

NSCI7915 Preliminary Analysis

2023-05-20

Hypothesis 1: The abundance of *H.perlevis* is influenced by month

Metadata: Nudibranch predation boosts sponge silicon cycling (2023)

Independent Variable (IV) = Month (Jun20, Nov20, Apr20, Jun21)

Dependent variable (D) = Abundance (n)

Null Hypothesis (H_0): There is no significant relationship between the abundance of *Hymeniacidon perlevis* and the month.

Alternative Hypothesis (H_A): There is a relationship between the abundance of *Hymeniacidon perlevis* and the month.

Due to the predator *Doris Verrucosa* and its peak abundance season being late June/early July we expect to see a relationship of June 2021 = June 2020 > April 2020 > November 2021 in regards to biomass.

Table 1: Statistical summary of data from the 2023 study

```
##      N.quadrat      N.H..perlevis      Volumen.Hym..mL.      Date
##  Min.       : 1.000    Min.       : 3.00    Min.       : 1.90    Length:71
##  1st Qu.: 5.000    1st Qu.:12.00    1st Qu.: 27.35    Class :character
##  Median : 9.000    Median :20.00    Median : 62.60    Mode  :character
##  Mean   : 9.789    Mean   :23.58    Mean   :120.08
##  3rd Qu.:14.000    3rd Qu.:31.00    3rd Qu.:167.85
##  Max.    :22.000    Max.    :73.00    Max.    :690.50
##
##      X              row_id
##  Mode:logical  1          : 1
##  NA's:71      2          : 1
##              3          : 1
##              4          : 1
##              5          : 1
##              6          : 1
##              (Other):65
```

Table 2: Standard deviation of abundance (n) and volume (mL) data from the 2023 study for hypothesis 1
§ 2

```
## [1] "Standard Deviation (N.H..perlevis): 16.8342015567212 Standard Deviation (Volumen.Hym..mL.): 141
```

Relationship between abundance (n) per quadrant and month

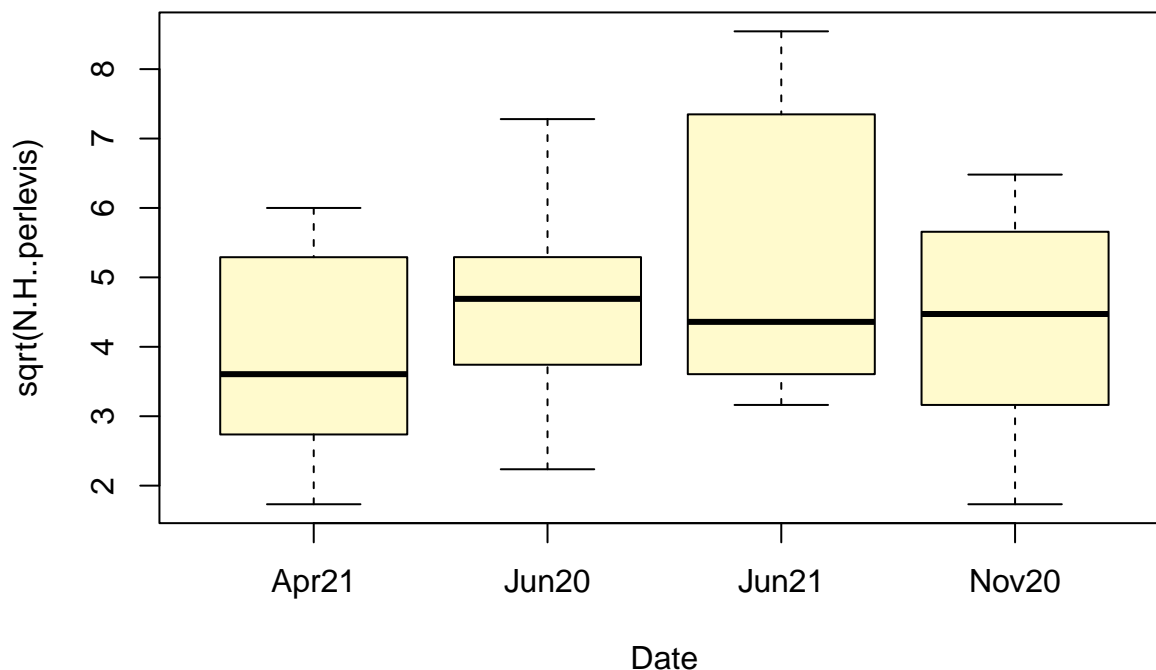


Figure 1: Boxplot of the relationship between abundance (n) per quadrant and the month the measurement was taken. (x-axis): The month the abundance (n) was measured and (y-axis): the square root of the abundance (n).

