Domestication via the commensal pathway in a fish-invertebrate mutualism

This R Markdown includes descriptions of all variables and the code used for the analysis of the following eight datasets:

1. Transects
2. Census of farms
3. Paired choice experiments
4. Predation experiment 1
5. Predation experiment 2
6. Timed observations
7. Farm algae composition
8. Longfin damselfish body condition

In addition, this R Markdown includes descriptions of all variables for two additional datasets:

9. Estimates of mysid swarm density
10. Mysid waste excretion and nutrient availability

See methods section of Nature Communications manuscript for further information on data collection and analysis procedures.

```
# Set working directory and read the accompanying data files.
# This .Rmd file must be saved in the same folder as the data files.
# Do not use setwd(), but rather choose "Sessions/Set Working Directory/To Files Pa
ne Location"
# from the RStudio drop down menu

Transects<-read.csv('CBC2018_Transects_data.csv', strip.white=TRUE, header=TRUE)
Census<-read.csv('CBC2018_SiteFidelity_AM_data.csv', strip.white=TRUE, header=TRUE)
PairedChoice<- read.csv("CBC2018_PairedChoice_data.csv", strip.white=TRUE, header=T
RUE)
PredatorOne<-read.csv('CBC2018_PredationExperiment1_data.csv', strip.white=TRUE, he
ader=TRUE)
PredatorTwo<-read.csv('CBC2018_PredationExperiment2_data.csv', strip.white=TRUE, he
ader=TRUE)
Observation<-read.csv('CBC2018_Observations_data.csv', strip.white=TRUE, header=TRU
E)
FarmComp <- read.csv("CBC2018_Ochrophyta_data.csv")
BodyCondition<-read.csv('CBC2018_FishBodyCondition_data.csv', strip.white=TRUE, hea
der=TRUE)
```

1. Transects

Description of data

Transects conducted to determine whether mysid swarms were associated with the territories of farming *Stegastes* damselfish.

The dataset (CBC2018_Transects_data.csv) contains the following variables (L-R):

- Transect: transect number (1-30).
- FishSpecies: identity of *Stegastes* spp. ((BG) *S. leucostictus*, (BC) *S. partitus*, (CC) *S. variabilis*, (DD) *S. adustus*, (LF) *S. diencaeus*, (TS) *S. planifrons*).
- IntensiveOrExtensive: whether *Stegastes* spp. is an intensive or extensive farmer ((0) extensive, (1) intensive).
- MysidsPresent: whether mysids were present within farm ((0) absent, (1) present).

Analysis

Test for difference in preference with a chi-square test

```
chisq.test(Transects$IntensiveOrExtensive, Transects$MysidsPresent)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  Transects$IntensiveOrExtensive and Transects$MysidsPresent
## X-squared = 29.874, df = 1, p-value = 4.611e-08
```

Conclusion: There is a difference (i.e. mysids are more likely to be found in association with intensive farming damselfish species)

2. Census of farms

Description of data

Observations of swarm presence/ absence to determine whether mysids remain associated with the same longfin damselfish territories over an extended period. Analysis was conducted on a subset of data (CBC2018_SiteFidelity_AM_data.csv).

The full dataset (CBC2018_SiteFidelity_All_data.csv) contains the following variables (L-R):

- Territory: farm number (1-60).
- ObservationNumber: observation number (1-6).
- TimeOfDay: time of observation ((AM) 1-hr post sunrise, (PM) 1-hr post sunset).
- TerritoryType: whether mysids were present within farm at first observation ((Mysid) present, (NoMysid) absent).
- Presence1Absence0: whether mysids were present within farm at observation ((0) absent, (1) present).

The dataset used for analysis (CBC2018_SiteFidelity_AM_data.csv) contains the following variables (L-R):

- Territory: farm number (1-30).

- ObservationNumber: observation number (1, 3, 5).
- TimeOfDay: time of observation ((AM) 1-hr post sunrise).
- TerritoryType: whether mysids were present within farm at first observation ((Mysid) present).
- Presence1Absence0: whether mysids were present within farm at observation ((0) absent, (1) present).

Analysis

Test for difference with a Friedman test

```
friedman.test((Presence1Absence0~ObservationNumber|Territory),data=Census)
```

```
##
##  Friedman rank sum test
##
## data:  Presence1Absence0 and ObservationNumber and Territory
## Friedman chi-squared = 2, df = 2, p-value = 0.3679
```

Conclusion: There were no difference (i.e. if mysids were associated with a farm, they remained so through time)

3. Olfactory choice experiments

Description of data

A series of choice experiments to examine the responses of mysids to different olfactory cues.

The dataset (CBC2018_PairedChoice_data.csv) contains the following variables (L-R):

- Date: date of replicate.
- Time: time of replicate.
- Tester: identity of tester ((ZLC) Zara-Louise Cowan).
- Cue: olfactory cue being tested ((Turf) farmed algae, (Stegastes) longfin damselfish, (Dick) slippery dick wrasse, (Bicolor) bicolor damselfish).
- Control: water source of control water ((CBC) Carrie Bow Cay Research Station seawater system).
- SideOfCue: header bucket in which cue was places (Left or Right)
- Replicate: replicate number (1-30 as standard, plus additional trials as needed)
- LS: number of time points (out of 24) an individual mysid was recorded on the left hand side of flume during test period 1.
- RS: number of time points (out of 24) an individual mysid was recorded on the right hand side of flume during test period 1.
- RS: number of time points (out of 24) an individual mysid was recorded on the right hand side of flume during test period 2.
- LS: number of time points (out of 24) an individual mysid was recorded on the left hand side of flume during test period 2.
- Total: total number of time points across full replicate (48).
- Complete: whether a full replicate was completed ((y) yes or (n) no).
- Total_LS_Cue: total number of time points (out of 48) an individual mysid was recorded on the left hand side of flume over the entire replicate.

- Total_RS_Cue: total number of time points (out of 48) an individual mysid was recorded on the right hand side of flume over the entire replicate.
- Total_In_Cue: total number of time points (out of 48) an individual mysid was recorded in water containing the test cue over the entire replicate.
- Total_In_Control: total number of time points (out of 48) an individual mysid was recorded in control water over the entire replicate.
- %_In_Cue: percent of total replicate time (based on time points) an individual mysid spent in water containing the test cue.
- %_In_Control: percent of total replicate time (based on time points) an individual mysid spent in control water.
- %_LS: percent of total replicate time (based on time points) an individual mysid spent on left side of flume.
- %_RS: percent of total replicate time (based on time points) an individual mysid spent on right side of flume.
- Side_Preference: whether an individual mysid was determined to exhibit a side preference in the flume ((y) yes or (n) no).
- Note: additional information regarding replicates that were excluded from the analysis.

Analysis

Remove incomplete and side preferences from all data

```
d <- subset(PairedChoice,Complete == 'y')
d <- subset(d,Side_Preference == 'n')

library(Rmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: plyr
```

```
summarySE(d,  measurevar=c("X._In_Cue"),groupvars=c("Cue"))
```

```
##              Cue  N X._In_Cue       sd       se       ci
## 1        Bicolor 30  48.81944 18.33989 3.348391 6.848228
## 2           Dick 30  41.38889 21.06007 3.845025 7.863959
## 3      Stegastes 30  62.01389 20.36605 3.718315 7.604807
## 4 StegastesMysid 30  44.93056 23.51099 4.292500 8.779149
## 5           Turf 30  50.97222 19.23256 3.511369 7.181557
```

Subset bicolor damselfish vs. seawater data for analysis

```
d.Bicolor <- subset(d,Cue=='Bicolor')
d.Bicolor.se <- summarySE(d.Bicolor,measurevar="X._In_Cue")
d.Bicolor.se
```

```
##    .id  N X._In_Cue       sd       se       ci
## 1 <NA> 30  48.81944 18.33989 3.348391 6.848228
```

Test for difference in preference with a paired sample t-test

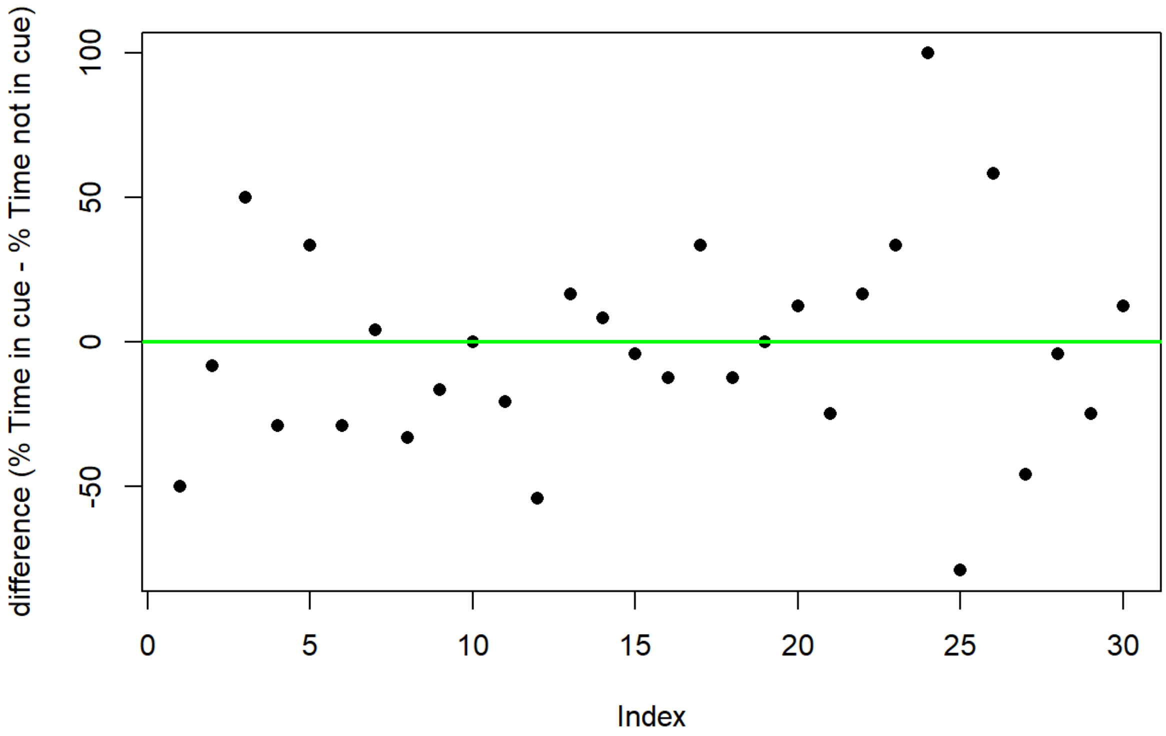
```
t.test(d.Bicolor$X._In_Cue,
       d.Bicolor$X._In_Control,
       paired=TRUE,
       comf.level=.95)
```

```
## 
##  Paired t-test
## 
## data:  d.Bicolor$X._In_Cue and d.Bicolor$X._In_Control
## t = -0.35257, df = 29, p-value = 0.727
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.05757  11.33534
## sample estimates:
## mean of the differences
##               -2.361111
```

Conclusion: There is no difference (i.e. mysids showed no preference for odor of bicolor damselfish vs. seawater)

Confirm the validity of assumptions of this test:

```
Diff.d.Bicolor=d.Bicolor$X._In_Cue-d.Bicolor$X._In_Control

plot(Diff.d.Bicolor,
     pch=16,
     ylab="difference (% Time in cue - % Time not in cue)")
abline(0,0,col="green",lwd = 2)
```



Test for normality

```
shapiro.test(x=Diff.d.Bicolor)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Diff.d.Bicolor
## W = 0.97762, p-value = 0.7594
```

All assumptions are valid

Subset longfish damselfish that were associated with mysids vs. longfish damselfish that were not associated with mysids

```
d.StegastesMysid <- subset(d,Cue=='StegastesMysid')
d.StegastesMysid.se <- summarySE(d.StegastesMysid,measurevar="X._In_Cue")
d.StegastesMysid.se
```

```
##    .id  N X._In_Cue       sd     se       ci
## 1 <NA> 30  44.93056 23.51099 4.2925 8.779149
```

Test for difference in preference with a paired sample t-test

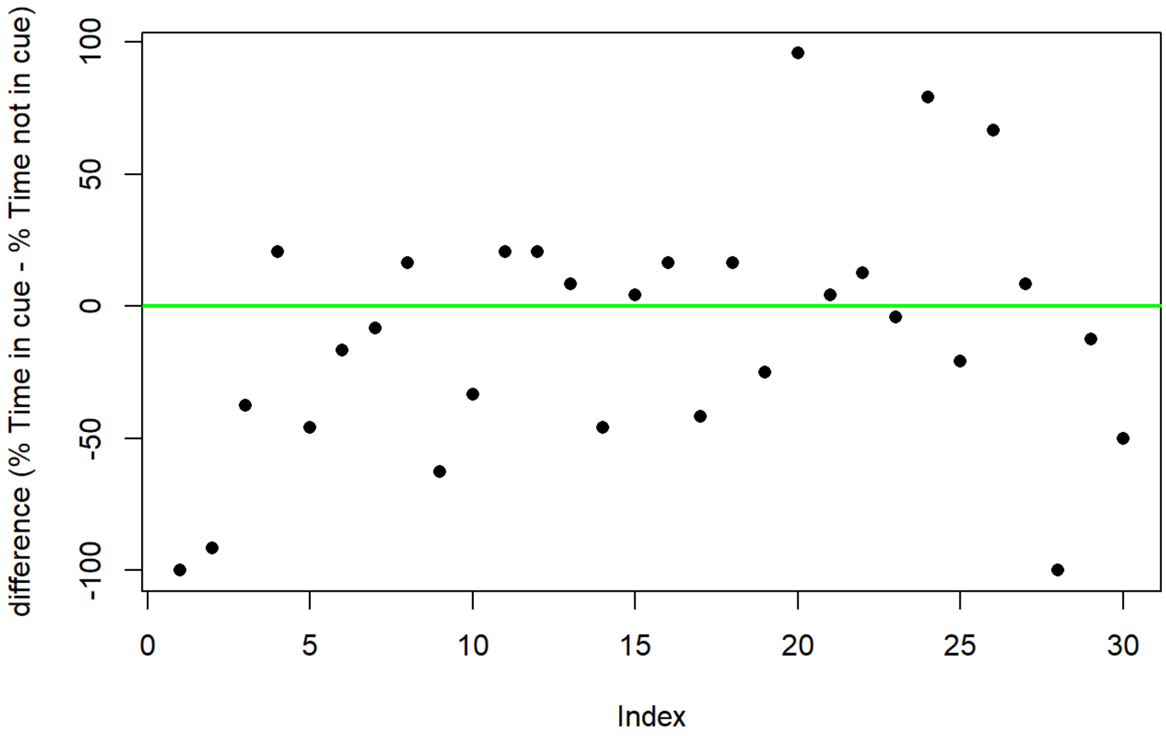
```
t.test(d.StegastesMysid$X._In_Cue,
       d.StegastesMysid$X._In_Control,
       paired=TRUE,
       comf.level=.95)
```

```
##
##  Paired t-test
##
## data:  d.StegastesMysid$X._In_Cue and d.StegastesMysid$X._In_Control
## t = -1.181, df = 29, p-value = 0.2472
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -27.697187   7.419409
## sample estimates:
## mean of the differences
##               -10.13889
```

Conclusion: There is no difference (i.e. mysids showed no preference for odor of longfish damselfish that were associated with mysids vs. longfish damselfish that were not associated with mysids damselfish)

Confirm the validity of assumptions of this test:

```
Diff.d.StegastesMysid=d.StegastesMysid$X._In_Cue-d.StegastesMysid$X._In_Control

plot(Diff.d.StegastesMysid,
     pch=16,
     ylab="difference (% Time in cue - % Time not in cue)")
abline(0,0,col="green",lwd = 2)
```



Test for normality

```
shapiro.test(x=Diff.d.StegastesMysid)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Diff.d.StegastesMysid
## W = 0.96169, p-value = 0.3418
```

All assumptions are valid

Subset farmed algae vs. seawater

```
d.Turf <- subset(d,Cue=='Turf')
d.Turf.se <- summarySE(d.Turf,measurevar="X._In_Cue")
d.Turf.se
```

```
##    .id  N X._In_Cue       sd       se       ci
## 1 <NA> 30  50.97222 19.23256 3.511369 7.181557
```

Test for difference in preference with a paired sample t-test

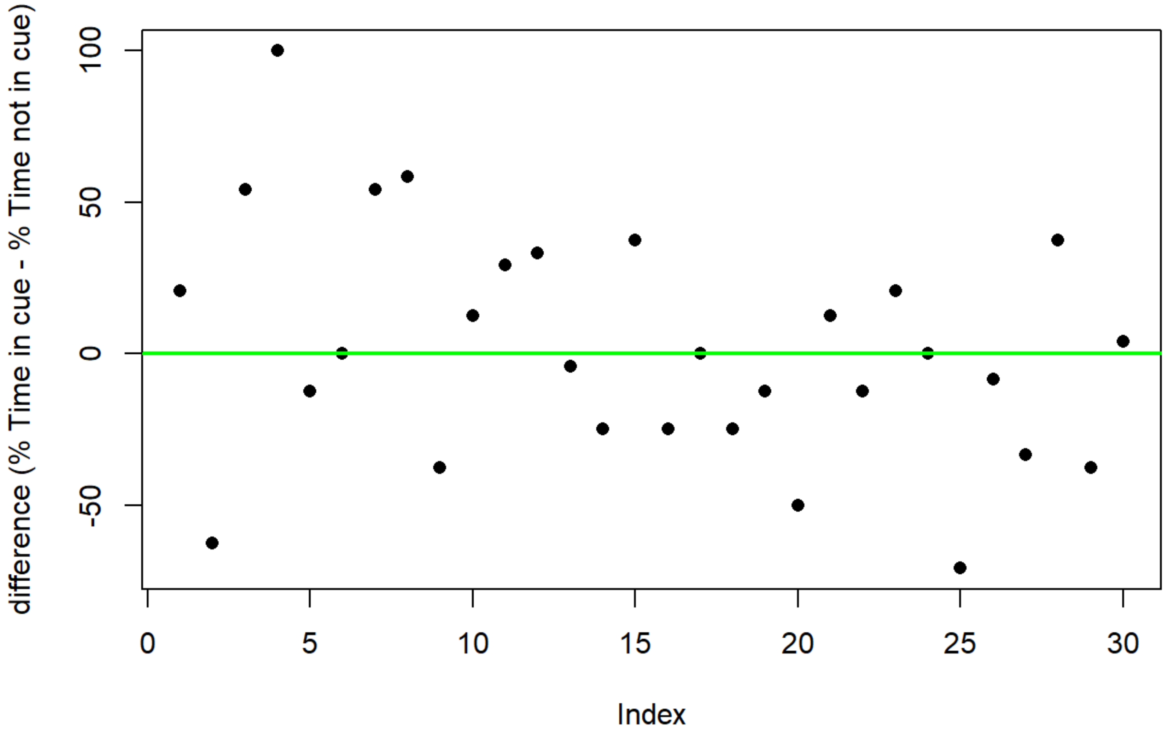
```
t.test(d.Turf$X._In_Cue,
       d.Turf$X._In_Control,
       paired=TRUE,
       comf.level=.95)
```

```
##
##  Paired t-test
##
## data:  d.Turf$X._In_Cue and d.Turf$X._In_Control
## t = 0.27688, df = 29, p-value = 0.7838
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.41867  16.30756
## sample estimates:
## mean of the differences
##                1.944444
```

Conclusion: There is no difference (i.e. mysids showed no preference for the odor of farm algae vs. seawater)

Confirm the validity of assumptions of this test:

```
Diff.d.Turf=d.Turf$X._In_Cue-d.Turf$X._In_Control

plot(Diff.d.Turf,
     pch=16,
     ylab="difference (% Time in cue - % Time not in cue)")
abline(0,0,col="green",lwd = 2)
```



Test for normality

```
shapiro.test(x=Diff.d.Turf)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Diff.d.Turf
## W = 0.98493, p-value = 0.9361
```

All assumptions are valid

Subset slippery dick wrasse vs. seawater

```
d.Dick <- subset(d,Cue=='Dick')
d.Dick.se <- summarySE(d.Dick,measurevar="X._In_Cue")
d.Dick.se
```

```
##    .id  N X._In_Cue       sd       se       ci
## 1 <NA> 30  41.38889 21.06007 3.845025 7.863959
```

Test for difference in preference with a paired sample t-test

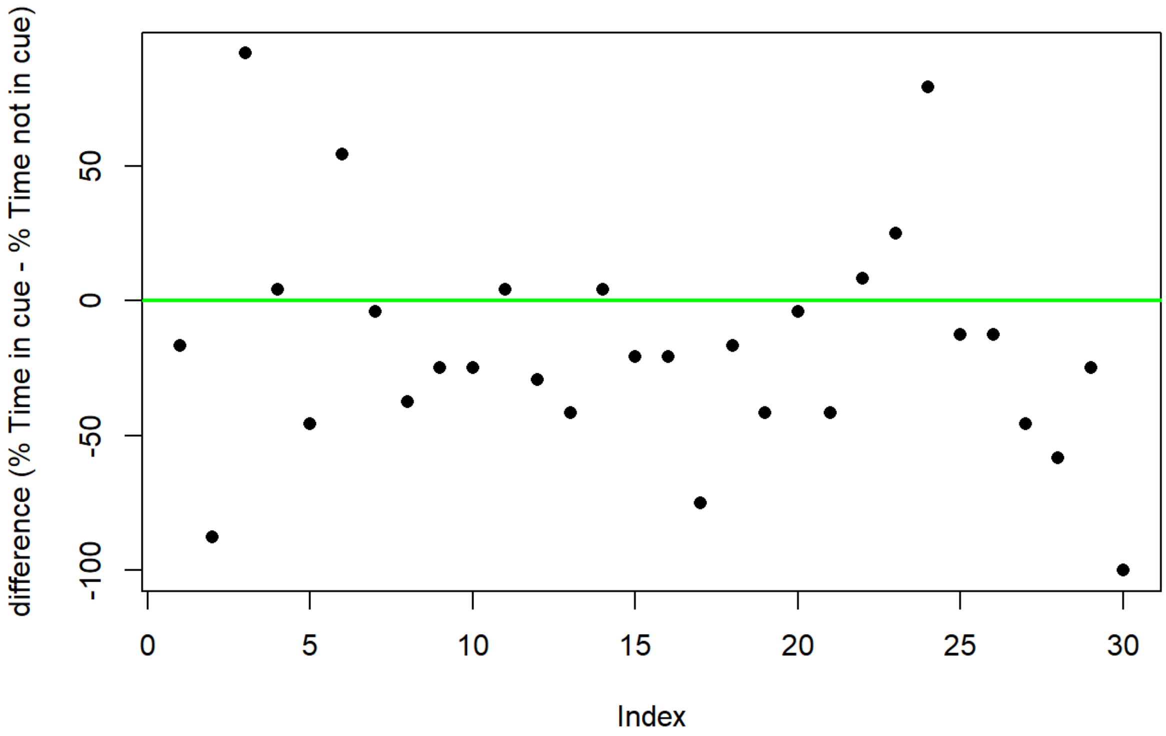
```
t.test(d.Dick$X._In_Cue,
       d.Dick$X._In_Control,
       paired=TRUE,
       comf.level=.95)
```

```
##
##  Paired t-test
##
## data:  d.Dick$X._In_Cue and d.Dick$X._In_Control
## t = -2.2395, df = 29, p-value = 0.03295
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -32.950140  -1.494304
## sample estimates:
## mean of the differences
##               -17.22222
```

Conclusion: There is a difference (i.e. mysids avoid the odor of slippery dick wrasse)

Confirm the validity of assumptions of this test:

```
Diff.d.Dick=d.Dick$X._In_Cue-d.Dick$X._In_Control

plot(Diff.d.Dick,
     pch=16,
     ylab="difference (% Time in cue - % Time not in cue)")
abline(0,0,col="green",lwd = 2)
```



Test for normality

```
shapiro.test(x=Diff.d.Dick)
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  Diff.d.Dick
## W = 0.94173, p-value = 0.1013
```

All assumptions are valid

Subset longfin damselfish vs. seawater

```
d.Stegastes <- subset(d,Cue=='Stegastes')
d.Stegastes.se <- summarySE(d.Stegastes,measurevar="X._In_Cue")
d.Stegastes.se
```

```
##    .id  N X._In_Cue       sd       se       ci
## 1 <NA> 30  62.01389 20.36605 3.718315 7.604807
```

Test for difference in preference with a paired sample t-test

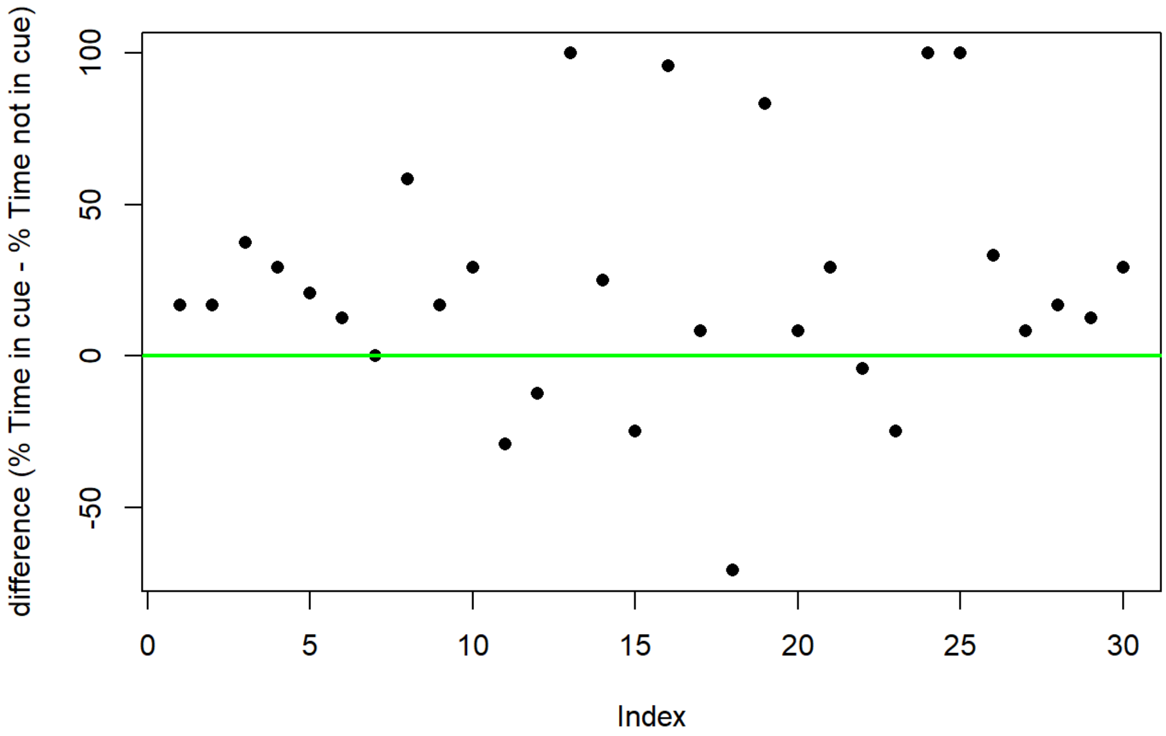
```
t.test(d.Stegastes$X._In_Cue,
       d.Stegastes$X._In_Control,
       paired=TRUE,
       comf.level=.95)
```

```
##
##  Paired t-test
##
## data:  d.Stegastes$X._In_Cue and d.Stegastes$X._In_Control
## t = 3.231, df = 29, p-value = 0.003065
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   8.818163 39.237392
## sample estimates:
## mean of the differences
##                24.02778
```

Conclusion: There is a difference (i.e. mysids are attracted to the odor of longfin damselfish)

Confirm the validity of assumptions of this test:

```
Diff.d.Stegastes=d.Stegastes$X._In_Cue-d.Stegastes$X._In_Control

plot(Diff.d.Stegastes,
     pch=16,
     ylab="difference (% Time in cue - % Time not in cue)")
abline(0,0,col="green",lwd = 2)
```



Test for normality

```
shapiro.test(x=Diff.d.Stegastes)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Diff.d.Stegastes
## W = 0.92277, p-value = 0.03167
```

Assumptions not met, so use non-parametric Wilcoxon test

```
wilcox.test(d.Stegastes$X._In_Cue, d.Stegastes$X._In_Control, paired=TRUE)
```

```
## Warning in wilcox.test.default(d.Stegastes$X._In_Cue,
## d.Stegastes$X._In_Control, : cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(d.Stegastes$X._In_Cue,
## d.Stegastes$X._In_Control, : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  d.Stegastes$X._In_Cue and d.Stegastes$X._In_Control
## V = 358, p-value = 0.002435
## alternative hypothesis: true location shift is not equal to 0
```

Conclusion: There is a difference (i.e. mysids are attracted to the odor of longfin damselfish)

4. Predation experiment 1

Description of data

A field-based experiment using captive mysid swarms to examine whether associating with longfin damselfish reduces predation risk to mysids.

