Stat 415 Regression: Classwork/Lab 4

Frankie Tyndall

2022-07-11

## Electtricty Consumption

y = household electricity comsumption

x = Number of rooms in home

## The data

x <- c(2,3, 4, 5, 6, 7, 8, 9, 10, 11)  
x

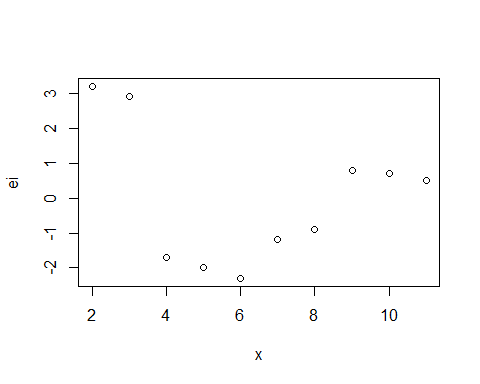
## [1] 2 3 4 5 6 7 8 9 10 11

ei <- c(3.2, 2.9, -1.7, -2.0, -2.3, -1.2, -0.9, 0.8, 0.7, 0.5)  
ei

## [1] 3.2 2.9 -1.7 -2.0 -2.3 -1.2 -0.9 0.8 0.7 0.5

## Plot Residual Vs Predictor.

plot(ei ~ x)



The problem that exist is that the plot of then residual vs x is not linear and the variance is not constant. The data is skewed to the right and the transformation that could be used is the X^2 or exp(X) transformations of the predictor variable.

## For the bivariate data given below, the residual plots suggest problems involving non-normality or non-constant error variance or both. Use Box – Cox method as indicated in class to produce a lamda power transformation that will best normalize the data. Your work should include all graphs and plots that support your final answer. Use a series of steps and reasoning demonstrated in class.

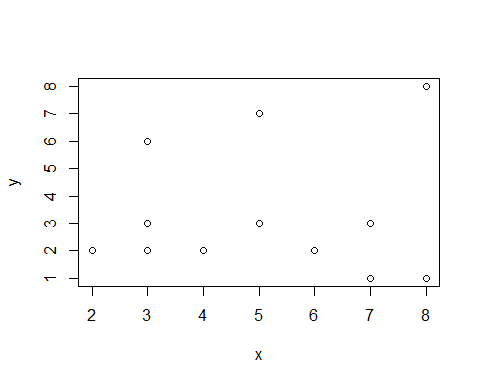
x <- c(7, 7, 8, 3, 2, 4, 4, 6, 6, 7, 5, 3, 3, 5, 8)  
x

## [1] 7 7 8 3 2 4 4 6 6 7 5 3 3 5 8

y <- c(1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 6, 7, 8)  
y

## [1] 1 1 1 2 2 2 2 2 2 3 3 3 6 7 8

plot(y ~ x)



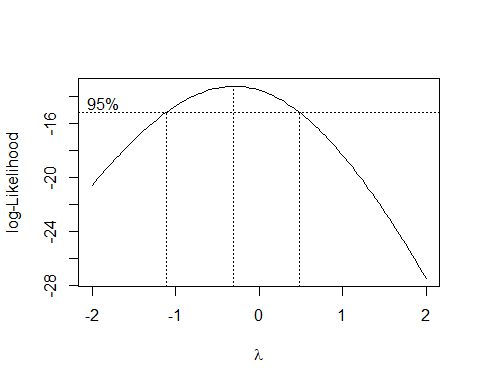
model <- lm(y ~ x)  
model

##   
## Call:  
## lm(formula = y ~ x)  
##   
## Coefficients:  
## (Intercept) x   
## 3.000e+00 2.709e-16

box\_cox <- boxcox(y ~ y)