Table 3: Overdispersion function for glm model

```
##          chisq          ratio          rdf          p
## 7.146177e+02 1.066594e+01 6.700000e+01 4.572501e-109
```

Table 4: Overdispersion function for glmer model

```
##          chisq          ratio          rdf          p
## 9.7369477 0.1475295 66.0000000 1.0000000
```

Table 5: Chisq drop1 test results

```
## Single term deletions
##
## Model:
## N.H..perlevis ~ Date + (1 | row_id)
##          npar      AIC      LRT Pr(Chi)
## <none>      574.87
## Date       3 574.94 6.0749 0.108
```

Table 6: Generalised linear mixed effects model (GLMM) and posthoc tukey test

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: glmer(formula = N.H..perlevis ~ Date + (1 | row_id), data = prey_ab_vol,
## family = "poisson")
##
## Linear Hypotheses:
##           Estimate Std. Error z value Pr(>|z|)
## Jun20 - Apr21 == 0    0.4080    0.2295   1.778   0.2825
## Jun21 - Apr21 == 0    0.5396    0.2312   2.333   0.0899 .
## Nov20 - Apr21 == 0    0.1873    0.2610   0.718   0.8895
## Jun21 - Jun20 == 0    0.1316    0.2051   0.642   0.9179
## Nov20 - Jun20 == 0   -0.2207    0.2382  -0.927   0.7895
## Nov20 - Jun21 == 0   -0.3523    0.2399  -1.469   0.4547
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

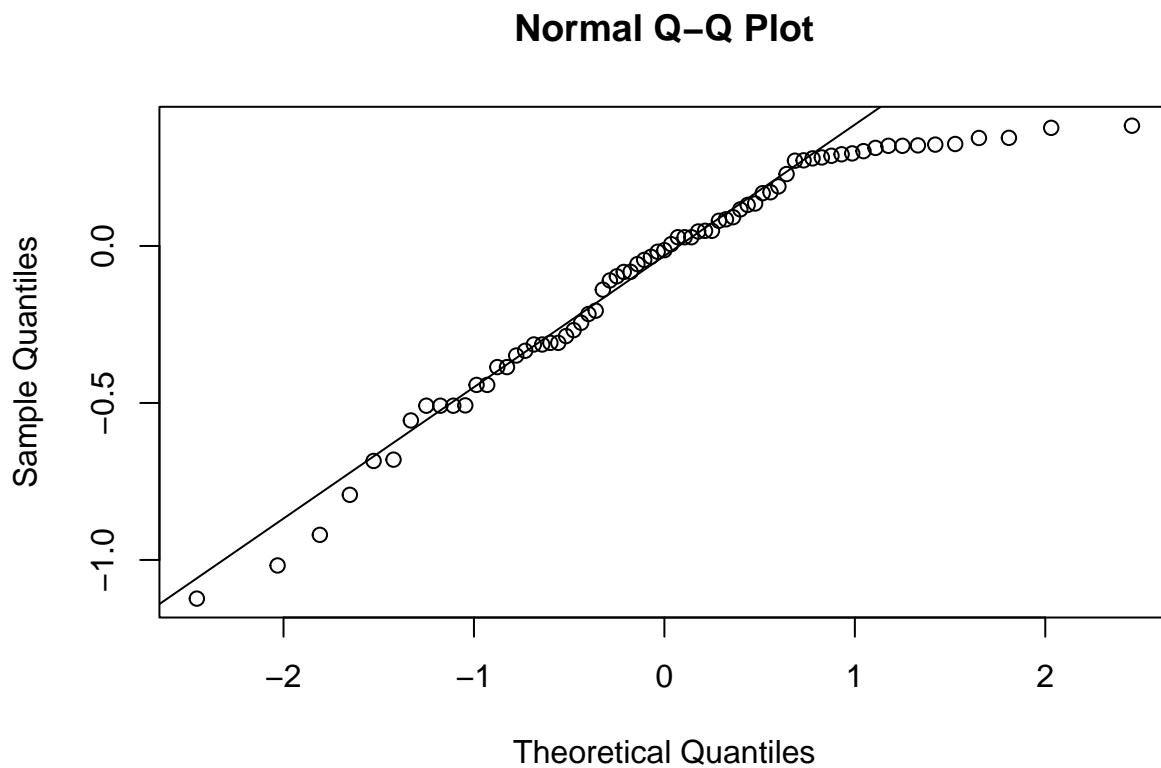


Figure 2: QQ plot for GLMM model fixed effects.