The dataset (CBC2018_PredationExperiment1_data.csv) contains the following variables (L-R):

- Trial: trial number (1-30).
- Date: date of trial.
- Habitat: benthic habitat ((NS) outside longfin damselfish farm, (ST) within longfin damselfish farm).
- Treatment: treatment type ((ANM) bag with seawater only, (FM) bag with imitation mysids, (M) bag with live mysids).
- Order: order in which a trial took place (1 to 3).
- NumberOfSpecies_LessThan0.5m: number of fish species that came within 0.5 m of bag during trial.
- NumberOfFishIndividuals_LessThan0.5m: number of individual fish that came within 0.5 m of bag during trial.
- TimeOfFirstBite_S: time of first strike on bag (in seconds), or (na) if no strikes taken.
- NumberOfFishThatMadeStrikes: number of individual fish that struck at bag during trial.
- NumberOfSpeciesThatMadeStrikes: number of fish species that struck at bag during trial.
- TotalNumberOfStrikes: total number of strikes made at bag during trial.
- NumberOfStegastesChases: total number of chases directed at other fishes by resident longfin damselfish during trial.

Analysis

```
library(glmmTMB)
```

```
## Warning in checkMatrixPackageVersion(): Package version inconsistency detected.
## TMB was built with Matrix version 1.2.15
## Current Matrix version is 1.2.14
## Please re-install 'TMB' from source using install.packages('TMB', type = 'sourc
e') or ask CRAN for a binary version of 'TMB' matching CRAN's 'Matrix' package
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```r
library(DHARMa)
```

```
## This is DHARMa 0.3.3.0. For overview type '?DHARMa'. For recent changes, type ne
ws(package = 'DHARMa') Note: Syntax of plotResiduals has changed in 0.3.0, see ?plo
tResiduals for details
```

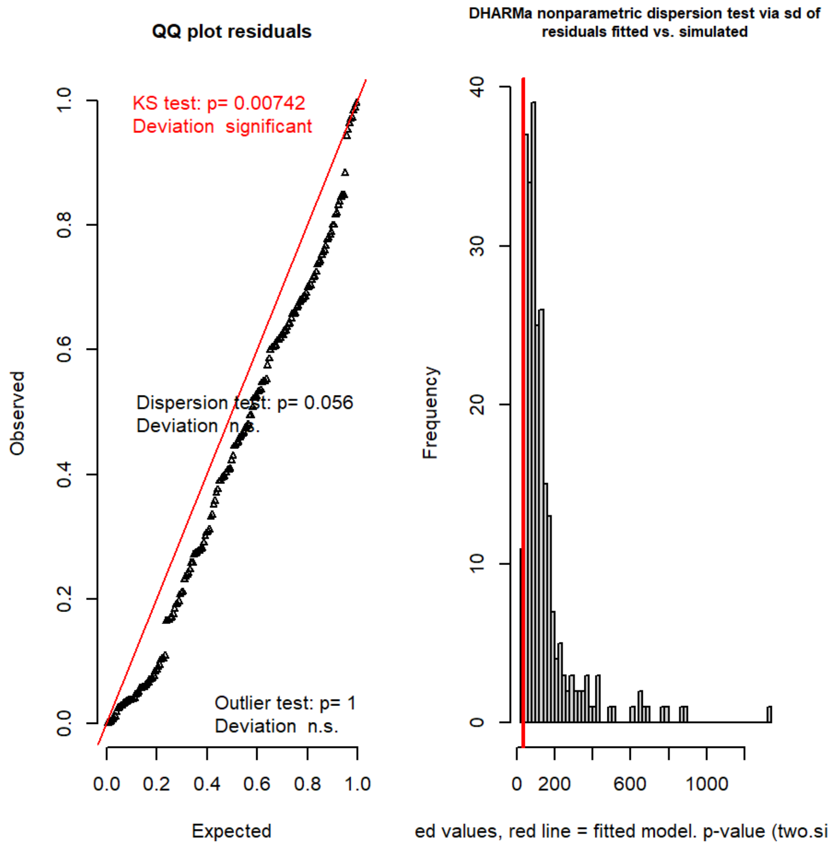
```r
library(emmeans)

PredatorOne$Trial<-factor(PredatorOne$Trial)
```

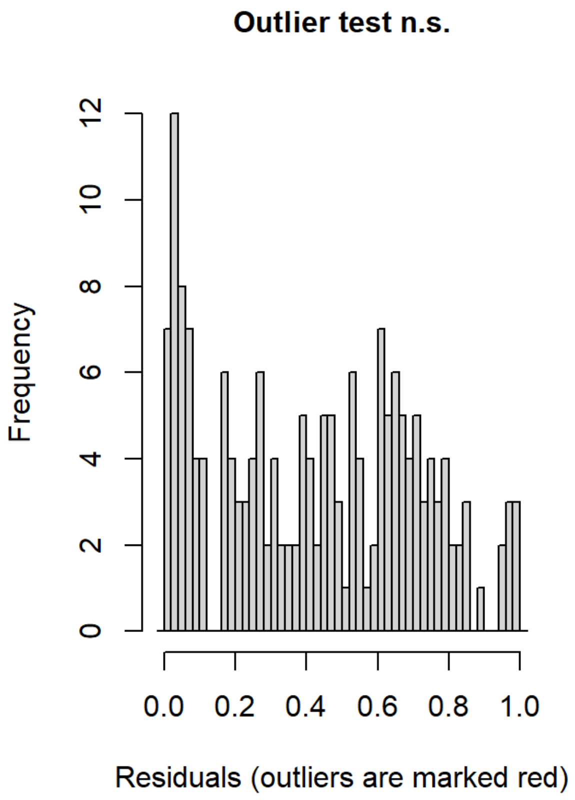
Test the effect of habitat (inside or outside farm) and treatment (mysid, imitation mysid [control 1], or empty bag [control 2]) on number of strikes by predators with a glmm

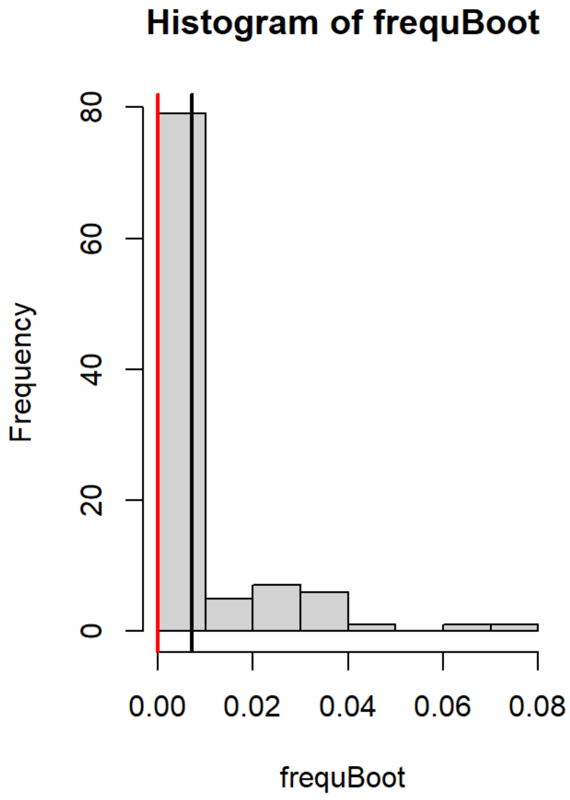
```r
mod1<-glmer(TotalNumberOfStrikes~Habitat*Treatment+(1|Trial),family=poisson,data=Pr
edatorOne)
```

Test model assumptions

```r
testResiduals(mod1)
```




Outlier test n.s.
Frequency
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)

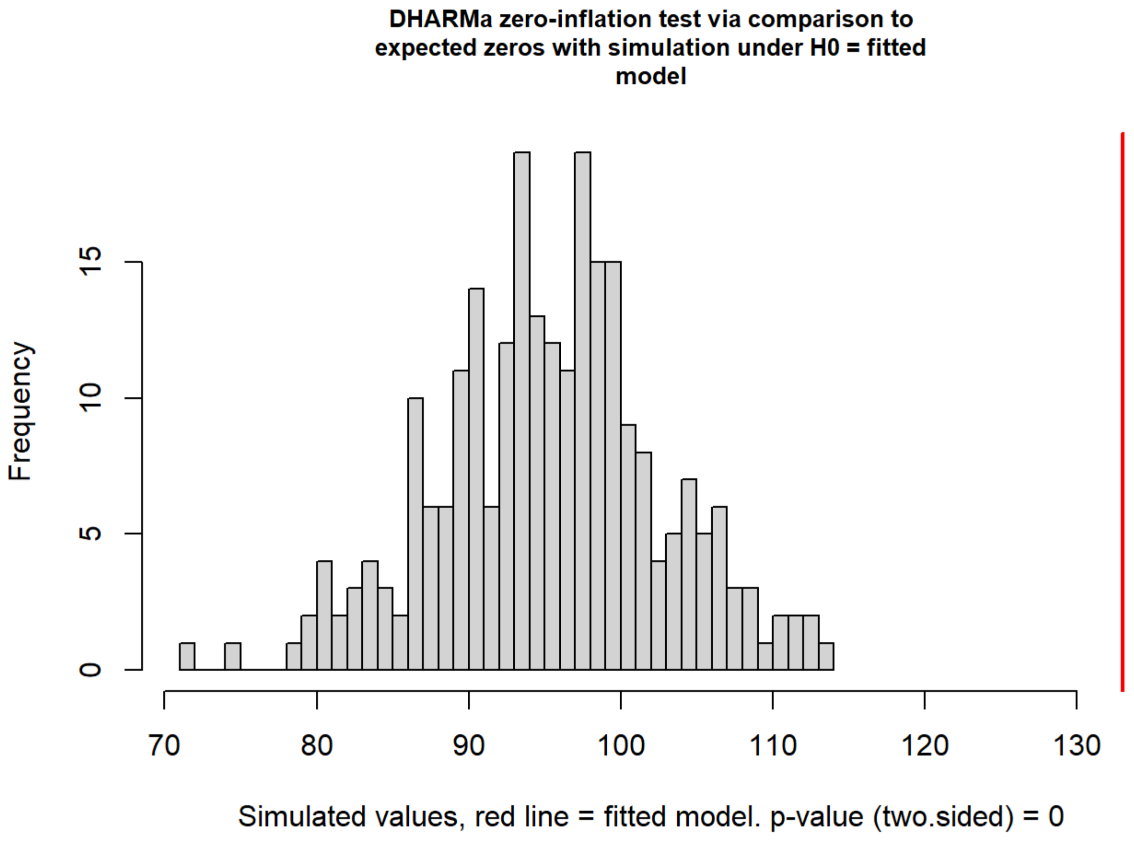

Histogram of frequBoot
Frequency
0
20
40
60
80
0.00
0.02
0.04
0.06
0.08
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.12468, p-value = 0.007424
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.23377, p-value = 0.056
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04180556
## sample estimates:
## outlier frequency (expected: 0.00711111111111111 )
##                                                   0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.12468, p-value = 0.007424
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.23377, p-value = 0.056
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04180556
## sample estimates:
## outlier frequency (expected: 0.00711111111111111 )
##                                                   0
```

Model assumptions not met, investigate further

```
testZeroInflation(mod1)
```



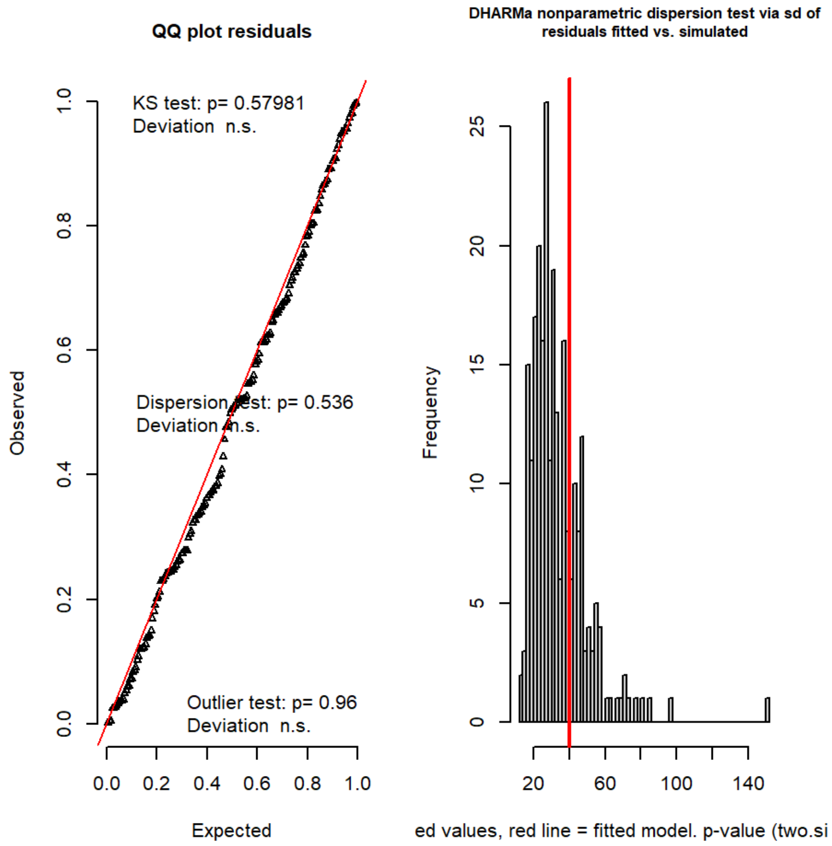
```
##
##  DHARMa zero-inflation test via comparison to expected zeros with
##  simulation under H0 = fitted model
##
## data:  simulationOutput
## ratioObsSim = 1.3885, p-value < 2.2e-16
## alternative hypothesis: two.sided
```

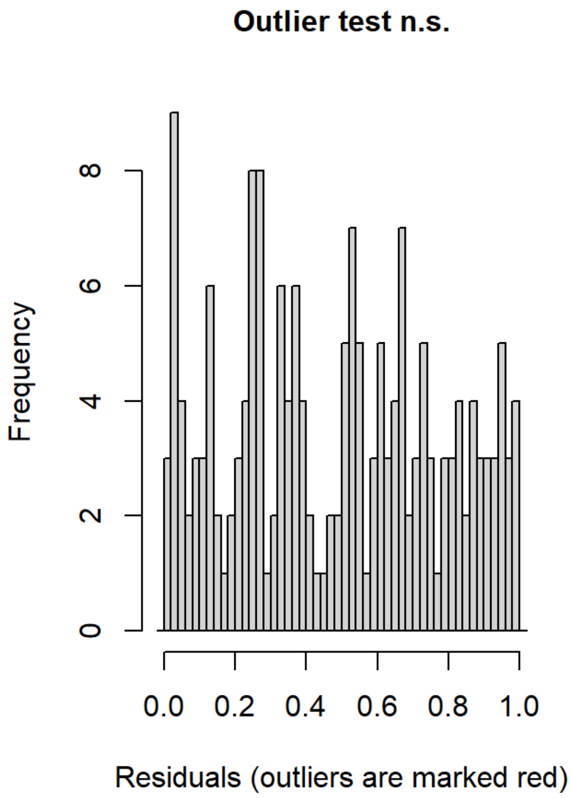
Data is zero-inflated, so use glmmTMB

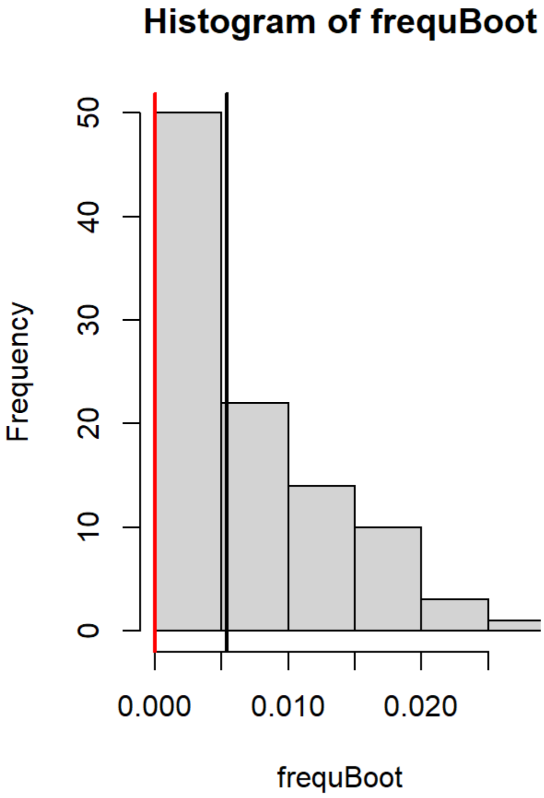
```
mod1<-glmmTMB(TotalNumberOfStrikes~Habitat*Treatment+(1|Trial),family=poisson,data=
PredatorOne,ziformula=~1)
```

Test model assumptions

```
testResiduals(mod1)
```


QQ plot residuals
KS test: p= 0.57981
Deviation n.s.
Dispersion test: p= 0.536
Deviation n.s.
Outlier test: p= 0.96
Deviation n.s.
Observed
Expected
0.0
0.2
0.4
0.6
0.8
1.0
DHARMa nonparametric dispersion test via sd of residuals fitted vs. simulated
Frequency
0
5
10
15
20
25
20
60
100
140
ed values, red line = fitted model. p-value (two.si


Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
0.000
0.010
0.020
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.058011, p-value = 0.5798
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 1.1782, p-value = 0.536
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.02222222
## sample estimates:
## outlier frequency (expected: 0.00538888888888889 )
##                                                   0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.058011, p-value = 0.5798
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 1.1782, p-value = 0.536
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.02222222
## sample estimates:
## outlier frequency (expected: 0.00538888888888889 )
##                                                  0
```

Assumptions are met

Test effect of interaction

```
mod11<-update(mod1,.~.-Habitat:Treatment)

anova(mod1,mod11,test="F")
```

```
## Data: PredatorOne
## Models:
## mod11: TotalNumberOfStrikes ~ Habitat + Treatment + (1 | Trial), zi=~1, disp=~1
## mod1: TotalNumberOfStrikes ~ Habitat * Treatment + (1 | Trial), zi=~1, disp=~1
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod11  6 854.06 873.22 -421.03   842.06
## mod1   8 850.87 876.42 -417.44   834.87 7.1881      2    0.02749 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod1)
```

```
##  Family: poisson  ( log )
## Formula:          TotalNumberOfStrikes ~ Habitat * Treatment + (1 | Trial)
## Zero inflation:                        ~1
## Data: PredatorOne
## 
##      AIC      BIC   logLik deviance df.resid 
##    850.9    876.4   -417.4    834.9      172 
## 
## Random effects:
## 
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  Trial  (Intercept) 0.7412   0.8609  
## Number of obs: 180, groups:  Trial, 30
## 
## Conditional model:
##                       Estimate Std. Error z value Pr(>|z|)    
## (Intercept)             2.0550     0.3940   5.215 1.84e-07 ***
## HabitatST              -2.2449     0.7157  -3.137  0.00171 ** 
## TreatmentFM            -2.6725     0.5492  -4.866 1.14e-06 ***
## TreatmentM              2.2841     0.3528   6.475 9.49e-11 ***
## HabitatST:TreatmentFM   2.4313     0.9524   2.553  0.01069 *  
## HabitatST:TreatmentM    0.8936     0.7168   1.247  0.21249    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Zero-inflation model:
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept)   0.5623     0.2228   2.524   0.0116 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant interaction between habitat and treatment, with overall more strikes made at live mysids outside of farms)

Conduct post-hoc test

```
HabTreTotStrikes <- emmeans(mod1, pairwise ~ Habitat | Treatment,adjust="mvt")
HabTreTotStrikes
```

```
## $emmeans
## Treatment = ANM:
##  Habitat emmean    SE  df lower.CL upper.CL
##  NS       2.055 0.394 172     1.17    2.943
##  ST      -0.190 0.653 172    -1.66    1.283
##
## Treatment = FM:
##  Habitat emmean    SE  df lower.CL upper.CL
##  NS      -0.618 0.457 172    -1.65    0.411
##  ST      -0.431 0.557 172    -1.69    0.824
##
## Treatment = M:
##  Habitat emmean    SE  df lower.CL upper.CL
##  NS       4.339 0.169 172     3.98    4.693
##  ST       2.988 0.177 172     2.62    3.358
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
## Conf-level adjustment: mvt method for 2 estimates
##
## $contrasts
## Treatment = ANM:
##  contrast estimate     SE  df t.ratio p.value
##  NS - ST     2.245 0.7157 172  3.137  0.0020
##
## Treatment = FM:
##  contrast estimate     SE  df t.ratio p.value
##  NS - ST    -0.186 0.6545 172 -0.285  0.7762
##
## Treatment = M:
##  contrast estimate     SE  df t.ratio p.value
##  NS - ST     1.351 0.0591 172 22.851  <.0001
##
## Results are given on the log (not the response) scale.
```

