Data analysis project

Jess Qualley and Rebecca Hansen

March 27, 2020

##Clear code (**delete before submitting**\*)

rm(list = ls(all = TRUE))

#Preparation ##Loading and citing packages

#Load faraway package  
library(faraway)  
  
#Load plyr package  
library(plyr)

##   
## Attaching package: 'plyr'

## The following object is masked from 'package:faraway':  
##   
## ozone

#Load tidyverse package  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✔ ggplot2 3.2.1 ✔ purrr 0.3.3  
## ✔ tibble 2.1.3 ✔ dplyr 0.8.3  
## ✔ tidyr 1.0.2 ✔ stringr 1.4.0  
## ✔ readr 1.3.1 ✔ forcats 0.4.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::arrange() masks plyr::arrange()  
## ✖ purrr::compact() masks plyr::compact()  
## ✖ dplyr::count() masks plyr::count()  
## ✖ dplyr::failwith() masks plyr::failwith()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::id() masks plyr::id()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ dplyr::mutate() masks plyr::mutate()  
## ✖ dplyr::rename() masks plyr::rename()  
## ✖ dplyr::summarise() masks plyr::summarise()  
## ✖ dplyr::summarize() masks plyr::summarize()

#Load lmer package  
library(lme4)

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

#Cite packages  
citation("faraway")

##   
## To cite package 'faraway' in publications use:  
##   
## Julian Faraway (2016). faraway: Functions and Datasets for Books  
## by Julian Faraway. R package version 1.0.7.  
## https://CRAN.R-project.org/package=faraway  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {faraway: Functions and Datasets for Books by Julian Faraway},  
## author = {Julian Faraway},  
## year = {2016},  
## note = {R package version 1.0.7},  
## url = {https://CRAN.R-project.org/package=faraway},  
## }  
##   
## ATTENTION: This citation information has been auto-generated from  
## the package DESCRIPTION file and may need manual editing, see  
## 'help("citation")'.

citation("plyr")

##   
## To cite plyr in publications use:  
##   
## Hadley Wickham (2011). The Split-Apply-Combine Strategy for Data  
## Analysis. Journal of Statistical Software, 40(1), 1-29. URL  
## http://www.jstatsoft.org/v40/i01/.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Article{,  
## title = {The Split-Apply-Combine Strategy for Data Analysis},  
## author = {Hadley Wickham},  
## journal = {Journal of Statistical Software},  
## year = {2011},  
## volume = {40},  
## number = {1},  
## pages = {1--29},  
## url = {http://www.jstatsoft.org/v40/i01/},  
## }

citation("tidyverse")

##   
## Wickham et al., (2019). Welcome to the tidyverse. Journal of  
## Open Source Software, 4(43), 1686,  
## https://doi.org/10.21105/joss.01686  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Article{,  
## title = {Welcome to the {tidyverse}},  
## author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani},  
## year = {2019},  
## journal = {Journal of Open Source Software},  
## volume = {4},  
## number = {43},  
## pages = {1686},  
## doi = {10.21105/joss.01686},  
## }

##Read in and organize data

#Read from csv file  
herring <- read.csv("HerringOtolithDatabase 15 Jan 2020\_JQ.csv", header=TRUE, fileEncoding="UTF-8-BOM", stringsAsFactors = FALSE)  
  
#Create a data frame  
herring\_data <- data.frame(sal.id = herring$FishCode,  
 oto.width = herring$AverageWidth,  
 digestion = herring$Digestion,  
 sal.sp = herring$SalmonSpecies,  
 sal.sex = herring$SalmonSex,  
 sal.length = (herring$SalmonLength),  
 sal.weight = herring$SalmonWeight,  
 coll.month = as.factor(herring$CollectionMonth),  
 coll.year = herring$CollectionYear,  
 coll.doy = herring$CollectionDayofYear,  
 coll.season = herring$Season,  
 lat = herring$Latitude,  
 long = herring$Longitude,  
 stat.area = as.factor(herring$StatArea))  
  
#figure out which stat areas to remove  
levels(as.factor(herring$StatArea))

## [1] "1" "13" "14" "15" "16" "17" "18" "19" "20" "23" "25"   
## [12] "28" "29" "101" "125"

#filter for 2018 data, chinook salmon only, omit northern BC areas (Haid Gwaii) and west coast Vancouver Island  
herring\_data\_raw <- herring\_data %>% filter(herring\_data$sal.sp == "ch" & herring\_data$coll.year == "2018" & herring\_data$stat.area != "1" & herring\_data$stat.area != "101" & herring\_data$stat.area != "125" & herring\_data$stat.area != "23" & herring\_data$stat.area != "25")  
  
#filter for 2018 data, chinook salmon only, omit northern BC areas (Haid Gwaii) and west coast Vancouver Island, excluding rows with NAs for continuous variables  
herring\_data <- herring\_data %>% filter(herring\_data$sal.sp == "ch" & herring\_data$coll.year == "2018" & herring\_data$stat.area != "1" & herring\_data$stat.area != "101" & herring\_data$stat.area != "125" & herring\_data$stat.area != "23" & herring\_data$stat.area != "25") %>% na.omit()  
  
#Add transformed length and weight columns  
herring\_data$sal.length.trans <- log(herring\_data$sal.length)  
herring\_data$sal.weight.trans <- log(herring\_data$sal.weight)

#Section 1: Introduction, Question, Goals and Hypotheses (400 words) ##A) Introduce relevant concepts: - State of salmon in BC - Salmon diet - Herring as a prey item for salmon - Gape size theory - Yearly availbaility of herring of different sizes - Regional differences

How the dataset was collected: - Salmon stomachs donated to Juanes lab by anglers - Salmon stomach contents sorted and classfied based on digestion - Otoliths collected and measured

Background needed to understand the dataset: - Otolith width as the best proxy for herring length

At least two scientiﬁc papers that are related to your topic:

##B) Research question: - What physiological, environmental and temporal factors determine the size of herring eaten by Chinook salmon in the Salish Sea

##C) Goal of research: - Description, because we want to know what variables best explain the patterns in the size of herring eaten by Chinook salmon

##D) Hypothesis: - Salmon Length: Positive non-linear relationship, logarithmic curve where asymptote is maximum herring size - Salmon Weight: Positive non-linear relationship, logarithmic curve where asymptote is maximum herring size - Collection day of year: Non-linear, positive quadratic function where vertex is midway through the year - Month: Greater otilith width for October-May greater, September to be greater than August which is greater than July which is greater than June

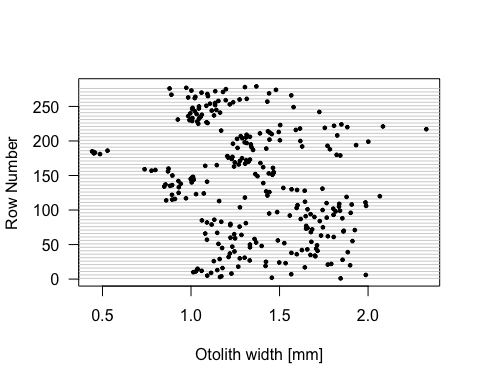
##E) Model selection: - We want to determine what factors explain patterns in the size of herring eaten by salmon - So we want to choose a model which best describes this pattern - Our possible models are not nested, we aren’t testing a null model - We have a fairly small sample size - Use the AICc