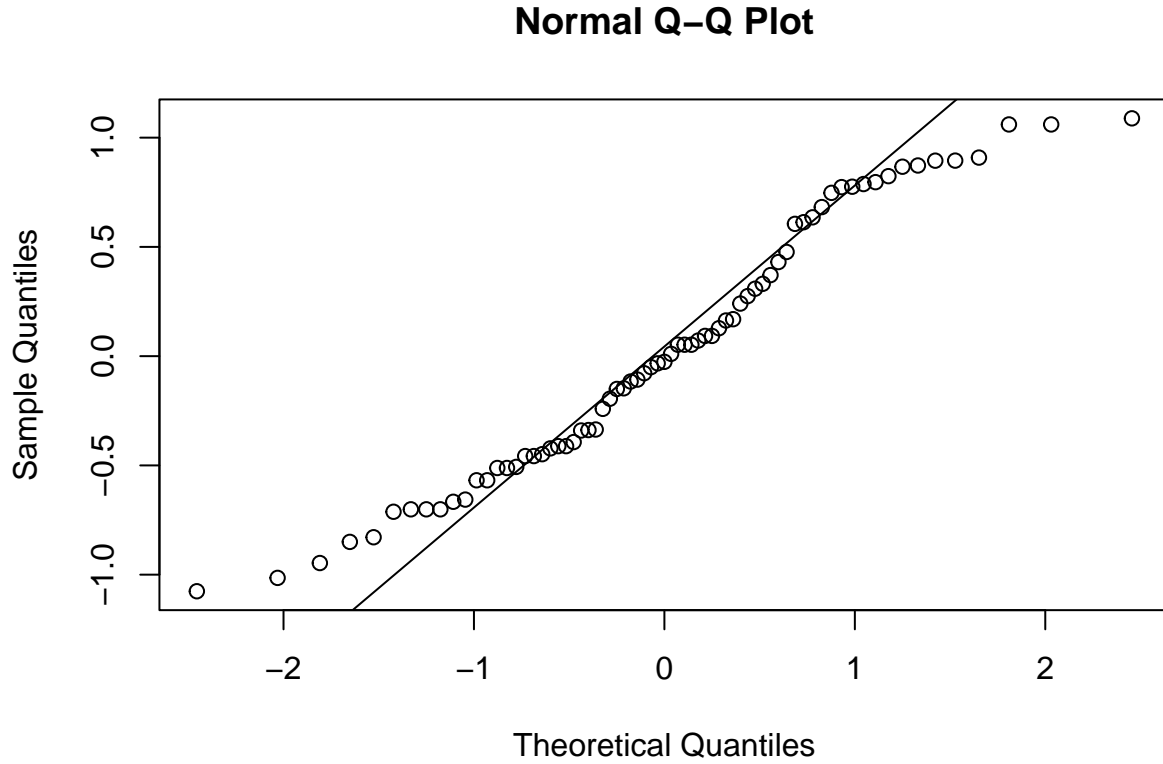


Figure 3: QQ plot for GLMM model random effects.

Table 7: Levenes test for homogeneity of variance

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  1.5264 0.2157
##      67
```

Results

A boxplot was used to visualise the relationship between the abundance of *H. perlevis* over the months. The single terms deletions analysis was used to assess the significance of the 'date' variable in predicting the volume of *H. perlevis*. The initial model fitted with the variable 'date' as the predictor had an Akaike information criterion (AIC) of 574.87 and npara of 3. Removing 'date' from the model slightly increased AIC to 574.94. The likelihood ratio test (LRT) statistic was 6.0749, with a $p = 0.108$. The removal of the 'date' variable did not significantly impact the model fit. The 'date' variable is not a significant predictor of the abundance (n) of *H. perlevis*.

Using row ID as the intercept, results from the Generalized Linear Mixed-effects Model (glmer) and posthoc Tukey test showed there was not sufficient evidence ($p\text{-value} < 0.05$) to reject the null hypothesis while accounting for the random effects; Jun20-Apr 21 ($Pr = 0.2826$), Jun21-Apr21 ($Pr = 0.0901$), Nov20-Apr21 ($Pr = 0.8895$), Jun21-Jun20 ($Pr = 0.9179$), Nov20-Jun20 ($Pr = 0.7895$), Nov20-Jun21 ($Pr = 0.4547$). Levenes test supports the null hypothesis of homogeneity of variance using pearson distribution $F_{3,67} = 1.5264$, $p = 0.2157$.

Assumptions of Analyses

Independence of observations: No two observations in the dataset are related to each other or affect each other in any way.

Poisson distribution: The GLMM model does not suffer from overdispersion (refer to table 4). Residuals are normally distributed (refer to figure 2)

Accounting for under/overdispersion of variance: Poisson distribution assumed that the variance of the abundance is equal across all levels for the variable date. For pearson distribution the assumption of homogeneity of variance was met (refer to table 7)

Random effects assumption: Random effects are normally distributed (refer to figure 3).

Hypothesis 2: Is there a relationship between the number of sponges and the biomass (mL)

Metadata from: Nudibranch predation boosts sponge silicon cycling (2023)

IV= Abundance of sponges (n)

D= H.perlevis volume (mL)

Null Hypothesis (H_0): There is no significant relationship between the abundance of Hymeniacion perlevis and volume Alternative Hypothesis (H_A): There is a significant relationship between the abundance of Hymeniacion perlevis and volume

Analyses: pearsons correlation test

Table 8: Statistical summary of abundance and volume data from the 2023 study

