform



Shapiro-Wilk test

Shapiro-Wilk normality test

data: y\_thetamax

W = 0.98113, p-value = 0.5256

Use the transform to get a Tolerance interval and retransform back

# Box Cox

(mean(y\_thetamax)-(2.555\*sd(y\_thetamax)))^(1/theta\_max) #U p,gamma with Box Cox

(mean(y\_thetamax)+(2.555\*sd(y\_thetamax)))^(1/theta\_max) #L p,gamma with Box Cox

#51.28032 24834.69