R Notebook

Exploring the newly emerging relationships between native seagrasses and non-native juvenile clams

Cristina Galván, Araceli Puente

August 01, 2023

1. Loading R libraries

```
library(readxl) # Import excel files into R.
library(dplyr) # A tool for working with data frames.
library(plyr) # A tool to combine data.
library(car) # Regression analysis.
library(ggplot2) # Create data visualizations.
library(ggpubr) # Create data visualizations.
library(sjPlot) # Visualization of model residuals
library(gridExtra) # Arrange multiple graphs on the same figure
library(lmerTest) # Tests in Linear Mixed Effects Model
library(emmeans) # Realize pairwise comparisons
library(multcomp) # Add letters display for pairwise comparison
```

2. Loading and checking the data

```
field_data <- read_excel("RawData_FieldExp_Clam.xlsx")
str(field_data)
```

```
## tibble [408 x 9] (S3: tbl_df/tbl/data.frame)
##  $ PLOT             : num [1:408] 1 1 1 1 1 1 1 1 1 1 ...
##  $ REPLICATE        : num [1:408] 1 1 1 1 1 1 1 1 1 1 ...
##  $ SEAGRASS_COVERAGE: num [1:408] 0 0 0 0 0 0 0 0 0 0 ...
##  $ PREDATION        : num [1:408] 1 1 0 0 0 0 0 0 0 0 ...
##  $ ID_SURVIVAL_CLAM : num [1:408] 1 2 1 2 3 4 5 6 7 8 ...
##  $ LENGTH_mm        : num [1:408] 12.6 11.7 14.5 14.7 12.9 ...
##  $ SHELL_WEIGHT_g   : num [1:408] 0.226 0.164 0.277 0.309 0.198 0.281 0.295 0.201 0.24 0.274 ...
##  $ FLESH_WEIGHT_g   : num [1:408] 0.021 0.015 0.028 0.026 0.015 0.028 0.024 0.018 0.017 0.024 ...
##  $ CONDITION_INDEX  : num [1:408] 9.29 9.15 10.11 8.41 7.58 ...
```

3. Defining factor levels

Two factor levels are considered:

1. Seagrass coverage

- Low coverage: < 1%.
- Moderate coverage: 40 - 60%.
- High coverage: > 90%.

2. Predation from above

- No: restricted access of predators from above. With protective mesh.
- Yes: free access of predators from above. Without protective mesh.

```
field_data$SEAGRASS_COVERAGE <- factor(field_data$SEAGRASS_COVERAGE,
                                    levels = c(0, 40, 90),
                                    labels = c("Low", "Medium", "High"))
field_data$PREDATION <- factor(field_data$PREDATION,
                                  levels = c(0, 1),
                                  labels = c("No", "Yes"))
str(field_data)
```

```
## tibble [408 x 9] (S3: tbl_df/tbl/data.frame)
##  $ PLOT             : num [1:408] 1 1 1 1 1 1 1 1 1 1 ...
##  $ REPLICATE        : num [1:408] 1 1 1 1 1 1 1 1 1 1 ...
##  $ SEAGRASS_COVERAGE: Factor w/ 3 levels "Low","Medium",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ PREDATION        : Factor w/ 2 levels "No","Yes": 2 2 1 1 1 1 1 1 1 1 ...
##  $ ID_SURVIVAL_CLAM : num [1:408] 1 2 1 2 3 4 5 6 7 8 ...
##  $ LENGTH_mm        : num [1:408] 12.6 11.7 14.5 14.7 12.9 ...
##  $ SHELL_WEIGHT_g   : num [1:408] 0.226 0.164 0.277 0.309 0.198 0.281 0.295 0.201 0.24 0.274 ...
##  $ FLESH_WEIGHT_g   : num [1:408] 0.021 0.015 0.028 0.026 0.015 0.028 0.024 0.018 0.017 0.024 ...
##  $ CONDITION_INDEX  : num [1:408] 9.29 9.15 10.11 8.41 7.58 ...
```

4. Preparing data

Three response variables are considered for further analysis: Shell Length (‘LENGTH_mm’), Condition Index (‘CONDITION_INDEX’) and Instantaneous Survival Rate (‘SURVIVAL_RATE’). Mean values of shell lenght and condition index were calculated at the subplot level. The monthly instantaneous survival rate, which is not included in the dataset uploaded above, is calculated as the logarithm of the proportion of alive clam individuals collected at the end of the experiment. The initial number of clams was 30 in each plot.