#Section 2: The Response Variable (150 words) ##A)

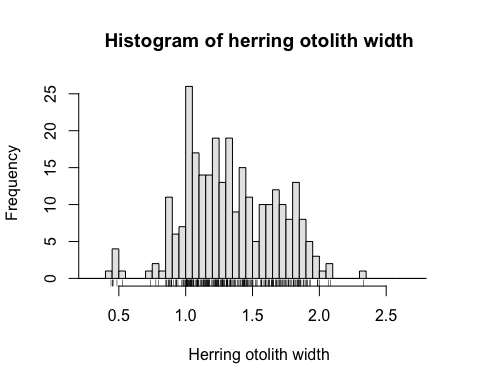
#Checking for zeroes and NAs in otilith width  
table(herring\_data\_raw$oto.width > 0, useNA = "always")

##   
## TRUE <NA>   
## 508 1

#Add column for number of rows  
n <- nrow(herring\_data)  
herring\_data$row <- 1:n  
  
#Dotchart of otolith width (response variable)  
with(herring\_data, plot(oto.width, row, las = 1, type = "n", xlab = "Otolith width [mm]", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(oto.width, row, las = 1, type = "p", pch = 19, cex = 0.5))



#Histogram of response variable with rugmarks  
hist(herring\_data$oto.width, xlim = c(0.3, 2.8), breaks = seq(0.2, 2.8, by = 0.05), col = grey(0.9), main = "Histogram of herring otolith width", xlab = "Herring otolith width")  
rug(herring\_data$oto.width)



##B) Qualitative description of response variable: - Continuous - Bounded by 0, can’t take negative values

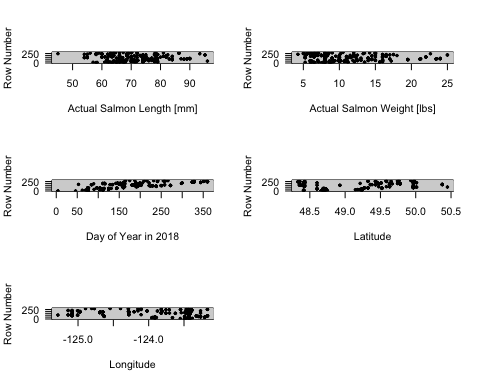
##C) Potential issues with response variable - Intially no zeroes and one missing value - We removed the single NA value for all subsequent analyses - Outliers at > 2.2 mm and < 0.6 mm, bu they’re biologically reasonable and could represent a few individuals at extreme age classes - - Non-independence is expected from repeated measurements of multiple herring that occurred within the same salmon stomach/experimental unit. It is likely that similar sized herring will occur in the same school targeted by a salmon.

* We expect non-idependence of herring size both from temporal and spatial autocorrelation. Sources of temporal autocorrelation could be that herring consumed by salmon at the same time of year have similar growth trajectories. Sources of spatial autocorrelation of herring size may include the fact that herring in a similar area have similar feeding resources and grow rates in addition to smaller fish that occur closer to subtidal habitats compared to larger fish that occur in deeper, pelagic habitats which may contribute to spatial autocorrelation.

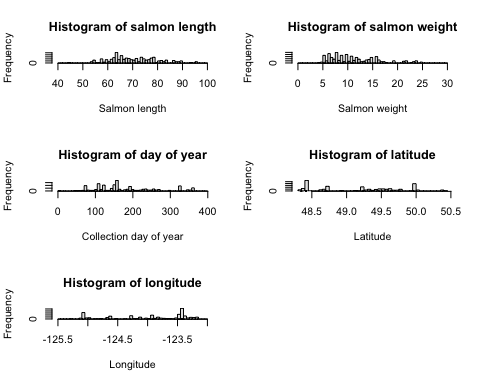
##D) Candidate distributions for variable - We are most likely to use the normal distribution. While the distribution is bound by 0 we have continuous data values (non-integer) and it is mostly symmetrical around the mean.

#Section 3: The Explanatory Variables (150 words) ##A) ###Continuous explanatory variables

#Multipanel of dotcharts for continuous explanatory variables  
par(mfrow = c(3, 2))  
  
#Salmon length  
with(herring\_data, plot(sal.length, row, las = 1, type = "n", xlab = "Actual Salmon Length [mm]", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(sal.length, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Salmon weight  
with(herring\_data, plot(sal.weight, row, las = 1, type = "n", xlab = "Actual Salmon Weight [lbs]", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(sal.weight, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Collection day of year  
with(herring\_data, plot(coll.doy, row, las = 1, type = "n", xlab = "Day of Year in 2018", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(coll.doy, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Latitude  
with(herring\_data, plot(lat, row, las = 1, type = "n", xlab = "Latitude", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(lat, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Longitude  
with(herring\_data, plot(long, row, las = 1, type = "n", xlab = "Longitude", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(long, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Multipanel of histograms for continuous explanatory variables  
par(mfrow = c(3, 2))

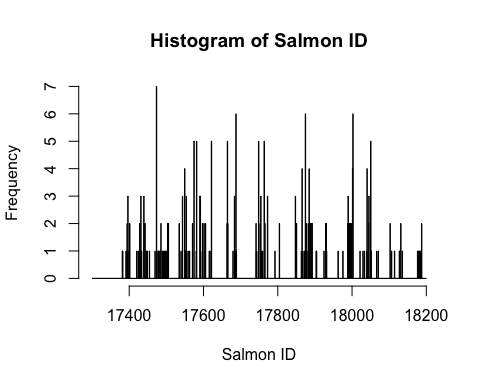


#Salmon length  
hist(herring\_data$sal.length, col = grey(0.9), breaks = seq(40, 100, by = 1), main = "Histogram of salmon length", xlab = "Salmon length")  
rug(herring\_data$sal.length)  
  
#Salmon weight  
hist(herring\_data$sal.weight, col = grey(0.9), breaks = seq(0, 30, by = 0.5), main = "Histogram of salmon weight", xlab = "Salmon weight")  
rug(herring\_data$sal.weight)  
  
#Collection day of year  
hist(herring\_data$coll.doy, col = grey(0.9), breaks = seq(0, 400, by = 7), main = "Histogram of day of year", xlab = "Collection day of year")  
rug(herring\_data$coll.doy)  
  
#Latitude  
hist(herring\_data$lat, col = grey(0.9), breaks = seq(48, 52, by = 10), main = "Histogram of latitude", xlab = "Latitude")  
rug(herring\_data$lat)  
  
#Longitude  
hist(herring\_data$long, col = grey(0.9), breaks = seq(-125.5, -123, by = 0.05), main = "Histogram of longitude", xlab = "Longitude")  
rug(herring\_data$long)

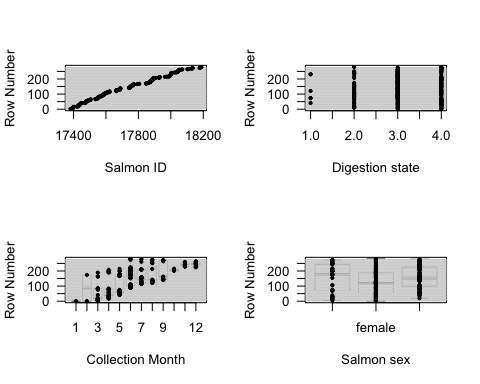


###Categorical explanatory variables

#Make a histogram of salmon id vs digestion  
hist(herring\_data$sal.id, col = grey(0.9), breaks = seq(17300, 18200, by = 1), main = "Histogram of Salmon ID", xlab = "Salmon ID")



