**STAT 462 – Applied Regression Analysis**

**Fall 2017, Lab 8**

Prepare a short report with relevant output, your comments, and answers to the questions (this does not need to be exhaustive or polished, but should contain enough to show that you completed all tasks and analyses).

Submit the report at the end of the lab session.

Consider again the dataset *bears.txt* used in previous labs.

This contains several variables measured on n=141 “bear capturing” occasions, with the following variables:

*ID:* Identification number

*Age:* Bear's age, in months

*Month:* Month when the measurement was made. Sex. 1 = male 2 = female

*Head.L:* Length of the head, in inches

*Head.W:* Width of the head, in inches

*Neck.G:* Girth (distance around) the neck, in inches

*Length:* Body length, in inches

*Chest.G:* Girth (distance around) the chest, in inches

*Weight:* Weight of the bear, in pounds

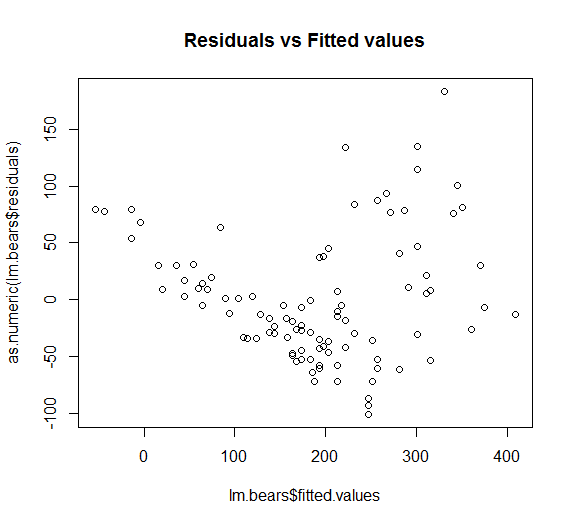
*Obs.No:* Observation number for this bear. For example, the bear with ID=41 (Bertha) was measured on four occasions. The value of Obs.No goes from 1 to 4 for these observations

*Name:* The names of the bears given to them by the researchers.

As you did in previous labs, consider only the first observation for each bear (bears\_indep=bears[bears$Obs.No==1,]).

Consider a single linear regression model with response y=“Weight” and predictor x=“Length”.

* After fitting the model, draw a scatterplot of the residuals versus the fitted values.



* Do you have any evidence against the linearity of the model? If yes, what can you try to improve the model?

**Yes. The data show no linearity because it has a non-linear trend.**

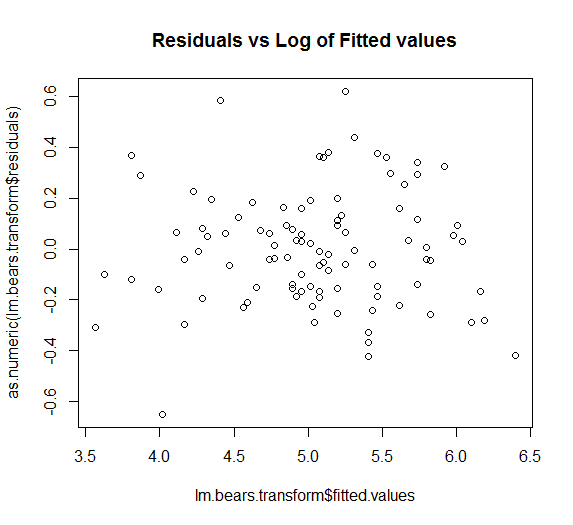
**We can transform x to have to improve the model.**

* Do you have any evidence against the equivariance of errors? If yes, what can you try to improve the model?

**Yes. The data show no equivariance of errors because the data on the right of the graph are more spread-out than the data on the left of the graph.**

**We can transform y to have**

* Fit a new model after transforming the response, in particular use response y=log(Weight) and predictor x=“Length”. Draw a scatterplot of the residuals versus the fitted values.



* Do you have any evidence against the linearity of the model? If yes, what can you try to improve the model?

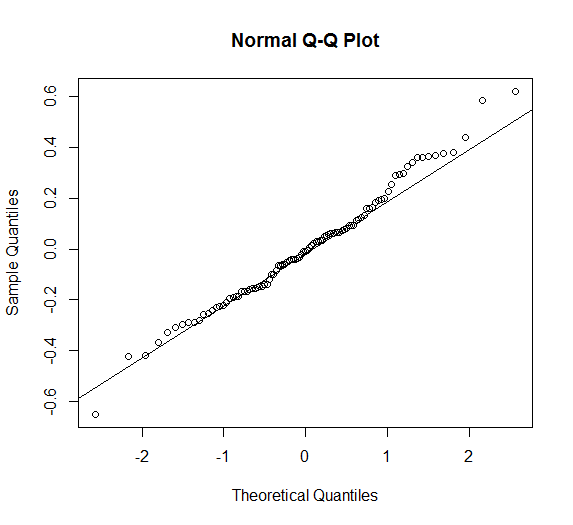
**No, the data shows the linearity because the data are evenly spread out and show no trend.**

* Do you have any evidence against the equivariance of errors? If yes, what can you try to improve the model?

**No, the data shows the equivariance of errors because data are evenly**

**spread out.**

* For the model with the transformed response, draw a Q-Q plot of the residuals and perform a Shapiro-Wilk test. What can you conclude?



Shapiro-Wilk normality test

data: as.numeric(lm.bears.transform$residuals)

W = 0.98805, p-value = 0.519

**By the Q-Q plot, the residuals are normally distributed because all data lie on the line.**

**By the Shapiro-Wilk test, the p-value = 0.519 > α, so we fail to reject null hypothesis that residuals are normally distributed.**

**Thus, residuals are normally distributed.**

**R code:**

setwd("//udrive.win.psu.edu/Users/j/q/jql5883/Desktop/math462")

getwd()

bears=read.csv("bears.txt", header=T, sep="")

bears=bears[bears$Obs.No==1,]

head(bears)

attach(bears)

lm.bears=lm(Weight~Length)

plot(lm.bears$fitted.values,as.numeric(lm.bears$residuals),main="Residuals vs Fitted values")

lm.bears.transform=lm(log(Weight)~Length)

plot(lm.bears.transform$fitted.values,as.numeric(lm.bears.transform$residuals),main="Residuals vs Log of Fitted values")

qqnorm(as.numeric(lm.bears.transform$residuals))

qqline(as.numeric(lm.bears.transform$residuals))

shapiro.test(as.numeric(lm.bears.transform$residuals))