```
TreHabTotStrikes <- emmeans(mod1, pairwise ~ Treatment | Habitat)
TreHabTotStrikes
```

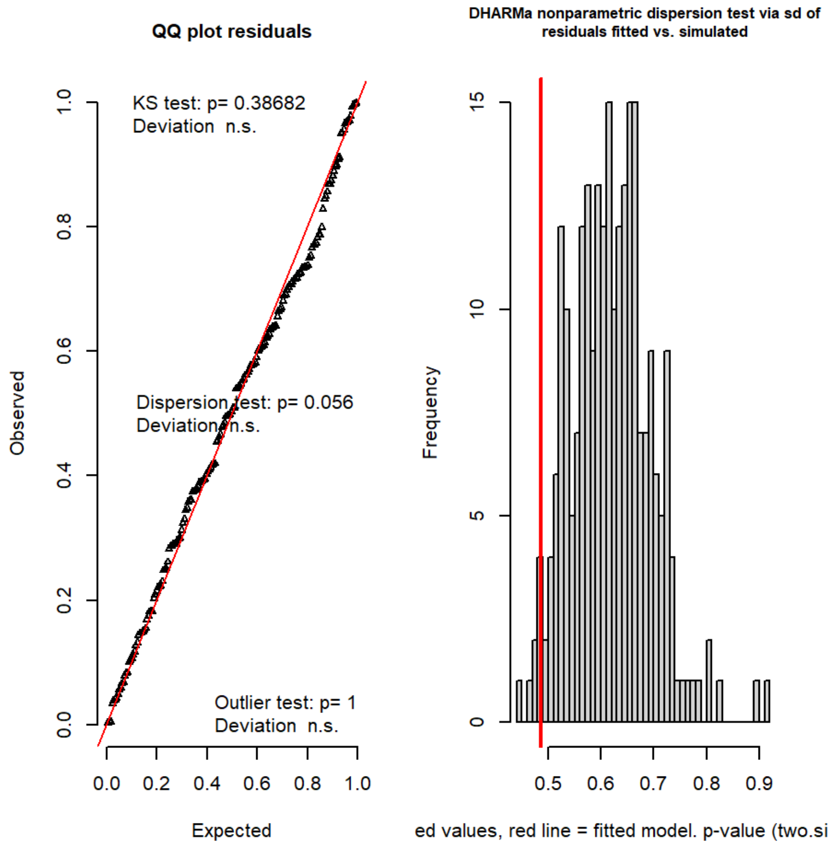
```
## $emmeans
## Habitat = NS:
##  Treatment emmean    SE  df lower.CL upper.CL
##  ANM        2.055 0.394 172     1.28    2.833
##  FM        -0.618 0.457 172    -1.52    0.284
##  M          4.339 0.169 172     4.01    4.673
##
## Habitat = ST:
##  Treatment emmean    SE  df lower.CL upper.CL
##  ANM       -0.190 0.653 172    -1.48    1.100
##  FM        -0.431 0.557 172    -1.53    0.669
##  M          2.988 0.177 172     2.64    3.337
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
## Habitat = NS:
##  contrast estimate    SE  df t.ratio p.value
##  ANM - FM    2.673 0.549 172   4.866 <.0001
##  ANM - M    -2.284 0.353 172  -6.475 <.0001
##  FM - M     -4.957 0.426 172 -11.644 <.0001
##
## Habitat = ST:
##  contrast estimate    SE  df t.ratio p.value
##  ANM - FM    0.241 0.770 172   0.313 0.9473
##  ANM - M    -3.178 0.632 172  -5.026 <.0001
##  FM - M     -3.419 0.534 172  -6.407 <.0001
##
## Results are given on the log (not the response) scale.
## P value adjustment: tukey method for comparing a family of 3 estimates
```

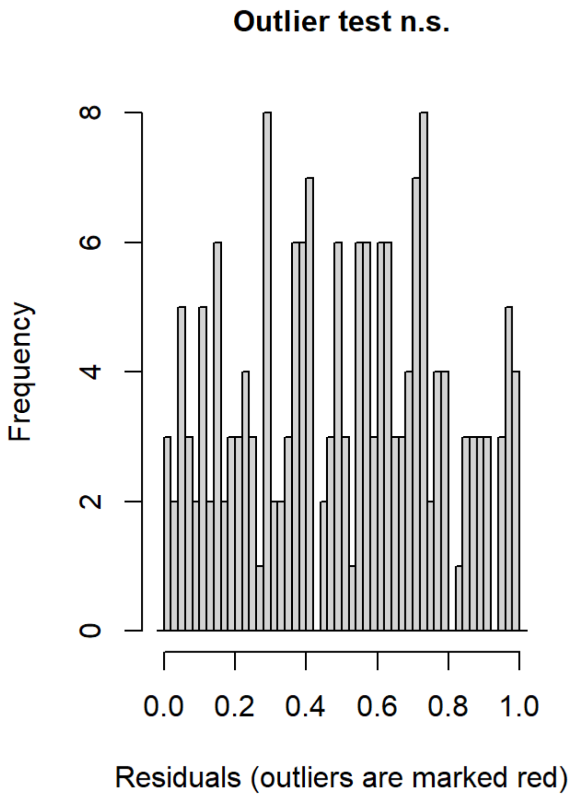
Test effect of habitat (inside or outside farm) and treatment (mysid, imitation mysid [control 1], or empty bag [control 2]) on number of predator species that struck during trials using a glmm

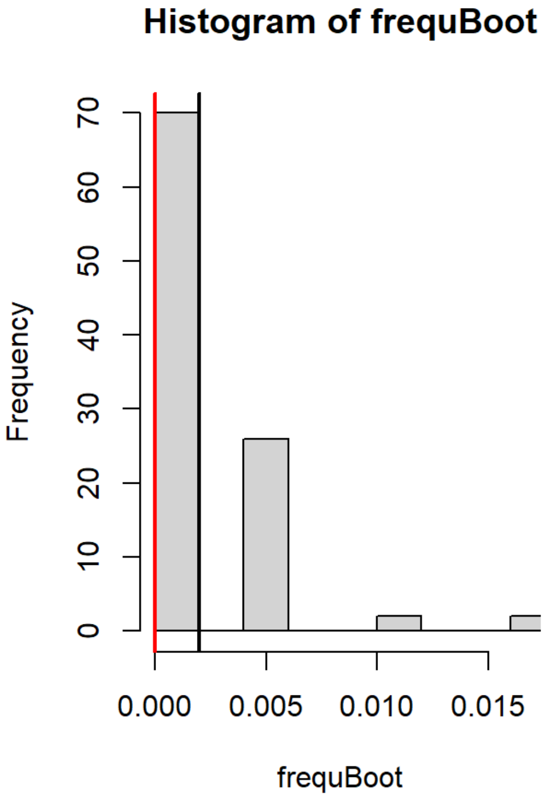
```
mod23<-glmer(NumberOfSpeciesThatMadeStrikes~Habitat*Treatment+(1|Trial),family=pois
son,data=PredatorOne)
```

Test model assumptions

```
testResiduals(mod23)
```


QQ plot residuals
KS test: p= 0.38682
Deviation n.s.
Dispersion test: p= 0.056
Deviation n.s.
Outlier test: p= 1
Deviation n.s.
Observed
0.0
0.2
0.4
0.6
0.8
1.0
Expected
DHARMa nonparametric dispersion test via sd of residuals fitted vs. simulated
Frequency
0
5
10
15
0.5
0.6
0.7
0.8
0.9
ed values, red line = fitted model. p-value (two.si


Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
60
70
0.000
0.005
0.010
0.015
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.067403, p-value = 0.3868
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.78474, p-value = 0.056
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01111111
## sample estimates:
## outlier frequency (expected: 0.002 )
##                                    0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.067403, p-value = 0.3868
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.78474, p-value = 0.056
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01111111
## sample estimates:
## outlier frequency (expected: 0.002 )
##                                     0
```

Assumptions are met

Test effect of interaction

```
mod223<-update(mod23,.~.-Habitat:Treatment)

anova(mod23,mod223)
```

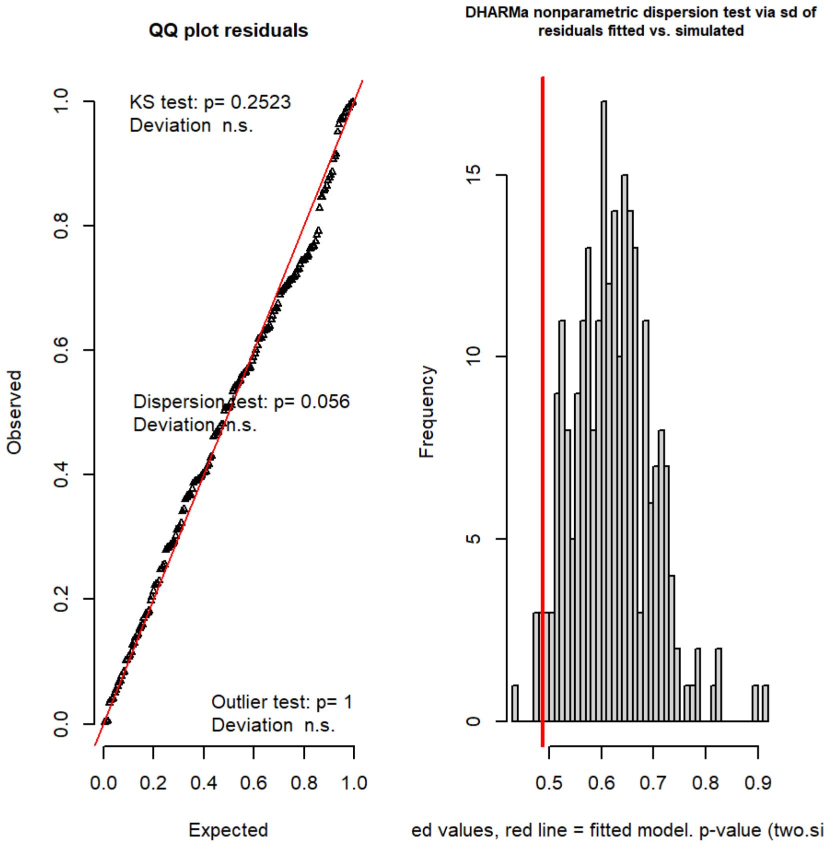
```
## Data: PredatorOne
## Models:
## mod223: NumberOfSpeciesThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
## mod23: NumberOfSpeciesThatMadeStrikes ~ Habitat * Treatment + (1 | Trial)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod223    5 202.89 218.86 -96.447   192.89
## mod23     7 205.84 228.19 -95.920   191.84 1.0534  2     0.5906
```

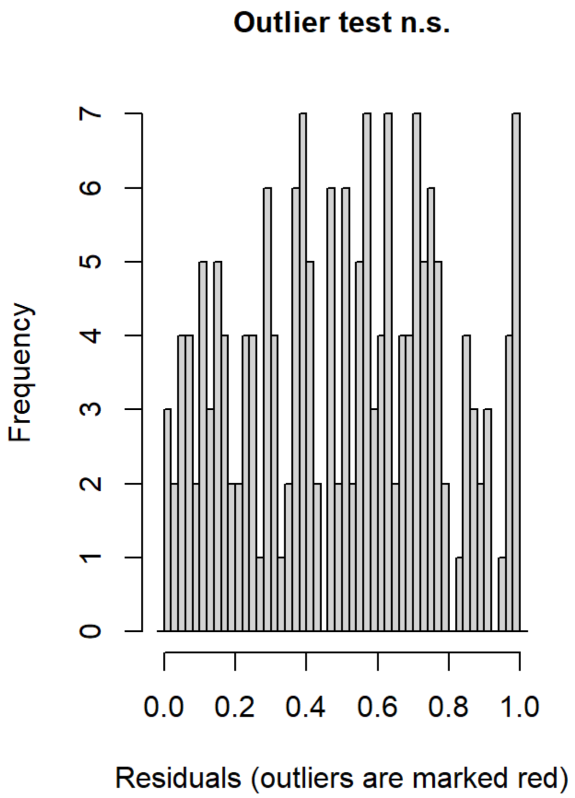
Interaction not significant, so removed

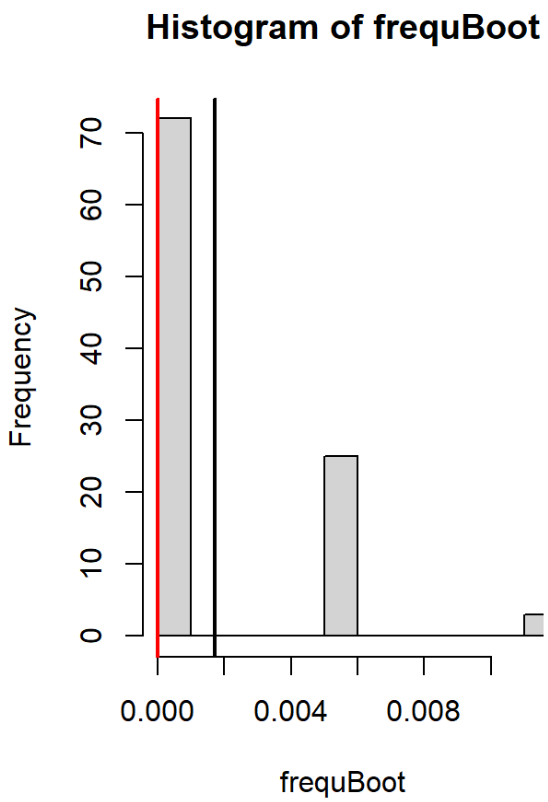
```
mod23<-glmer(NumberOfSpeciesThatMadeStrikes~Habitat+Treatment+(1|Trial),family=pois
son,data=PredatorOne)
```

Test model assumptions

```
testResiduals(mod23)
```




Outlier test n.s.
Frequency
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
60
70
0.000
0.004
0.008
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.075797, p-value = 0.2523
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.78476, p-value = 0.056
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.000000000 0.008472222
## sample estimates:
## outlier frequency (expected: 0.00172222222222222 )
##                                                  0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.075797, p-value = 0.2523
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.78476, p-value = 0.056
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.000000000 0.008472222
## sample estimates:
## outlier frequency (expected: 0.00172222222222222 )
##                                                  0
```

Assumptions are met

Test the effect of Habitat

```
mod223<-update(mod23,.~.-Habitat)

anova(mod23,mod223)
```

```
## Data: PredatorOne
## Models:
## mod223: NumberOfSpeciesThatMadeStrikes ~ Treatment + (1 | Trial)
## mod23: NumberOfSpeciesThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
##        npar    AIC    BIC   logLik deviance Chisq Df Pr(>Chisq)
## mod223    4 218.45 231.22 -105.227   210.45
## mod23     5 202.89 218.86  -96.447   192.89 17.56  1  2.784e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant effect of habitat on the number of species that made strikes, with more species making strikes outside of farms)

Conduct post-hoc test

```
HabitatNumSp<-emmeans(mod23, "Habitat")
pairs(HabitatNumSp,adjust="mvt")
```

```
##  contrast estimate    SE  df z.ratio p.value
##  NS - ST      1.12 0.287 Inf 3.895   0.0001
## 
## Results are averaged over the levels of: Treatment 
## Results are given on the log (not the response) scale.
```

Test the effect of treatment

```
mod223<-update(mod23,.~.-Treatment)

anova(mod23,mod223)
```

```
## Data: PredatorOne
## Models:
## mod223: NumberOfSpeciesThatMadeStrikes ~ Habitat + (1 | Trial)
## mod23: NumberOfSpeciesThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
##        npar    AIC    BIC   logLik deviance  Chisq Df Pr(>Chisq)    
## mod223    3 279.90 289.48 -136.949   273.90                         
## mod23     5 202.89 218.86  -96.447   192.89 81.004  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant effect of treatment on the number of species that made strikes, with more species striking at the live mysids than the two controls)

Conduct post-hoc test

```
TreatmentNumSp<-emmeans(mod23, "Treatment")
pairs(TreatmentNumSp,adjust="mvt")
```

```
##  contrast estimate    SE  df z.ratio p.value
##  ANM - FM    -1.25 0.800 Inf -1.566  0.2469
##  ANM - M     -3.33 0.718 Inf -4.640  <.0001
##  FM - M      -2.08 0.400 Inf -5.197  <.0001
## 
## Results are averaged over the levels of: Habitat 
## Results are given on the log (not the response) scale. 
## P value adjustment: mvt method for 3 tests
```

```
summary(mod23)
```

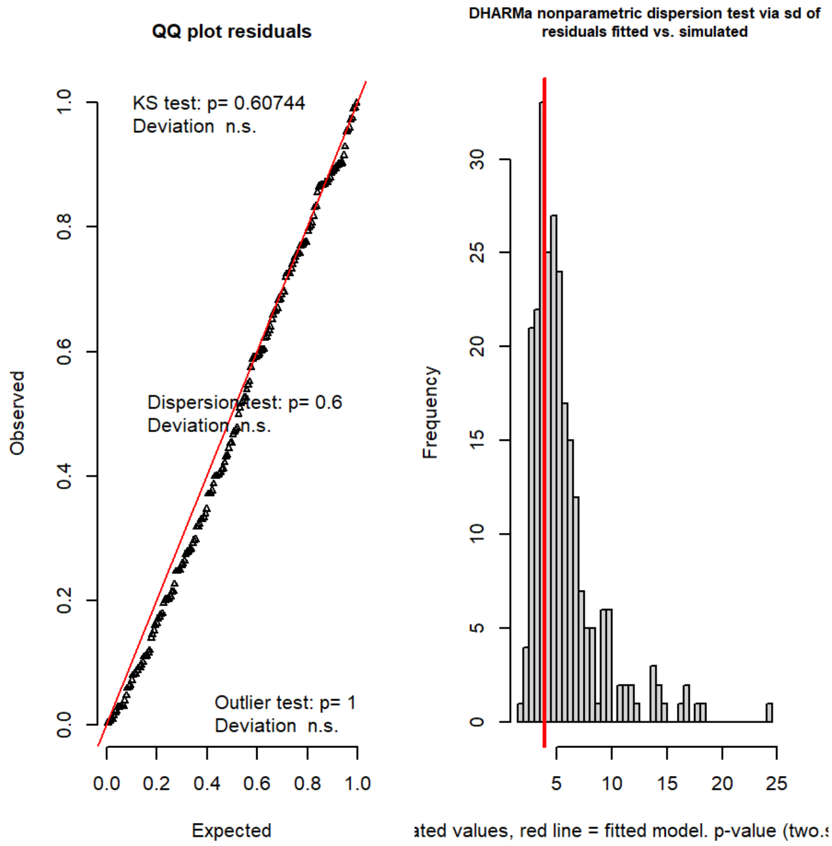
```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: NumberOfSpeciesThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
##    Data: PredatorOne
## 
##      AIC      BIC   logLik deviance df.resid 
##    202.9    218.9    -96.4    192.9      175 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -1.1275 -0.3895 -0.2278 -0.1255  6.7137 
## 
## Random effects:
##  Groups Name        Variance Std.Dev.
##  Trial  (Intercept) 0.06992  0.2644  
## Number of obs: 180, groups:  Trial, 30
## 
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept)  -3.0253     0.7139  -4.237 2.26e-05 ***
## HabitatST    -1.1192     0.2874  -3.895 9.83e-05 ***
## TreatmentFM   1.2528     0.8002   1.566    0.117    
## TreatmentM    3.3322     0.7182   4.640 3.49e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) HbttST TrtmFM
## HabitatST   -0.099              
## TreatmentFM -0.872  0.000       
## TreatmentM  -0.971  0.000  0.867
```

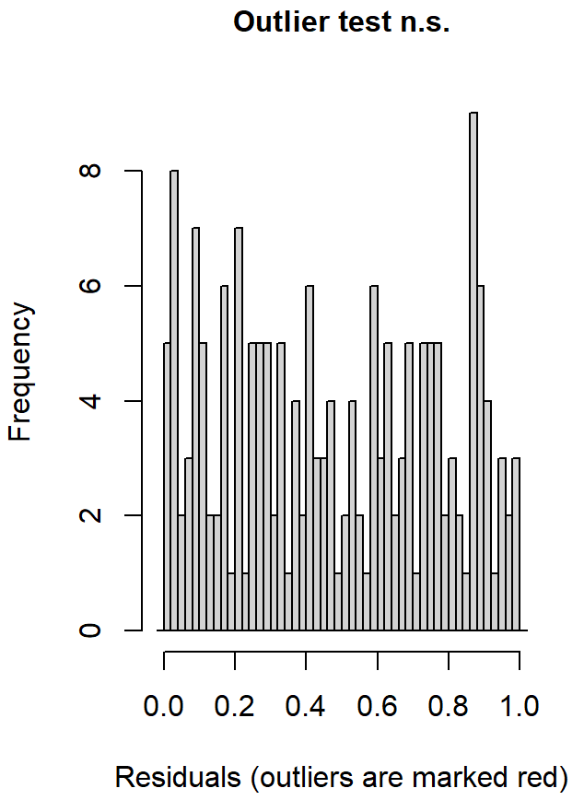
Test effect of habitat (inside or outside farm) and treatment (mysid, imitation mysid [control 1], or empty bag [control 2]) on number of fish that made strikes using glmm

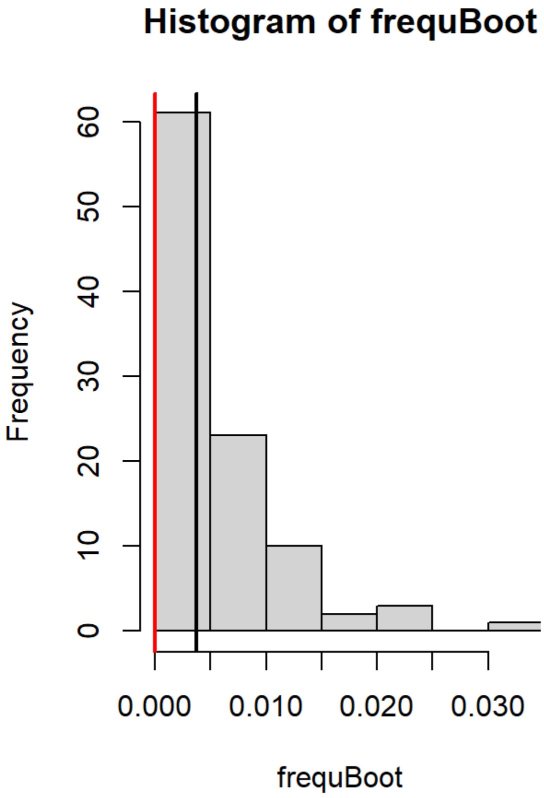
```
mod2<-glmer(NumberOfFishThatMadeStrikes~Habitat*Treatment+(1|Trial),family=poisson,
data=PredatorOne)
```

Test model assumptions

```
testResiduals(mod2)
```


QQ plot residuals
KS test: p= 0.60744
Deviation n.s.
Dispersion test: p= 0.6
Deviation n.s.
Outlier test: p= 1
Deviation n.s.
Observed
Expected
0.0
0.2
0.4
0.6
0.8
1.0
DHARMa nonparametric dispersion test via sd of residuals fitted vs. simulated
Frequency
0
5
10
15
20
25
30
ated values, red line = fitted model. p-value (two.s


Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
60
0.000
0.010
0.020
0.030
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.056777, p-value = 0.6074
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.68121, p-value = 0.6
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.02222222
## sample estimates:
## outlier frequency (expected: 0.00372222222222222 )
##                                                  0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.056777, p-value = 0.6074
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.68121, p-value = 0.6
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.02222222
## sample estimates:
## outlier frequency (expected: 0.00372222222222222 )
##                                                  0
```

Model assumptions met

Test effect of interaction

```
mod22<-update(mod2,.~.-Habitat:Treatment)

anova(mod2,mod22)
```

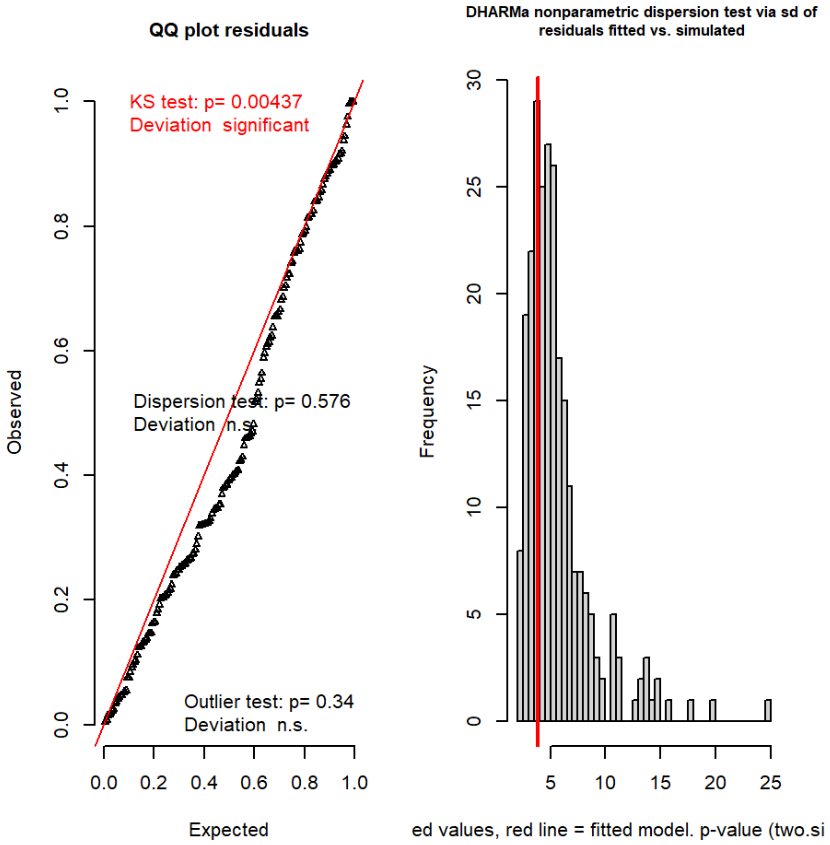
```
## Data: PredatorOne
## Models:
## mod22: NumberOfFishThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
## mod2: NumberOfFishThatMadeStrikes ~ Habitat * Treatment + (1 | Trial)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod22    5 392.18 408.15 -191.09   382.18
## mod2     7 391.73 414.08 -188.87   377.73 4.4518  2      0.108
```

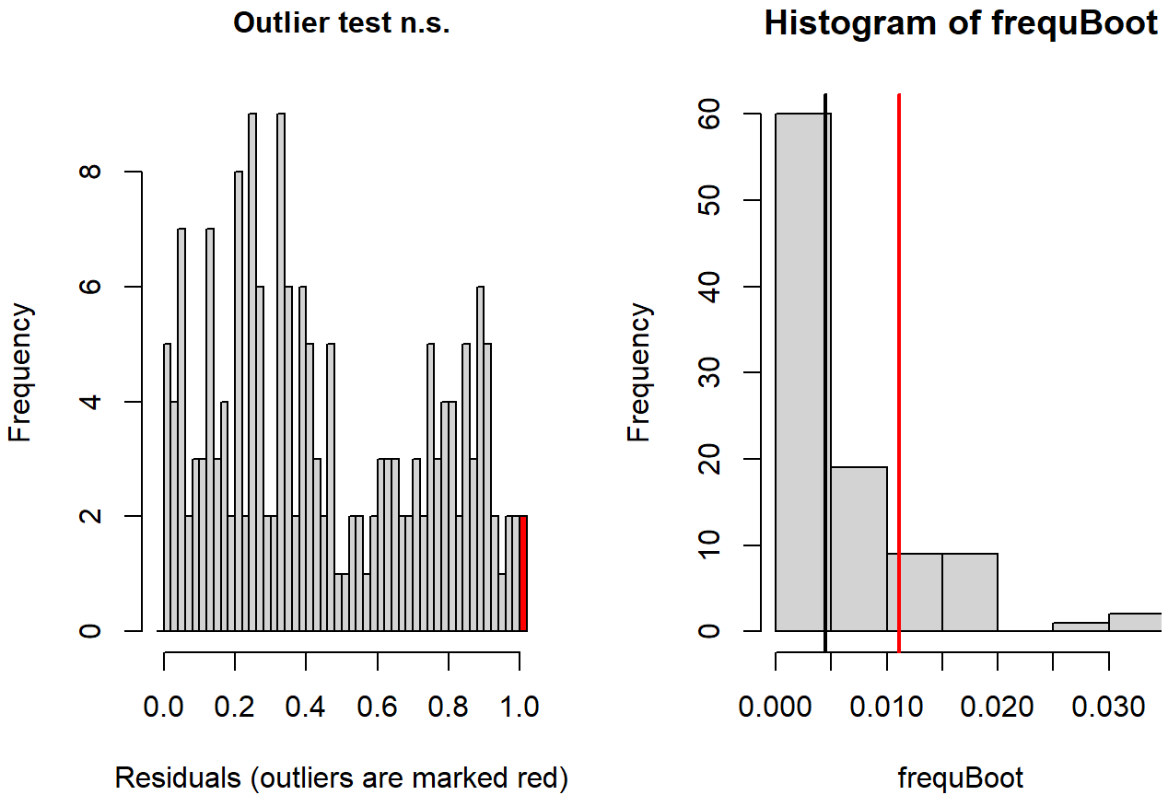
Interaction not significant, so removed

```
mod2<-glmer(NumberOfFishThatMadeStrikes~Habitat+Treatment+(1|Trial),family=poisson,
data=PredatorOne)
```

