Zuur Ch 02 part 2

Ed Harris

2020.04.08

Ch 02 outline

-Learn as you go philosophy

-Data exploration

-Linear regression modelling

-Linear regression assumptions

-Examples

2.3 Violate those assumptions

Vanilla linear model full assumptions

-Gaussian residuals

-Homogeneous variance

-"fixed" X (discuss briefly)

-Independence

-Correct model specification…

2.3.6 wedge clams

Donax hanleyanus
Uruguay, Rocha, Aguas-Dulces
NMR 17634. Actual size 26 mm

2.3.6 wedge clams

```
Clams <- read.table("Clams.txt", header = T)
str(Clams)
```

```
## 'data.frame':    398 obs. of  5 variables:
##  $ MONTH   : num  11 11 11 11 11 11 11 11 11 11 ...
##  $ LENGTH  : num  28.4 16.6 13.7 17.4 11.8 ...
##  $ AFD     : num  0.248 0.052 0.028 0.07 0.022 0.187 0.361 0.05 0.087 0.128 ...
##  $ LNLENGTH: num  3.35 2.81 2.62 2.86 2.47 ...
##  $ LNAFD   : num  -1.39 -2.96 -3.57 -2.65 -3.83 ...
```
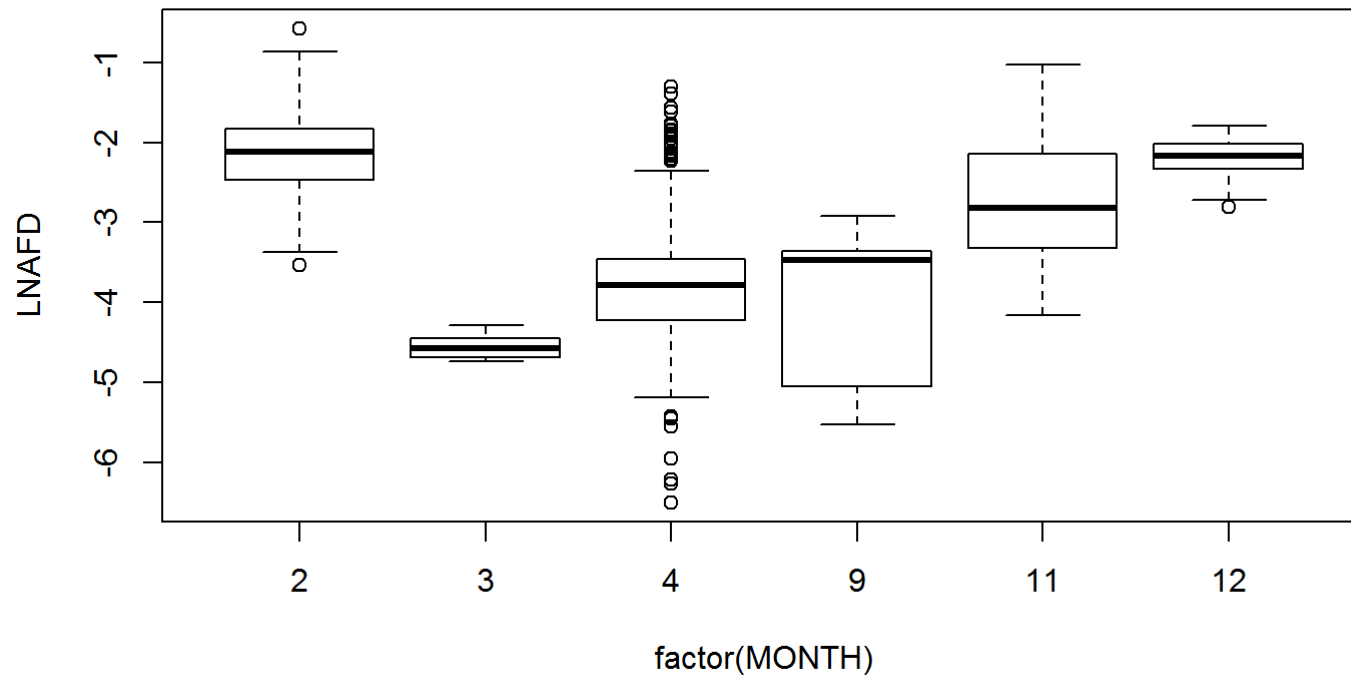
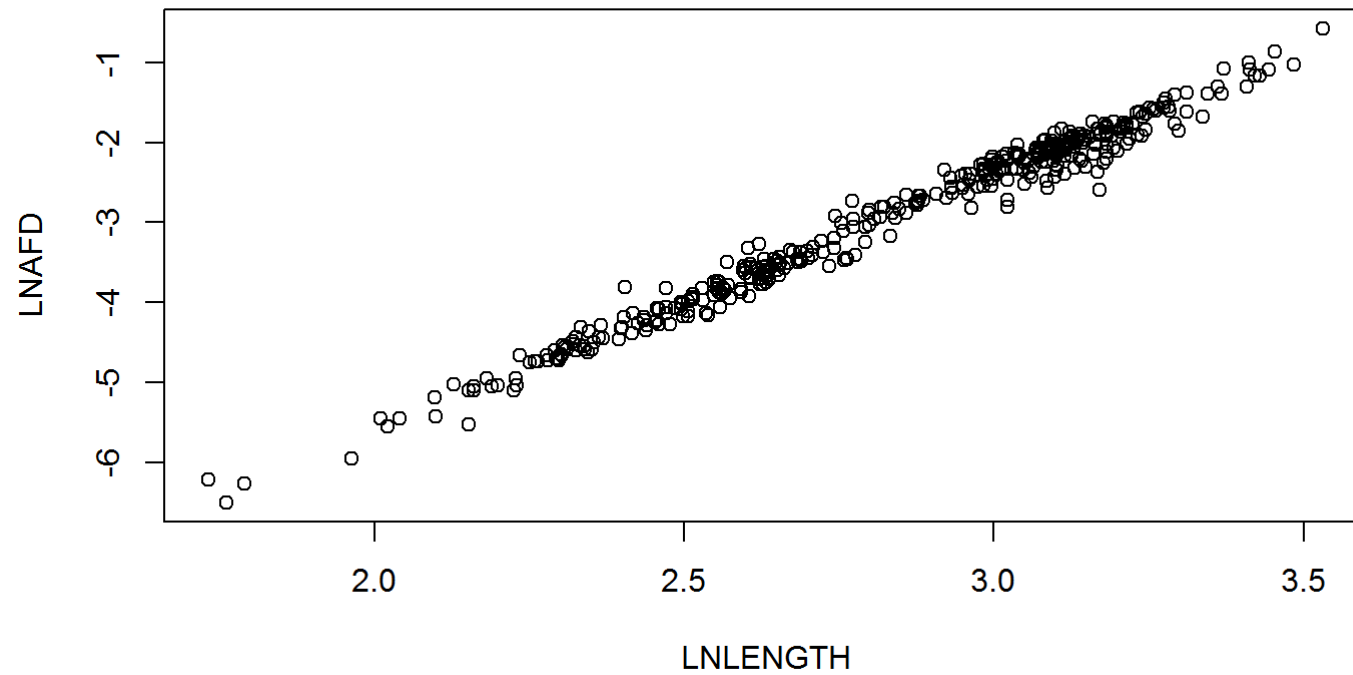Month - month of measurement
Length - length (mm?)
AFD - weight
LNLENGTH - log(Length)
LMAFD - log(AFD)

2.3.6 wedge clams

2.3.6 wedge clams

2.3.6 wedge clams

"possible models":