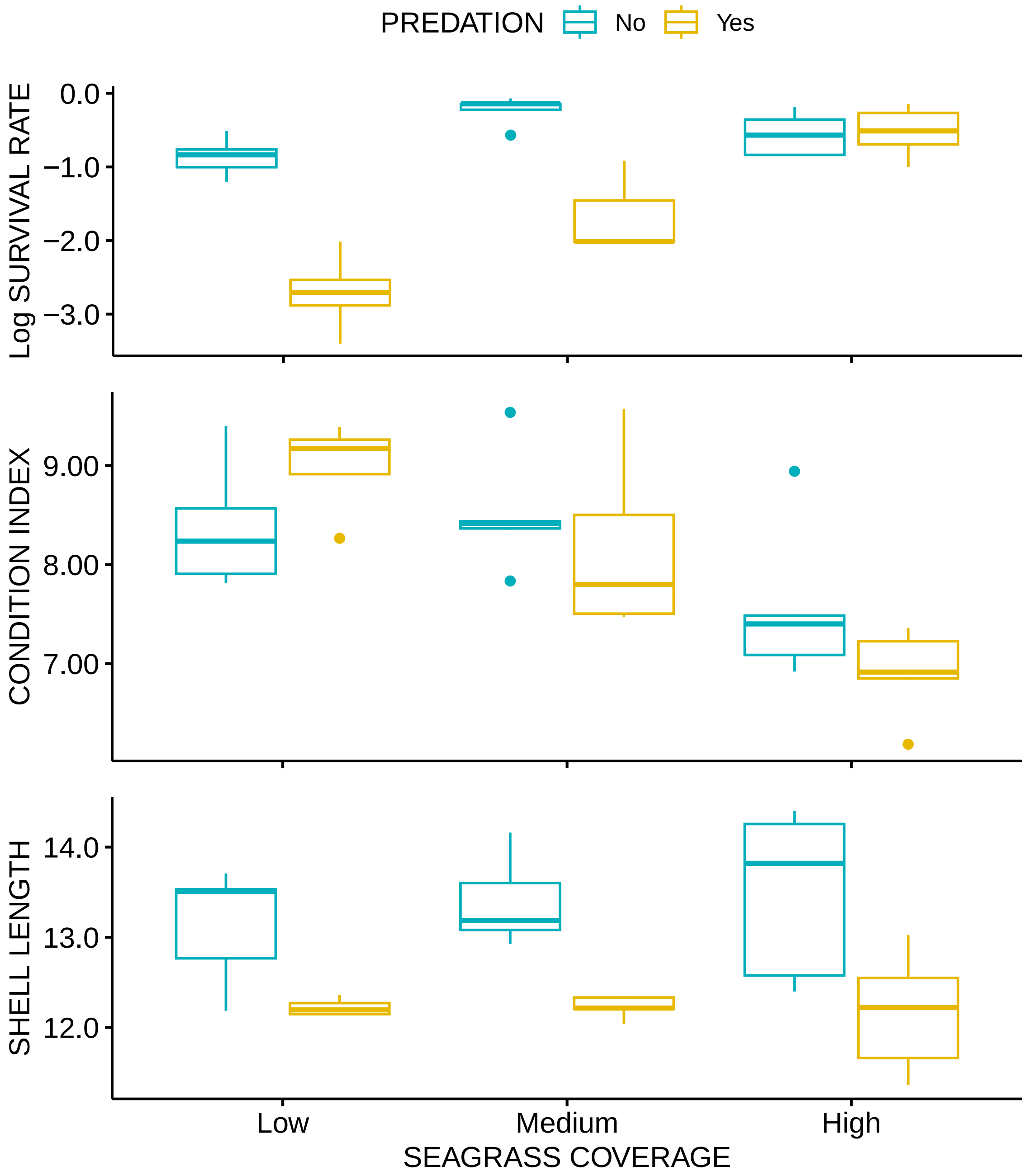
```
data_table <- field_data %>%
  group_by(PLOT, SEAGRASS_COVERAGE, PREDATION) %>%
  dplyr::summarise(LENGTH_mm = mean(LENGTH_mm),
                   CONDITION_INDEX = mean(CONDITION_INDEX, na.rm = T),
                   SURVIVAL_RATE = log(n()/30)) # Instantaneous survival rate

head(data_table)
```

```
## # A tibble: 6 x 6
## # Groups:   PLOT, SEAGRASS_COVERAGE [3]
##    PLOT SEAGRASS_COVERAGE PREDATION LENGTH_mm CONDITION_INDEX SURVIVAL_RATE
##   <dbl> <fct>             <fct>         <dbl>           <dbl>         <dbl>
## 1     1 Low               No             13.5            8.24       -0.762
## 2     1 Low               Yes            12.1            9.22       -2.71
## 3     2 Medium            No             13.1            8.44       -0.143
## 4     2 Medium            Yes            12.3            9.57       -2.01
## 5     3 Medium            No             13.2            7.83       -0.0690
## 6     3 Medium            Yes            12.3            7.47       -1.46
```

5. Performing exploratory data analysis

5.1. Normality and homogeneity of variance

```
f1 <- ggboxplot(data_table, x = "SEAGRASS_COVERAGE", y = "SURVIVAL_RATE",
                 color = "PREDATION", palette = c("#00AFBB", "#E7B800"),
                 ylab = 'Log SURVIVAL RATE', xlab = FALSE) +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.1)) +theme(axis.text.x=element_blank()
f2 <- ggboxplot(data_table, x = "SEAGRASS_COVERAGE", y = "CONDITION_INDEX",
                 color = "PREDATION", palette = c("#00AFBB", "#E7B800"),
                 ylab = 'CONDITION INDEX', xlab = FALSE) +
  theme(legend.position = "none") +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.01)) +
  theme(axis.text.x=element_blank())
f3 <- ggboxplot(data_table, x = "SEAGRASS_COVERAGE", y = "LENGTH_mm",
                 color = "PREDATION", palette = c("#00AFBB", "#E7B800"),
                 xlab = 'SEAGRASS COVERAGE', ylab = 'SHELL LENGTH') +
  theme(legend.position = "none") +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.1))
grid.arrange(f1,f2,f3)
```



There may be evidences that the Condition Index is non-normal across populations (symmetry of each boxplot). There is no evidence that the variance of survival, condition and length of clams differs between the populations (height of the boxplots), and there is no evidence of a relationship between mean and variance.

6. Model fitting

A split plot design was applied. The levels of the main treatment factor (Seagrass Coverage) are grouped in blocks (Plot). Within each plot, two subplots randomly distributed were randomly assigned to different treatments of a second factor (Predation). Linear mixed-effect models were adjusted to consider fixed

(Seagrass coverage and Predation) and random (Plot as blocking factor) effects on Instantaneous Survival Rate, Condition Index and Shell Length.