#Multipanel of dotcharts for categorical explanatory variables  
par(mfrow = c(2, 2))  
  
#Salmon id  
with(herring\_data, plot(sal.id, row, las = 1, type = "n", xlab = "Salmon ID", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(sal.id, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Digestion  
with(herring\_data, plot(digestion, row, las = 1, type = "n", xlab = "Digestion state", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(digestion, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Collection Month  
with(herring\_data, plot(coll.month, row, las = 1, type = "n", xlab = "Collection Month", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(coll.month, row, las = 1, type = "p", pch = 19, cex = 0.5))  
  
#Salmon sex (note - label the first categorical NA variable)  
with(herring\_data, plot(sal.sex, row, las = 1, type = "n", xlab = "Salmon sex", ylab = "Row Number"))  
abline (h = seq(1, n, by = 5), col = "lightgrey")  
with(herring\_data, points(sal.sex, row, las = 1, type = "p", pch = 19, cex = 0.5))



#Tables for categorical explanatory variables  
#Make a table of salmon sex versus collection month  
with(herring\_data, table(sal.sex, coll.month))

## coll.month  
## sal.sex 1 2 3 4 5 6 7 8 9 10 11 12  
## 0 1 3 10 12 4 3 0 5 1 0 8  
## female 1 1 15 19 24 36 20 9 3 1 9 2  
## male 0 0 9 9 4 25 7 11 10 0 9 8

#Make a table of salmon sex versus digestion  
with(herring\_data, table(sal.sex, digestion))

## digestion  
## sal.sex 1 2 3 4  
## 2 4 20 21  
## female 2 24 60 54  
## male 2 12 48 30

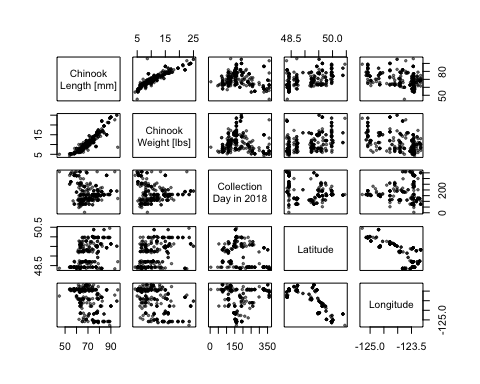
#Make a table of collection month versus digestion  
with(herring\_data, table(coll.month, digestion))

## digestion  
## coll.month 1 2 3 4  
## 1 0 0 1 0  
## 2 0 1 0 1  
## 3 1 7 9 10  
## 4 0 5 16 17  
## 5 2 3 14 21  
## 6 0 10 35 20  
## 7 0 7 12 11  
## 8 1 0 11 8  
## 9 0 1 8 9  
## 10 0 0 1 1  
## 11 2 5 8 3  
## 12 0 1 13 4

##B) - Outliers - For the continuous explanatory variable salmon length, there is one outlier at 45cm and three outliers between 90-100cm; however, these are reasonable sizes for recreationally caught chinook salmon. While not considered outliers, there are few larger salmon that have large weights. - Transformations why or why not? We may want to transform weight and length as they appear slightly right-skewed. \*\*note: we must decide this later.

#Section 4: Collinearity, Balance, and Variance Inﬂation Factors (200 words) ##A)

#create data frame for pairwise scatter with only continuous explanatory variables - salmon lenght, weight, collection day of year, latitude and longitude  
herring\_collinearity <- data.frame("sal.length" = herring\_data$sal.length, "sal.weight" = herring\_data$sal.weight, "coll.doy" = herring\_data$coll.doy, "lat" = herring\_data$lat, "long" = herring\_data$long)  
  
#Pairwise scatterplot of explanatory variables (salmon length, weight, day of year, latitude and longitude)  
plot(herring\_collinearity[1:5], cex = 0.5, pch = 19, col = rgb(0, 0, 0, 0.5),   
 labels = c("Chinook\nLength [mm]", "Chinook\nWeight [lbs]", "Collection\nDay in 2018", "Latitude", "Longitude"))



#Pairwise correlation coefficients for explanatory variables  
print(cor(na.omit(herring\_data[ ,c("sal.length", "sal.weight", "coll.doy", "lat", "long")])), digits = 2)

## sal.length sal.weight coll.doy lat long  
## sal.length 1.00 0.91 -0.16 0.46 -0.41  
## sal.weight 0.91 1.00 -0.16 0.43 -0.46  
## coll.doy -0.16 -0.16 1.00 -0.26 0.13  
## lat 0.46 0.43 -0.26 1.00 -0.83  
## long -0.41 -0.46 0.13 -0.83 1.00

#Variance Inflation Factors  
print(vif(na.omit(herring\_data[ ,c("sal.length", "sal.weight", "coll.doy", "lat", "long")])))

## sal.length sal.weight coll.doy lat long   
## 6.818335 6.932405 1.113278 3.954376 3.823013

#Balance of dataset (if appropriate)

Comment on balance - Referring to tables in section 3: we observe more salmon samples collected between March(3) to July(7) compared to other months and there are more females than males, although there is a high number of unsexed fish in this dataset. There is a higher number of herring categorized as digestion level 2, 3 and 4 compared to level 1. The difficulties we may encounter is to include sex if only 2/3 of the fish are sexed, therefore, it will be harder to evaluate whether sex is a useful predictor or not. The unbalanced nature of temporal sampling, it will be harder to evaluate the relationship during winter months. There isn’t simultaneous unbalancing across multiple variables, we only see holes, so this is less likely to be confounding but rather limiting in making any generalized conclusions about how sex and the winter period relates to herring size consumed by salmon.

Comment on pairwise scatterplots - there is a strong positive relationship between Chinook length and weight (0.91) in addition to a strong negative relationship between latitude and longitude (-0.83). While there isn’t an obvious linear relationship between salmon length and weigth vs collection day, there is evidence for a concave up relationship between collection day of year and length and weight perhaps due to seasonality of salmon caught during the spring and summer months.

Comment on VIFs - The VIF values of salmon length (6.818335), weight (6.932405), latitude (3.954376) and longitude (3.823013) are greater than 3 and therefore strongly colinear.

Comment on potential difficulties from correlations and benefits - the difficulty will be including both length and weight or both latitude and longitute in our statistical models. We would want to include either length or weight and either latitude or longitude so that they don’t confound each other. Seasonality in salmon size (length and weigth) may confound the relationship between these two variables.