LNAFN ~ LNLENGTH + MONTH
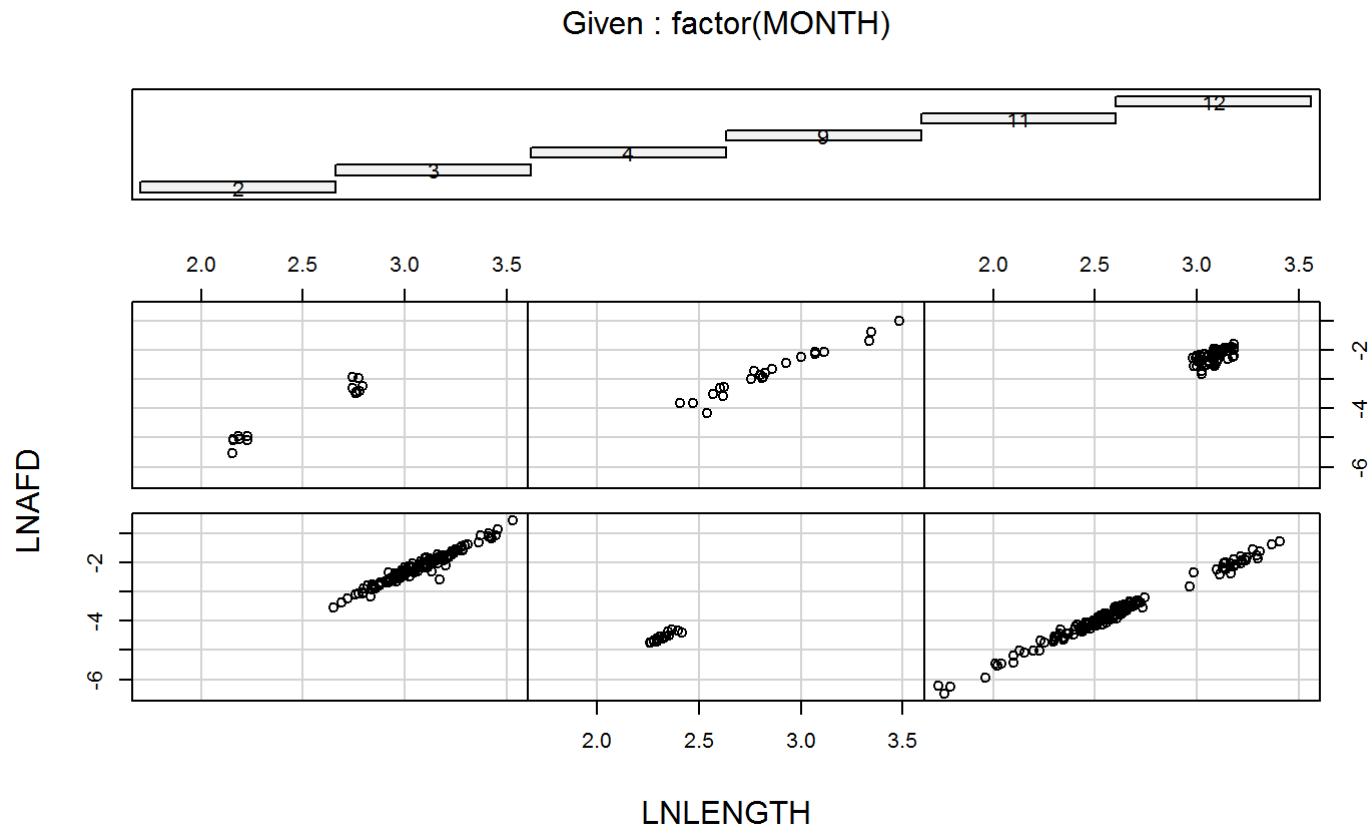LNAFN ~ LNLENGTH * MONTH

…#isThisOK? **Discuss**

Reading for model selection

Aho, K., Derryberry, D., Peterson, T., 2014. Model selection for ecologists: the worldviews of AIC and BIC. Ecology 95, 631-636. https://doi.org/10.1890/13-1452.1