Test model assumptions

```
testResiduals(mod2)
```



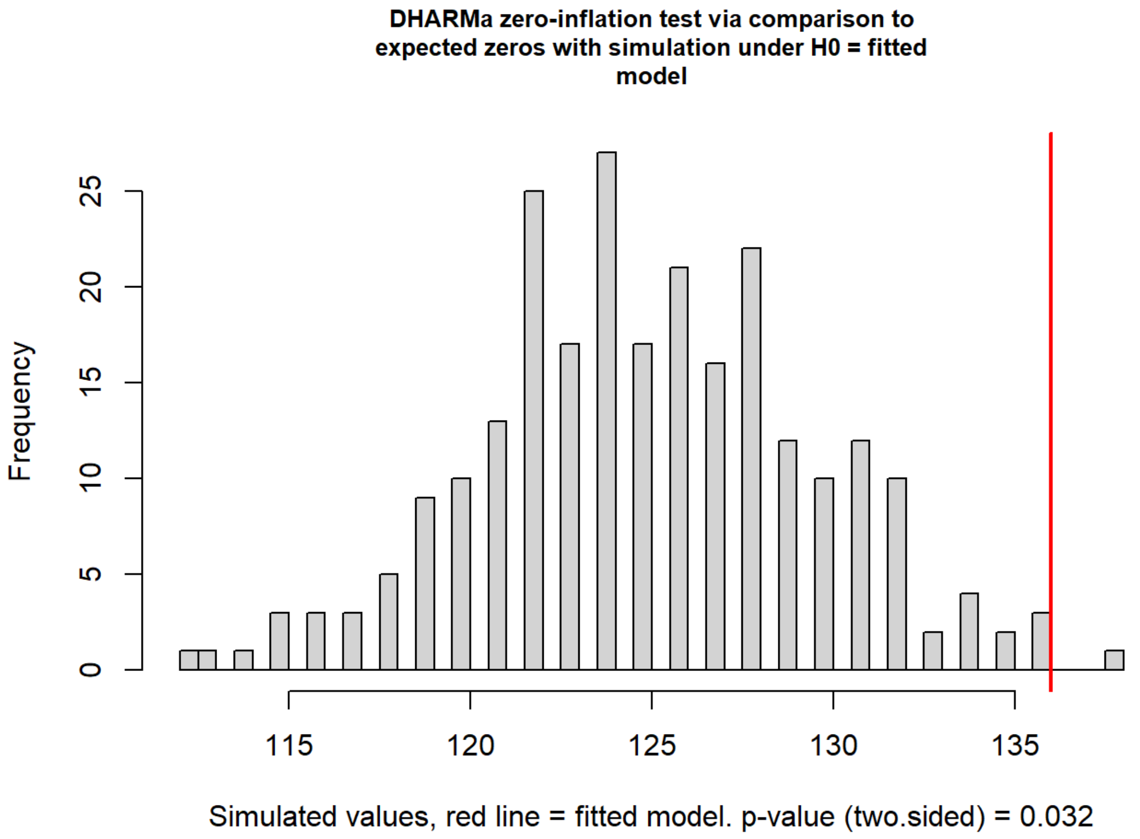

Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)
Histogram of frequBoot
Frequency
0
10
20
30
40
50
60
0.000
0.010
0.020
0.030
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.13044, p-value = 0.004375
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.68248, p-value = 0.576
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 180, p-value = 0.42
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.0000 0.0225
## sample estimates:
## outlier frequency (expected: 0.0045 )
##                            0.01111111
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.13044, p-value = 0.004375
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.68248, p-value = 0.576
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 180, p-value = 0.42
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.0000 0.0225
## sample estimates:
## outlier frequency (expected: 0.0045 )
##                            0.01111111
```

Assumptions not met, investigate further

```
testZeroInflation(mod2)
```



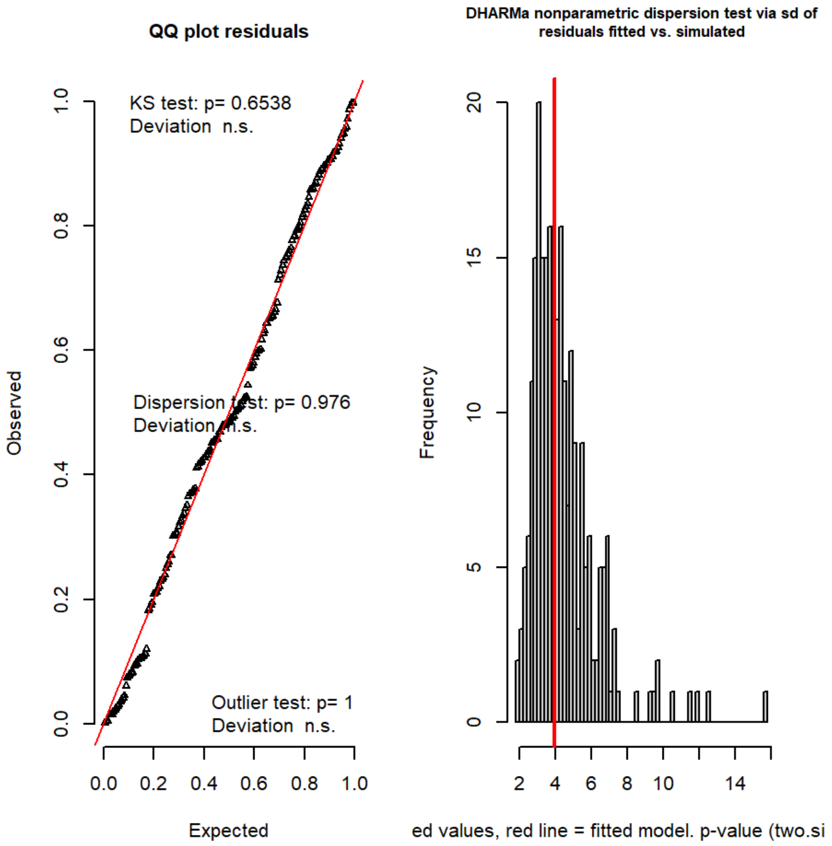
```
##
##  DHARMa zero-inflation test via comparison to expected zeros with
##  simulation under H0 = fitted model
##
## data:  simulationOutput
## ratioObsSim = 1.0869, p-value = 0.032
## alternative hypothesis: two.sided
```

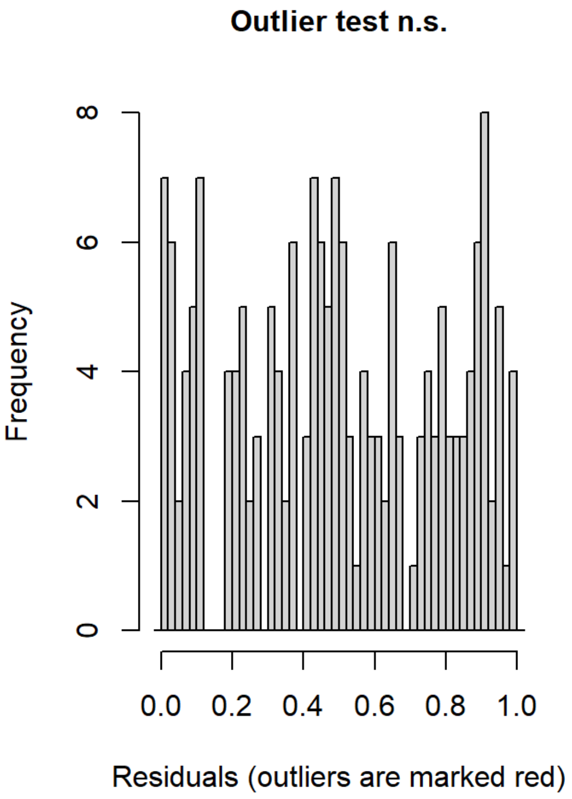
Model zero-inflated, use glmmTMB

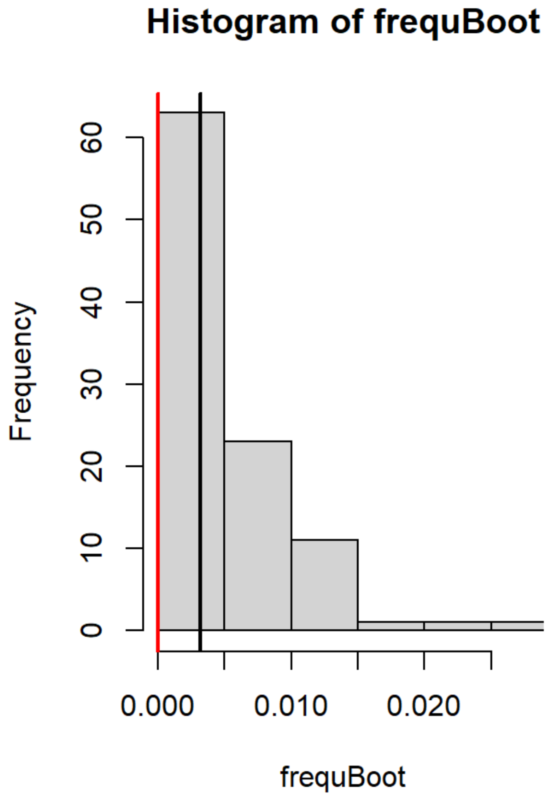
```
mod2<-glmmTMB(NumberOfFishThatMadeStrikes~Habitat*Treatment+(1|Trial),family=poisso
n,data=PredatorOne,ziformula=~1)
```

Test model assumptions

```
testResiduals(mod2)
```


QQ plot residuals
KS test: p= 0.6538
Deviation n.s.
Dispersion test: p= 0.976
Deviation n.s.
Outlier test: p= 1
Deviation n.s.
Observed
0.0
0.2
0.4
0.6
0.8
1.0
Expected
0.0
0.2
0.4
0.6
0.8
1.0
DHARMa nonparametric dispersion test via sd of
residuals fitted vs. simulated
Frequency
0
5
10
15
20
2
4
6
8
10
14
ed values, red line = fitted model. p-value (two.si


Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
60
0.000
0.010
0.020
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.054724, p-value = 0.6538
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.89347, p-value = 0.976
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01402778
## sample estimates:
## outlier frequency (expected: 0.00316666666666667 )
##                                                  0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.054724, p-value = 0.6538
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.89347, p-value = 0.976
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01402778
## sample estimates:
## outlier frequency (expected: 0.00316666666666667 )
##                                                  0
```

Model assumption met

Test effect of interaction

```
mod22<-update(mod2,.~.-Habitat:Treatment)

anova(mod2,mod22)
```

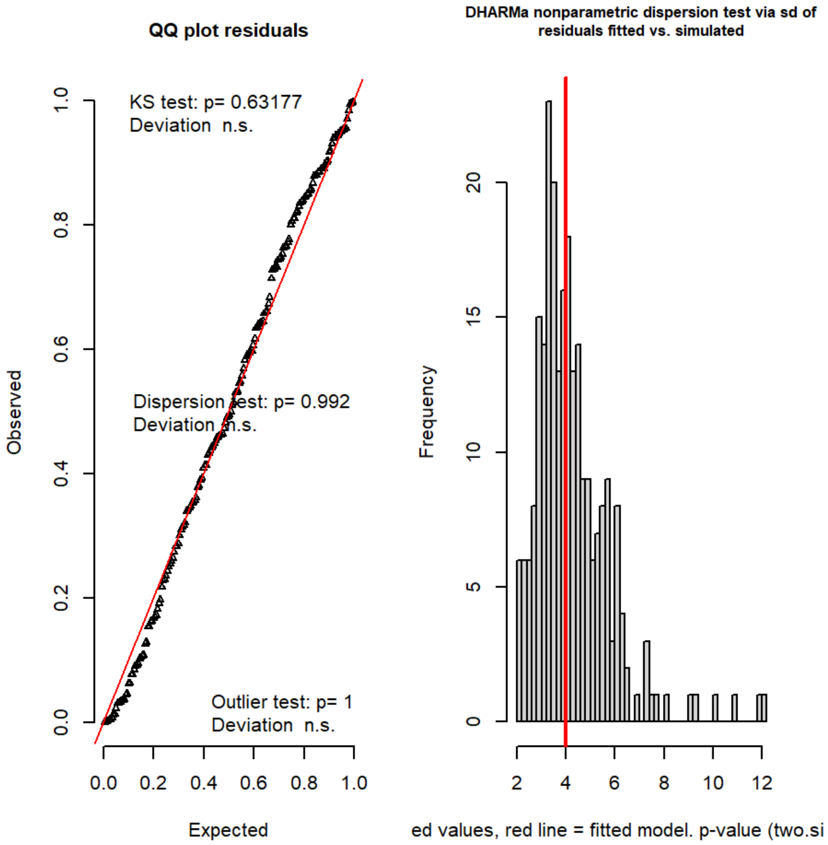
```
## Data: PredatorOne
## Models:
## mod22: NumberOfFishThatMadeStrikes ~ Habitat + Treatment + (1 | Trial), zi=~1, d
isp=~1
## mod2: NumberOfFishThatMadeStrikes ~ Habitat * Treatment + (1 | Trial), zi=~1, di
sp=~1
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod22  6 370.59 389.74 -179.29   358.59
## mod2   8 371.91 397.46 -177.96   355.91 2.6718      2     0.2629
```

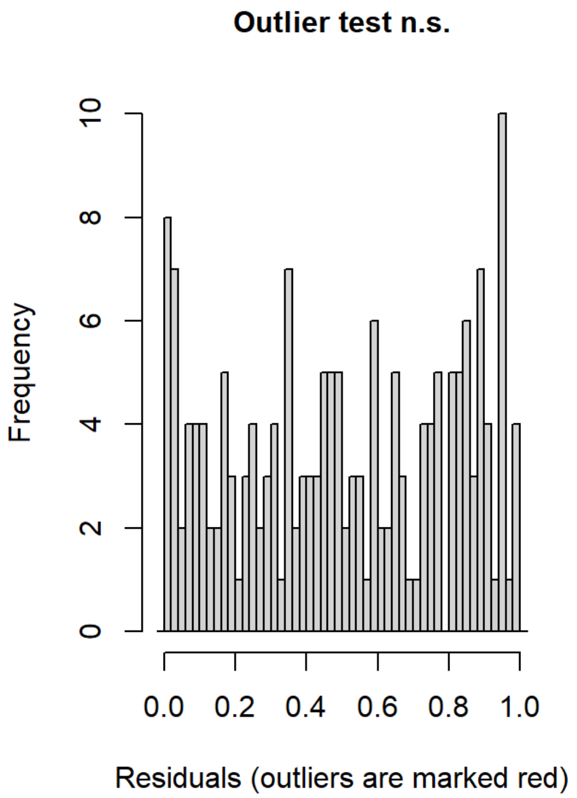
Interaction not significant, so removed

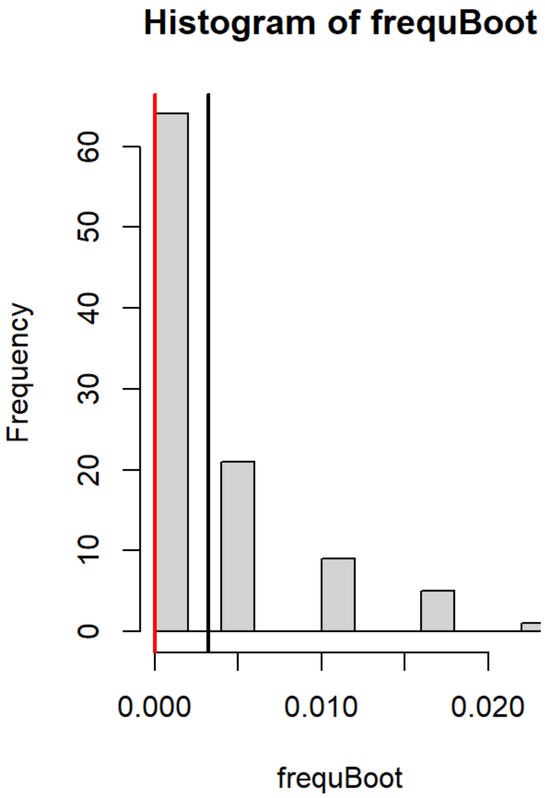
```
mod2<-glmmTMB(NumberOfFishThatMadeStrikes~Habitat+Treatment+(1|Trial),family=poisso
n,data=PredatorOne,ziformula=~1)
```

Test model assumptions

```
testResiduals(mod2)
```




Outlier test n.s.
Frequency
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
60
0.000
0.010
0.020
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.055698, p-value = 0.6318
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.93192, p-value = 0.992
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01666667
## sample estimates:
## outlier frequency (expected: 0.00322222222222222 )
##                                                  0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.055698, p-value = 0.6318
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.93192, p-value = 0.992
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01666667
## sample estimates:
## outlier frequency (expected: 0.00322222222222222 )
##                                                  0
```

Model assumptions met

Test effect of habitat

```
mod22<-update(mod2,.~.-Habitat)
```

```
## Warning in (function (start, objective, gradient = NULL, hessian = NULL, : NA/
## NaN function evaluation
```

```
## Warning in fitTMB(TMBStruc): Model convergence problem; non-positive-definite
## Hessian matrix. See vignette('troubleshooting')
```

```
## Warning in fitTMB(TMBStruc): Model convergence problem; false convergence (8).
## See vignette('troubleshooting')
```

Model failed to converge, investigate data further

```
with(PredatorOne, tapply(NumberOfFishThatMadeStrikes, Habitat, mean))
```

```
##        NS        ST
## 3.3111111 0.5555556
```

```
with(PredatorOne, tapply(NumberOfFishThatMadeStrikes, Habitat, var))
```

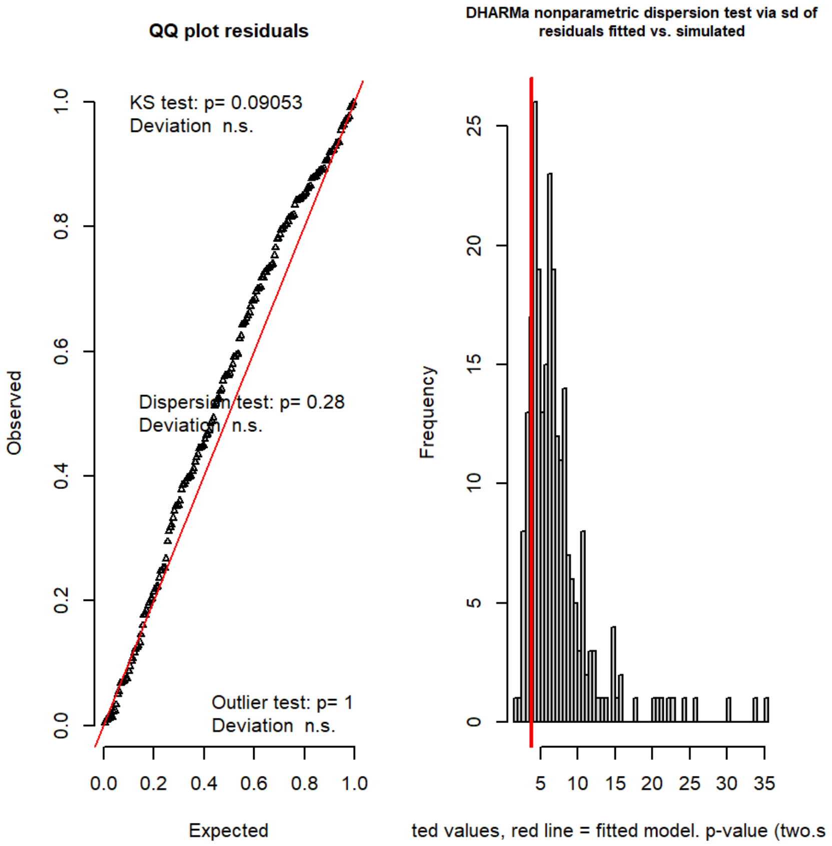
```
##        NS        ST
## 43.407740  4.968789
```

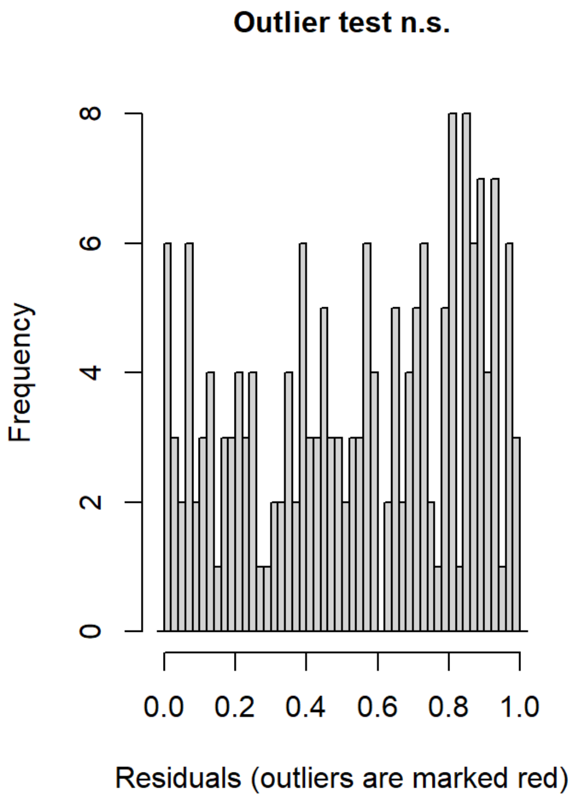
Appears we have an overdispersion problem, use glmer.nb

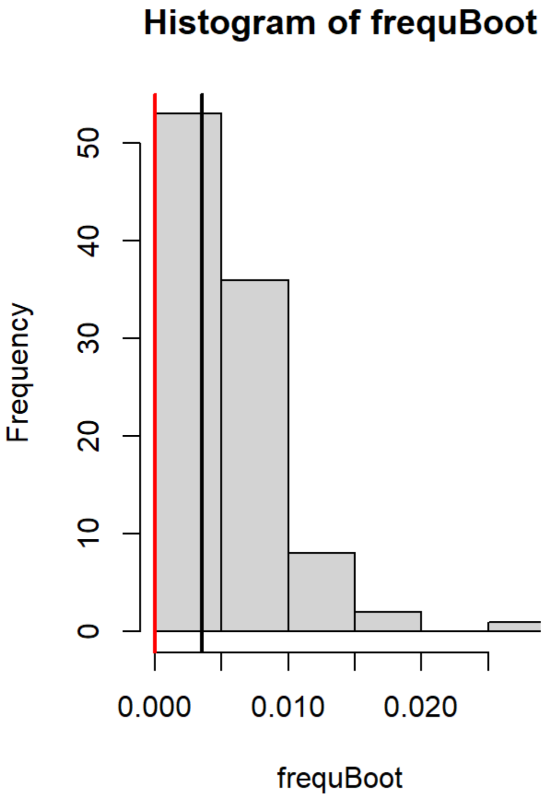
```
mod2<-glmer.nb(NumberOfFishThatMadeStrikes~Habitat*Treatment+(1|Trial),data=Predato
rOne)
```

Test model assumptions

```
testResiduals(mod2)
```




Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
0.000
0.010
0.020
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.092722, p-value = 0.09053
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.51244, p-value = 0.28
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01402778
## sample estimates:
## outlier frequency (expected: 0.0035 )
##                                     0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.092722, p-value = 0.09053
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.51244, p-value = 0.28
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01402778
## sample estimates:
## outlier frequency (expected: 0.0035 )
##                                     0
```

Model assumptions met

Test effect of interaction

```
mod22<-update(mod2,.~.-Habitat:Treatment)

anova(mod2,mod22)
```

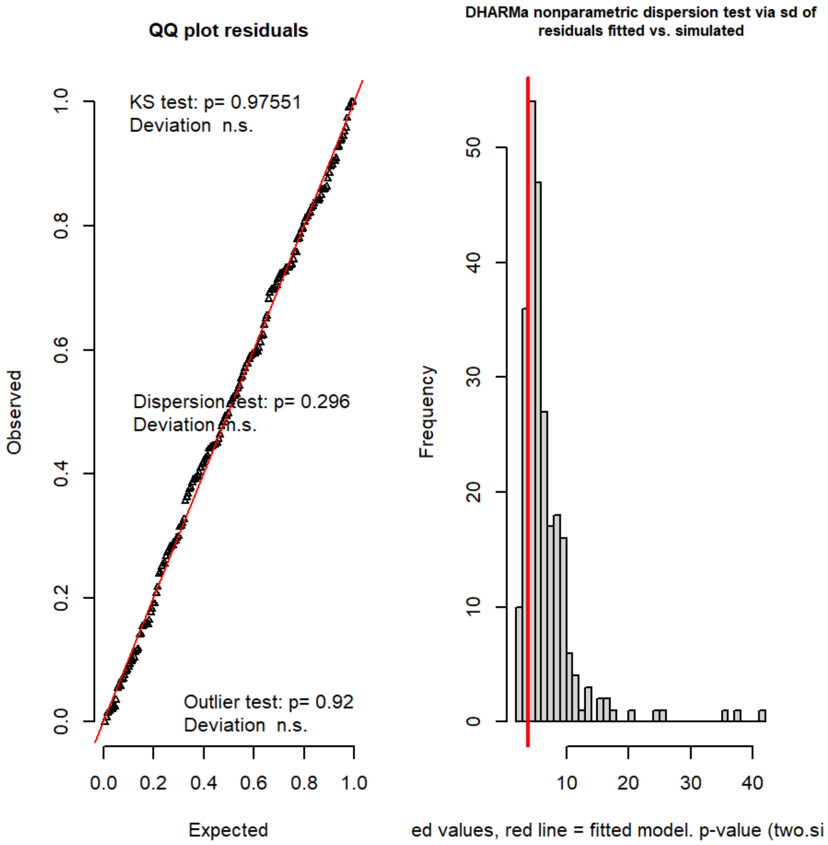
```
## Data: PredatorOne
## Models:
## mod22: NumberOfFishThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
## mod2: NumberOfFishThatMadeStrikes ~ Habitat * Treatment + (1 | Trial)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod22    6 358.04 377.20 -173.02   346.04
## mod2     8 357.18 382.72 -170.59   341.18 4.8688  2    0.08765 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

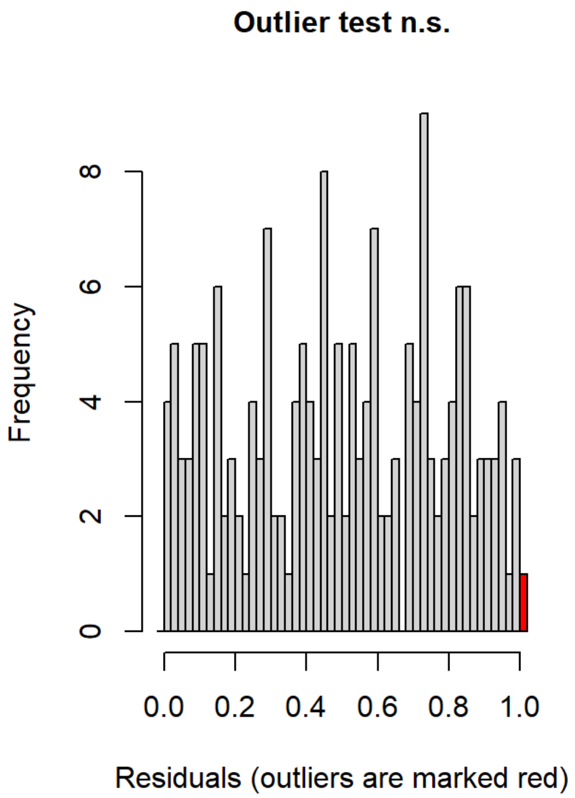
Interaction not significant, so removed

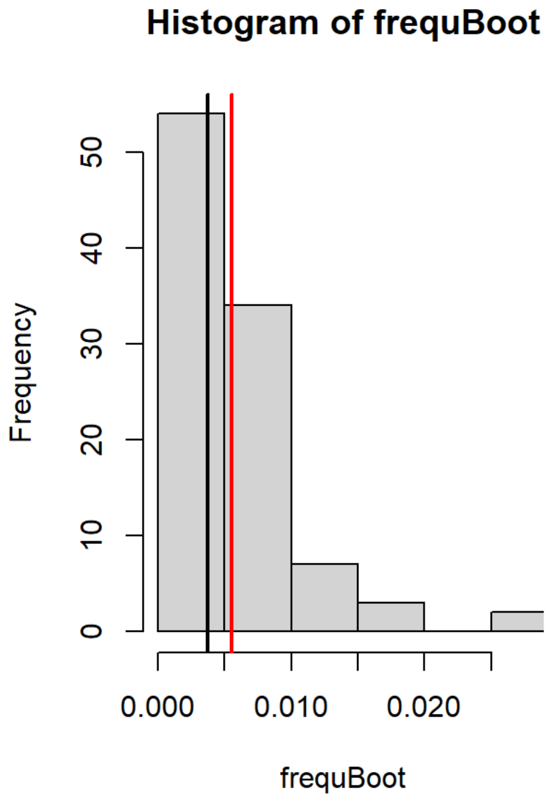
```
mod2<-glmer.nb(NumberOfFishThatMadeStrikes~Habitat+Treatment+(1|Trial),data=Predato
rOne)
```

Test model assumptions

