

Assignment 2 – Linear Models

Sandra Emry

2016-10-28

Read in and look at the data

```
#Load packages
library(nlme)
library(visreg)
suppressMessages(library(tidyverse))

# read in df: aging
aging <- read_csv("~/Documents/Biol501/biol501/data/aging.csv")

## Parsed with column specification:
## cols(
##   Order = col_character(),
##   Species = col_character(),
##   rawdistance = col_double(),
##   rawstrain = col_double(),
##   rawstress = col_double(),
##   rawinitial = col_double(),
##   rawfinal = col_double(),
##   standistance = col_double(),
##   stanstrain = col_double(),
##   stanstress = col_double(),
##   staninitial = col_double(),
##   stanfinal = col_double()
## )

# structure of data frame
str(aging)

## Classes 'tbl_df', 'tbl' and 'data.frame':   161 obs. of  12
## variables:
##   $ Order      : chr  "Red" "Red" "Red" "Red" ...
##   $ Species    : chr  "Polyneura" "Polyneura" "Polyneura"
##   "Polyneura" ...
##   $ rawdistance : num  4.5 7 11.5 15 5.5 13 16.5 23.5 29.5 28.5 ...
##   $ rawstrain  : num  0.104 0.089 0.104 0.149 0.289 0.38 0.368 0.235
##   0.165 0.283 ...
##   $ rawstress  : num  1.93 1.52 1.44 1.26 4.45 ...
##   $ rawinitial : num  17.02 24.07 15.7 7.04 31.2 ...
##   $ rawfinal   : num  12.93 12.11 9.79 2.79 10.08 ...
##   $ standistance: num  -0.526 -0.263 0.211 0.579 -0.717 ...
##   $ stanstrain : num  -0.06726 -0.20179 -0.06726 0.33632 0.00814 ...
##   $ stanstress : num  0.2547 -0.0106 -0.0665 -0.1776 0.1788 ...
```

```
## $ staninitial : num  0.0666 0.5084 -0.0161 -0.5588 0.9239 ...
## $ stanfinal   : num  0.3748 0.2876 0.0409 -0.7033 -0.1162 ...
## - attr(*, "spec")=List of 2
## ..$ cols      :List of 12
## .. ..$ Order      : list()
## .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
## .. ..$ Species     : list()
## .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
## .. ..$ rawdistance : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ rawstrain   : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ rawstress   : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ rawinitial  : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ rawfinal    : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ standistance: list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ stanstrain  : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ stanstress  : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ staninitial : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## .. ..$ stanfinal   : list()
## .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
## ..$ default: list()
## .. ..- attr(*, "class")= chr  "collector_guess" "collector"
## ..- attr(*, "class")= chr "col_spec"
```

```
aging <- aging %>%
  na.omit() %>% # delete rows with NAs
  select(Order, Species, standistance, stanstrain) # select only the
  variables used in this analysis
```

```
# first few rows of dataframe
head(aging)
```