Testing alternative models

Month:
9 11 12
2 3 4

Given : factor(MONTH)

Testing alternative models

```
## Single term deletions
##
## Model:
## LNAFD ~ LNLENGTH * factor(MONTH)
##                            Df Sum of Sq    RSS     AIC F value  Pr(>F)
## <none>                                  6.4490 -1616.8
## LNLENGTH:factor(MONTH)  5   0.20328 6.6523 -1614.4  2.4334 0.03444 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Which model is best?

2.3.6 wedge clam model validation
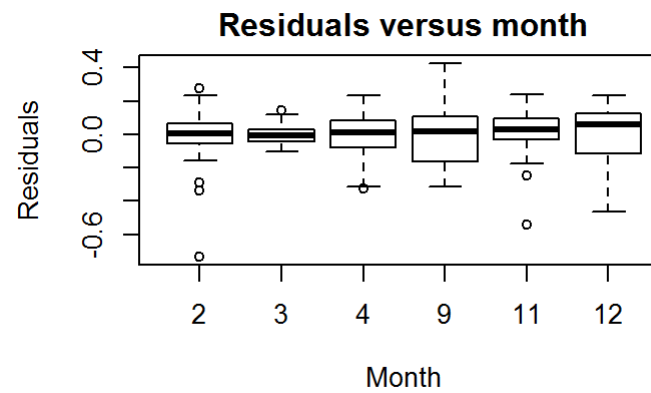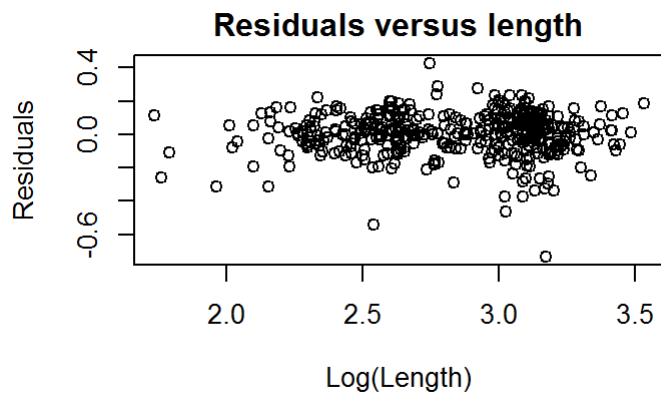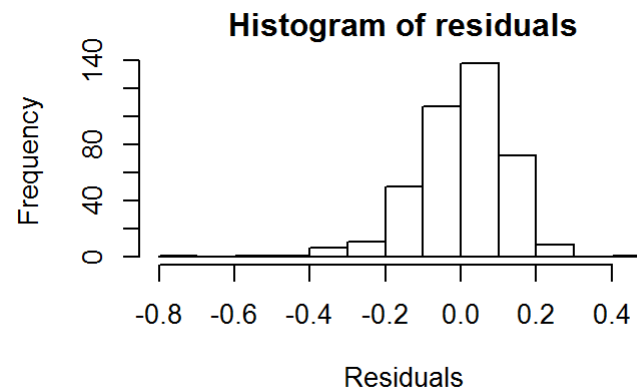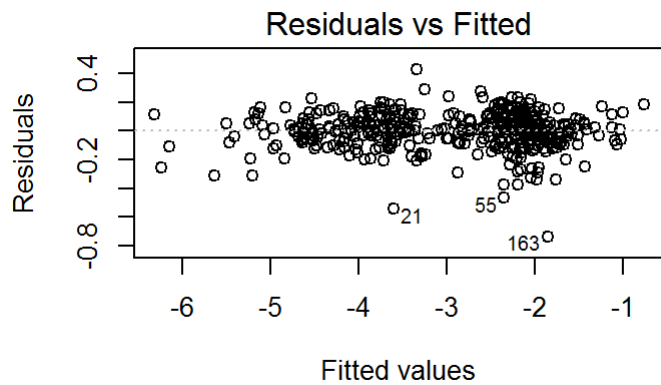
Do this for your models:

1. Residual versus fitted values
   (looking for homogeneity of variance)
   ((heteroskedasticity)) *spelling

2. Residual distribution
   (qq plot or histogram, etc.)

3. Residuals versus explanatory variables
   (is residual variation independent of
   variation in explanatory vars?)

Heteroskedasticity

Paloyo, A.R., 2011. When Did We Begin to Spell "Heteros*Edasticity" Correctly? (SSRN Scholarly Paper No. ID 1973444). Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.1973444

Test of assumptions

Discuss

Test of assumptions

Formal tests of homogeneity of residuals

E.g. F-test (Faraway 2005) does variance differ by clam length?