```
testResiduals(mod2)
```



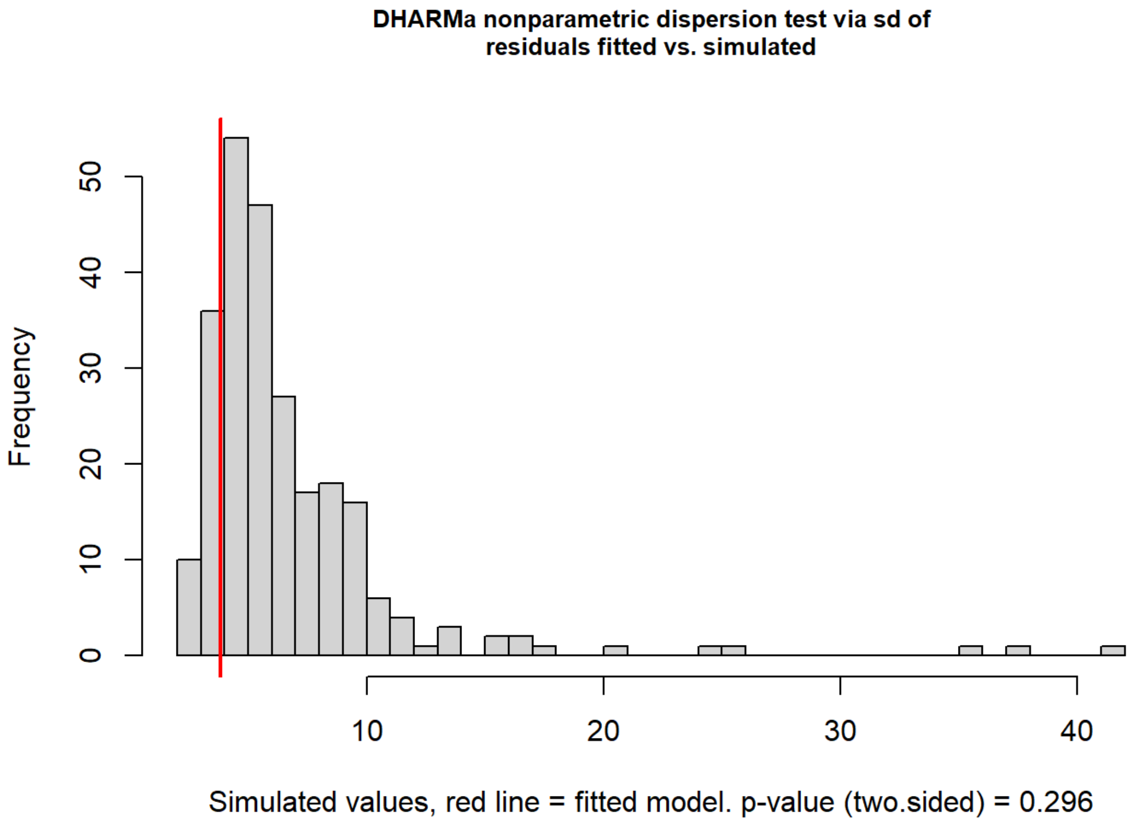

Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
0.000
0.010
0.020
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.035749, p-value = 0.9755
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.56321, p-value = 0.296
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 180, p-value = 0.92
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01666667
## sample estimates:
## outlier frequency (expected: 0.00372222222222222 )
##                                         0.005555556
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.035749, p-value = 0.9755
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.56321, p-value = 0.296
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 180, p-value = 0.92
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01666667
## sample estimates:
## outlier frequency (expected: 0.00372222222222222 )
##                                        0.005555556
```

```
testDispersion(mod2)
```



```
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.56321, p-value = 0.296
## alternative hypothesis: two.sided
```

Model assumptions met

Test effect of habitat

```
mod22<-update(mod2,.~.-Habitat)

anova(mod2,mod22)
```

```
## Data: PredatorOne
## Models:
## mod22: NumberOfFishThatMadeStrikes ~ Treatment + (1 | Trial)
## mod2: NumberOfFishThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
##       npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## mod22    5 407.04 423.01 -198.52   397.04
## mod2     6 358.03 377.19 -173.02   346.03 51.01  1  9.189e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant effect of habitat, with more fish making strikes made at live mysids outside of farms)

Conduct post-hoc test

```
HabitatNumFish<-emmeans(mod2, "Habitat")
pairs(HabitatNumFish,adjust="mvt")
```

```
##  contrast estimate  SE  df z.ratio p.value
##  NS - ST      2.05 0.3 Inf 6.841   <.0001
##
## Results are averaged over the levels of: Treatment
## Results are given on the log (not the response) scale.
```

Test effect of treatment

```
mod22<-update(mod2,.~.-Treatment)

anova(mod2,mod22)
```

```
## Data: PredatorOne
## Models:
## mod22: NumberOfFishThatMadeStrikes ~ Habitat + (1 | Trial)
## mod2: NumberOfFishThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod22    4 607.74 620.51 -299.87   599.74
## mod2     6 358.03 377.19 -173.02   346.03 253.71  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant effect of treatment, with more fish making strikes made at live mysids compared to the control treatments)

Conduct post-hoc test

```
TreatmentNumFish<-emmeans(mod2, "Treatment")
pairs(TreatmentNumFish,adjust="mvt")
```

```
##  contrast estimate    SE  df z.ratio p.value
##  ANM - FM   -0.408 0.779 Inf -0.524  0.8553
##  ANM - M    -4.749 0.634 Inf -7.495  <.0001
##  FM - M     -4.341 0.522 Inf -8.316  <.0001
##
## Results are averaged over the levels of: Habitat
## Results are given on the log (not the response) scale.
## P value adjustment: mvt method for 3 tests
```

```
summary(mod2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Negative Binomial(1.397)  ( log )
## Formula: NumberOfFishThatMadeStrikes ~ Habitat + Treatment + (1 | Trial)
##    Data: PredatorOne
##
##      AIC      BIC   logLik deviance df.resid
##    358.0    377.2   -173.0    346.0      174
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0035 -0.2914 -0.1422 -0.0712  6.6698
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  Trial  (Intercept) 0.8998   0.9486
## Number of obs: 180, groups:  Trial, 30
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.8457     0.6510  -4.372 1.23e-05 ***
## HabitatST    -2.0527     0.3001  -6.841 7.88e-12 ***
## TreatmentFM   0.4080     0.7793   0.524    0.601
## TreatmentM    4.7489     0.6336   7.495 6.63e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) HbttST TrtmFM
## HabitatST   -0.015
## TreatmentFM -0.726  0.014
## TreatmentM  -0.924 -0.121  0.746
```

Test effect of habitat (inside or outside farm) and treatment (mysid, imitation mysid [control 1], or empty bag [control 2]) on number of defensive chases by longfin damselfish using glmm

```
mod4<-glmer(NumberOfStegastesChases~Habitat*Treatment+(1|Trial),family=poisson,data
=PredatorOne)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues
```

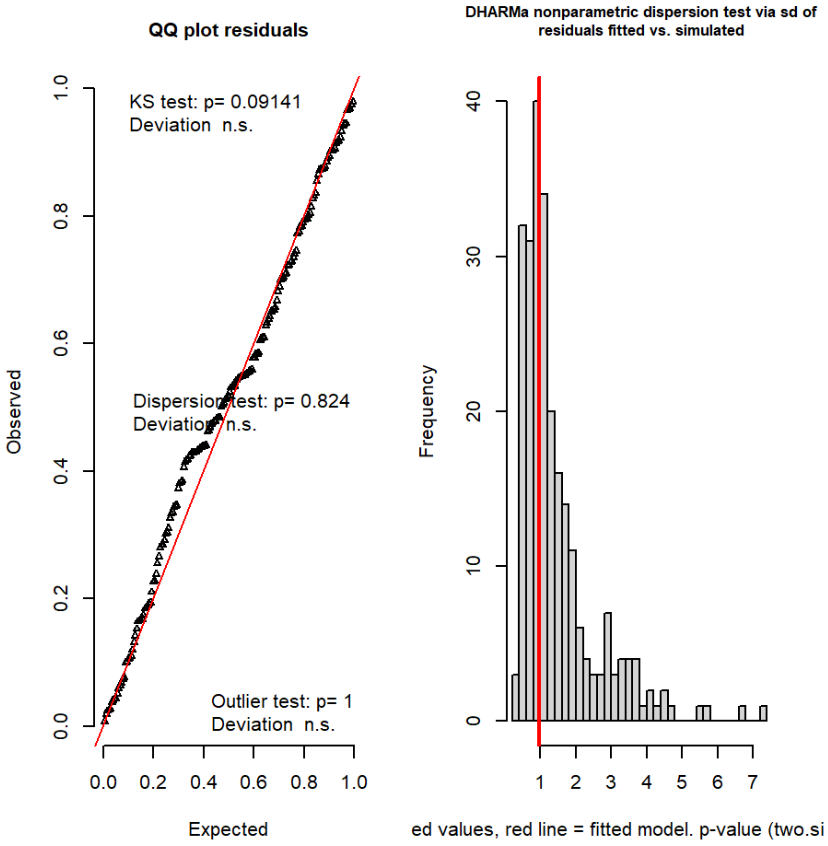
Glmer fails to converge, so we used glmmTMB

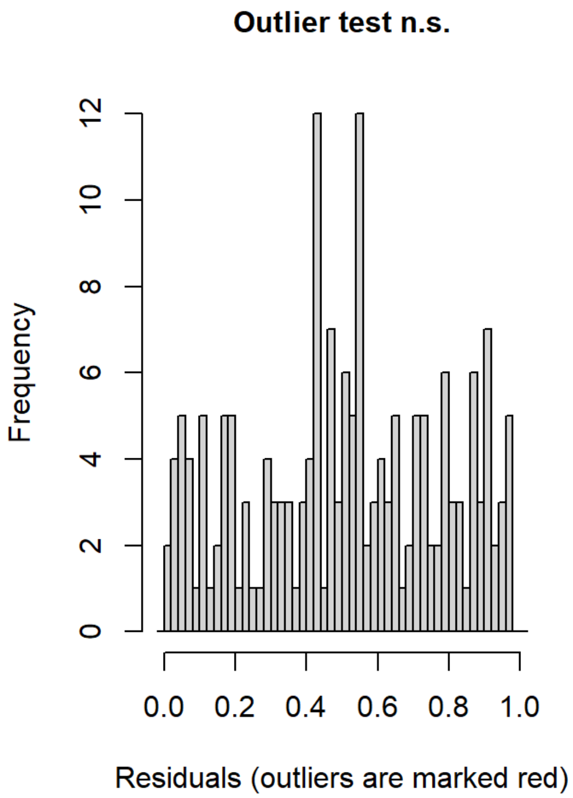
```
mod4<-glmmTMB(NumberOfStegastesChases~Habitat*Treatment+(1|Trial),family=poisson,da
ta=PredatorOne)
```

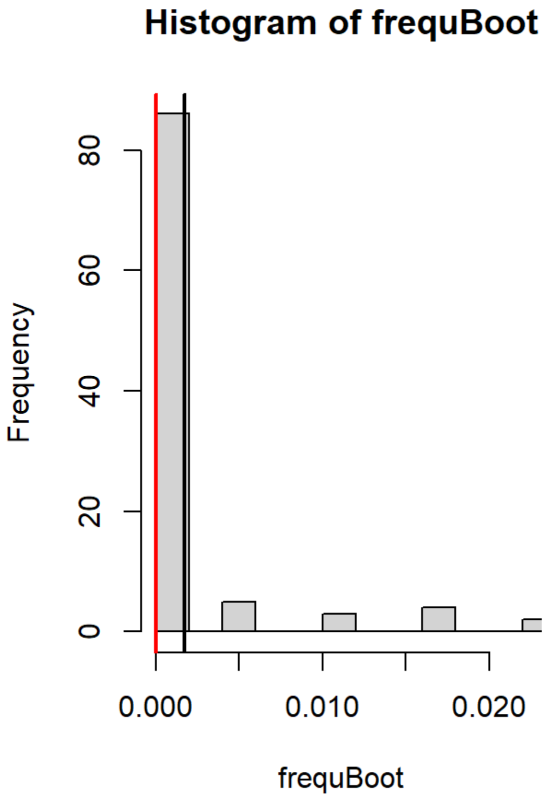
Model converges successfully

Test model assumptions

```
testResiduals(mod4)
```




Outlier test n.s.
Frequency
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
20
40
60
80
0.000
0.010
0.020
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.092579, p-value = 0.09141
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.65729, p-value = 0.824
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01666667
## sample estimates:
## outlier frequency (expected: 0.00172222222222222 )
##                                                   0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.092579, p-value = 0.09141
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.65729, p-value = 0.824
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01666667
## sample estimates:
## outlier frequency (expected: 0.00172222222222222 )
##                                                  0
```

Test effect of interaction

```
mod44<-update(mod4,.~.-Habitat:Treatment)

anova(mod4,mod44,test="F")
```

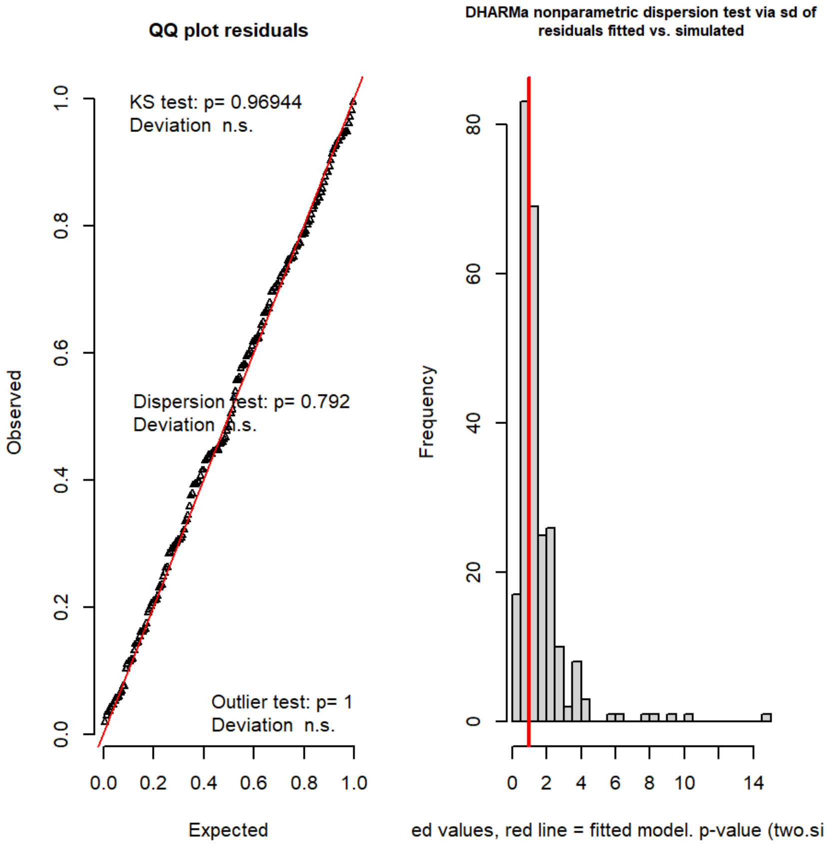
```
## Data: PredatorOne
## Models:
## mod44: NumberOfStegastesChases ~ Habitat + Treatment + (1 | Trial), zi=~0, disp=
~1
## mod4: NumberOfStegastesChases ~ Habitat * Treatment + (1 | Trial), zi=~0, disp=~
1
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod44  5 193.04 209.01 -91.521   183.04
## mod4   7 195.60 217.95 -90.801   181.60 1.4402      2     0.4867
```

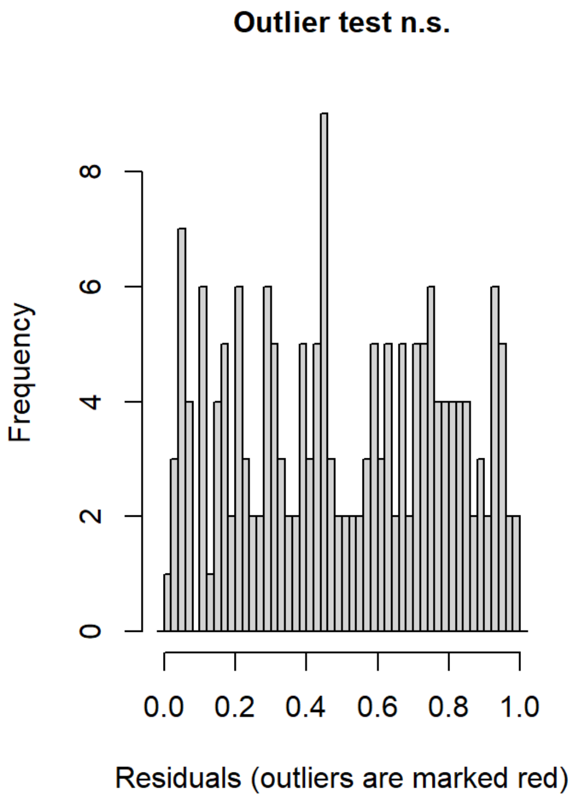
Interaction not significant, so removed

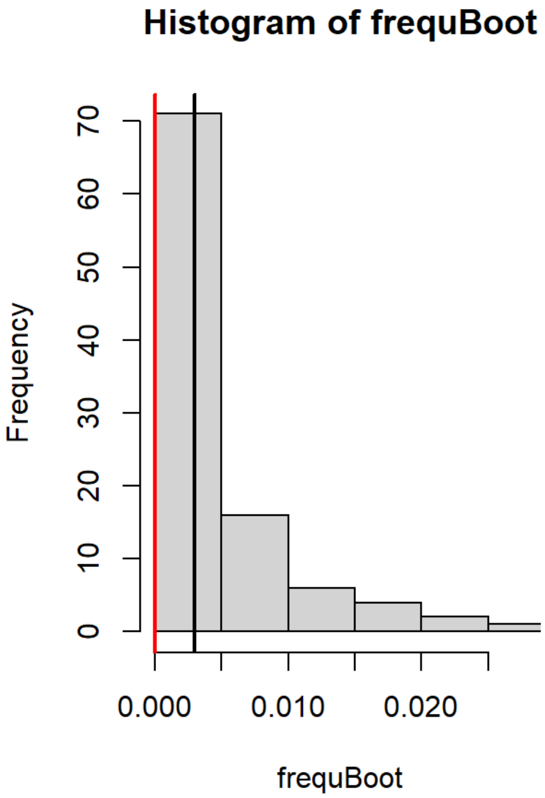
```
mod4<-glmmTMB(NumberOfStegastesChases~Habitat+Treatment+(1|Trial),family=poisson,da
ta=PredatorOne)
```

Test model assumptions

```
testResiduals(mod4)
```




Outlier test n.s.
Frequency
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
60
70
0.000
0.010
0.020
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.036593, p-value = 0.9694
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.62304, p-value = 0.792
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01958333
## sample estimates:
## outlier frequency (expected: 0.00294444444444444 )
##                                                  0
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.036593, p-value = 0.9694
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.62304, p-value = 0.792
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 0, observations = 180, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.01958333
## sample estimates:
## outlier frequency (expected: 0.00294444444444444 )
##                                                  0
```

Model assumptions met

Test effect of habitat

```
mod44<-update(mod4,.~.-Habitat)

anova(mod4,mod44,test="F")
```

```
## Data: PredatorOne
## Models:
## mod44: NumberOfStegastesChases ~ Treatment + (1 | Trial), zi=~0, disp=~1
## mod4: NumberOfStegastesChases ~ Habitat + Treatment + (1 | Trial), zi=~0, disp=~
1
##       Df    AIC    BIC   logLik deviance Chisq Chi Df Pr(>Chisq)
## mod44  4 259.99 272.76 -125.996   251.99
## mod4   5 193.04 209.01  -91.521   183.04 68.95      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant effect of habitat, with longfin damselfish making more defensive chases inside versus outside farms)

Conduct post-hoc test

```
HabitatNumChases<-emmeans(mod4, "Habitat")
pairs(HabitatNumChases,adjust="mvt")
```

```
##  contrast estimate   SE  df t.ratio p.value
##  NS - ST     -4.03 1.01 175 -3.990  0.0001
##
## Results are averaged over the levels of: Treatment
## Results are given on the log (not the response) scale.
```

Test effect of treatment

```
mod44<-update(mod4,.~.-Treatment)

anova(mod4,mod44,test="F")
```

```
## Data: PredatorOne
## Models:
## mod44: NumberOfStegastesChases ~ Habitat + (1 | Trial), zi=~0, disp=~1
## mod4: NumberOfStegastesChases ~ Habitat + Treatment + (1 | Trial), zi=~0, disp=~
1
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod44  3 195.12 204.69 -94.558   189.12
## mod4   5 193.04 209.01 -91.521   183.04 6.0725      2    0.04801 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant effect of treatment, with more fish making defensive chases during the live mysid trials compared to the two controls)

Conduct post-hoc test

```
TreatmentNumChases<-emmeans(mod4, "Treatment")
pairs(TreatmentNumChases,adjust="mvt")
```

```
##  contrast estimate    SE  df t.ratio p.value
##  ANM - FM    0.069 0.372 175  0.186  0.9811
##  ANM - M    -0.624 0.320 175 -1.951  0.1269
##  FM - M     -0.693 0.327 175 -2.118  0.0886
##
## Results are averaged over the levels of: Habitat
## Results are given on the log (not the response) scale.
## P value adjustment: mvt method for 3 tests
```

```
summary(mod4)
```

```
##  Family: poisson  ( log )
## Formula:          NumberOfStegastesChases ~ Habitat + Treatment + (1 | Trial)
## Data: PredatorOne
## 
##      AIC      BIC   logLik deviance df.resid 
##    193.0    209.0    -91.5    183.0      175 
## 
## Random effects:
## 
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  Trial  (Intercept) 2.769    1.664   
## Number of obs: 180, groups:  Trial, 30
## 
## Conditional model:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept) -5.88672    1.13057  -5.207 1.92e-07 ***
## HabitatST    4.02535    1.00889   3.990 6.61e-05 ***
## TreatmentFM -0.06899    0.37161  -0.186   0.8527    
## TreatmentM   0.62415    0.31997   1.951   0.0511 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Predation experiment 2

Description of data

A field-based experiment using naturally occurring mysid swarms to examine whether the presence of longfin damselfish reduces predation risk to mysids.

The dataset (CBC2018_PredationExperiment2_data.csv) contains the following variables (L-R):

- Territory: farm number (1-30).
- Date: date of trial.
- Treatment: farm type ((Removal) damselfish removed in P2, (Control) damselfish present in P2.
- StrikesP1: Number of strikes by predatory fishes on swarm in Period 1 (baseline predation rate).
- StrikesP2: Number of strikes by predatory fishes on swarm in Period 2.

Analysis

Create subsets of data based on treatment (control or removal)

```
P2control <- subset(PredatorTwo, Treatment=='Control')
P2removal <- subset(PredatorTwo, Treatment=='Removal')
```

Test control treatment data for normality

```
C <- with(data=P2control, StrikesP1 - StrikesP2)
shapiro.test(C)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  C
## W = 0.55711, p-value = 1.056e-05
```

Data non-normal, use non-parametric Wilcoxon signed-rank test

```
wilcox.test(P2control$StrikesP1, P2control$StrikesP2, paired = TRUE)
```

```
## Warning in wilcox.test.default(P2control$StrikesP1, P2control$StrikesP2, :
## cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(P2control$StrikesP1, P2control$StrikesP2, :
## cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  P2control$StrikesP1 and P2control$StrikesP2
## V = 0, p-value = 0.08897
## alternative hypothesis: true location shift is not equal to 0
```

Conclusion: There is no difference (i.e. no difference in the number of strikes by predators in before vs. after conditions)

Test treatment data for normality