6.1. Selection of the random structure of the model

The optimal random structure of the model is selected. For that, a simple random intercepts model and a random intercepts and slope model were fitted and explored which model fits the data best.

```
## [1] "INSTANTANEOUS SURVIVAL RATE"
```

```
mod_lmer_sr <- lmerTest::lmer(SURVIVAL_RATE ~ SEAGRASS_COVERAGE * PREDATION
                              + (1 | PLOT), # random intercept
                              data_table, REML = TRUE)
mod_lmer_sr1 <- lmerTest::lmer(SURVIVAL_RATE ~ SEAGRASS_COVERAGE * PREDATION
                               + (SEAGRASS_COVERAGE | PLOT), # random intercept/slope
                               data_table, REML = TRUE,
                               control = lmerControl(check.nobs.vs.nRE = "ignore"))
anova(mod_lmer_sr, mod_lmer_sr1)
```

```
## Data: data_table
## Models:
## mod_lmer_sr: SURVIVAL_RATE ~ SEAGRASS_COVERAGE * PREDATION + (1 | PLOT)
## mod_lmer_sr1: SURVIVAL_RATE ~ SEAGRASS_COVERAGE * PREDATION + (SEAGRASS_COVERAGE | PLOT)
##              npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod_lmer_sr     8 34.051 44.990 -9.0257   18.052
## mod_lmer_sr1   13 43.932 61.707 -8.9661   17.932 0.1192  5     0.9997
```

```
## [1] "CONDITION INDEX"
```

```
mod_lmer_ci <- lmerTest::lmer(CONDITION_INDEX ~ SEAGRASS_COVERAGE * PREDATION
                              + (1 | PLOT), # random intercept
                              data_table, REML = TRUE)
mod_lmer_ci1 <- lmerTest::lmer(CONDITION_INDEX ~ SEAGRASS_COVERAGE * PREDATION
                               + (SEAGRASS_COVERAGE | PLOT), # random intercept/slope
                               data_table, REML = TRUE,
                               control = lmerControl(check.nobs.vs.nRE = "ignore"))
anova(mod_lmer_ci, mod_lmer_ci1)
```

```
## Data: data_table
## Models:
## mod_lmer_ci: CONDITION_INDEX ~ SEAGRASS_COVERAGE * PREDATION + (1 | PLOT)
## mod_lmer_ci1: CONDITION_INDEX ~ SEAGRASS_COVERAGE * PREDATION + (SEAGRASS_COVERAGE | PLOT)
##              npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod_lmer_ci     8 66.195 77.133 -25.097   50.195
## mod_lmer_ci1   13 75.945 93.720 -24.973   49.945 0.2492  5     0.9985
```

```
## [1] "SHELL LENGTH"
```

```
mod_lmer_le <- lmerTest::lmer(LENGTH_mm ~ SEAGRASS_COVERAGE * PREDATION
                              + (1 | PLOT), # random intercept
                              data_table, REML = TRUE)
mod_lmer_le1 <- lmerTest::lmer(LENGTH_mm ~ SEAGRASS_COVERAGE * PREDATION
                               + (SEAGRASS_COVERAGE | PLOT), # random intercept/slope
                               data_table, REML = TRUE,
                               control = lmerControl(check.nobs.vs.nRE = "ignore"))
anova(mod_lmer_le, mod_lmer_le1)
```

```
## Data: data_table
## Models:
## mod_lmer_le: LENGTH_mm ~ SEAGRASS_COVERAGE * PREDATION + (1 | PLOT)
## mod_lmer_le1: LENGTH_mm ~ SEAGRASS_COVERAGE * PREDATION + (SEAGRASS_COVERAGE | PLOT)
##              npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod_lmer_le     8 55.051 65.989 -19.525   39.051
## mod_lmer_le1   13 58.758 76.533 -16.379   32.758 6.2929  5     0.2787
```

Conclusion for all response variables: the more complex random intercepts and slopes model does not fit the data significantly better than the simpler random intercepts model, and thus the simpler model will be used.