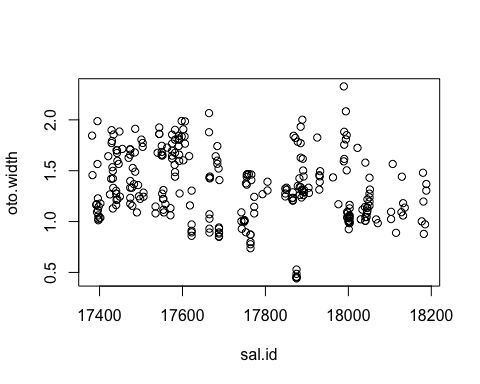
#Section 5: Statistical Methods and Model Fitting (200 words) Statistical methods - Which you will use and why - Cite packages used for modelling - What family of probability distributions (what link function) - Which explantory variables are we including - If and why data points are excluded - If transformations are used and why - If random-effects are used and why

Models

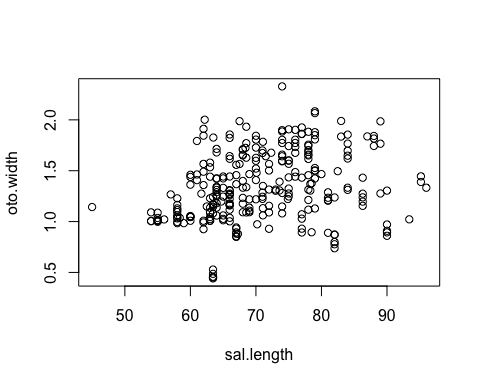
1. sal.id + doy + long + length + digestion
2. sal.id + doy + long + length
3. sal.id + doy + long + weight + digestion
4. sal.id + doy + long + weight
5. sal.id + doy + lat + length + digestion
6. sal.id + doy + lat + length
7. sal.id + doy + lat + weight + digestion
8. sal.id + doy + lat + weight
9. sal.id + month + long + length + digestion
10. sal.id + month + long + length
11. sal.id + month + long + weight + digestion
12. sal.id + month + long + weight
13. sal.id + month + lat + length + digestion
14. sal.id + month + lat + length
15. sal.id + month + lat + weight + digestion
16. sal.id + month + lat + weight

##Exploratory plotting

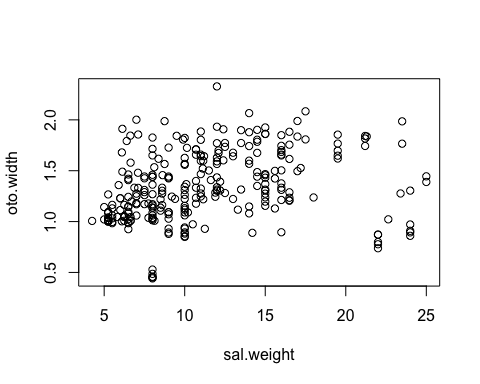
#Sal.ID  
plot(oto.width ~ sal.id, data = herring\_data)



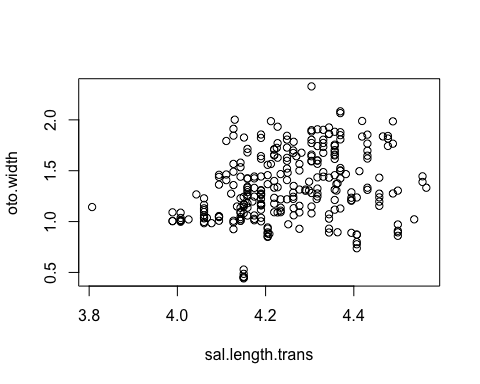
#Sal.length  
plot(oto.width ~ sal.length, data = herring\_data)



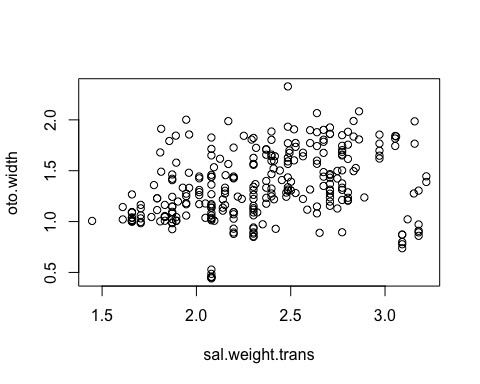
#Sal.weight  
plot(oto.width ~ sal.weight, data = herring\_data)



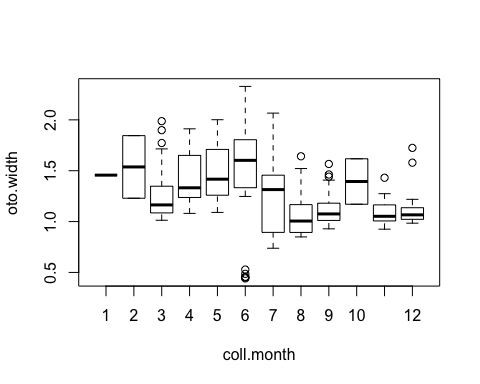
#Sal.length.trans  
plot(oto.width ~ sal.length.trans, data = herring\_data)



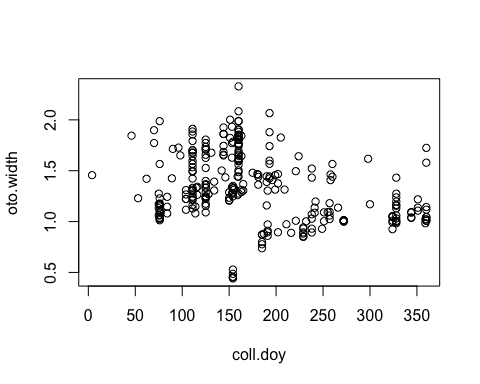
#Sal.weight.trans  
plot(oto.width ~ sal.weight.trans, data = herring\_data)



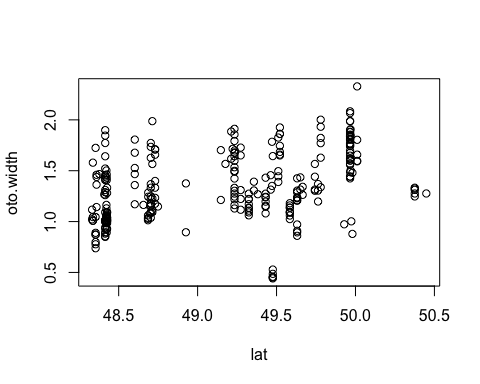
#Coll.month  
plot(oto.width ~ coll.month, data = herring\_data)



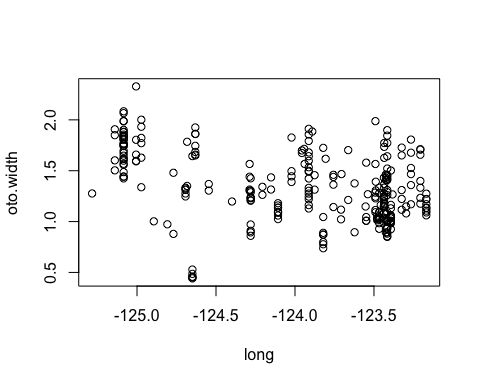
#Coll.doy  
plot(oto.width ~ coll.doy, data = herring\_data)



#Lat  
plot(oto.width ~ lat, data = herring\_data)



#Long  
plot(oto.width ~ long, data = herring\_data)

 ##Untransformed models

herring\_data <- herring\_data %>% filter(coll.month == "3" | coll.month == "4" | coll.month == "5" | coll.month == "6" | coll.month == "7")  
  
#Fit model 1  
model\_1 <- lmer(oto.width ~ coll.month + long + sal.length + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 2  
model\_2 <- lmer(oto.width ~ coll.month + long + sal.length + (1|sal.id), data = herring\_data)  
  