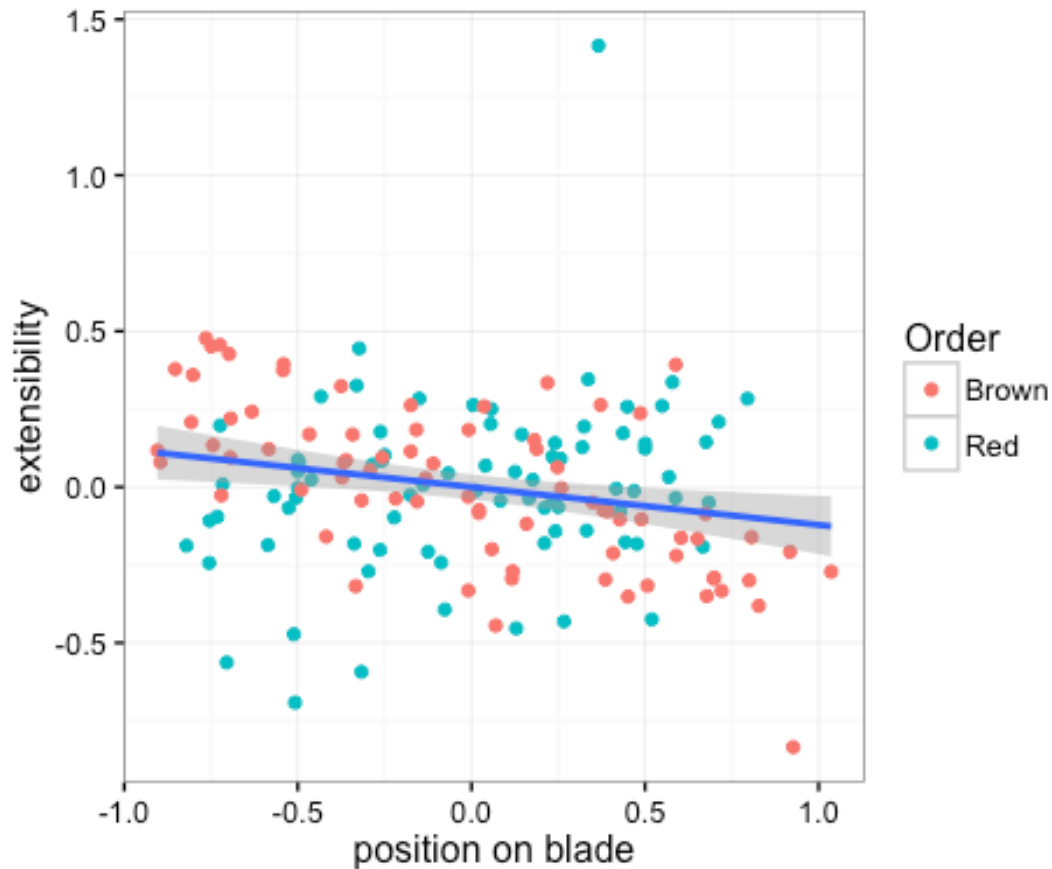
```
## # A tibble: 6 × 4
##   Order    Species standistance  stanstrain
##   <chr>    <chr>         <dbl>        <dbl>
## 1 Red    Polyneura  -0.5263158 -0.067264574
## 2 Red    Polyneura  -0.2631579 -0.201793722
## 3 Red    Polyneura   0.2105263 -0.067264574
## 4 Red    Polyneura   0.5789474  0.336322870
## 5 Red    Sparlingia -0.7167382  0.008139535
## 6 Red    Sparlingia -0.3304721  0.325581395
```

Explain the purpose of the study that yielded the data.

The dataset used in the assignment is from a paper published by a former graduate student in my supervisor's lab. The paper, entitled "Divergent growth strategies between red algae and kelps influence biomechanical properties", was published in 2015 in *American Journal of Botany*. This study examined how material properties, such as strength and extensibility, that are associated with aging, differ across several species of brown and red algae. The variables used in this analysis are standistance, stanstrain and species. Macroalgae form new tissue in one location, so as distance from this position along the blade increases, so does the age of tissue. The variable standistance represents the position along the thallus, standardized to the mean length of blades. This is to control for differences in the length of individuals, as well as to account for how material properties differ along the thallus in different species. Stanstrain describes how extensible the tissue is, calculated by the change in length divided by original length. Species is included in the model to control for the possibility that differences in material properties are phylogenetic in nature and not how aging alters tissues.

Plot the data

```
ggplot(data = aging, aes(x = standistance, y = stanstrain)) +  
  geom_point(aes(color = Order)) +  
  xlab("position on blade") +  
  ylab("extensibility") +  
  geom_smooth(se = TRUE, method = "lm") +  
  theme_bw() +  
  theme(legend.position="right")
```



Parameters estimated with these data

I will fit a mixed effects model that will estimate the y-intercept (β_0) and the slope of the line explained by standdistance (β_1).

Formula: $Y = \beta_0 + \beta_1 X$

This data will test the hypothesis that as distance increases, strain will decrease due to aging of the tissues.