6.2. Parameter estimates of the fixed structure

The parameters of the fixed structure of the models for the selected random structure are estimated.

```
## [1] "INSTANTANEOUS SURVIVAL RATE"
```

```
summary(mod_lmer_sr, ddf="Kenward-Roger")
```

```
## Linear mixed model fit by REML. t-tests use Kenward-Roger's method [
## lmerModLmerTest]
## Formula: SURVIVAL_RATE ~ SEAGRASS_COVERAGE * PREDATION + (1 | PLOT)
##    Data: data_table
##
## REML criterion at convergence: 29.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8688 -0.7559  0.0165  0.5371  2.0678
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  PLOT     (Intercept) 0.0000   0.0000
##  Residual             0.1376   0.3709
## Number of obs: 29, groups:  PLOT, 15
##
## Fixed effects:
##                                     Estimate Std. Error      df t value
## (Intercept)                          -0.8633     0.1659 23.0000  -5.205
## SEAGRASS_COVERAGEMedium               0.6340     0.2346 23.0000   2.703
## SEAGRASS_COVERAGEHigh                 0.3074     0.2346 23.0000   1.310
## PREDATIONYes                         -1.8448     0.2513 12.5956  -7.340
```

```
## SEAGRASS_COVERAGEMedium:PREDATIONYes   0.3908     0.3438 11.9414   1.137
## SEAGRASS_COVERAGEHigh:PREDATIONYes     1.8774     0.3438 11.9414   5.461
##                                      Pr(>|t|)
## (Intercept)                          2.81e-05 ***
## SEAGRASS_COVERAGEMedium              0.012697 *
## SEAGRASS_COVERAGEHigh                0.202987
## PREDATIONYes                         6.80e-06 ***
## SEAGRASS_COVERAGEMedium:PREDATIONYes 0.277982
## SEAGRASS_COVERAGEHigh:PREDATIONYes   0.000148 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Correlation of Fixed Effects:
##                     (Intr) SEAGRASS_COVERAGEMd SEAGRASS_COVERAGEHg PREDAT
## SEAGRASS_COVERAGEMd -0.707
## SEAGRASS_COVERAGEHg -0.707  0.500
## PREDATIONYs         -0.667  0.471               0.471
## SEAGRASS_COVERAGEM:  0.485 -0.686              -0.343              -0.728
## SEAGRASS_COVERAGEH:  0.485 -0.343              -0.686              -0.728
##                     SEAGRASS_COVERAGEM:
## SEAGRASS_COVERAGEMd
## SEAGRASS_COVERAGEHg
## PREDATIONYs
## SEAGRASS_COVERAGEM:
## SEAGRASS_COVERAGEH:  0.529
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help(’isSingular’)
```

```
## [1] "CONDITION INDEX"
```

```
summary(mod_lmer_ci, ddf="Kenward-Roger")
```

```
## Linear mixed model fit by REML. t-tests use Kenward-Roger’s method [
## lmerModLmerTest]
## Formula: CONDITION_INDEX ~ SEAGRASS_COVERAGE * PREDATION + (1 | PLOT)
##    Data: data_table
##
## REML criterion at convergence: 54.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9059 -0.5397 -0.2472  0.3768  1.9658
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  PLOT     (Intercept) 0.1892   0.435
##  Residual             0.2673   0.517
## Number of obs: 29, groups:  PLOT, 15
##
## Fixed effects:
##                                      Estimate Std. Error      df t value
## (Intercept)                            8.3856     0.3022 19.7741  27.753
## SEAGRASS_COVERAGEMedium                0.1331     0.4273 19.7741   0.312
```

```
## SEAGRASS_COVERAGEHigh                 -0.8183     0.4273 19.7741  -1.915
## PREDATIONYes                           0.5571     0.3587 12.1205   1.553
## SEAGRASS_COVERAGEMedium:PREDATIONYes  -0.9054     0.4853 11.6320  -1.865
## SEAGRASS_COVERAGEHigh:PREDATIONYes    -1.2170     0.4853 11.6320  -2.508
##                                      Pr(>|t|)
## (Intercept)                            <2e-16 ***
## SEAGRASS_COVERAGEMedium                0.7586
## SEAGRASS_COVERAGEHigh                  0.0701 .
## PREDATIONYes                           0.1460
## SEAGRASS_COVERAGEMedium:PREDATIONYes   0.0875 .
## SEAGRASS_COVERAGEHigh:PREDATIONYes     0.0281 *
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Correlation of Fixed Effects:
##                     (Intr) SEAGRASS_COVERAGEMd SEAGRASS_COVERAGEHg PREDAT
## SEAGRASS_COVERAGEMd -0.707
## SEAGRASS_COVERAGEHg -0.707  0.500
## PREDATIONYs         -0.499  0.353               0.353
## SEAGRASS_COVERAGEM:  0.367 -0.519              -0.259              -0.735
## SEAGRASS_COVERAGEH:  0.367 -0.259              -0.519              -0.735
##                     SEAGRASS_COVERAGEM:
## SEAGRASS_COVERAGEMd
## SEAGRASS_COVERAGEHg
## PREDATIONYs
## SEAGRASS_COVERAGEM:
## SEAGRASS_COVERAGEH:  0.541
```

```
## [1] "SHELL LENGTH"
```

```
summary(mod_lmer_le, ddf="Kenward-Roger")
```

```
## Linear mixed model fit by REML. t-tests use Kenward-Roger’s method [
## lmerModLmerTest]
## Formula: LENGTH_mm ~ SEAGRASS_COVERAGE * PREDATION + (1 | PLOT)
##    Data: data_table
##
## REML criterion at convergence: 45.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4318 -0.6216  0.1503  0.5202  1.2951
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  PLOT     (Intercept) 0.2123   0.4608
##  Residual             0.1404   0.3746
## Number of obs: 29, groups:  PLOT, 15
##
## Fixed effects:
##                                      Estimate Std. Error      df t value
## (Intercept)                           13.1400     0.2656 17.1473  49.475
## SEAGRASS_COVERAGEMedium                0.2511     0.3756 17.1473   0.669
```

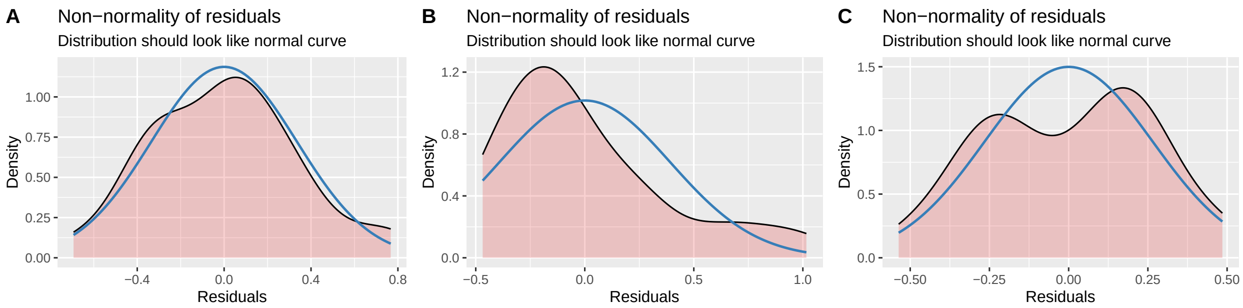
```
## SEAGRASS_COVERAGEHigh                  0.3508     0.3756 17.1473   0.934
## PREDATIONYes                          -0.8319     0.2620 11.8063  -3.175
## SEAGRASS_COVERAGEMedium:PREDATIONYes  -0.3321     0.3533 11.4504  -0.940
## SEAGRASS_COVERAGEHigh:PREDATIONYes    -0.4957     0.3533 11.4504  -1.403
##                                      Pr(>|t|)
## (Intercept)                           < 2e-16 ***
## SEAGRASS_COVERAGEMedium               0.51269
## SEAGRASS_COVERAGEHigh                 0.36325
## PREDATIONYes                          0.00815 **
## SEAGRASS_COVERAGEMedium:PREDATIONYes  0.36655
## SEAGRASS_COVERAGEHigh:PREDATIONYes    0.18714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##                     (Intr) SEAGRASS_COVERAGEMd SEAGRASS_COVERAGEHg PREDAT
## SEAGRASS_COVERAGEMd -0.707
## SEAGRASS_COVERAGEHg -0.707  0.500
## PREDATIONYs         -0.407  0.288               0.288
## SEAGRASS_COVERAGEM:  0.301 -0.425              -0.213              -0.739
## SEAGRASS_COVERAGEH:  0.301 -0.213              -0.425              -0.739
##                     SEAGRASS_COVERAGEM:
## SEAGRASS_COVERAGEMd
## SEAGRASS_COVERAGEHg
## PREDATIONYs
## SEAGRASS_COVERAGEM:
## SEAGRASS_COVERAGEH:  0.546
```