#Fit model 3  
model\_3 <- lmer(oto.width ~ coll.month + long + sal.weight + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 4  
model\_4 <- lmer(oto.width ~ coll.month + long + sal.weight + (1|sal.id), data = herring\_data)  
  
#Fit model 5  
model\_5 <- lmer(oto.width ~ coll.month + lat + sal.length + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 6  
model\_6 <- lmer(oto.width ~ coll.month + lat + sal.length + (1|sal.id), data = herring\_data)  
  
#Fit model 7  
model\_7 <- lmer(oto.width ~ coll.month + lat + sal.weight + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 8  
model\_8 <- lmer(oto.width ~ coll.month + lat + sal.weight + (1|sal.id), data = herring\_data)  
  
#Fit model 9  
model\_9 <- lmer(oto.width ~ coll.doy + long + sal.length + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 10  
model\_10 <- lmer(oto.width ~ coll.doy + long + sal.length + (1|sal.id), data = herring\_data)  
  
#Fit model 11  
model\_11 <- lmer(oto.width ~ coll.doy + long + sal.weight + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 12  
model\_12 <- lmer(oto.width ~ coll.doy + long + sal.weight + (1|sal.id), data = herring\_data)  
  
#Fit model 13  
model\_13 <- lmer(oto.width ~ coll.doy + lat + sal.length + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 14  
model\_14 <- lmer(oto.width ~ coll.doy + lat + sal.length + (1|sal.id), data = herring\_data)  
  
#Fit model 15  
model\_15 <- lmer(oto.width ~ coll.doy + lat + sal.weight + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 16  
model\_16 <- lmer(oto.width ~ coll.doy + lat + sal.weight + (1|sal.id), data = herring\_data)

##Transformed models

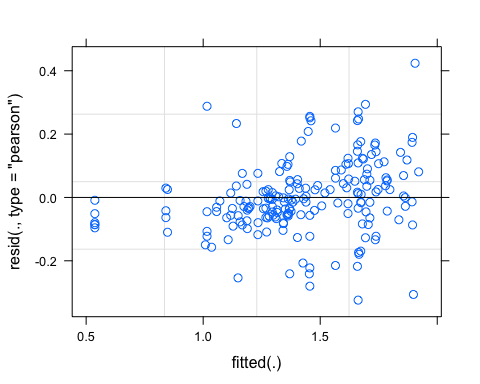
#Fit model 1t  
model\_1t <- lmer(oto.width ~ coll.month + long + sal.length.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 2t  
model\_2t <- lmer(oto.width ~ coll.month + long + sal.length.trans + (1|sal.id), data = herring\_data)  
  
#Fit model 3t  
model\_3t <- lmer(oto.width ~ coll.month + long + sal.weight.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 4t  
model\_4t <- lmer(oto.width ~ coll.month + long + sal.weight.trans + (1|sal.id), data = herring\_data)  
  
#Fit model 5t  
model\_5t <- lmer(oto.width ~ coll.month + lat + sal.length.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 6t  
model\_6t <- lmer(oto.width ~ coll.month + lat + sal.length.trans + (1|sal.id), data = herring\_data)  
  
#Fit model 7t  
model\_7t <- lmer(oto.width ~ coll.month + lat + sal.weight.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 8t  
model\_8t <- lmer(oto.width ~ coll.month + lat + sal.weight.trans + (1|sal.id), data = herring\_data)  
  
#Fit model 9t  
model\_9t <- lmer(oto.width ~ coll.doy + long + sal.length.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 10t  
model\_10t <- lmer(oto.width ~ coll.doy + long + sal.length.trans + (1|sal.id), data = herring\_data)  
  
#Fit model 11t  
model\_11t <- lmer(oto.width ~ coll.doy + long + sal.weight.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 12t  
model\_12t <- lmer(oto.width ~ coll.doy + long + sal.weight.trans + (1|sal.id), data = herring\_data)  
  
#Fit model 13t  
model\_13t <- lmer(oto.width ~ coll.doy + lat + sal.length.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 14t  
model\_14t <- lmer(oto.width ~ coll.doy + lat + sal.length.trans + (1|sal.id), data = herring\_data)  
  
#Fit model 15t  
model\_15t <- lmer(oto.width ~ coll.doy + lat + sal.weight.trans + digestion + (1|sal.id), data = herring\_data)  
  
#Fit model 16t  
model\_16t <- lmer(oto.width ~ coll.doy + lat + sal.weight.trans + (1|sal.id), data = herring\_data)

#Section 6: Model Checking ##All models summarized (delete before submitting?)

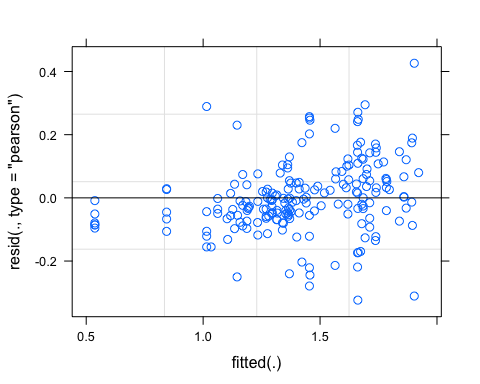
#Summary of model 1  
summary\_1 <- summary(model\_1)$coef  
  
#Summary of model 2  
summary\_2 <- summary(model\_2)$coef  
  
#Summary of model 3  
summary\_3 <- summary(model\_3)$coef  
  
#Summary of model 4  
summary\_4 <- summary(model\_4)$coef  
  
#Summary of model 5  
summary\_5 <- summary(model\_5)$coef  
  
#Summary of model 6  
summary\_6 <- summary(model\_6)$coef  
  
#Summary of model 7  
summary\_7 <- summary(model\_7)$coef  
  
#Summary of model 8  
summary\_8 <- summary(model\_8)$coef  
  
#Summary of model 9  
summary\_9 <- summary(model\_9)$coef  
  
#Summary of model 10  
summary\_10 <- summary(model\_10)$coef  
  
#Summary of model 11  
summary\_11 <- summary(model\_11)$coef  
  
#Summary of model 12  
summary\_12 <- summary(model\_12)$coef  
  
#Summary of model 13  
summary\_13 <- summary(model\_13)$coef  
  
#Summary of model 14  
summary\_14 <- summary(model\_14)$coef  
  
#Summary of model 15  
summary\_15 <- summary(model\_15)$coef  
  
#Summary of model 16  
summary\_16 <- summary(model\_16)$coef

##All models plotted against data (delete before submitting?)