```
R <- with(data=P2removal, StrikesP1 - StrikesP2)
shapiro.test(R)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R
## W = 0.64647, p-value = 7.031e-05
```

Data non-normal, use non-parametric Wilcoxon signed-rank test

```
wilcox.test(P2removal$StrikesP1, P2removal$StrikesP2, paired = TRUE)
```

```
## Warning in wilcox.test.default(P2removal$StrikesP1, P2removal$StrikesP2, :
## cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(P2removal$StrikesP1, P2removal$StrikesP2, :
## cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  P2removal$StrikesP1 and P2removal$StrikesP2
## V = 0, p-value = 0.003702
## alternative hypothesis: true location shift is not equal to 0
```

Conclusion: There is a difference (i.e. significantly more strikes at mysids following removal of the farming damselfish)

6. Timed observations

Description of data

Field observations of longfin damselfish with and without associated mysid swarms.

The dataset (CBC2018_Observations_data.csv) contains the following variables (L-R):

- StegastesGroupNumber: observation number (1-60).
- TerritoryType: farm type ((Mysids) with associated mysids, (NoMysids) without associated mysids.
- TerritoryTypeOneOrZero: farm type as either (1) with associated mysids, (0) without associated mysids.
- StegastesStrikesAtMysids: total number of strikes by focal longfin damselfish on mysid swam during observational period.
- StegastesBitesOnFarm: total number of bites by focal longfin damselfish on farmed substrate during observational period.
- ChasesFromMysids: total number of chases by focal longfin damselfish directed at other fish species approaching mysid swarm during observational period.
- ChasesFromTurf: total number of chases by focal longfin damselfish directed at other fish species within its farm but not approaching mysid swarm during observational period.
- TotalChases: total number of chases by focal longfin damselfish during observational period.
- MysidInteractionEvents: total number of direct non-aggressive interactions between focal longfin damselfish and mysid swarm during observational period.
- TotalPredatorStrikes: total number of strikes by predatory fishes on mysid swarm during observational period.
- StegastesSizeCM: estimated total length of focal longfin damselfish in cm.
- StegastesGroupSize: total number of longfin damselfish associated with focal farm.
- EstimatedFarmSizeM2: estimated size of focal farm in m2.

Analysis

Test effect of farm type and/or longfin damselfish group size on number of chases by longfin damselfish using glm

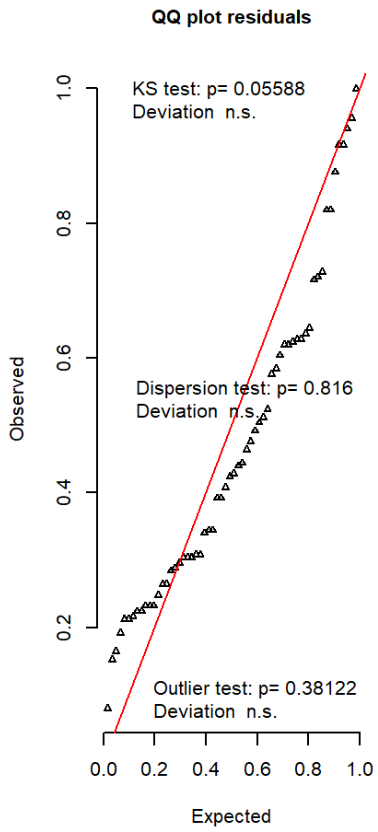
```
Observation1=data.frame(Observation)
summary(Observation1)
```

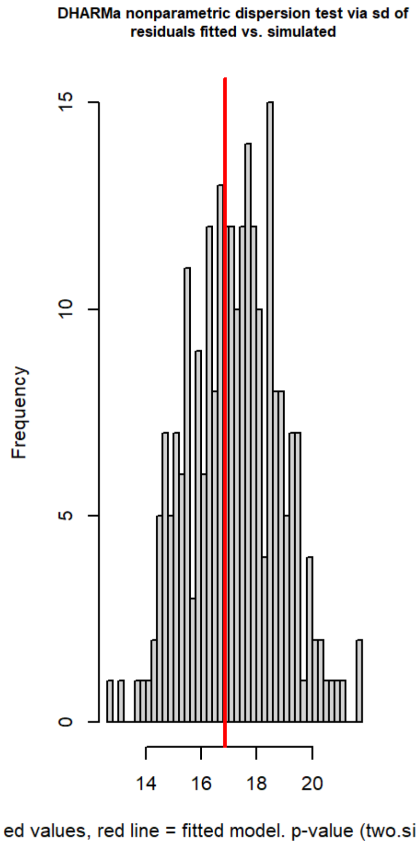
```
##  StegastesGroupNumber  TerritoryType TerritoryTypeOneOrZero
##  Min.   : 1.00        Mysids  :30    Min.   :0.0
##  1st Qu.:15.75        NoMysids:30    1st Qu.:0.0
##  Median :30.50                       Median :0.5
##  Mean   :30.50                       Mean   :0.5
##  3rd Qu.:45.25                       3rd Qu.:1.0
##  Max.   :60.00                       Max.   :1.0
##
##  StegastesStrikesAtMysids StegastesBitesOnFarm ChasesFromMysids ChasesFromTurf
##  0 :27                    Min.   : 15.0        na     :30       Min.   : 1.00
##  1 : 2                    1st Qu.:143.8        0      : 6       1st Qu.: 5.00
##  2 : 1                    Median :181.5        1      : 5       Median :12.00
##  na:30                    Mean   :193.0        2      : 4       Mean   :15.57
##                           3rd Qu.:240.0        3      : 3       3rd Qu.:19.25
##                           Max.   :438.0        7      : 3       Max.   :99.00
##                                                (Other): 9
##   TotalChases     MysidInteractionEvents TotalPredatorStrikes StegastesSizeCM
##  Min.   :  1.00   na     :30             0 :15                Min.   :4.500
##  1st Qu.:  5.75   0      : 7             1 : 7                1st Qu.:7.000
##  Median : 13.00   2      : 4             2 : 4                Median :8.000
##  Mean   : 17.72   4      : 4             3 : 2                Mean   :7.692
##  3rd Qu.: 25.00   1      : 3             4 : 1                3rd Qu.:8.000
##  Max.   :121.00   3      : 3             6 : 1                Max.   :9.000
##                   (Other): 9             na:30
##  StegastesGroupSize EstimatedFarmSizeM2
##  Min.   :1.000      Min.   :0.420
##  1st Qu.:1.000      1st Qu.:1.000
##  Median :1.000      Median :1.725
##  Mean   :1.717      Mean   :2.296
##  3rd Qu.:2.000      3rd Qu.:3.000
##  Max.   :7.000      Max.   :9.000
##
```

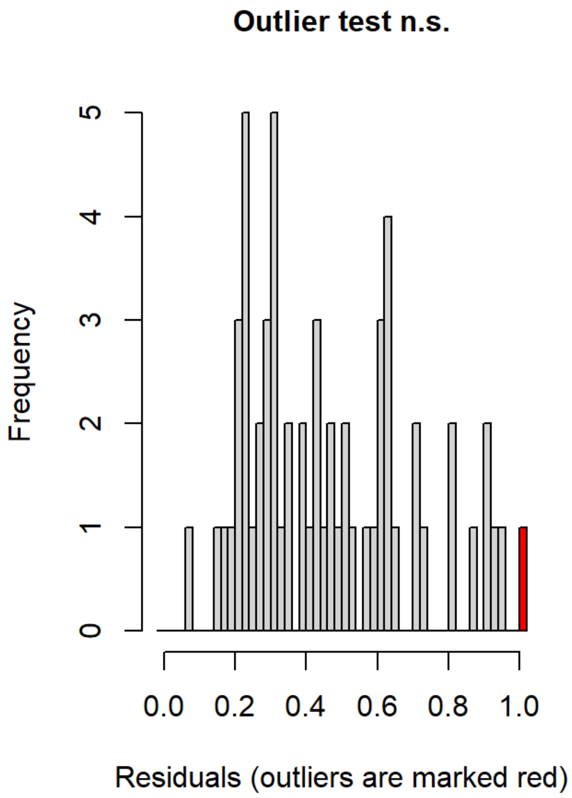
```
m1<-glm(TotalChases~TerritoryType*StegastesGroupSize,data=Observation1)
```

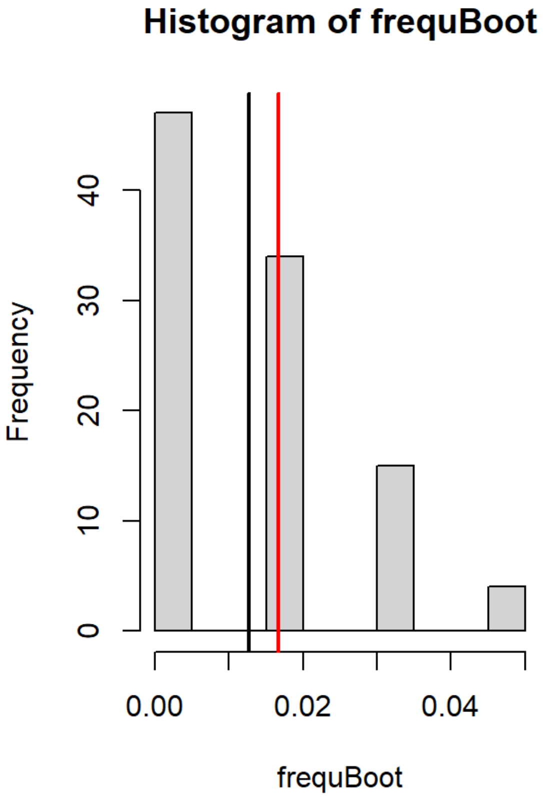
Test model assumptions

```
testResiduals(m1)
```


QQ plot residuals
KS test: p= 0.05588
Deviation n.s.
Dispersion test: p= 0.816
Deviation n.s.
Outlier test: p= 0.38122
Deviation n.s.
Observed
Expected
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


DHARMa nonparametric dispersion test via sd of
residuals fitted vs. simulated
Frequency
0
5
10
15
14
16
18
20
ed values, red line = fitted model. p-value (two.si


Outlier test n.s.
Frequency
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)

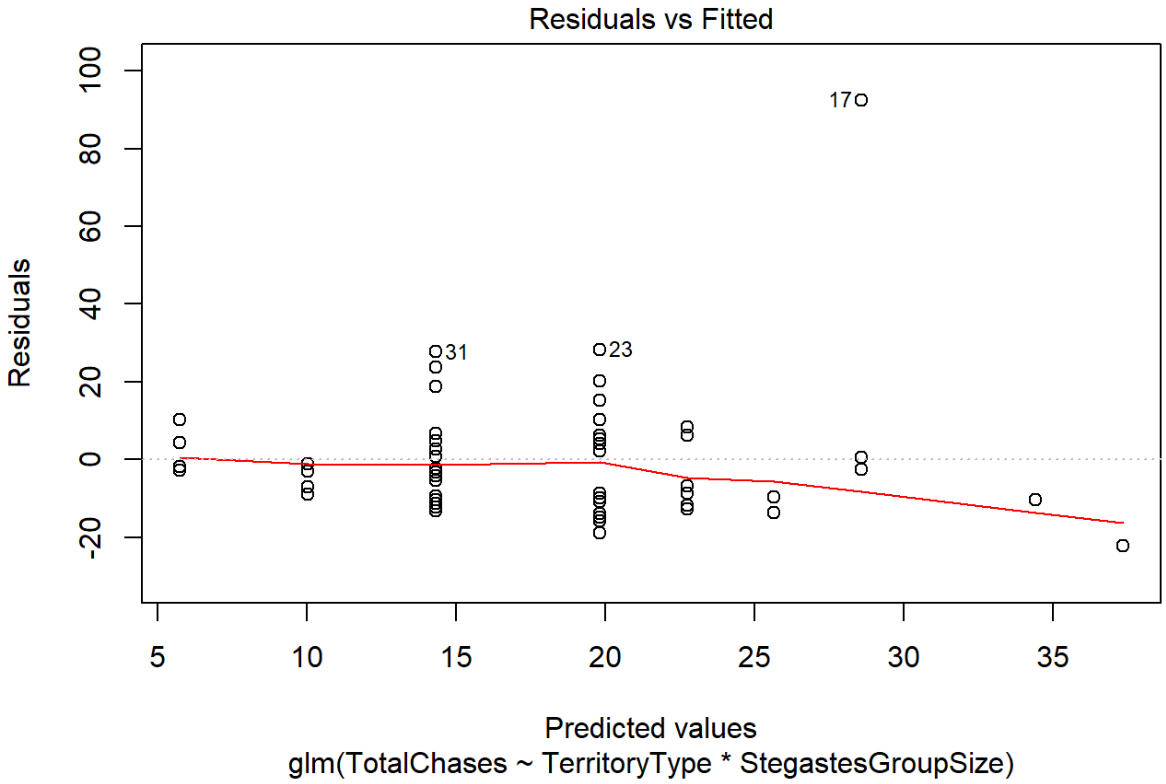

Histogram of frequBoot
Frequency
0
10
20
30
40
0.00
0.02
0.04
frequBoot

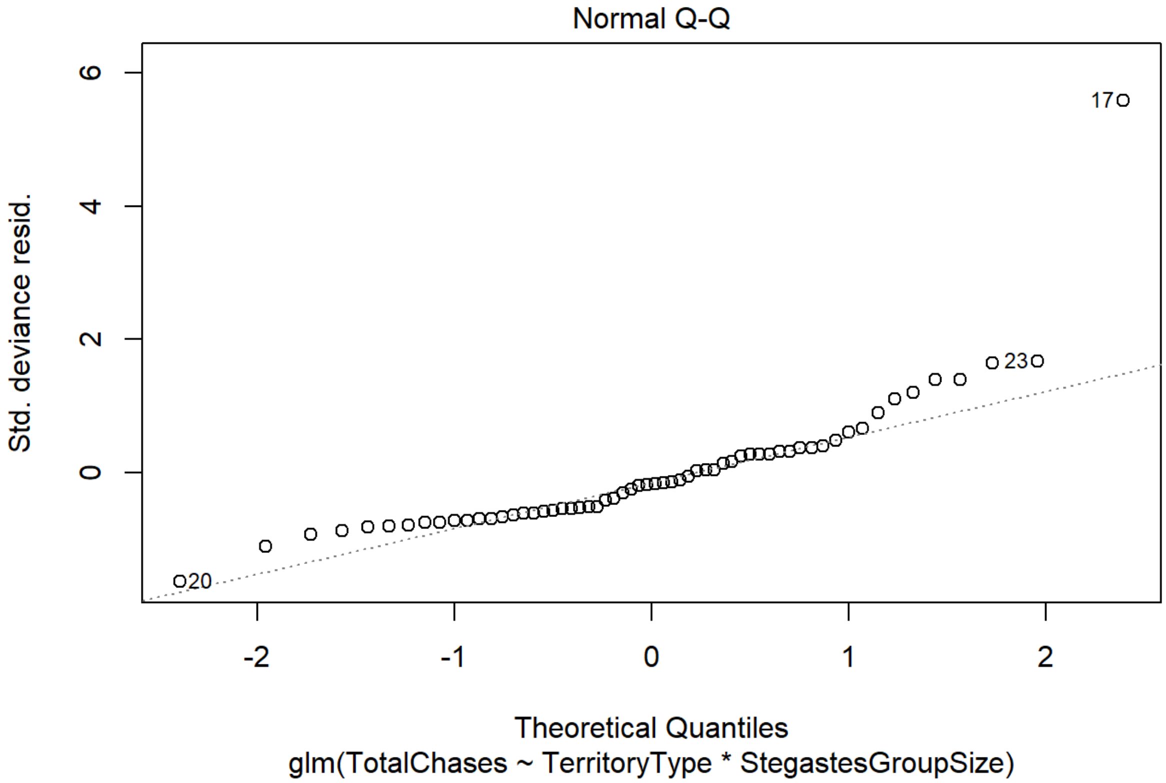
```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.17267, p-value = 0.05588
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.97848, p-value = 0.816
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00 0.05
## sample estimates:
## outlier frequency (expected: 0.0126666666666667 )
##                                        0.01666667
```

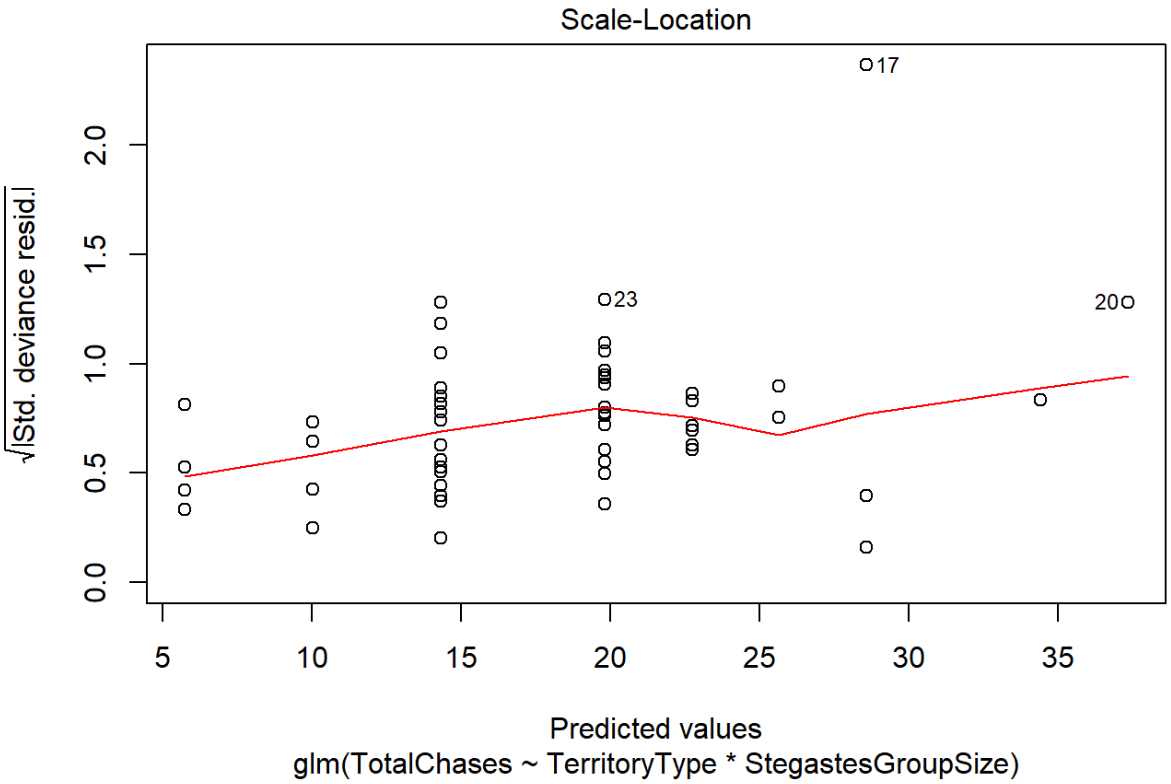
```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.17267, p-value = 0.05588
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.97848, p-value = 0.816
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00 0.05
## sample estimates:
## outlier frequency (expected: 0.0126666666666667 )
##                                        0.01666667
```

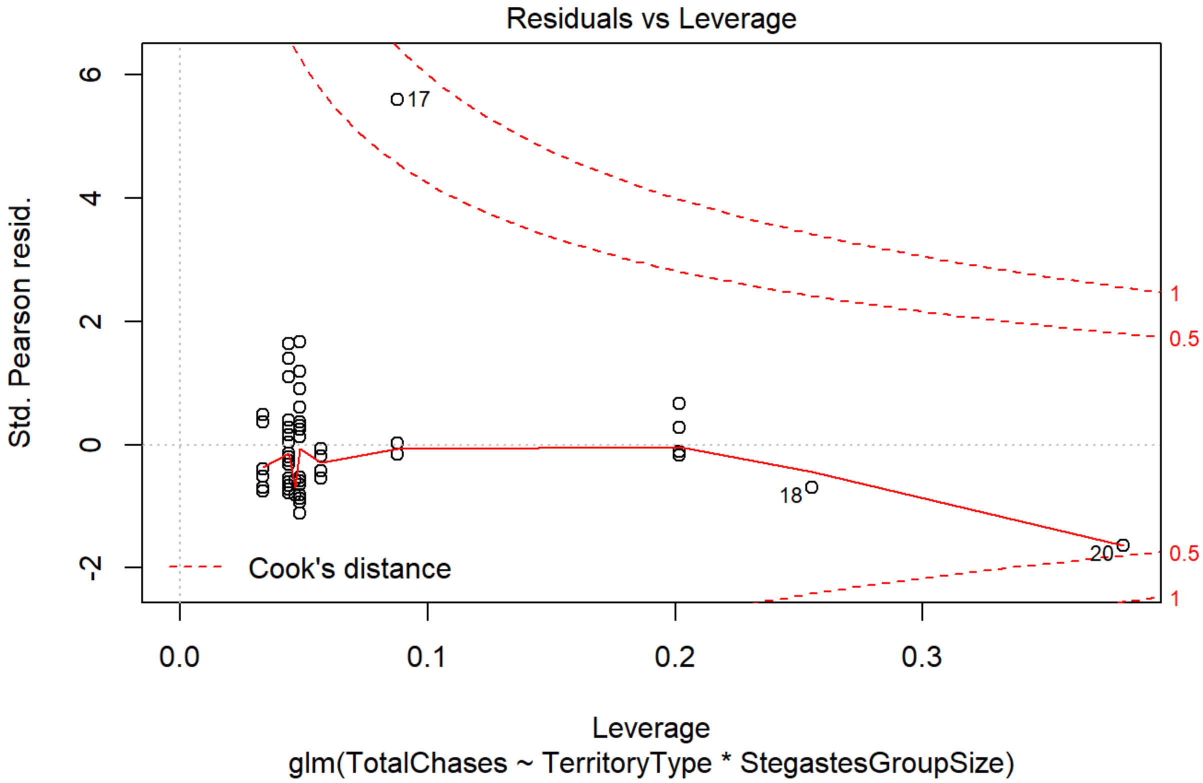
Model assumption not met, investigate data further

```
plot(m1)
```


Residuals vs Fitted
Residuals
100
80
60
40
20
0
-20
17
31
23
5
10
15
20
25
30
35
Predicted values
glm(TotalChases ~ TerritoryType * StegastesGroupSize)


Normal Q-Q
Std. deviance resid.
6
4
2
0
17
23
20
-2
-1
0
1
2
Theoretical Quantiles
glm(TotalChases ~ TerritoryType * StegastesGroupSize)





Outlier at row 17. While not significant according to DHARMA outlier test, further investigation indicates it is an issue (see plots above), and based on this is removed

```
Observation2=Observation1[-c(17),]
summary(Observation2)
```

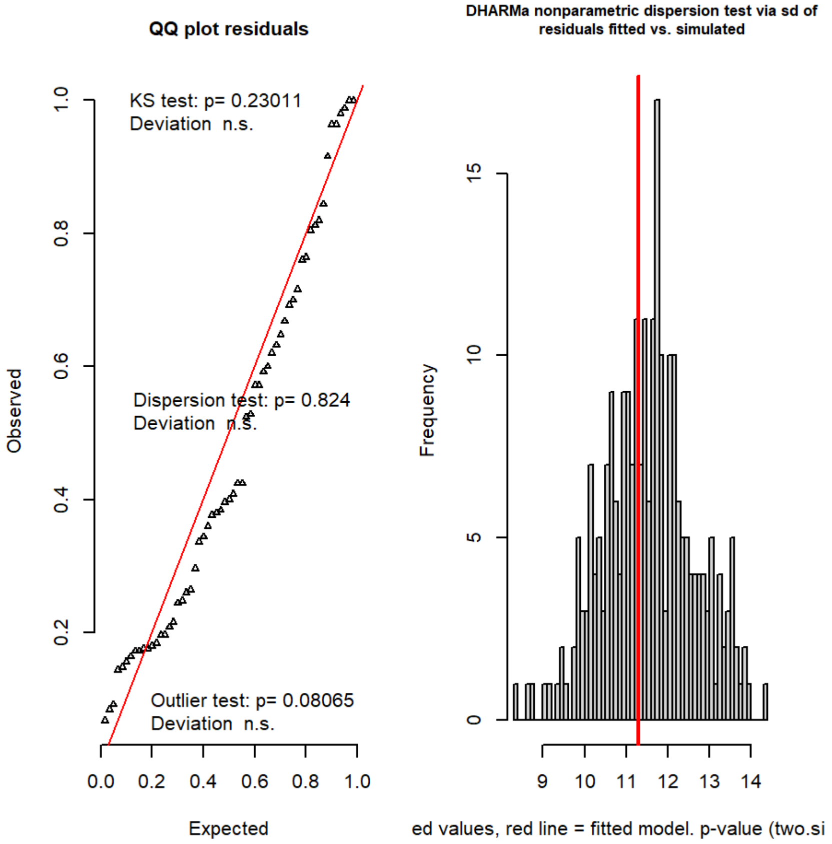
```
##  StegastesGroupNumber  TerritoryType TerritoryTypeOneOrZero
##  Min.   : 1.00        Mysids  :29    Min.   :0.0000       
##  1st Qu.:15.50        NoMysids:30    1st Qu.:0.0000       
##  Median :31.00                       Median :0.0000       
##  Mean   :30.73                       Mean   :0.4915       
##  3rd Qu.:45.50                       3rd Qu.:1.0000       
##  Max.   :60.00                       Max.   :1.0000       
##                                                           
##  StegastesStrikesAtMysids StegastesBitesOnFarm ChasesFromMysids ChasesFromTurf
##  0 :26                    Min.   : 15.0        na     :30       Min.   : 1.00 
##  1 : 2                    1st Qu.:142.5        0      : 6       1st Qu.: 5.00 
##  2 : 1                    Median :183.0        1      : 5       Median :12.00 
##  na:30                    Mean   :193.6        2      : 4       Mean   :14.15 
##                           3rd Qu.:240.0        3      : 3       3rd Qu.:19.00 
##                           Max.   :438.0        7      : 3       Max.   :42.00 
##                                                (Other): 8                     
##   TotalChases    MysidInteractionEvents TotalPredatorStrikes StegastesSizeCM
##  Min.   : 1.00   na     :30             0 :15                Min.   :4.50   
##  1st Qu.: 5.50   0      : 7             1 : 7                1st Qu.:7.00   
##  Median :12.00   2      : 4             2 : 4                Median :8.00   
##  Mean   :15.97   4      : 4             3 : 1                Mean   :7.72   
##  3rd Qu.:24.50   1      : 3             4 : 1                3rd Qu.:8.00   
##  Max.   :48.00   5      : 3             6 : 1                Max.   :9.00   
##                  (Other): 8             na:30                               
##  StegastesGroupSize EstimatedFarmSizeM2
##  Min.   :1.000      Min.   :0.420      
##  1st Qu.:1.000      1st Qu.:1.000      
##  Median :1.000      Median :1.500      
##  Mean   :1.678      Mean   :2.234      
##  3rd Qu.:2.000      3rd Qu.:3.000      
##  Max.   :7.000      Max.   :9.000      
## 
```

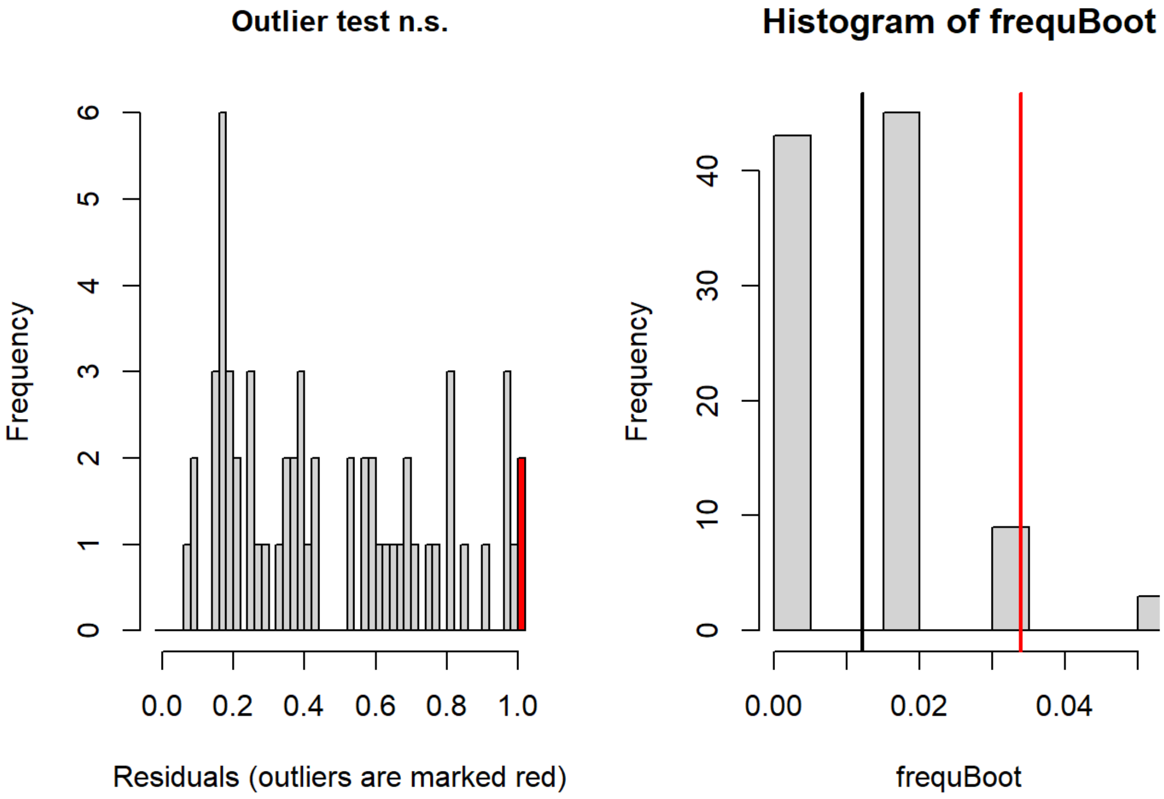
Outlier removed