```
## 
##  F test to compare two variances
## 
## data:  E1 and E2
## F = 0.67571, num df = 75, denom df = 321, p-value = 0.04211
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4821187 0.9858005
## sample estimates:
## ratio of variances 
##          0.6757124
```

Test of assumptions

Formal tests of homogeneity of residuals

Bartlett test does variance differ by month?

```
##
##  Bartlett test of homogeneity of variances
##
## data:  E and factor(Clams$MONTH)
## Bartlett's K-squared = 37.017, df = 5, p-value = 5.942e-07
```

Results

We have some reasons to be unsatisfied with the "plain old" linear model

```
## Analysis of Variance Table
##
## Response: LNAFD
##                         Df Sum Sq Mean Sq     F value  Pr(>F)
## LNLENGTH                 1 480.86  480.86 28781.3185 < 2e-16 ***
## factor(MONTH)            5   2.04    0.41    24.4282 < 2e-16 ***
## LNLENGTH:factor(MONTH)   5   0.20    0.04     2.4334 0.03444 *
## Residuals              386   6.45    0.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.3.7 Moby's Teeth

Sperm whales beach and die. Famous Moby on Firth of Forth. Dataset looking at Nitrogen isotope accumulation in teeth by age.

2.3.7 Moby's Teeth

2.3.7 Moby's Teeth

```
Teeth <- read.table("TeethNitrogen.txt", header = T)
str(Teeth)
```

```
## 'data.frame':     307 obs. of  3 variables:
##  $ X15N : num  11.7 11.7 11.5 11.7 11.7 ...
##  $ Age  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Tooth: Factor w/ 11 levels "I1/98","M143/96D",..: 7 7 7 7 7 7 7 7 7 7 ...
```