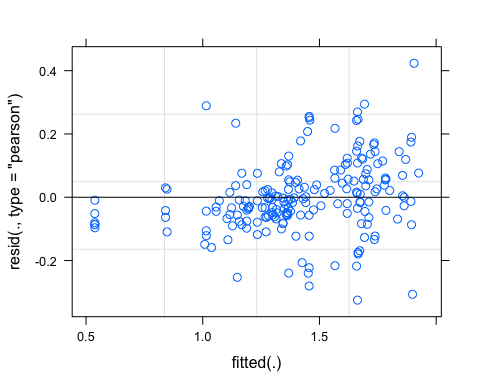
#Plot model 1 residuals  
plot(model\_1, which = 1)



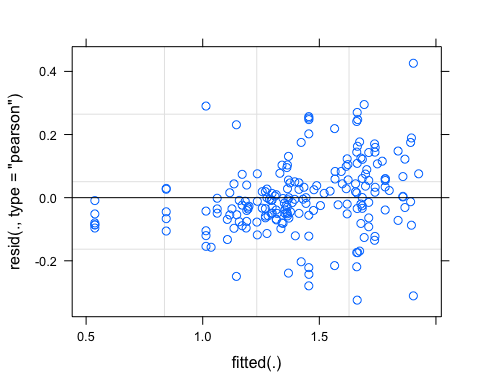
#Plot model 2 residuals  
plot(model\_2, which = 1)



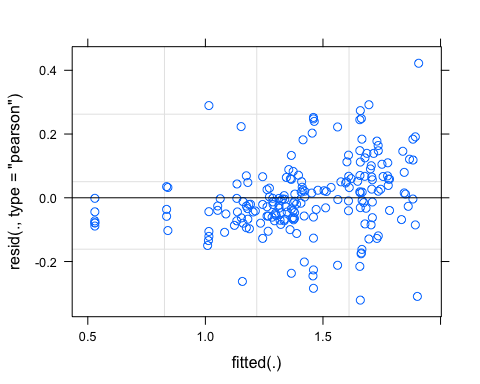
#Plot model 3 residuals  
plot(model\_3, which = 1)



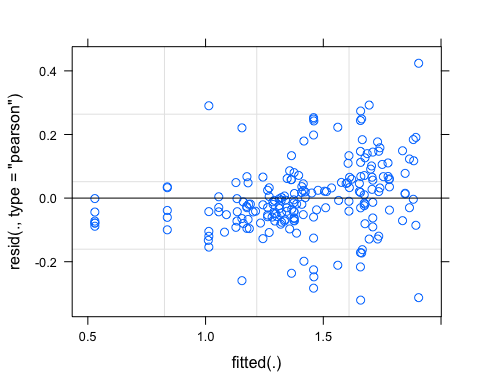
#Plot model 4 residuals  
plot(model\_4, which = 1)



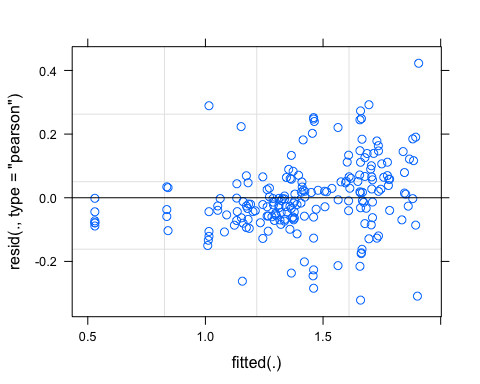
#Plot model 5 residuals  
plot(model\_5, which = 1)



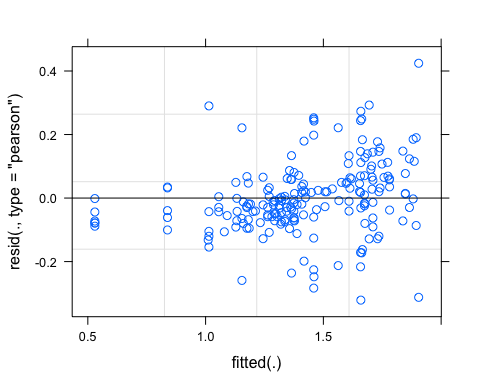
#Plot model 6 residuals  
plot(model\_6, which = 1)



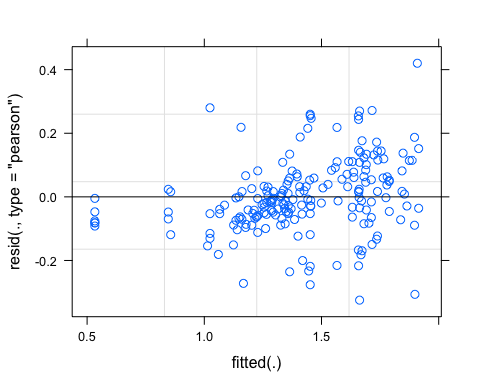
#Plot model 7 residuals  
plot(model\_7, which = 1)



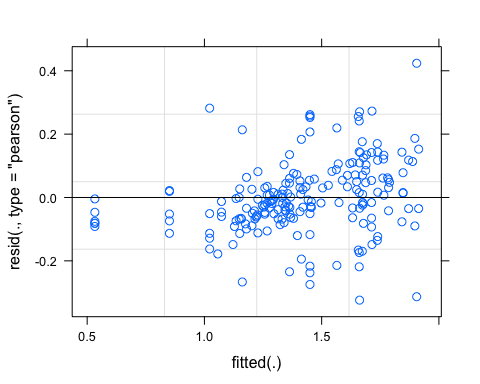
#Plot model 8 residuals  
plot(model\_8, which = 1)



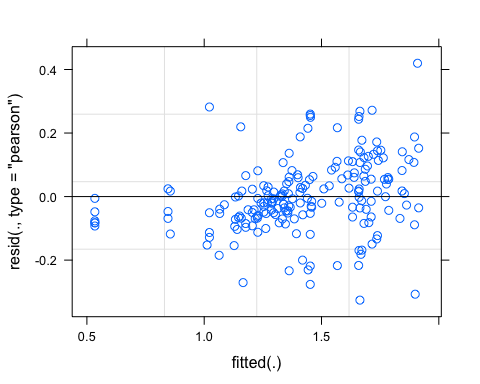
#Plot model 9 residuals  
plot(model\_9, which = 1)



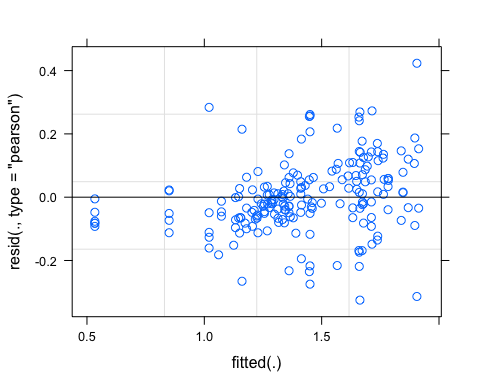
#Plot model 10 residuals  
plot(model\_10, which = 1)



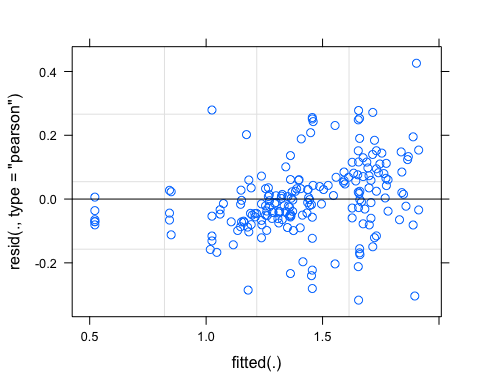
#Plot model 11 residuals  
plot(model\_11, which = 1)