## Warning in model.matrix.default(mt, mf, contrasts): the response appeared on the  
## right-hand side and was dropped

## Warning in model.matrix.default(mt, mf, contrasts): problem with term 1 in  
## model.matrix: no columns are assigned



box\_cox

## $x  
## [1] -2.00000000 -1.95959596 -1.91919192 -1.87878788 -1.83838384 -1.79797980  
## [7] -1.75757576 -1.71717172 -1.67676768 -1.63636364 -1.59595960 -1.55555556  
## [13] -1.51515152 -1.47474747 -1.43434343 -1.39393939 -1.35353535 -1.31313131  
## [19] -1.27272727 -1.23232323 -1.19191919 -1.15151515 -1.11111111 -1.07070707  
## [25] -1.03030303 -0.98989899 -0.94949495 -0.90909091 -0.86868687 -0.82828283  
## [31] -0.78787879 -0.74747475 -0.70707071 -0.66666667 -0.62626263 -0.58585859  
## [37] -0.54545455 -0.50505051 -0.46464646 -0.42424242 -0.38383838 -0.34343434  
## [43] -0.30303030 -0.26262626 -0.22222222 -0.18181818 -0.14141414 -0.10101010  
## [49] -0.06060606 -0.02020202 0.02020202 0.06060606 0.10101010 0.14141414  
## [55] 0.18181818 0.22222222 0.26262626 0.30303030 0.34343434 0.38383838  
## [61] 0.42424242 0.46464646 0.50505051 0.54545455 0.58585859 0.62626263  
## [67] 0.66666667 0.70707071 0.74747475 0.78787879 0.82828283 0.86868687  
## [73] 0.90909091 0.94949495 0.98989899 1.03030303 1.07070707 1.11111111  
## [79] 1.15151515 1.19191919 1.23232323 1.27272727 1.31313131 1.35353535  
## [85] 1.39393939 1.43434343 1.47474747 1.51515152 1.55555556 1.59595960  
## [91] 1.63636364 1.67676768 1.71717172 1.75757576 1.79797980 1.83838384  
## [97] 1.87878788 1.91919192 1.95959596 2.00000000  
##   
## $y  
## [1] -20.61572 -20.31909 -20.02625 -19.73732 -19.45245 -19.17176 -18.89538  
## [8] -18.62346 -18.35613 -18.09355 -17.83587 -17.58324 -17.33584 -17.09383  
## [15] -16.85737 -16.62665 -16.40185 -16.18314 -15.97072 -15.76477 -15.56548  
## [22] -15.37306 -15.18770 -15.00958 -14.83892 -14.67592 -14.52076 -14.37365  
## [29] -14.23478 -14.10434 -13.98252 -13.86951 -13.76548 -13.67061 -13.58506  
## [36] -13.50899 -13.44255 -13.38587 -13.33910 -13.30233 -13.27569 -13.25927  
## [43] -13.25313 -13.25736 -13.27201 -13.29711 -13.33270 -13.37877 -13.43534  
## [50] -13.50237 -13.57985 -13.66772 -13.76593 -13.87441 -13.99306 -14.12181  
## [57] -14.26053 -14.40913 -14.56746 -14.73541 -14.91282 -15.09955 -15.29544  
## [64] -15.50034 -15.71409 -15.93651 -16.16743 -16.40669 -16.65410 -16.90950  
## [71] -17.17271 -17.44356 -17.72187 -18.00748 -18.30020 -18.59988 -18.90635  
## [78] -19.21944 -19.53900 -19.86487 -20.19689 -20.53492 -20.87881 -21.22842  
## [85] -21.58362 -21.94425 -22.31020 -22.68134 -23.05754 -23.43869 -23.82466  
## [92] -24.21535 -24.61064 -25.01044 -25.41463 -25.82313 -26.23583 -26.65264  
## [99] -27.07347 -27.49823

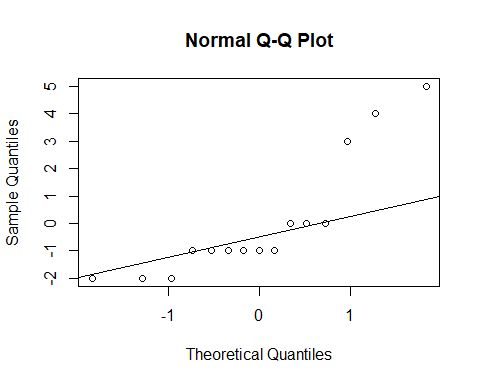
lambda <- box\_cox$x[which.max(box\_cox$y)]  
lambda

## [1] -0.3030303

# fit new linear regression  
  
new\_model <- lm(((y^lambda-1)/lambda) ~ x)  
new\_model

##   
## Call:  
## lm(formula = ((y^lambda - 1)/lambda) ~ x)  
##   
## Coefficients:  
## (Intercept) x   
## 1.06007 -0.06348

# old model  
  
qqnorm(model$residuals)   
qqline(model$residuals)



# new model   
  
qqnorm(new\_model$residuals)   
qqline(new\_model$residuals)