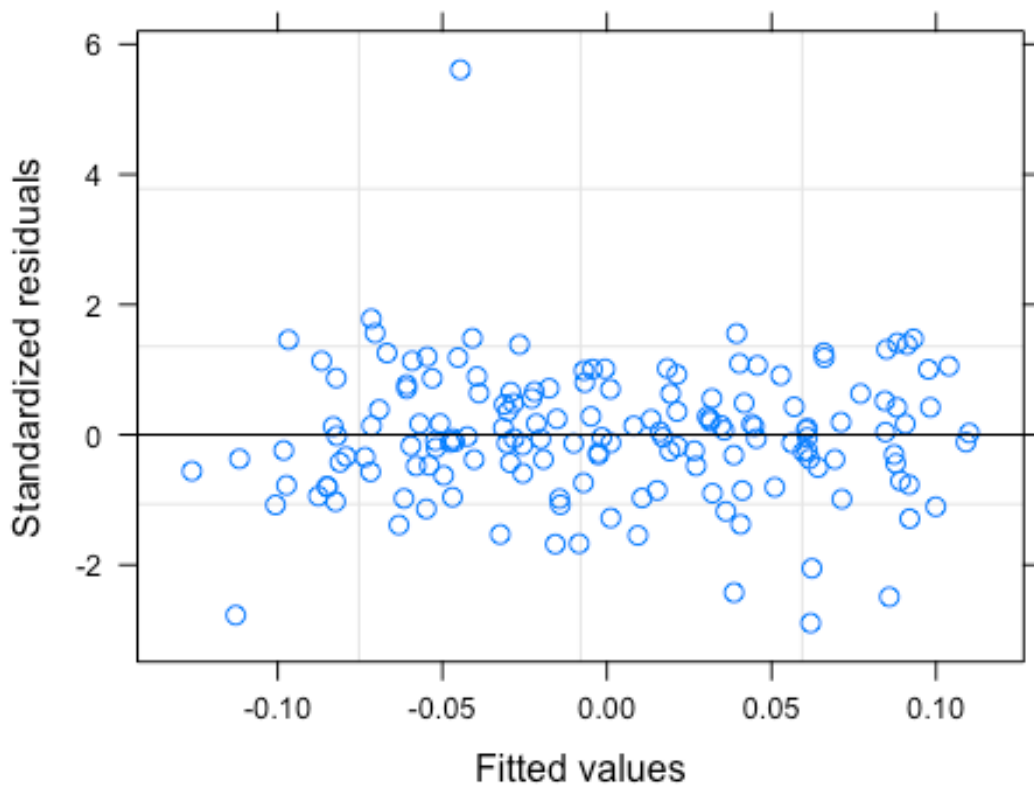
H₀: Distance has no effect on strain ($\beta_1 = 0$)

H_a: Distance has an effect on strain ($\beta_1 \neq 0$)

Fit a mixed effects model

Below I am fitting a model to the response variable stanstrain, with the predictor variable standistance as a fixed effect. Species is included as a random effect.