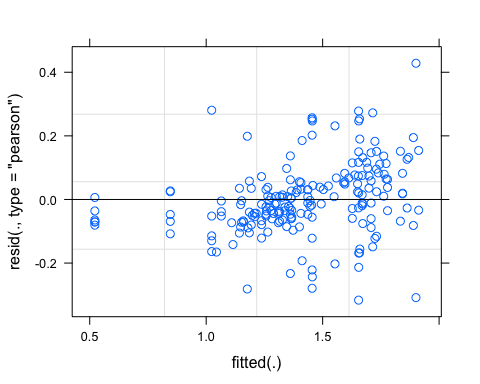
#Plot model 12 residuals  
plot(model\_12, which = 1)



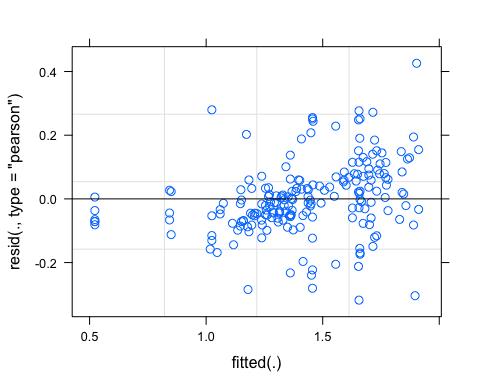
#Plot model 13 residuals  
plot(model\_13, which = 1)



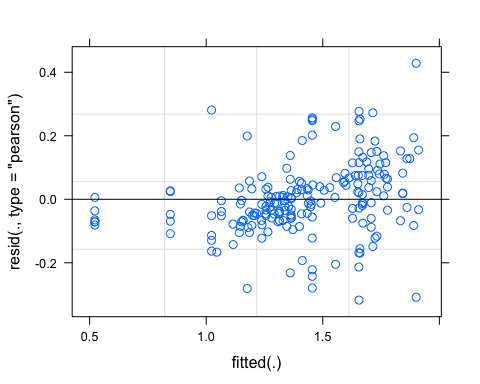
#Plot model 14 residuals  
plot(model\_14, which = 1)



#Plot model 15 residuals  
plot(model\_15, which = 1)

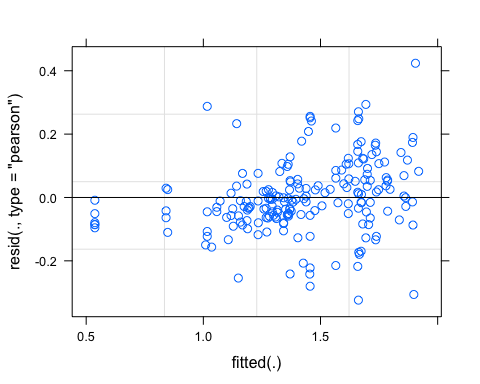


#Plot model 16 residuals  
plot(model\_16, which = 1)

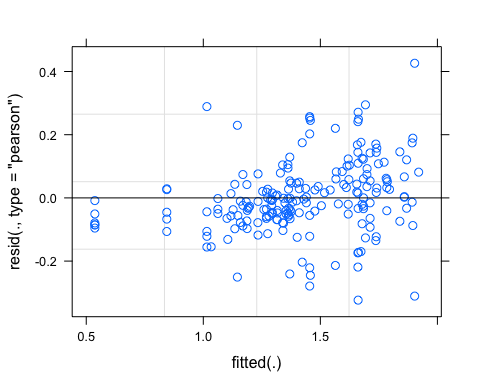


##All transformed models plotted against data (delete before submitting?)

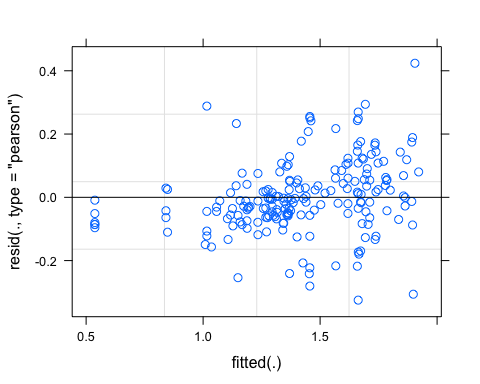
#Plot model 1t residuals  
plot(model\_1t, which = 1)



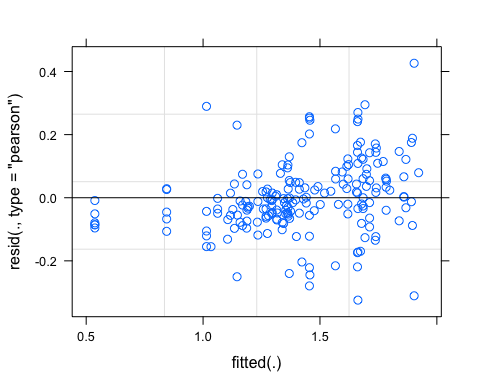
#Plot model 2t residuals  
plot(model\_2t, which = 1)



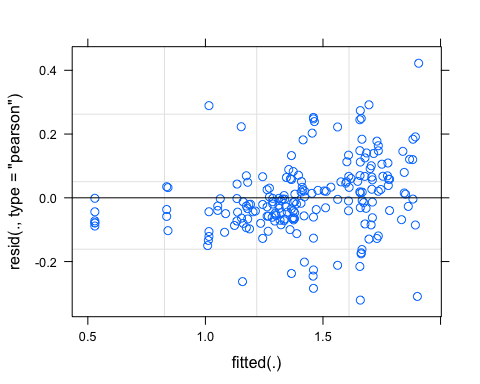
#Plot model 3t residuals  
plot(model\_3t, which = 1)



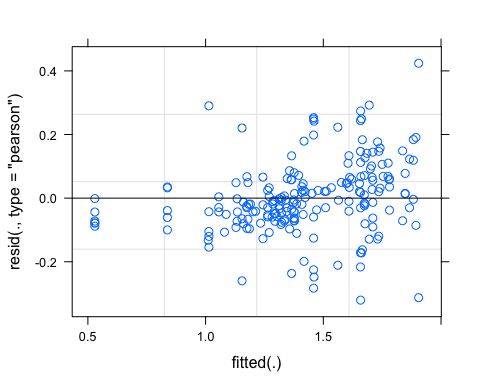
#Plot model 4t residuals  
plot(model\_4t, which = 1)



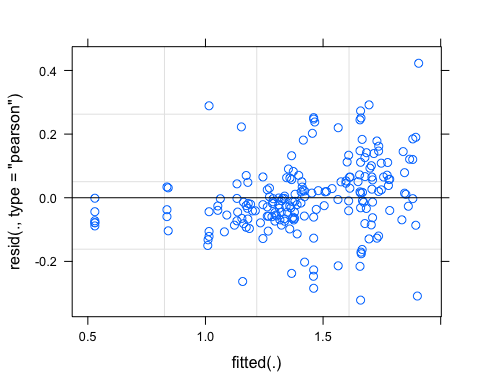
#Plot model 5t residuals  
plot(model\_5t, which = 1)



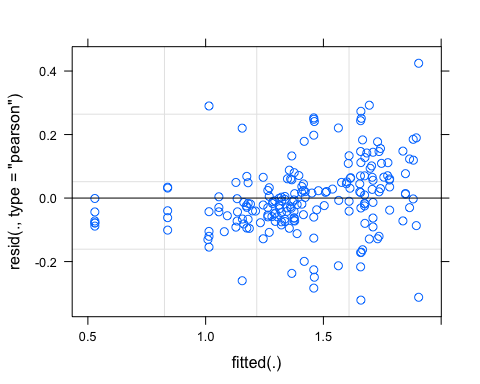
#Plot model 6t residuals  
plot(model\_6t, which = 1)