7. Checking the assumptions of the model

7.1. Normality in residuals

Visual inspection of normality in residuals for Instantaneous Survival Rate (A), Condition Index (B) and Shell Lenght (B)

```r
norm_sr <- plot_model(mod_lmer_sr, type = "diag", title = "")
norm_ci <- plot_model(mod_lmer_ci, type = "diag")
norm_le <- plot_model(mod_lmer_le, type = "diag")
ggarrange(norm_sr[[3]], norm_ci[[3]], norm_le[[3]],
          labels = c("A", "B", "C"),
          ncol=3, nrow=1)
```



The normality is checked using the Shapiro-Wilk test on the model residuals.

```
## [1] "INSTANTANEOUS SURVIVAL RATE"
```

```r
shapiro.test(residuals(mod_lmer_sr))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(mod_lmer_sr)
## W = 0.97189, p-value = 0.6121
```

```
## [1] "CONDITION INDEX"
```

```r
shapiro.test(residuals(mod_lmer_ci))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(mod_lmer_ci)
## W = 0.88714, p-value = 0.004888
```

```
## [1] "SHELL LENGTH"
```

```r
shapiro.test(residuals(mod_lmer_le))
```

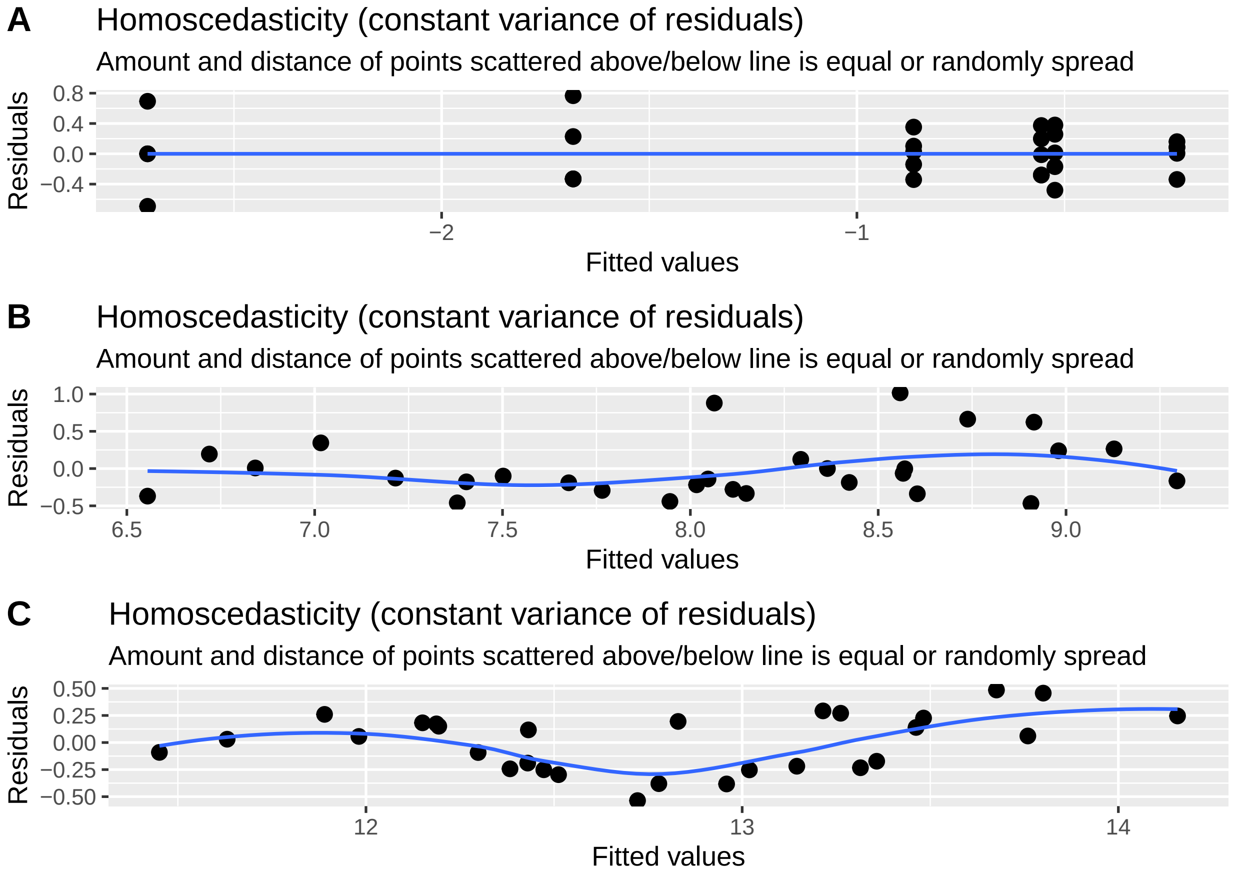
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(mod_lmer_le)
## W = 0.96107, p-value = 0.3491
```

From the output above we can see that the p-value is less than the significance level of 0.05 for the Condition Index. Therefore, we cannot assume the normality in the residuals.

7.2. Homogeneity of variances in residuals

Visual inspection of homocedasticity in residuals for Instantaneous Survival Rate (A), Condition Index (B) and Shell Lenght (B)

