> model <- lm(disp ~ mpg , data=mtcars)

> summary(model)

Call:

lm(formula = disp ~ mpg, data = mtcars)

Residuals:

Min 1Q Median 3Q Max

-103.05 -45.74 -8.17 46.65 153.75

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 580.884 41.740 13.917 1.26e-14 \*\*\*

mpg -17.429 1.993 -8.747 9.38e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 66.86 on 30 degrees of freedom

Multiple R-squared: 0.7183, Adjusted R-squared: 0.709

F-statistic: 76.51 on 1 and 30 DF, p-value: 9.38e-10

> check=function(lm.object,tests=FALSE){ ### Checking model assumptions

+ re=rstandard(lm.object) #standardized residuals

+ ylimits=c(-3,3)

+ ylimits[1]=ifelse(min(re)<(-3),min(re),-3)

+ ylimits[2]=ifelse(max(re)>(3),max(re),3)

+

+ par(mfrow=c(2,2))

+

+ # Normality

+ hist(re,main="Standardized residuals",xlab="std res.")

+ qqnorm(re,datax=TRUE)

+ qqline(re,datax=TRUE)

+

+ # Homogeneity of variance/Model Fit

+ plot(re~fitted.values(lm.object),xlab=expression(hat(y)),ylab="std res.",main="Homogeneity / Fit",ylim=ylimits)

+ abline(h=0)

+

+ # Independence

+ plot(re,type="o",pch=22,xlab="Order",ylab="std res.",main="Independence",ylim=ylimits)

+ abline(h=0)

+

+ par(mfrow=c(1,1))

+ if(tests==TRUE){

+ if(!require(lawstat)){

+ print("trying to install lawstat")

+ install.packages("lawstat",dependencies=TRUE)

+ if(!require(lawstat))stop("could not install lawstat")

+ }

+ if(!require(car)){

+ print("trying to install car")

+ install.packages("car",dependencies=TRUE)

+ if(!require(car))stop("could not install car")

+ }

+ ConstantVar=NULL

+ if(class(lm.object)[1]=="lm")ConstantVar=ncvTest(lm.object)

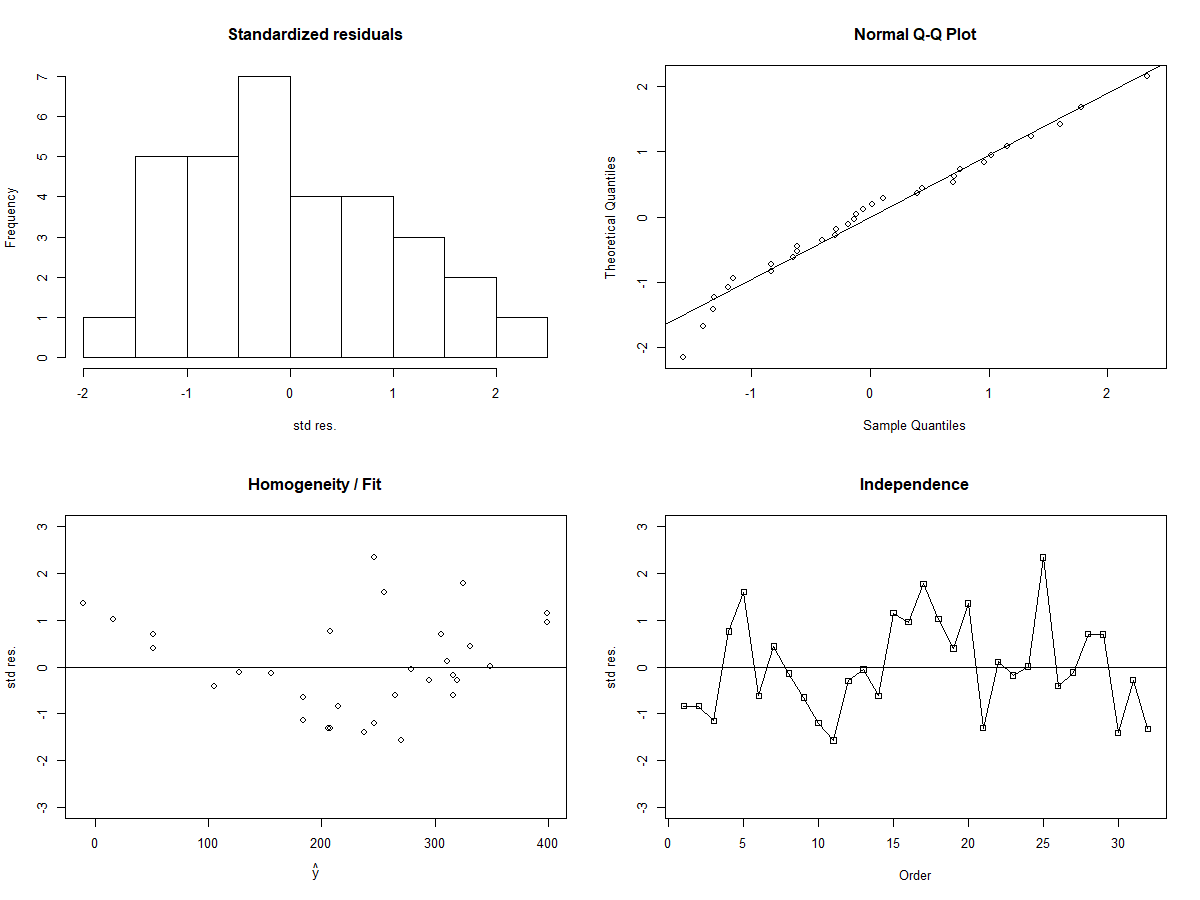
+ if(class(lm.object)[1]=="aov")ConstantVar=leveneTest(lm.object$model[,1],lm.object$model[,2])