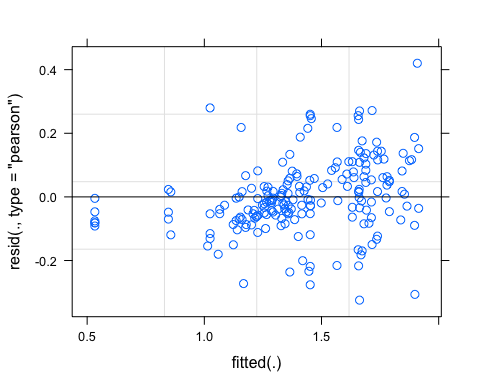
#Plot model 7t residuals  
plot(model\_7t, which = 1)



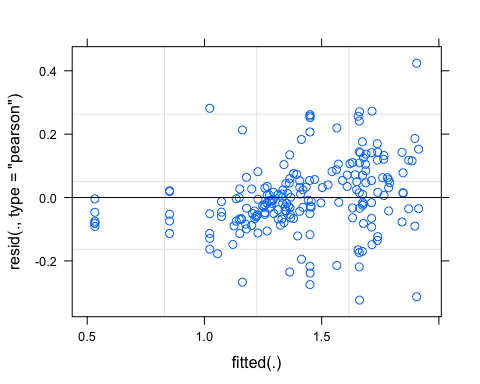
#Plot model 8t residuals  
plot(model\_8t, which = 1)



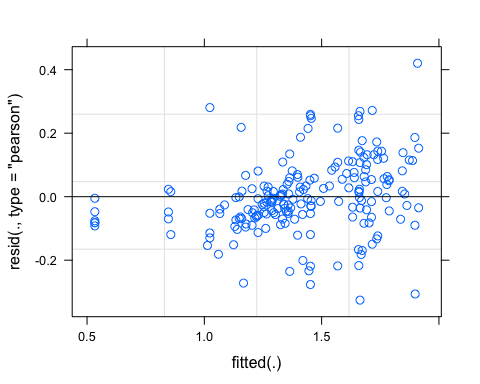
#Plot model 9t residuals  
plot(model\_9t, which = 1)



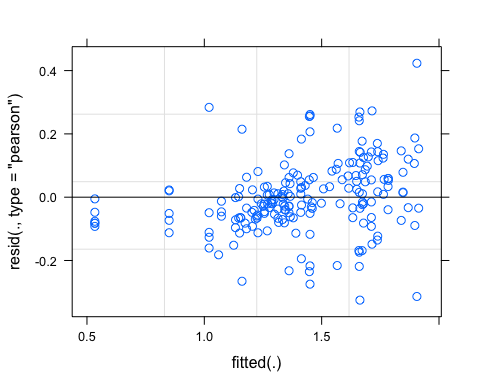
#Plot model 10t residuals  
plot(model\_10t, which = 1)



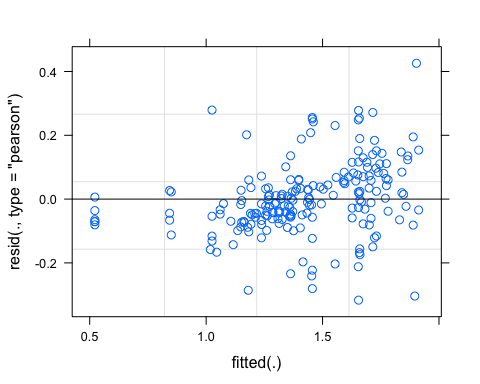
#Plot model 11t residuals  
plot(model\_11t, which = 1)



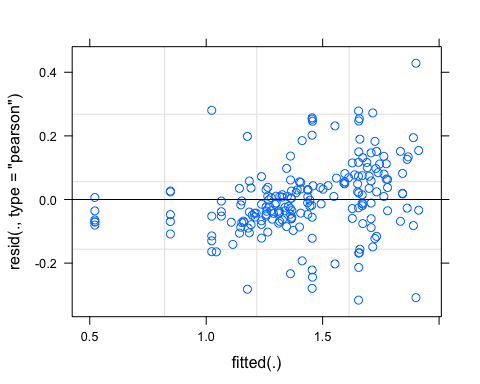
#Plot model 12t residuals  
plot(model\_12, which = 1)



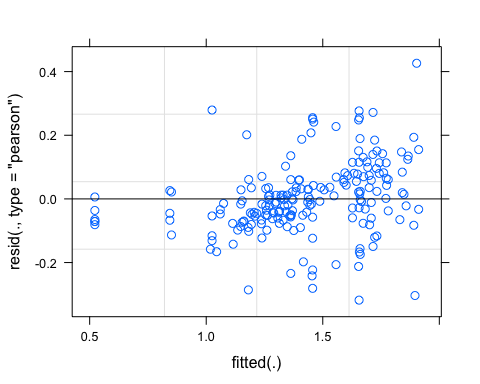
#Plot model 13t residuals  
plot(model\_13t, which = 1)



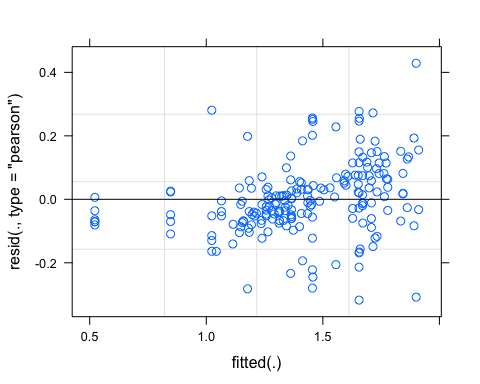
#Plot model 14t residuals  
plot(model\_14t, which = 1)



#Plot model 15t residuals  
plot(model\_15t, which = 1)



#Plot model 16t residuals  
plot(model\_16t, which = 1)

 ##Chosen models

#Check the assumptions of at least one model  
#- plotting and interpreting the residuals of the model  
#- check for correlation in the residuals due to grouping variables

Summary of the plausibility of the model - Potential outliers

#Section 7: Model Summary, Conﬁdence Intervals and Model Comparison

#Model comparison and/or model selection and/or hypothesis testing  
#- AIC (or AICc) values  
#- AIC (or AICc) weights  
#- R-squared (or adjusted R-squared) values  
#- p-values.   
#Summary for at least one ﬁtted model  
#- coeﬃcient estimates and their standard errors  
#- conﬁdence intervals for coeﬃcient estimate

#Section 8: Plotting a Model with the Data

#Plot at least one model with data  
#- use method that shows strengths and weaknesses  
#- include measurement of uncertainty

#Section 9: Discussion (400 words) ##A) Assessment of how well the model seems to ﬁt the data - model checking - visualization of the model with the data - ##B) Qualitative interpretation of the model - interpretation of results in relation to your research questions and hypotheses - general interpretation of conﬁdence intervals or standard errors for coeﬃcient estimates - ##C) Limitations of the model(s) that you have ﬁt - Limitations - Possible options for improvement - Possible approaches for further data analysis

##D) Results into context - citie at least two relevant scientiﬁc papers