```r
ggarrange(norm_sr[[4]], norm_ci[[4]], norm_le[[4]],
          labels = c("A", "B", "C"),
          ncol=1, nrow=3)
```



The homogeneity of variances is checked using the Levene's test.

```
## [1] "INSTANTANEOUS SURVIVAL RATE"
```

```
leveneTest(residuals(mod_lmer_sr) ~ data_table$SEAGRASS_COVERAGE * data_table$PREDATION)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.4394 0.8164
##       23
```

```
## [1] "CONDITION INDEX"
```

```
leveneTest(residuals(mod_lmer_ci) ~ data_table$SEAGRASS_COVERAGE * data_table$PREDATION)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.0884 0.9933
##       23
```

```
## [1] "SHELL LENGTH"
```

```r
leveneTest(residuals(mod_lmer_le) ~ data_table$SEAGRASS_COVERAGE * data_table$PREDATION)
```

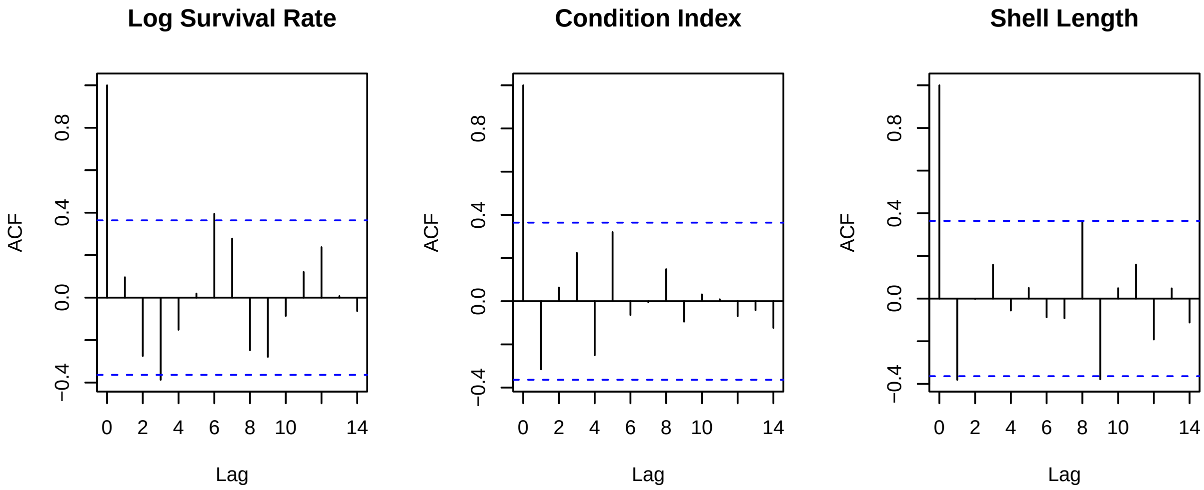
```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.4738 0.7919
##       23
```

From the output above we can assume the homogeneity of variances in the different treatment groups.

7.3. Independence of residuals

The spatial autocorrelation is calculated to assess the independence of residuals.

```r
par(mfrow=c(1,3))

acf_sr <- acf(resid(mod_lmer_sr), plot = FALSE)
plot(acf_sr, main = "Log Survival Rate")

acf_ci <- acf(resid(mod_lmer_ci), plot = FALSE)
plot(acf_ci, main = "Condition Index")

acf_le <- acf(resid(mod_lmer_le), plot = FALSE)
plot(acf_le, main = "Shell Length")
```



The Autocorrelation Factor (ACF) indicates that there is not a strong pattern of a contagious structure running through the residuals.

8. Transforming data to fullfill assumptions

Condition Index does not accomplish the assumption of normality of residuals. Therefore, data was transformed and normality and homocedasticity was checked.

8.1. Data transformation

```
bcfun <- function(lambda, resp = "CONDITION_INDEX") {
  y <- data_table[[resp]]
  data_table$newy <- if (lambda==0) log(y) else (y^lambda -1)/lambda
  log_jac <- sum((lambda-1)*log(y))
  newfit <- update(mod_lmer_ci, newy ~ SEAGRASS_COVERAGE * PREDATION +
                   (1 | PLOT), data = data_table)
  return(-2*(c(logLik(newfit))+ log_jac))
}
lambdavec <- seq(-2, 2, by = 0.2)
boxcox <- vapply(lambdavec, bcfun, FUN.VALUE = numeric(1))

LV<-lambdavec[1][as.numeric(which.max(boxcox))]
data_table$ci <- data_table$CONDITION_INDEX^LV-1/LV
```

8.3. Model fitting to transformed data

```
mod_lmer_ci2 <- lmerTest::lmer(ci ~ SEAGRASS_COVERAGE * PREDATION + (1 | PLOT),
                    data_table, REML = TRUE)
```

8.2. Checking the assumptions

```
print("NORMALITY")
```

```
## [1] "NORMALITY"
```

```
shapiro.test(residuals(mod_lmer_ci2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(mod_lmer_ci2)
## W = 0.96621, p-value = 0.462
```

```
print("HOMOCEDASTICITY")
```

```
## [1] "HOMOCEDASTICITY"
```

```
leveneTest(residuals(mod_lmer_ci2) ~ data_table$SEAGRASS_COVERAGE * data_table$PREDATION)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.2844 0.9168
##       23
```

Normality, homocedasticity and independence of residuals is checked for the transformed data of Condicion Index.

9. INSTANTANEOUS SURVIVAL RATE

9.1. Model evaluation

An F-test based on the Kenward-Roger approach is computed to identify significant factors in the model of Instantaneous Survival Rate.