+ list(Independence=list(runs.test(re),durbinWatsonTest(lm.object)),Normality=shapiro.test(re),c("Constant Variance only valid if data are in groups"),ConstantVar=ConstantVar)

+ }

+ }

> check(model)



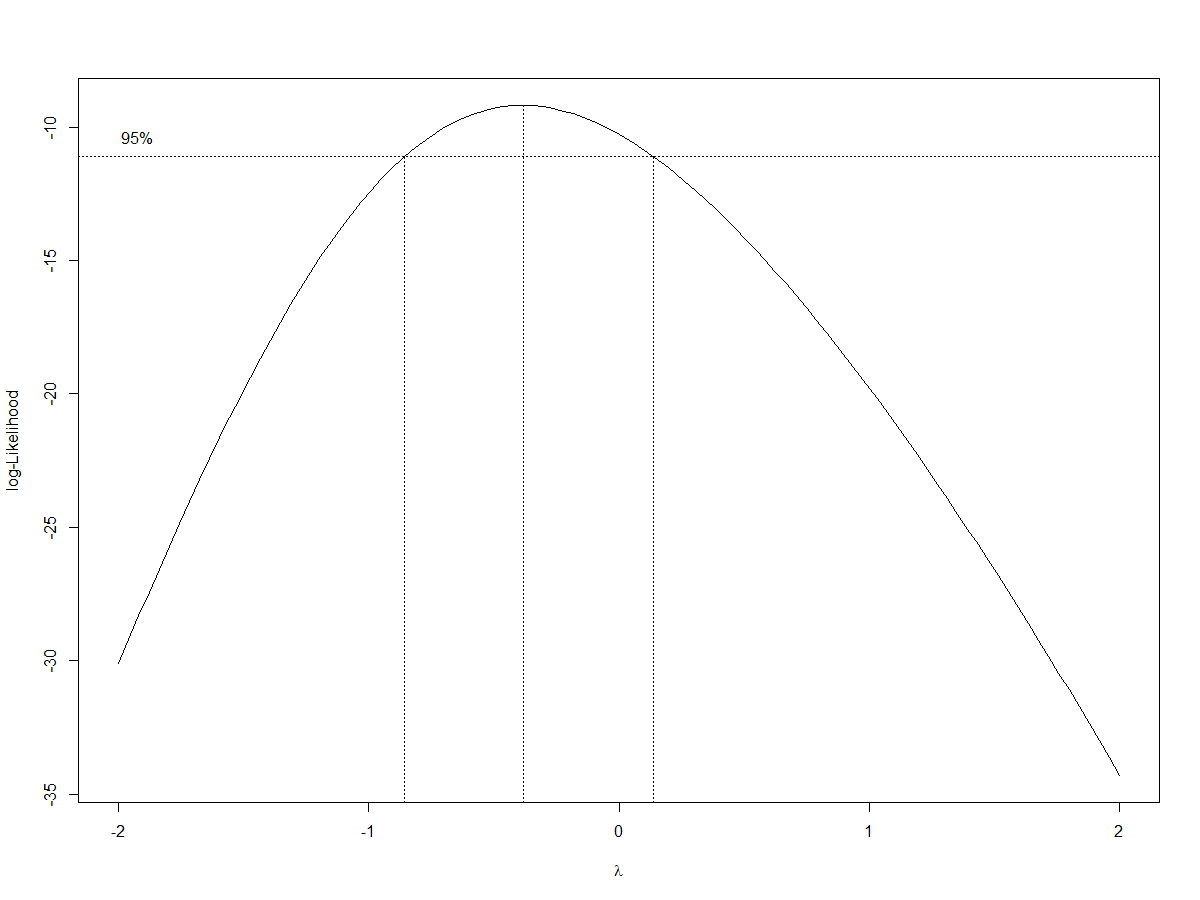
Normality - Based on the histogram and Q-Q plot, we cannot safely assume normality. In particular, the bell shape we are looking for in the histogram looks skewed. We will err on the side of caution and not make such an assumption.