2.3.7 Moby's Teeth

Model (just Moby's) N ~ age...

```
M2 <- lm(X15N ~ Age, data = Teeth,
         subset = (Teeth$Tooth == "Moby"))
anova(M2)


## Analysis of Variance Table
##
## Response: X15N
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Age         1 79.902  79.902  338.43 < 2.2e-16 ***
## Residuals 40  9.444   0.236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
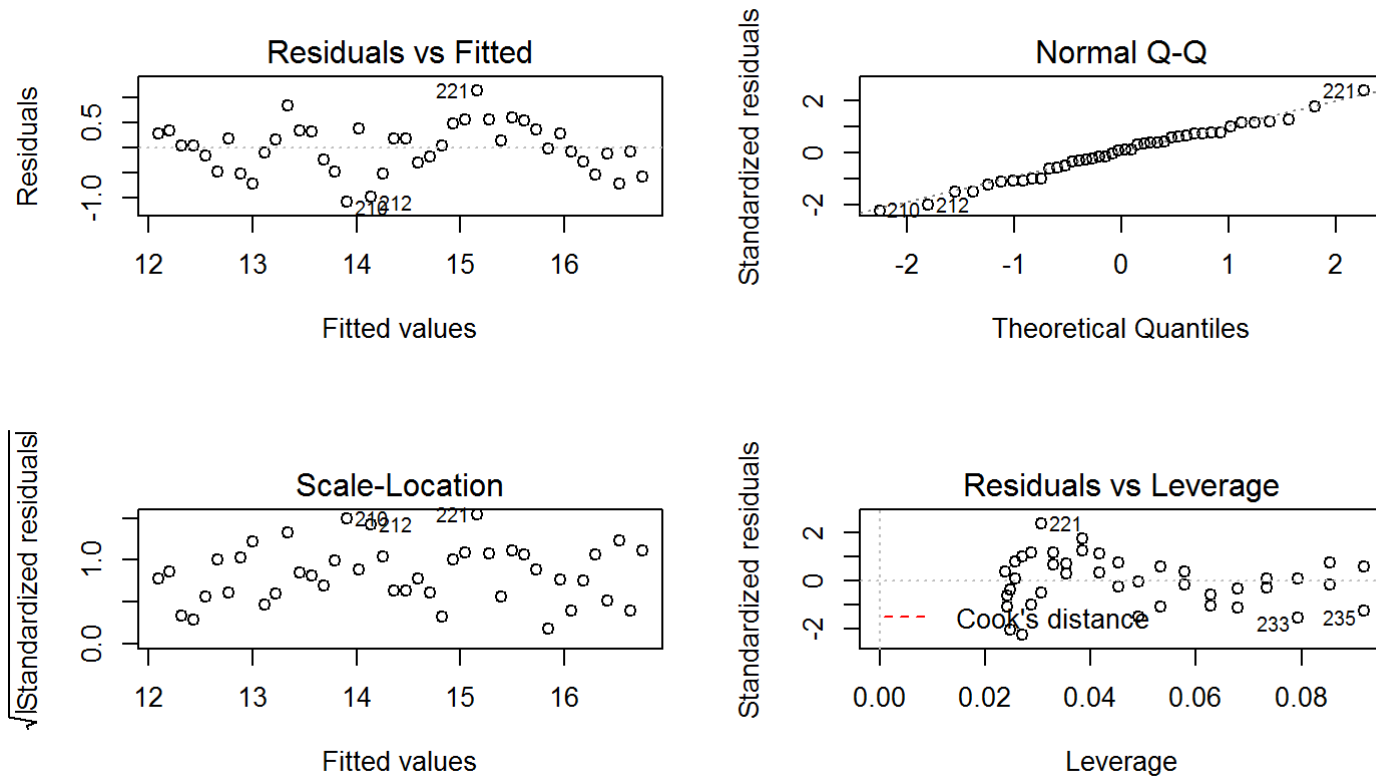
2.3.7 Moby's Teeth

Regular diagnostic plots available in R by using plot() on a linear model

object

2.3.7 Moby's Teeth

NB - Striking pattern of residuals by case! - Normal qq plot (discuss) looks okay - Cook's distance plot (looking for values outside red line - none here) ((measure of "high leverage" outliers to consider excluding))

Fox, J., 2015. Applied Regression Analysis and Generalized Linear Models, 3rd ed. SAGE Publications, Los Angeles.

2.3.7 Moby's Teeth

Violation of homogeneity and non-independence…

```
justMoby <- which(Teeth$Tooth == "Moby")
plot(X15N ~ Age, data = Teeth[justMoby, ])
abline(lm(X15N ~ Age, data = Teeth[justMoby, ]))
```

2.3.7 Moby's Teeth

Accept or reject this model?

$$y_i = 11.75 + 0.11 \times age_i$$

```
##
## Call:
## lm(formula = X15N ~ Age, data = Teeth[justMoby, ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.07102 -0.28706  0.04346  0.33820  1.13724
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.748940   0.163559   71.83   <2e-16 ***
## Age          0.113794   0.006186   18.40   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4859 on 40 degrees of freedom
```

Homework :)

For the Nereis dataset: -run the model concentration ~ biomass * nutrients -evaluate assumptions -come prepared to discuss it with evidence for next time