```
m1<-glm(TotalChases~TerritoryType*StegastesGroupSize,data=Observation2)
```

Test model assumptions

```
testResiduals(m1)
```


QQ plot residuals
KS test: p= 0.23011
Deviation n.s.
Dispersion test: p= 0.824
Deviation n.s.
Outlier test: p= 0.08065
Deviation n.s.
Observed
Expected
0.2
0.4
0.6
0.8
1.0
0.0
DHARMa nonparametric dispersion test via sd of residuals fitted vs. simulated
Frequency
0
5
10
15
9
10
11
12
13
14
ed values, red line = fitted model. p-value (two.si


Outlier test n.s.
Frequency
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)
Histogram of frequBoot
Frequency
0
10
20
30
40
0.00
0.02
0.04
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.13532, p-value = 0.2301
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.97811, p-value = 0.824
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 59, p-value = 0.24
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04279661
## sample estimates:
## outlier frequency (expected: 0.0122033898305085 )
##                                         0.03389831
```

```
## $uniformity
## 
##  One-sample Kolmogorov-Smirnov test
## 
## data:  simulationOutput$scaledResiduals
## D = 0.13532, p-value = 0.2301
## alternative hypothesis: two-sided
## 
## 
## $dispersion
## 
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
## 
## data:  simulationOutput
## ratioObsSim = 0.97811, p-value = 0.824
## alternative hypothesis: two.sided
## 
## 
## $outliers
## 
##  DHARMa bootstrapped outlier test
## 
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 59, p-value = 0.24
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04279661
## sample estimates:
## outlier frequency (expected: 0.0122033898305085 )
##                                           0.03389831
```

Model assumptions met

Test effect of interaction

```
m11<-update(m1,.~.-TerritoryType:StegastesGroupSize)

anova(m1,m11,test="F")
```

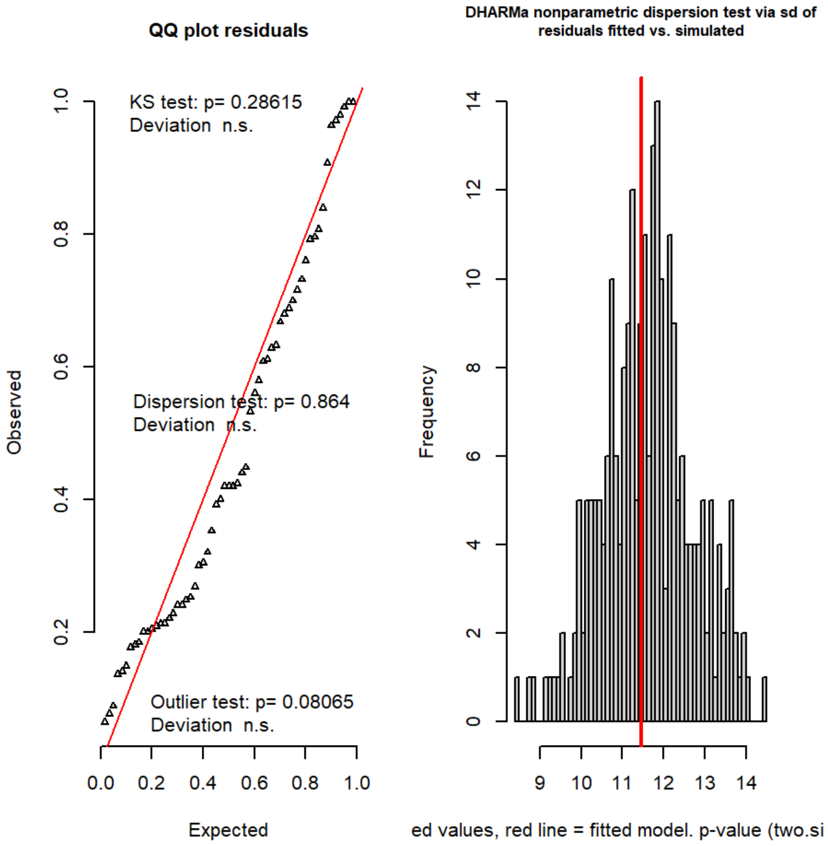
```
## Analysis of Deviance Table
## 
## Model 1: TotalChases ~ TerritoryType * StegastesGroupSize
## Model 2: TotalChases ~ TerritoryType + StegastesGroupSize
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
## 1        55     7385.9                          
## 2        56     7625.8 -1  -239.91 1.7865 0.1869
```

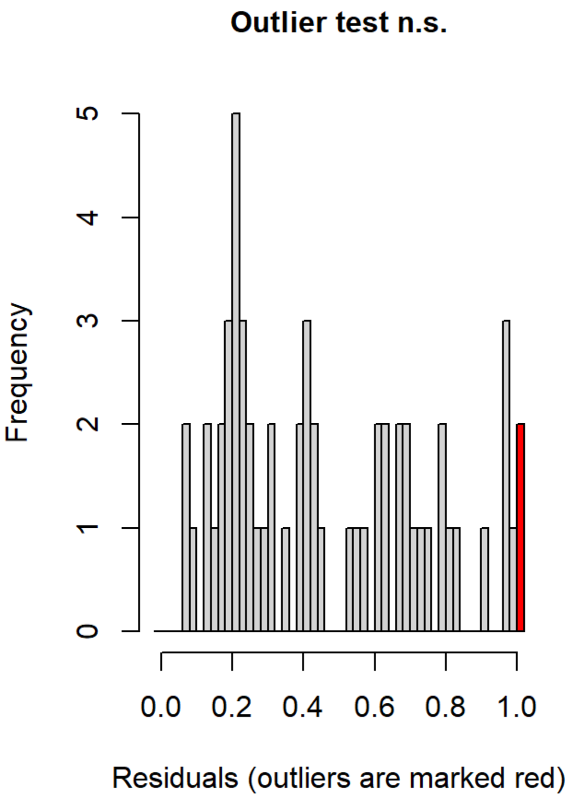
Interaction not significant, so removed

```
m1<-glm(TotalChases~TerritoryType+StegastesGroupSize,data=Observation2)
```

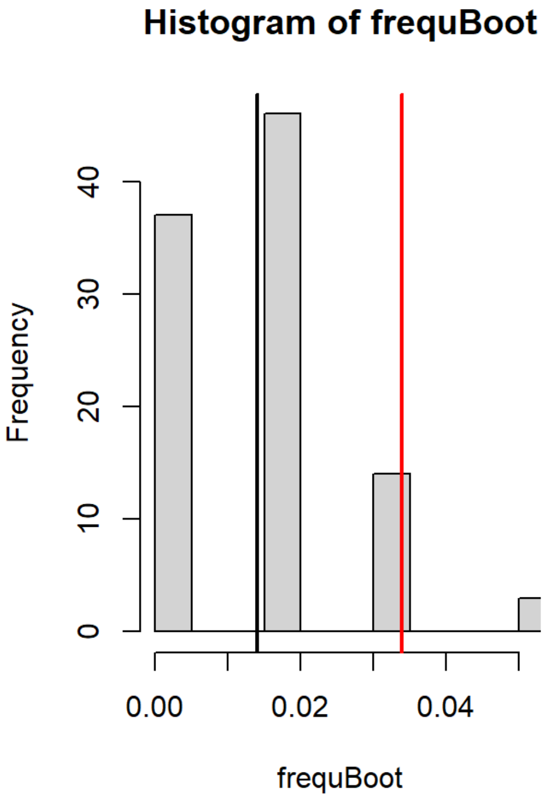
Test model assumptions

```
testResiduals(m1)
```





Outlier test n.s.



Histogram of frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.12827, p-value = 0.2862
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.98697, p-value = 0.864
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 59, p-value = 0.34
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04279661
## sample estimates:
## outlier frequency (expected: 0.0140677966101695 )
##                                         0.03389831
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.12827, p-value = 0.2862
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.98697, p-value = 0.864
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 59, p-value = 0.34
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04279661
## sample estimates:
## outlier frequency (expected: 0.0140677966101695 )
##                                        0.03389831
```

Model assumptions met

Test effect of longfin damselfish group size

```
m11<-update(m1,.~.-StegastesGroupSize)

anova(m1,m11,test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: TotalChases ~ TerritoryType + StegastesGroupSize
## Model 2: TotalChases ~ TerritoryType
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
## 1        56     7625.8
## 2        57     7666.4 -1  -40.592 0.2981 0.5873
```

Conclusion: There is no difference (i.e. no effect of group size on the total number of defensive chases)

Test the effect of farm type

```
m11<-update(m1,.~.-TerritoryType)

anova(m1,m11,test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: TotalChases ~ TerritoryType + StegastesGroupSize
## Model 2: TotalChases ~ StegastesGroupSize
##   Resid. Df Resid. Dev Df Deviance      F  Pr(>F)
## 1        56     7625.8
## 2        57     8357.9 -1  -732.15 5.3765 0.02408 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

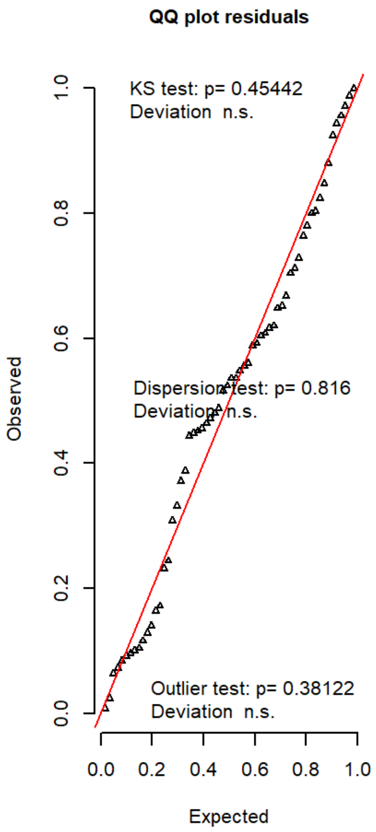
Conclusion: There is a difference (i.e. significant effect of farm type on the total number of defensive chases)

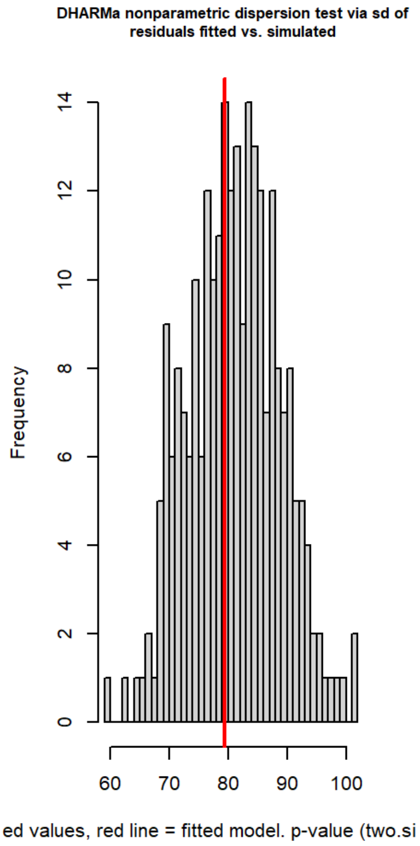
Test effect of farm type and/or longfin damselfish group size on number of bites on farm by longfin damselfish using glm

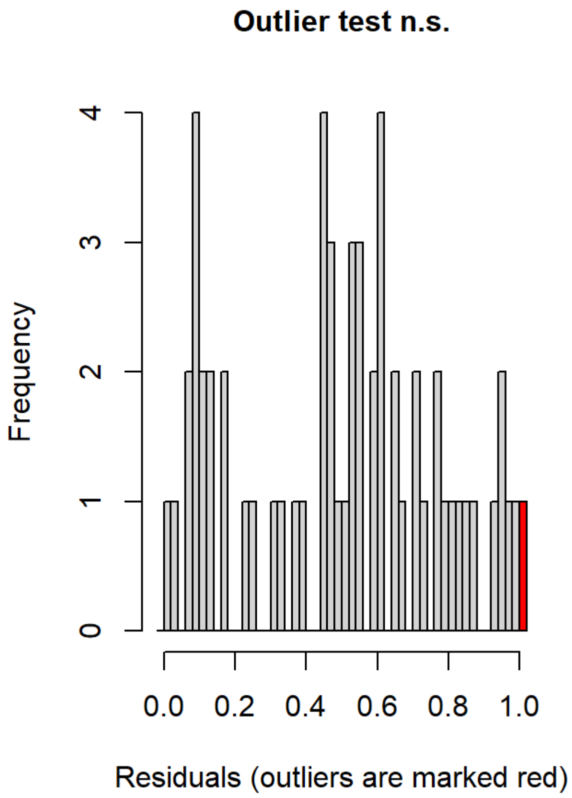
```
m2<-glm(StegastesBitesOnFarm~TerritoryType*StegastesGroupSize,data=Observation1)
```

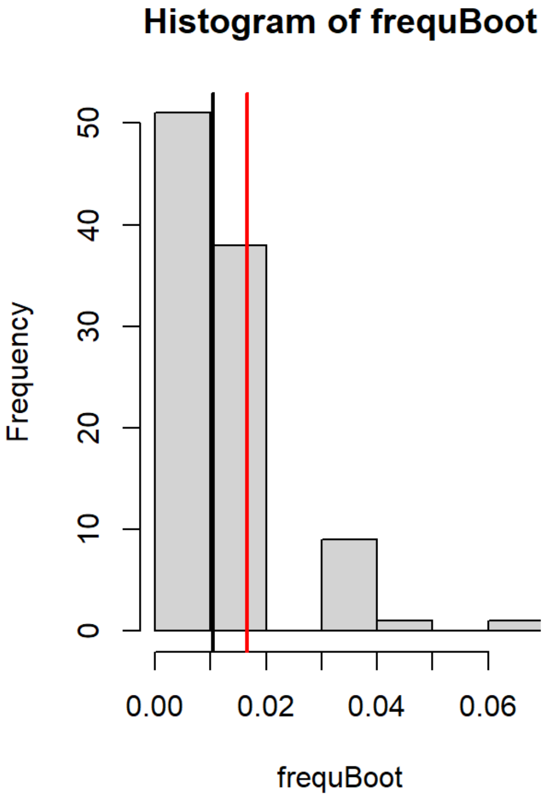
Test model assumptions

```
testResiduals(m2)
```


QQ plot residuals
KS test: p= 0.45442
Deviation n.s.
Dispersion test: p= 0.816
Deviation n.s.
Outlier test: p= 0.38122
Deviation n.s.
Observed
0.0
0.2
0.4
0.6
0.8
1.0
Expected
0.0
0.2
0.4
0.6
0.8
1.0


DHARMa nonparametric dispersion test via sd of residuals fitted vs. simulated
Frequency
0
2
4
6
8
10
12
14
60
70
80
90
100
ed values, red line = fitted model. p-value (two.si


Outlier test n.s.
Frequency
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Residuals (outliers are marked red)


Histogram of frequBoot
Frequency
0
10
20
30
40
50
0.00
0.02
0.04
0.06
frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.11067, p-value = 0.4544
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.97848, p-value = 0.816
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 0.98
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.03333333
## sample estimates:
## outlier frequency (expected: 0.0105 )
##                             0.01666667
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.11067, p-value = 0.4544
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.97848, p-value = 0.816
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 0.98
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.03333333
## sample estimates:
## outlier frequency (expected: 0.0105 )
##                            0.01666667
```

Model assumption met

Test effect of interaction

```
m22<-update(m2,.~.-TerritoryType:StegastesGroupSize)

anova(m2,m22,test="F")
```

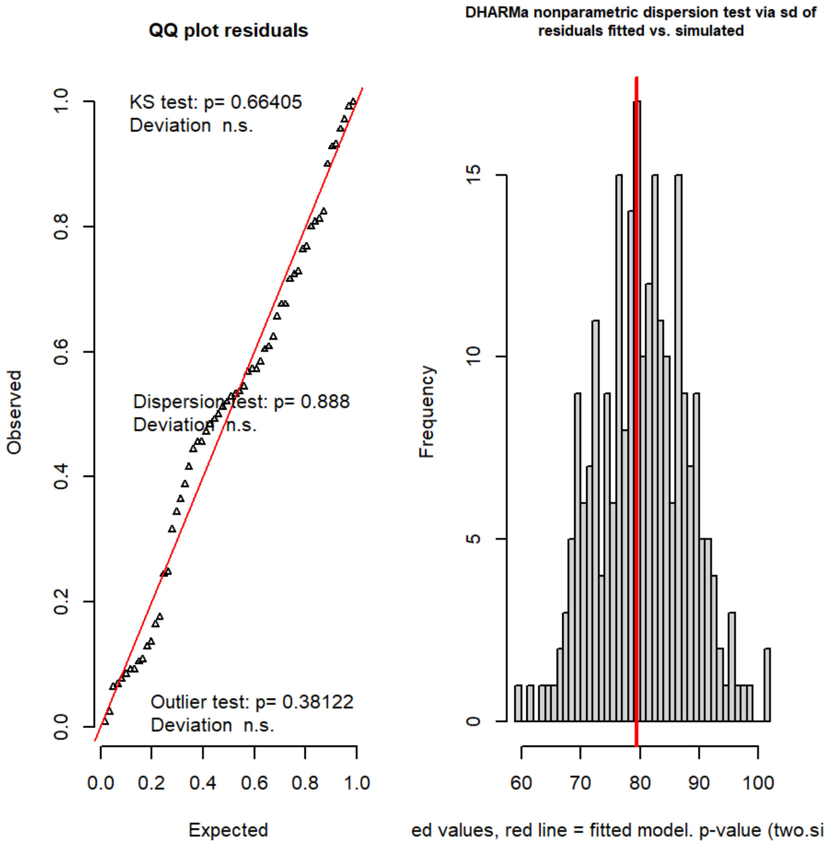
```
## Analysis of Deviance Table
##
## Model 1: StegastesBitesOnFarm ~ TerritoryType * StegastesGroupSize
## Model 2: StegastesBitesOnFarm ~ TerritoryType + StegastesGroupSize
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
## 1        56     371569
## 2        57     372613 -1  -1044.7 0.1574  0.693
```

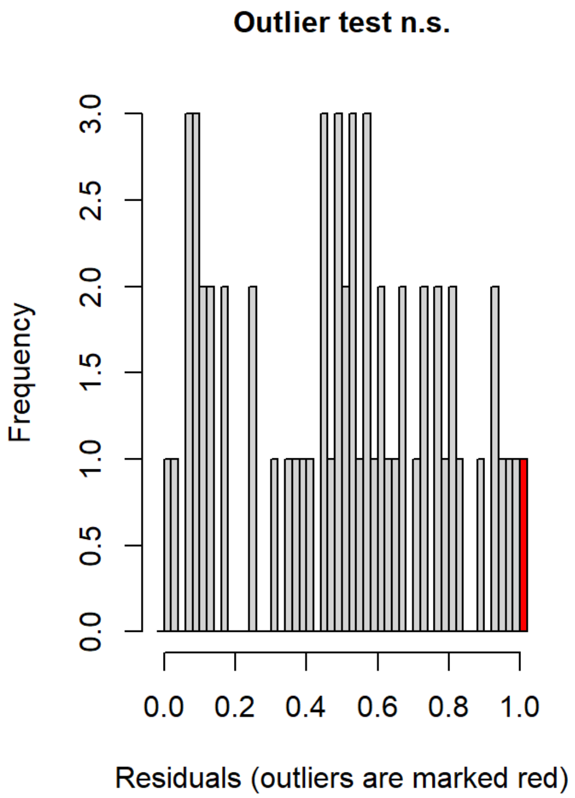
Interaction is not significant, so removed.

