Ex 1.50

Tanuj Guha

2023-02-25

library(Stat2Data)  
data("CountyHealth")  
head (CountyHealth)

## County MDs Hospitals Beds  
## 1 Bay, FL 351 3 605  
## 2 Beaufort, NC 95 2 134  
## 3 Beaver, PA 260 2 567  
## 4 Bernalillo, NM 2797 11 1435  
## 5 Bibb, GA 769 5 976  
## 6 Clinton, PA 42 2 245

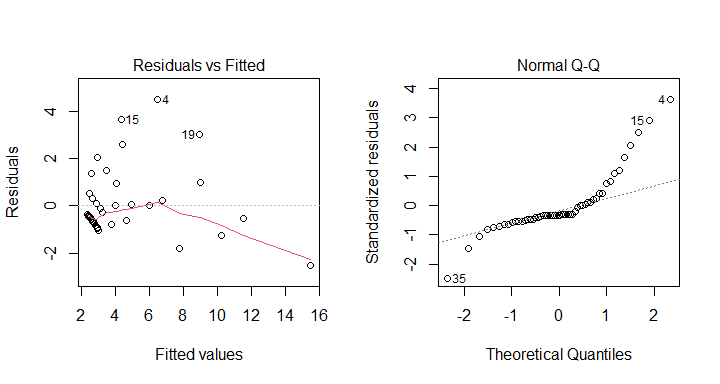
CountyHealth$LogMD = log(CountyHealth$MDs)

TestLM = lm (Hospitals ~ MDs, data=CountyHealth)  
summary (TestLM)

##   
## Call:  
## lm(formula = Hospitals ~ MDs, data = CountyHealth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5122 -0.5921 -0.4105 0.1258 4.5132   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.356e+00 2.034e-01 11.58 6.8e-16 \*\*\*  
## MDs 1.477e-03 9.587e-05 15.41 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.271 on 51 degrees of freedom  
## Multiple R-squared: 0.8231, Adjusted R-squared: 0.8197   
## F-statistic: 237.3 on 1 and 51 DF, p-value: < 2.2e-16

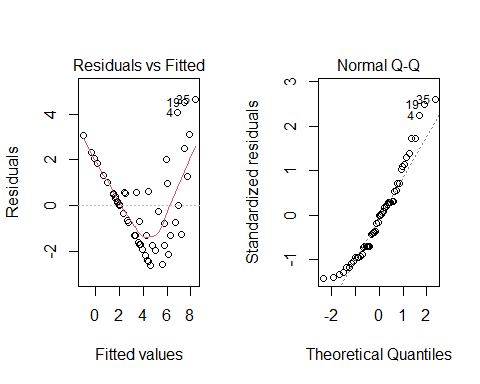
We know that the number of MDs without transformation don’t provide a good enough linear model. Let us see what it looks like:

par (mfrow = c(1, 2))  
plot (TestLM, which=1:2)



Now let us create a model iwth the log transformed MDs, and check if the tranformation helped.

TestLogLM = lm (Hospitals ~ LogMD, data=CountyHealth)  
par (mfrow = c(1, 2))  
plot (TestLogLM, which=1:2)



The log transformation is most certainly a better step: it decreases obvious outliers as indicated by the Q-Q plot meaning their distributions are now better matched. However, the Residuals/Fitted graph indicates the possibility of heteroskedasticity.