Independence - The time series plot appears to show no discernible pattern, so we will assume independence.

Homogeneity / Model Fit - The residuals vs. fitted values plot does not show a constant spread of residuals near the 0 line. We will not assume homogeneity of variance or appropriate model fit.

Since at least 2 assumptions are violated, we will transform the responses using “boxcox”:

> bc <- boxcox(model)



> (lambda <- bc$x[which.max(bc$y)])

[1] -0.3838384

> mtcars$disp2 = (mtcars$disp)^lambda

> model2 = lm(disp2 ~ mpg, data = mtcars)

> summary(model2)

Call:

lm(formula = disp2 ~ mpg, data = mtcars)

Residuals:

Min 1Q Median 3Q Max

-0.0303575 -0.0058812 0.0003705 0.0060982 0.0181120

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.0399082 0.0076684 5.204 1.31e-05 \*\*\*

mpg 0.0047257 0.0003661 12.909 8.80e-14 \*\*\*

---

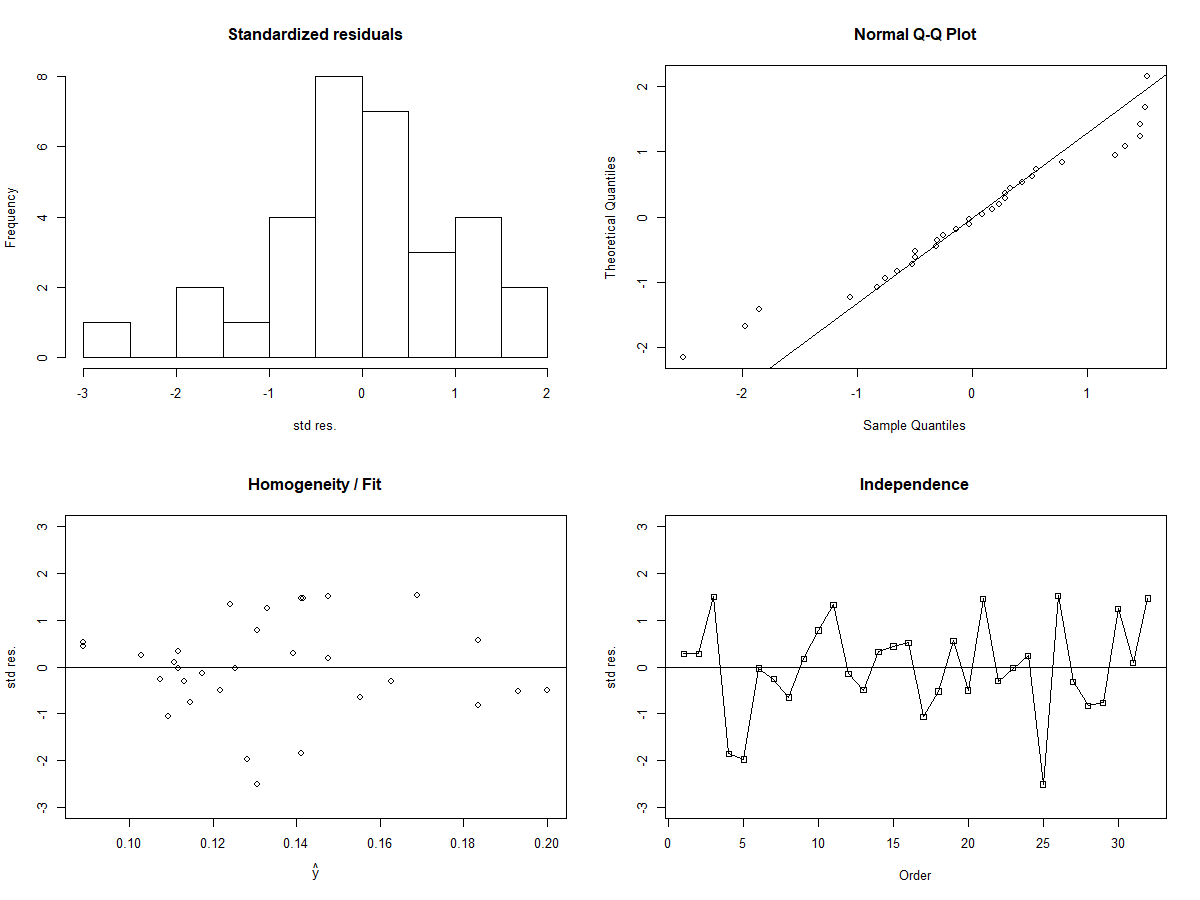
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01228 on 30 degrees of freedom

Multiple R-squared: 0.8474, Adjusted R-squared: 0.8424

F-statistic: 166.7 on 1 and 30 DF, p-value: 8.804e-14

> check(model2)



The transformation has not solved any of the previously addressed issues. If anything, we have worsened some of the previous graphical concerns. We can infer that simple linear regression does not appear to be a good model for the data we are working with.