```
m2<-glm(StegastesBitesOnFarm~TerritoryType+StegastesGroupSize,data=Observation1)
```

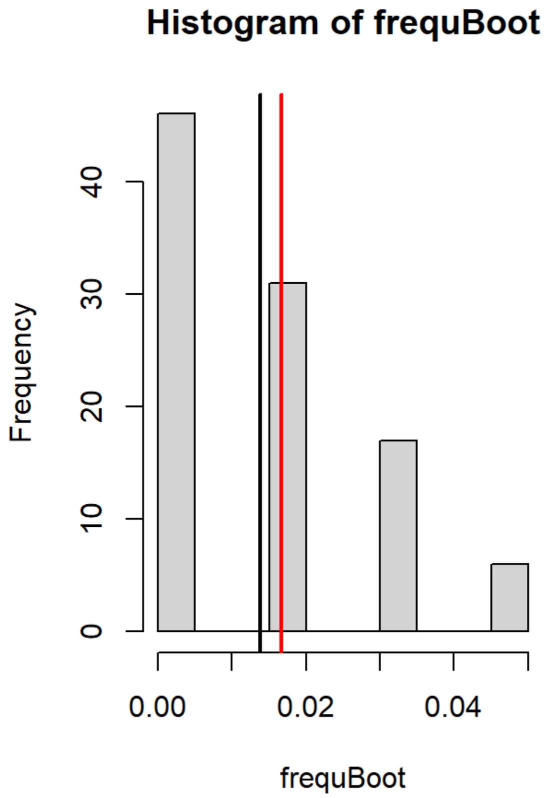
Test model assumptions

```
testResiduals(m2)
```





Outlier test n.s.



Histogram of frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.094, p-value = 0.6641
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.98718, p-value = 0.888
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00 0.05
## sample estimates:
## outlier frequency (expected: 0.0138333333333333 )
##                                        0.01666667
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.094, p-value = 0.6641
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.98718, p-value = 0.888
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00 0.05
## sample estimates:
## outlier frequency (expected: 0.0138333333333333 )
##                                         0.01666667
```

Model assumptions met

Test effect of longfin damselfish group size

```
m22<-update(m2,.~.-StegastesGroupSize)

anova(m2,m22,test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: StegastesBitesOnFarm ~ TerritoryType + StegastesGroupSize
## Model 2: StegastesBitesOnFarm ~ TerritoryType
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
## 1        57     372613
## 2        58     381339 -1  -8725.7 1.3348 0.2528
```

Conclusion: There is no difference (i.e. no effect of group size on the total number of bites on farm)

Test effect of farm type

```
m22<-update(m2,.~.-TerritoryType)

anova(m2,m22,test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: StegastesBitesOnFarm ~ TerritoryType + StegastesGroupSize
## Model 2: StegastesBitesOnFarm ~ StegastesGroupSize
##   Resid. Df Resid. Dev Df Deviance      F   Pr(>F)
## 1        57     372613
## 2        58     425483 -1   -52870 8.0877 0.006178 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare area of farms with and without mysids

shapiro.test (CBC2018_Observations_data$EstimatedFarmSizeM2)p < 0.05 - not normal. Use non-parametric test.

Wilcoxon test

```
wilcox.test(EstimatedFarmSizeM2 ~ TerritoryType, data=Observation)
```

```
## Warning in wilcox.test.default(x = c(2, 1, 2, 1, 0.48, 2, 2.5, 1.5, 1, 1.2, :
## cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  EstimatedFarmSizeM2 by TerritoryType
## W = 562.5, p-value = 0.09507
## alternative hypothesis: true location shift is not equal to 0
```

Conclusion: There is a difference (i.e. significant effect of farm type on the total number of bites on farm)

7. Farm algae composition

Description of data

Comparison of algal composition within longfin damselfish farms with and without associated mysids.

The dataset (CBC2018_Ochrophyta_data.csv) contains the following variables (L-R):

- Territory: farm number (1-60).
- Treatment: farm type ((WM) with associated mysids, (NM) without associated mysids.
- Ochrophyta: abundance of Ochrophyta within farm based on a classification scheme ((L) less than 10 percent coverage, (M) 10-30 percent coverage, (H) more than 30 percent coverage.

Analysis

Load required packages

```
library(nnet)
library(effects)
```

```
## Loading required package: carData
```

```
## Use the command
##     lattice::trellis.par.set(effectsTheme())
##   to customize lattice options for effects plots.
## See ?effectsTheme for details.
```

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------
------------- tidyverse 1.3.0 --
```

```
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------------------------------
------- tidyverse_conflicts() --
## x dplyr::arrange()   masks plyr::arrange()
## x purrr::compact()   masks plyr::compact()
## x dplyr::count()     masks plyr::count()
## x tidyr::expand()    masks Matrix::expand()
## x dplyr::failwith()  masks plyr::failwith()
## x dplyr::filter()    masks stats::filter()
## x dplyr::id()        masks plyr::id()
## x dplyr::lag()       masks stats::lag()
## x dplyr::mutate()    masks plyr::mutate()
## x tidyr::pack()      masks Matrix::pack()
## x dplyr::rename()    masks plyr::rename()
## x dplyr::summarise() masks plyr::summarise()
## x dplyr::summarize() masks plyr::summarize()
## x tidyr::unpack()    masks Matrix::unpack()
```

Define factors

```
FarmComp$Treatment <- factor(FarmComp$Treatment, levels = c('NM','WM'))
FarmComp$Ochrophyta <- factor(FarmComp$Ochrophyta, levels = c("L","M","H"))
```

Multinomial logistic regression model with nnet package

```
model <- multinom(Ochrophyta ~ Treatment, data = FarmComp)
```

```
## # weights:  9 (4 variable)
## initial  value 65.916737
## iter  10 value 52.565097
## iter  10 value 52.565097
## final  value 52.565097
## converged
```

```
summary(model)
```

```
## Call:
## multinom(formula = Ochrophyta ~ Treatment, data = FarmComp)
##
## Coefficients:
##   (Intercept) TreatmentWM
## M  -0.6419223    1.683299
## H  -2.9443274    3.098321
##
## Std. Errors:
##   (Intercept) TreatmentWM
## M   0.3906859   0.6149082
## H   1.0259142   1.1670567
##
## Residual Deviance: 105.1302
## AIC: 113.1302
```

Model summary

```
sum.model <- summary(model)
str(sum.model)
```

```
## List of 31
##  $ n            : num [1:3] 2 0 3
##  $ nunits       : int 6
##  $ nconn        : num [1:7] 0 0 0 0 3 6 9
##  $ conn         : num [1:9] 0 1 2 0 1 2 0 1 2
##  $ nsunits      : num 3
##  $ decay        : num 0
##  $ entropy      : logi FALSE
##  $ softmax      : logi TRUE
##  $ censored     : logi FALSE
##  $ value        : num 52.6
##  $ wts          : num [1:9] 0 0 0 0 -0.642 ...
##  $ convergence  : int 0
##  $ fitted.values : num [1:60, 1:3] 0.2 0.2 0.2 0.2 0.2 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:60] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:3] "L" "M" "H"
##  $ residuals    : num [1:60, 1:3] -0.2 -0.2 -0.2 -0.2 -0.2 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:60] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:3] "L" "M" "H"
##  $ lev          : chr [1:3] "L" "M" "H"
##  $ call         : language multinom(formula = Ochrophyta ~ Treatment, data = F
armComp)
##  $ terms        :Classes 'terms', 'formula'  language Ochrophyta ~ Treatment
##   .. ..- attr(*, "variables")= language list(Ochrophyta, Treatment)
##   .. ..- attr(*, "factors")= int [1:2, 1] 0 1
##   .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. ..$ : chr [1:2] "Ochrophyta" "Treatment"
##   .. .. .. ..$ : chr "Treatment"
##   .. ..- attr(*, "term.labels")= chr "Treatment"
##   .. ..- attr(*, "order")= int 1
##   .. ..- attr(*, "intercept")= int 1
##   .. ..- attr(*, "response")= int 1
##   .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. ..- attr(*, "predvars")= language list(Ochrophyta, Treatment)
##   .. ..- attr(*, "dataClasses")= Named chr [1:2] "factor" "factor"
##   .. .. ..- attr(*, "names")= chr [1:2] "Ochrophyta" "Treatment"
##  $ weights      : num [1:60, 1] 1 1 1 1 1 1 1 1 1 1 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:60] "1" "2" "3" "4" ...
##   .. ..$ : NULL
##  $ deviance     : num 105
##  $ rank         : int 2
##  $ lab          : chr [1:3] "L" "M" "H"
##  $ coefnames    : chr [1:2] "(Intercept)" "TreatmentWM"
##  $ vcoefnames   : chr [1:2] "(Intercept)" "TreatmentWM"
##  $ contrasts    :List of 1
##   ..$ Treatment: chr "contr.treatment"
##  $ xlevels      :List of 1
##   ..$ Treatment: chr [1:2] "NM" "WM"
##  $ edf          : num 4
```

```
##  $ AIC            : num 113
##  $ is.binomial    : logi FALSE
##  $ digits         : int 7
##  $ coefficients   : num [1:2, 1:2] -0.642 -2.944 1.683 3.098
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "M" "H"
##   .. ..$ : chr [1:2] "(Intercept)" "TreatmentWM"
##  $ standard.errors: num [1:2, 1:2] 0.391 1.026 0.615 1.167
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "M" "H"
##   .. ..$ : chr [1:2] "(Intercept)" "TreatmentWM"
##  - attr(*, "class")= chr "summary.multinom"
```

Calculate p-values

```
pt(abs(sum.model$coefficients / sum.model$standard.errors), df = nrow(FarmComp)-6,
lower = FALSE)
```

```
##   (Intercept) TreatmentWM
## M 0.053090266 0.004182880
## H 0.002924859 0.005202686
```

Obtain predicted probabilities and confidence intervals from model

```
summary(allEffects(model))
```

```
##  model: Ochrophyta ~ Treatment
##
## Treatment effect (probability) for L
## Treatment
##        NM        WM
## 0.6333454 0.2000161
##
## Treatment effect (probability) for M
## Treatment
##        NM        WM
## 0.3333169 0.5666685
##
## Treatment effect (probability) for H
## Treatment
##         NM         WM
## 0.03333769 0.23331549
##
##  Lower 95 Percent Confidence Limits for L
## Treatment
##         NM         WM
## 0.45116214 0.09272865
##
##  Lower 95 Percent Confidence Limits for M
## Treatment
##        NM        WM
## 0.1896446 0.3884447
##
##  Lower 95 Percent Confidence Limits for H
## Treatment
##          NM          WM
## 0.004676559 0.115499607
##
##  Upper 95 Percent Confidence Limits for L
## Treatment
##        NM        WM
## 0.7840063 0.3795111
##
##  Upper 95 Percent Confidence Limits for M
## Treatment
##        NM        WM
## 0.5164638 0.7291663
##
##  Upper 95 Percent Confidence Limits for H
## Treatment
##        NM        WM
## 0.2020040 0.4149324
## NULL
```

Summarize probabilities and confidence intervals in a tibble

```
data <- tibble(treatment = c("NM", "WM", "NM", "WM", "NM", "WM"),
               ochrophyta = c("Low coverage","Low coverage","Medium coverage","Medi
um coverage","High coverage","High coverage"),
               probability = c(0.633,0.2,0.333,0.567,0.033,0.233),
               lcl = c(0.451,0.093,0.19,0.388,0.005,0.115),
               ucl = c(0.784,0.38,0.516,0.729,0.202,0.415))
data
```

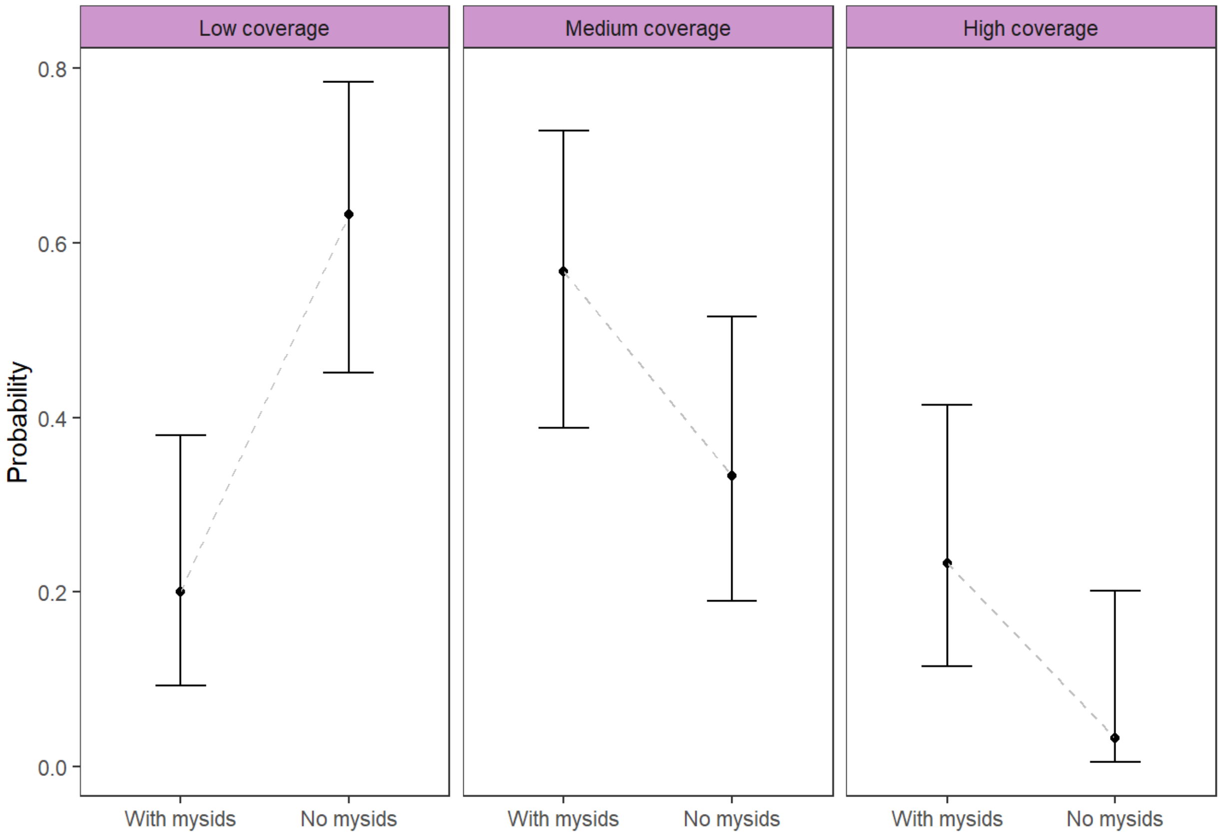
```
## # A tibble: 6 x 5
##   treatment ochrophyta      probability   lcl   ucl
##   <chr>     <chr>                 <dbl> <dbl> <dbl>
## 1 NM        Low coverage          0.633 0.451 0.784
## 2 WM        Low coverage          0.2   0.093 0.38
## 3 NM        Medium coverage       0.333 0.19  0.516
## 4 WM        Medium coverage       0.567 0.388 0.729
## 5 NM        High coverage         0.033 0.005 0.202
## 6 WM        High coverage         0.233 0.115 0.415
```

Define factors to reorder for ggplot

```
data$ochrophyta <- factor(data$ochrophyta, levels=c("Low coverage","Medium coverag
e","High coverage"))
data$treatment <- factor(data$treatment, levels=c("WM","NM"))
```

Plot predicted model probabilites and confidence intervals

```
gg <- ggplot(data, aes(treatment, probability, ymin = lcl, ymax = ucl, group = ochr
ophyta)) +
  geom_point() +
  geom_line(linetype = "dashed", color = "gray") +
  geom_errorbar(width = 0.3) +
  theme_bw() +
  scale_x_discrete (labels = c("With mysids", "No mysids")) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
  facet_grid(~ochrophyta) +
  theme(strip.background = element_rect(fill = "plum3")) +
  labs(y = "Probability", x = "")
gg
```



```
ggsave("ochrophyta.pdf", gg, width = 6, height = 3, useDingbats = FALSE)
```

Conclusion: Mysid-associated farms contain a significantly higher proportion of Ochrophyta (brown algae) than farms without associated mysids.

8. Longfin damselfish body condition

Description of data

Size and body condition of longfin damselfish with and without associated mysids.

The dataset (CBC2018_FishBodyCondition_data.csv) contains the following variables (L-R):

- Fish: replicate number (1-60).
- Date: date sample taken.
- Territory: farm identification code.
- TerritoryType: farm type that focal longfin damselfish was collected from ((Mysids) with associated mysids, (NoMysids) without associated mysids.
- TerritoryTypeYesOrNo: farm type as either (1) with associated mysids, (0) without associated mysids.
- TL: total length of focal longfin damselfish.
- SL: standard length of focal longfin damselfish.
- TW: total weight of focal longfin damselfish.
- LW: liver weight of focal longfin damselfish.
- HSI: calculated hepatosomatic index (HSI) value.

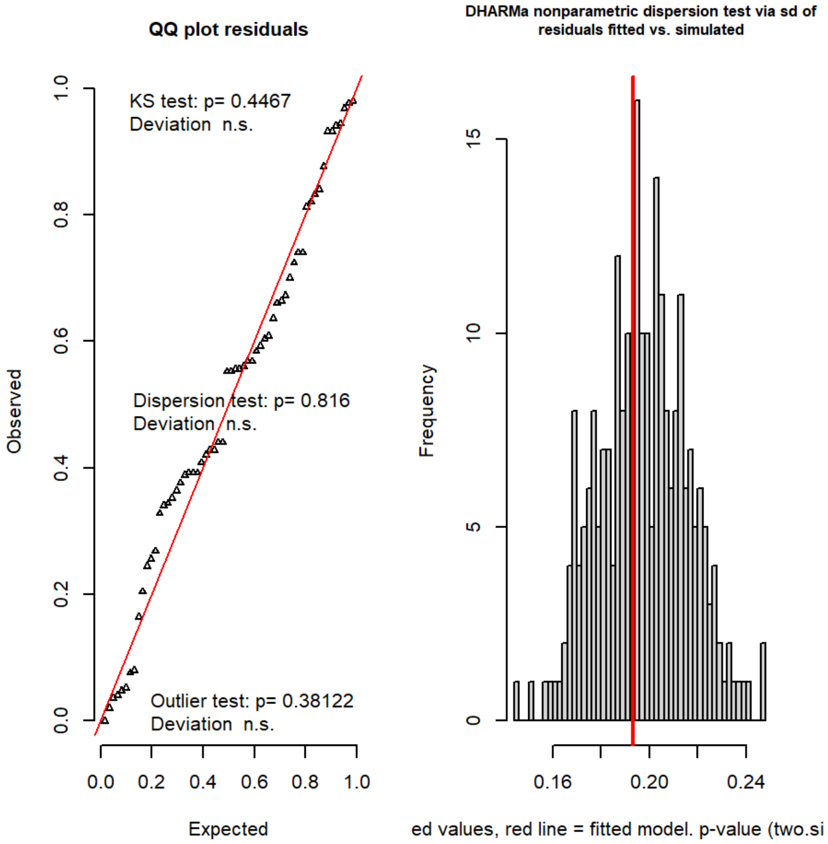
Analysis

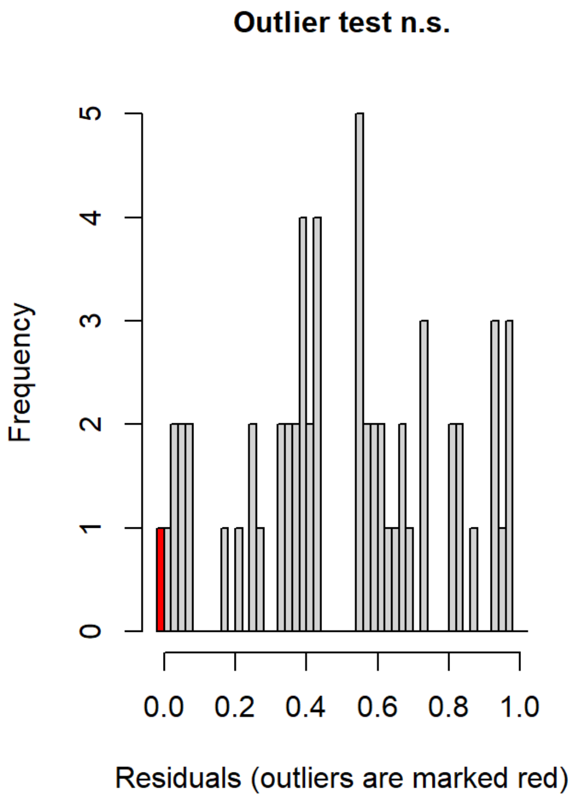
Test effect of farm type and fish SL on HSI using glm

```
m1<-glm(HSI~TerritoryType*SL,data=BodyCondition)
```

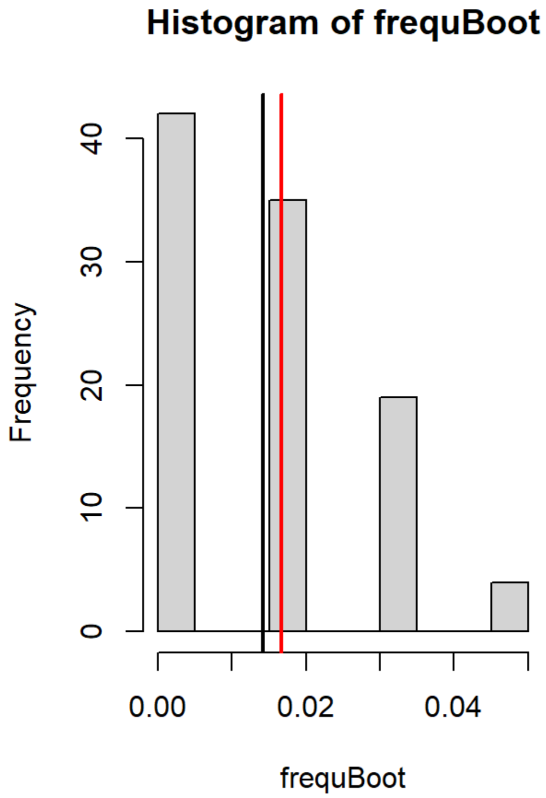
Test model assumptions

```
testResiduals(m1)
```





Outlier test n.s.



Histogram of frequBoot

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.11133, p-value = 0.4467
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.97848, p-value = 0.816
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00 0.05
## sample estimates:
## outlier frequency (expected: 0.0141666666666667 )
##                                        0.01666667
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.11133, p-value = 0.4467
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.97848, p-value = 0.816
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00 0.05
## sample estimates:
## outlier frequency (expected: 0.0141666666666667 )
##                                        0.01666667
```

Model assumptions met

Test effect of interaction

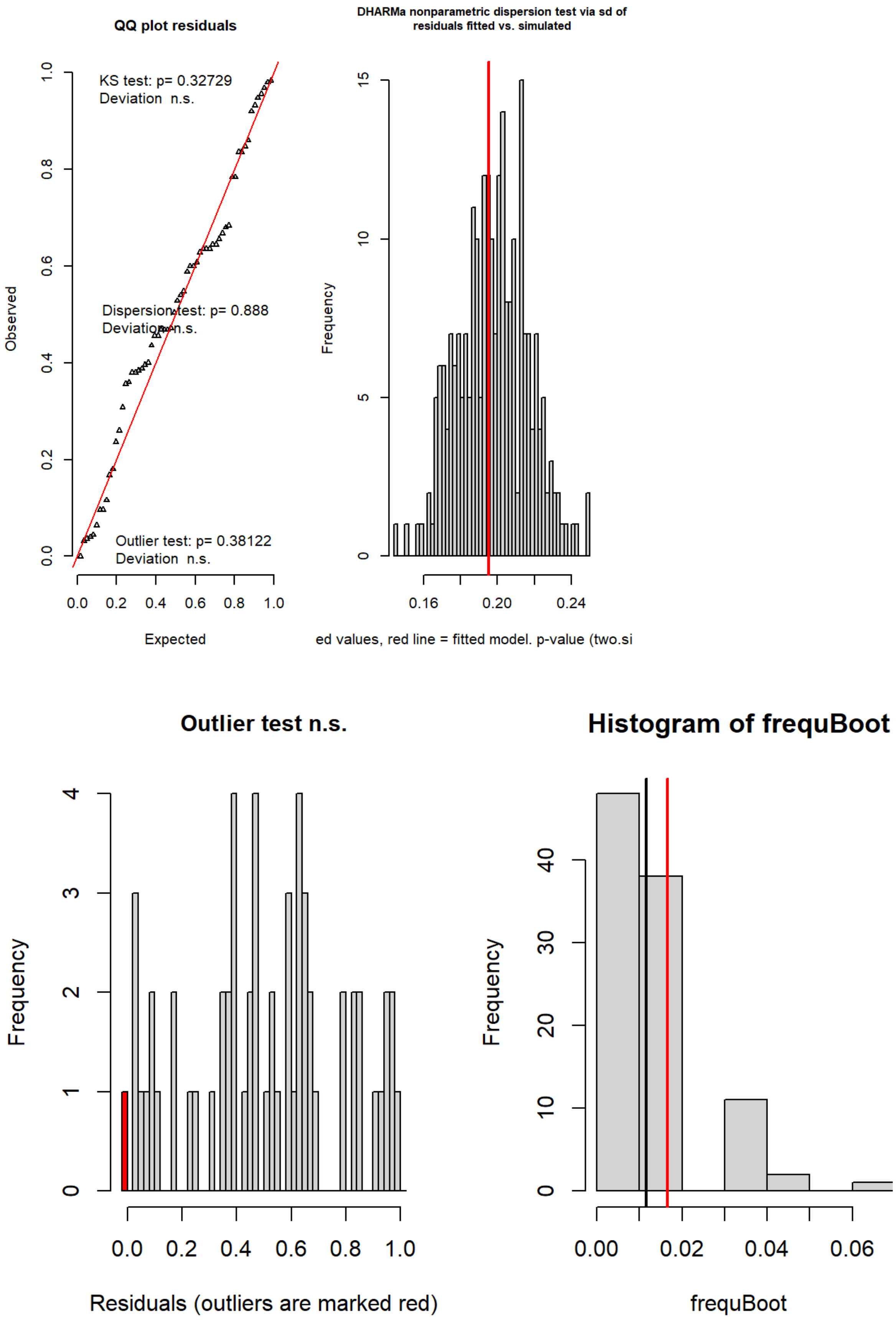
```
m11<-update(m1,.~.-TerritoryType:SL)

anova(m1,m11,test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: HSI ~ TerritoryType * SL
## Model 2: HSI ~ TerritoryType + SL
##   Resid. Df Resid. Dev Df  Deviance      F Pr(>F)
## 1        56     2.2000
## 2        57     2.2528 -1 -0.052842 1.3451 0.2511
```

No effect of interaction, so removed

Test model assumptions

```
m1<-glm(HSI~TerritoryType+SL,data=BodyCondition)

testResiduals(m1)
```



```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.12267, p-value = 0.3273
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.98718, p-value = 0.888
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04208333
## sample estimates:
## outlier frequency (expected: 0.0116666666666667 )
##                                         0.01666667
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.12267, p-value = 0.3273
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 0.98718, p-value = 0.888
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa bootstrapped outlier test
##
## data:  simulationOutput
## outliers at both margin(s) = 1, observations = 60, p-value = 1
## alternative hypothesis: two.sided
##  percent confidence interval:
##  0.00000000 0.04208333
## sample estimates:
## outlier frequency (expected: 0.0116666666666667 )
##                                        0.01666667
```

Model assumptions met

Test effect of farm type

```
m11<-update(m1,.~.-TerritoryType)

anova(m1,m11,test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: HSI ~ TerritoryType + SL
## Model 2: HSI ~ SL
##   Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
## 1        57     2.2528
## 2        58     5.5007 -1  -3.2479 82.177 1.209e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: There is a difference (i.e. significant effect of farm type on damselfish HSI)

Test effect of fish SL

```
m11<-update(m1,.~.-SL)

anova(m1,m11,test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: HSI ~ TerritoryType + SL
## Model 2: HSI ~ TerritoryType
##   Resid. Df Resid. Dev Df  Deviance      F Pr(>F)
## 1        57     2.2528
## 2        58     2.3307 -1 -0.077926 1.9717 0.1657
```

Conclusion: There is no difference (i.e. no effect of fish SL on HSI)

9. Estimates of mysid swarm density

Description of data

Surveys to determine the average size and density of swarms associated with longfin damselfish territories.

The dataset (CBC2018_SwarmCount_data.csv) contains the following variables (L-R):

- Territory: farm number (1-30).
- Elipsoid_vol: estimated volume of water occupied by swarm in cm3.
- Swarm_count: total number of mysids collected.
- Mysids_L: estimated number of mysids per L based on swarm count and ellipsoid volume.

Analysis

NA, dataset used to generate means and standard errors in manuscript.

10. Mysid waste excretion and nutrient availability

Description of data

An aquarium experiment examining rates of nutrient excretion by mysids.

The dataset (CBC2018_ExcretionExperiment_data.csv) contains the following variables (L-R):

- Date: date of trial.
- Replicate: number of replicate (1-30).
- Mysids_L: number of mysids in container ((0) 0 mysids, (100) 100 mysids, (200) 200 mysids)
- P_mgL: concentration of phosphorous (P) in mg/L recorded at end of trial.
- NH3N_mgL: concentration of nitrogen-ammonia (NH3-N) in mg/L recorded at end of trial.
- pH: pH of water in container at end of trial.
- Temp: temperature of water in Celsius in container at end of trial.
- Salinity_ppt: salinity of water in ppt in container at end of trial.

Analysis

NA, dataset used to generate means and standard errors in manuscript.

End of R Markdown.