```
anova(mod_lmer_sr, type = 3, ddf="Kenward-Roger")
```

```
## Type III Analysis of Variance Table with Kenward-Roger's method
##                             Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)    
## SEAGRASS_COVERAGE           7.4042  3.7021     2 11.690  26.909 4.219e-05 ***
## PREDATION                   8.4721  8.4721     1 11.714  61.584 5.320e-06 ***
## SEAGRASS_COVERAGE:PREDATION 4.7142  2.3571     2 11.690  17.132  0.000335 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The observed p-values are less than 0.05 in all cases, indicating that there is enough evidence to conclude that the levels of predation and seagrass coverage are associated with significant different instantaneous survival rates. The interaction between both factors is also significant, which indicates that the relationship between seagrass coverage and instantaneous survival rate depends on the predation.

9.2. Multiple pairwise comparisons

As the F-test is significant, we compute the post-hoc multiple pairwise-comparisons test to determine significant differences between specific treatments.

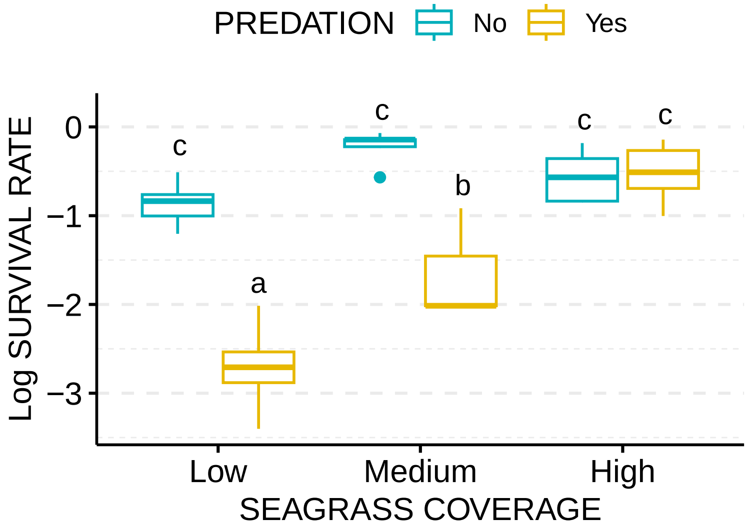
```
sr_posthoc <- mod_lmer_sr %>%
  emmeans(specs = ~ SEAGRASS_COVERAGE * PREDATION,
          lmer.df = "kenward-roger") %>%
  cld(adjust="sidak", Letters=letters)

sr_posthoc
```

```
##  SEAGRASS_COVERAGE PREDATION emmean    SE df lower.CL upper.CL .group
##  Low               Yes       -2.708 0.189 23   -3.251  -2.1649  a    
##  Medium            Yes       -1.683 0.166 23   -2.160  -1.2060   b   
##  Low               No        -0.863 0.166 23   -1.341  -0.3861    c  
##  High              No        -0.556 0.166 23   -1.033  -0.0787    c  
##  High              Yes       -0.523 0.166 23   -1.000  -0.0460    c  
##  Medium            No        -0.229 0.166 23   -0.707   0.2480    c  
## 
## Degrees-of-freedom method: kenward-roger 
## Confidence level used: 0.95 
## Conf-level adjustment: sidak method for 6 estimates 
## P value adjustment: sidak method for 15 tests 
## significance level used: alpha = 0.05 
## NOTE: If two or more means share the same grouping symbol,
##       then we cannot show them to be different.
##       But we also did not show them to be the same.
```

9.3. Results visualization

The compact letter display from post-hoc test was added to the box-plot of the ‘Instantaneous Survival Rate’ data grouped by the combinations of the levels of the two factors (‘Seagrass coverage’ and ‘Predation’)

```
p <- ggboxplot(data_table, x = "SEAGRASS_COVERAGE", y = "SURVIVAL_RATE",
               color = "PREDATION", palette = c("#00AFBB", "#E7B800"),
               xlab = 'SEAGRASS COVERAGE', ylab = 'Log SURVIVAL RATE')

p + annotate("text", x = 0.75, y = -0.2, label = sr_posthoc$.group[3]) +
  annotate("text", x = 1.22, y = -1.75, label = sr_posthoc$.group[1])+
  annotate("text", x = 1.75, y = 0.2, label = sr_posthoc$.group[6])+
  annotate("text", x = 2.19, y = -0.65, label = sr_posthoc$.group[2])+
  annotate("text", x = 2.75, y = 0.09, label = sr_posthoc$.group[4])+
  annotate("text", x = 3.15, y = 0.15, label = sr_posthoc$.group[5])+
  grids(linetype = "dashed", axis = "y")
```



10. CONDITION INDEX

10.1. Model evaluation

An F-test based on the Kenward-Roger approach is computed to identify significant factors in the model of Condition Index.

```
anova(mod_lmer_ci2, type = 3, ddf="Kenward-Roger")
```

```
## Type III Analysis of Variance Table with Kenward-Roger's method
##                                Sum Sq    Mean Sq NumDF  DenDF F value   Pr(>F)
## SEAGRASS_COVERAGE           6.7498e-05 3.3749e-05     2 11.912 10.1441 0.002678
## PREDATION                   7.0290e-06 7.0290e-06     1 11.399  2.1129 0.173033
## SEAGRASS_COVERAGE:PREDATION 2.9033e-05 1.4516e-05     2 11.377  4.3633 0.039222
##
```

```
## SEAGRASS_COVERAGE           **
## PREDATION
## SEAGRASS_COVERAGE:PREDATION *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The observed p-values are is less than 0.05 for seagrass coverage and the interaction between both predation and seagrass coverage factors, indicating that there are statistically significant differences between corresponding treatments.

10.2. Multiple pairwise comparisons

As the F-test is significant, we compute the post-hoc multiple pairwise-comparisons test to determine significant differences between specific treatments.