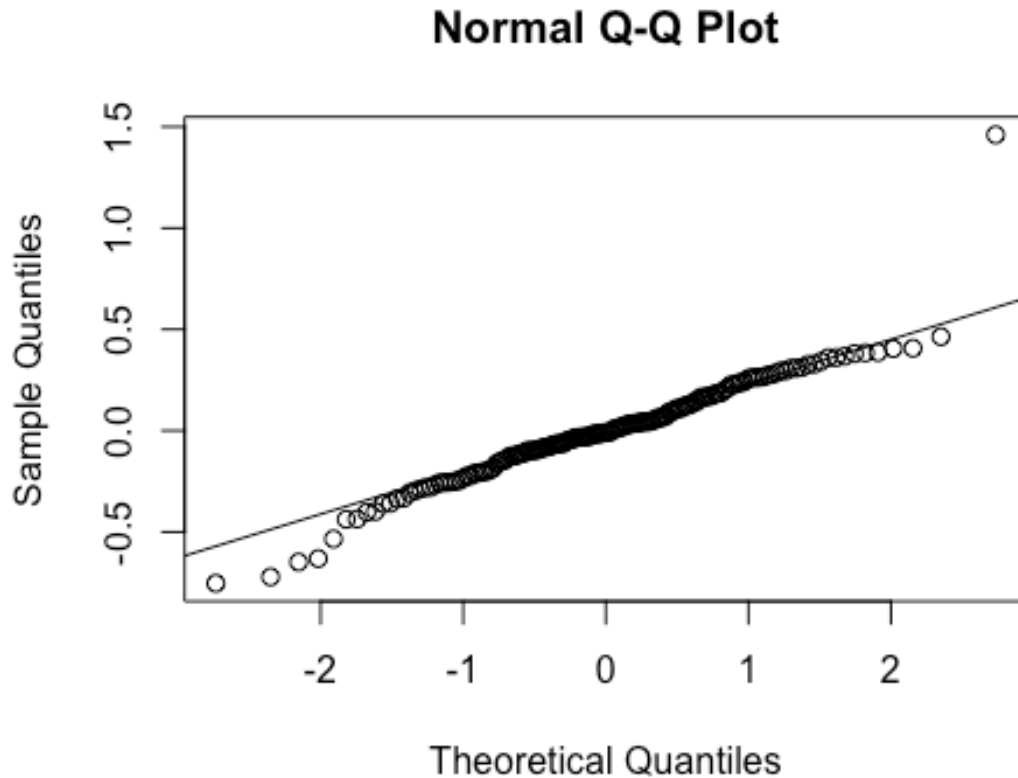
```
# fit a linear model with position along the blade as a fixed factor,  
species as random factor and breaking strain as the response variable  
z <- lme(stanstrain ~ standistance, random = ~1|Species, data = aging)  
  
## check assumptions of the model:  
# variances are homogenous: residuals equal for all values of x  
plot(z)
```



```
# errors are approximately normally distributed
```

```
qqnorm(residuals(z))
```

```
qqline(residuals(z))
```



```
# parameter estimates
```

```
summary(z)
```

```
## Linear mixed-effects model fit by REML
```

```
## Data: aging
```

```
##      AIC      BIC    logLik
```

```
## 39.8923 52.14268 -15.94615
```

```
##
```

```
## Random effects:
```

```
## Formula: ~1 | Species
```

```
##      (Intercept) Residual
```

```
## StdDev: 4.008748e-06 0.2604026
```

```
##
```

```
## Fixed effects: stanstrain ~ standistance
```

```
##              Value Std.Error DF   t-value p-value
```

```
## (Intercept)  0.00003715 0.02058664 133  0.0018044  0.9986
```

```
## standistance -0.12171240 0.04264009 133 -2.8544126  0.0050
```

```
## Correlation:
```

```
##      (Intr)
```

```
## standistance -0.001
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -2.89348662 -0.48511114 -0.03895269  0.63184296  5.60807519
##
## Number of Observations: 160
## Number of Groups: 26

# 95% confidence intervals for parameters
confint(z)

## [[1]]
## [1] -0.04031192 -0.08353589  0.04038622  0.08361018 -0.04031192
## [6] -0.08353589  0.04038622  0.08361018 -0.04031192 -0.08353589
## [11]  0.04038622  0.08361018 -0.04031192 -0.08353589  0.04038622
## [16]  0.08361018 -0.04031192 -0.08353589  0.04038622  0.08361018
## [21] -0.04031192 -0.08353589  0.04038622  0.08361018 -0.04031192
## [26] -0.08353589
##
## [[2]]
## [1] -0.08136333 -0.03813937 -0.16206147 -0.20528543 -0.08136333
## [6] -0.03813937 -0.16206147 -0.20528543 -0.08136333 -0.03813937
## [11] -0.16206147 -0.20528543 -0.08136333 -0.03813937 -0.16206147
## [16] -0.20528543 -0.08136333 -0.03813937 -0.16206147 -0.20528543
## [21] -0.08136333 -0.03813937 -0.16206147 -0.20528543 -0.08136333
## [26] -0.03813937
##
## [[3]]
## [1] -0.04031192 -0.08353589  0.04038622  0.08361018 -0.04031192
## [6] -0.08353589  0.04038622  0.08361018 -0.04031192 -0.08353589
## [11]  0.04038622  0.08361018 -0.04031192 -0.08353589  0.04038622
## [16]  0.08361018 -0.04031192 -0.08353589  0.04038622  0.08361018
## [21] -0.04031192 -0.08353589  0.04038622  0.08361018 -0.04031192
## [26] -0.08353589
##
## [[4]]
## [1] -0.08136333 -0.03813937 -0.16206147 -0.20528543 -0.08136333
## [6] -0.03813937 -0.16206147 -0.20528543 -0.08136333 -0.03813937
## [11] -0.16206147 -0.20528543 -0.08136333 -0.03813937 -0.16206147
## [16] -0.20528543 -0.08136333 -0.03813937 -0.16206147 -0.20528543
## [21] -0.08136333 -0.03813937 -0.16206147 -0.20528543 -0.08136333
## [26] -0.03813937

# test the null hypothesis
anova(z)

##           numDF denDF  F-value p-value
## (Intercept)      1   133 0.000000  1.000
## standistance      1   133 8.147671  0.005
```