```
ci_posthoc <- mod_lmer_ci2 %>%
  emmeans(specs = ~ SEAGRASS_COVERAGE * PREDATION,
          lmer.df = "kenward-roger") %>%
  cld(adjust="sidak", Letters=letters)

ci_posthoc
```

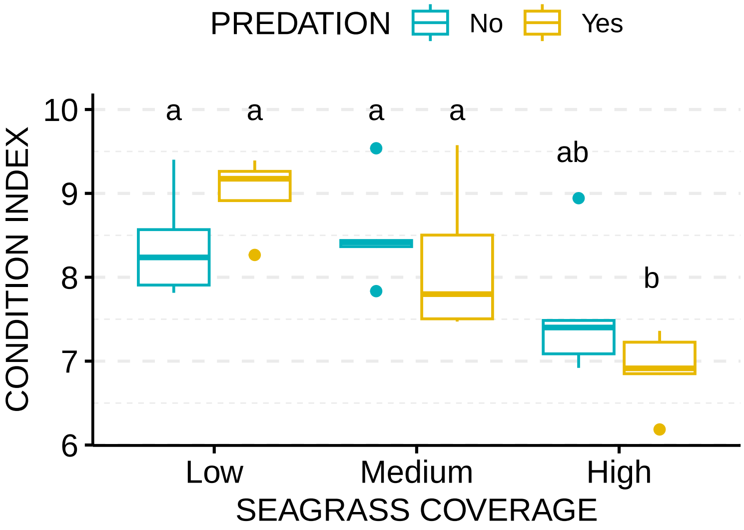
```
##  SEAGRASS_COVERAGE PREDATION emmean       SE    df lower.CL upper.CL .group
##  Low               Yes       0.5127 0.001264 20.85   0.5090   0.5163  a
##  Medium            No        0.5139 0.001147 18.68   0.5106   0.5173  a
##  Low               No        0.5144 0.001147 18.68   0.5110   0.5178  a
##  Medium            Yes       0.5154 0.001147 18.68   0.5120   0.5187  a
##  High              No        0.5179 0.001147 18.68   0.5145   0.5213  ab
##  High              Yes       0.5212 0.001147 18.68   0.5178   0.5246   b
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
## Conf-level adjustment: sidak method for 6 estimates
## P value adjustment: sidak method for 15 tests
## significance level used: alpha = 0.05
## NOTE: If two or more means share the same grouping symbol,
##       then we cannot show them to be different.
##       But we also did not show them to be the same.
```

10.3. Results visualization

The compact letter display from post-hoc test was added to the box-plot of the 'Condition Index' data grouped by the combinations of the levels of the two factors ('Seagrass coverage' and 'Predation')

```
p <- ggboxplot(data_table, x = "SEAGRASS_COVERAGE", y = "CONDITION_INDEX",
               color = "PREDATION", palette = c("#00AFBB", "#E7B800"),
               xlab = 'SEAGRASS COVERAGE', ylab = 'CONDITION INDEX')

p + annotate("text", x = 0.8, y = 10, label = ci_posthoc$.group[3]) +
  annotate("text", x = 1.2, y = 10, label = ci_posthoc$.group[1])+
  annotate("text", x = 1.8, y = 10, label = ci_posthoc$.group[2])+
```

```
  annotate("text", x = 2.2, y = 10, label = ci_posthoc$.group[4])+
  annotate("text", x = 2.75, y = 9.5, label = ci_posthoc$.group[5])+
  annotate("text", x = 3.12, y = 8, label = ci_posthoc$.group[6])+
  grids(linetype = "dashed", axis = "y")
```



11. SHELL LENGTH

11.1. Model evaluation

An F-test based on the Kenward-Roger approach is computed to identify significant factors in the model of Shell Length.

```
anova(mod_lmer_le, type = 3, ddf="Kenward-Roger")
```

```
## Type III Analysis of Variance Table with Kenward-Roger's method
##                              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## SEAGRASS_COVERAGE            0.0145  0.0073     2 11.944  0.0517    0.9498
## PREDATION                    8.5680  8.5680     1 11.320 61.0457 6.861e-06 ***
## SEAGRASS_COVERAGE:PREDATION 0.2820  0.1410     2 11.301  1.0045    0.3967
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results show that predation impacts significantly the mean shell length of clams.

11.2. Results visualization

The compact letter display from post-hoc test was added to the box-plot of the 'Shell Length' data grouped by the two levels of the factor 'Pedation'.

```
p <- ggboxplot(data_table, x = "PREDATION", y = "LENGTH_mm",
               color = "PREDATION", palette = c("#00AFBB", "#E7B800"),
               xlab = 'PREDATION', ylab = 'SHELL LENGTH')

p + annotate("text", x = 1, y = 15, label = "a") +
  annotate("text", x = 2, y = 13.5, label = "b") +
  grids(linetype = "dashed", axis = "y") +
  theme(legend.position="none")
```