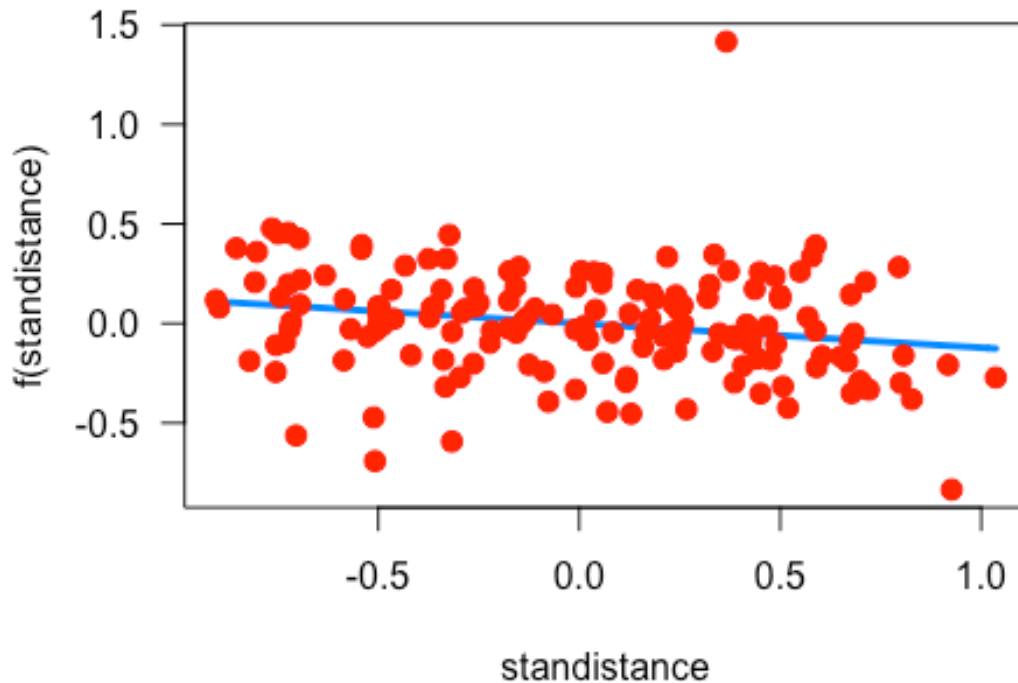
Interpret the output:

The results from the anova show a significant value of 0.005 for the parameter of standistance. This means that we can reject the null hypothesis that distance along the thallus has no effect on strain. The parameter for standistance is -0.12, which is equal to the slope of the line. As distance along the thallus increases, the amount of strain decreases, which is what was hypothesized. An intercept of ~ 0 , tells us that the newest tissue, or lowest standistance, has a strain of essentially 0. There are two standard deviation values reported for the random effects. The one labelled intercept explains the variability due to the random effect, Species, added to the model. The value is very small, so this tells us that Species is not explaining much of the variation in the response variable. The residual std dev explains that the variability that is not attributed to either the random or fixed effect. This can be visualized in the plot below, which includes the predicted values as well as the actual data points.

Create a graph to visualize the model fit to the data. Explain what the graph is showing.

The graph below shows the data points in red, and the predicted values in blue. We can see the model underestimates the amount of variance in the data. There is likely something we are missing, and this model does not fully explain the pattern of extensibility in tissues.

```
# plot predicted values from the model using visreg  
visreg(z, points.par = list(cex = 1.2, col = "red"))
```

Statistical assumptions of a mixed effects linear model were met by the data. Errors are approximately normally distributed, as seen in the q-q plot, and variances were homogenous. I assume that the experimental design included random sampling from the population, and that each data point is independent from others. Data meets the sphericity assumption of mixed effects models.

From the tests done, we can conclude that position on the blade does have an effect on extensibility, such that as you get further away from the point of new growth, the tissue is less extensible. However, there is variation in the response variable that is not explained by this model. Because older tissue is less extensible than newer

tissue, these sections of the blade will be less likely to withstand certain abiotic factors of a given environment such as high flows or waves.