```
##      N.quadrat      N.H..perlevis      Volumen.Hym..mL.      Date
##  Min.       : 1.000    Min.       : 3.00    Min.       : 1.90    Length:71
##  1st Qu.: 5.000    1st Qu.:12.00    1st Qu.: 27.35    Class :character
##  Median : 9.000    Median :20.00    Median : 62.60    Mode  :character
##  Mean   : 9.789    Mean   :23.58    Mean   :120.08
##  3rd Qu.:14.000    3rd Qu.:31.00    3rd Qu.:167.85
##  Max.   :22.000    Max.   :73.00    Max.   :690.50
##
##      X              row_id
##  Mode:logical    1        : 1
##  NA's:71         2        : 1
##                  3        : 1
##                  4        : 1
##                  5        : 1
##                  6        : 1
##                  (Other):65
```

Table 9: Pearsons correlation between the log of abundance (n) and volume (mL)

```
##
##  Pearson's product-moment correlation
##
##  data:  log(pre_y_ab_vol$N.H..perlevis) and log(pre_y_ab_vol$Volumen.Hym..mL.)
##  t = 4.9905, df = 69, p-value = 4.326e-06
##  alternative hypothesis: true correlation is not equal to 0
##  95 percent confidence interval:
##   0.3201512 0.6680298
##  sample estimates:
##           cor
## 0.5149891
```

Results

A Pearson correlation test was done between the abundance of *H.perlevis* (n) and the volume of *H.perlevis* (mL). The correlation coefficient was 0.5149891 indicating a moderate positive linear relationship (df= 69, p-value = 4.326×10^{-6} , confidence interval = 95%). This analysis provides evidence to reject the null hypothesis.

Assumption of analysis

Linearity: The relationship between Date and the volume of *H.perlevis* is linear

Continous data: Each variable is continous in nature

Independence of observations: No two observations in the dataset are related to each other or affect each other in any way.

Normal distribution: The residuals are normally distributed

Homogeneity of variance: The variance of the volume of *H.perlevis* is equal across all groups for the months

Hypothesis 3: Does the month have a effect on the volume of *H.perlevis*?

Metadata from: Nudibranch predation boosts sponge silicon cycling (2023)

IV= Month (Jun20,Nov20,Apr20,Jun21)

D= *H.perlevis* volume (mL)

Null Hypothesis (H_0): There is no significant relationship between the volume of *Hymeniacidon perlevis* and the month

Alternative Hypothesis (H_A): There is a significant relationship between the volume of *Hymeniacidon perlevis* and the month

Analyses: one-way ANOVA, Tukey post-hoc test, Boxplot, levenes test

Table 10: One way anova and posthoc tukey test between the volume (mL) of *H.perlevis* and the months

```
## Analysis of Variance Table
##
## Response: log(Volumen.Hym..mL.)
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Date        3  82.274  27.4246   51.381 < 2.2e-16 ***
## Residuals  67  35.762   0.5338
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = log(Volumen.Hym..mL.) ~ Date, data = prey_ab_vol)
##
## $Date
##           diff          lwr          upr          p adj
## Apr21-Nov20 1.1909930 0.46159661 1.920389 0.0003224
## Jun21-Nov20 1.8540258 1.17472826 2.533323 0.0000000
## Jun20-Nov20 3.0564074 2.38303878 3.729776 0.0000000
## Jun21-Apr21 0.6630328 0.01230726 1.313758 0.0442034
## Jun20-Apr21 1.8654144 1.22088054 2.509948 0.0000000
## Jun20-Jun21 1.2023817 0.61514235 1.789621 0.0000057
```

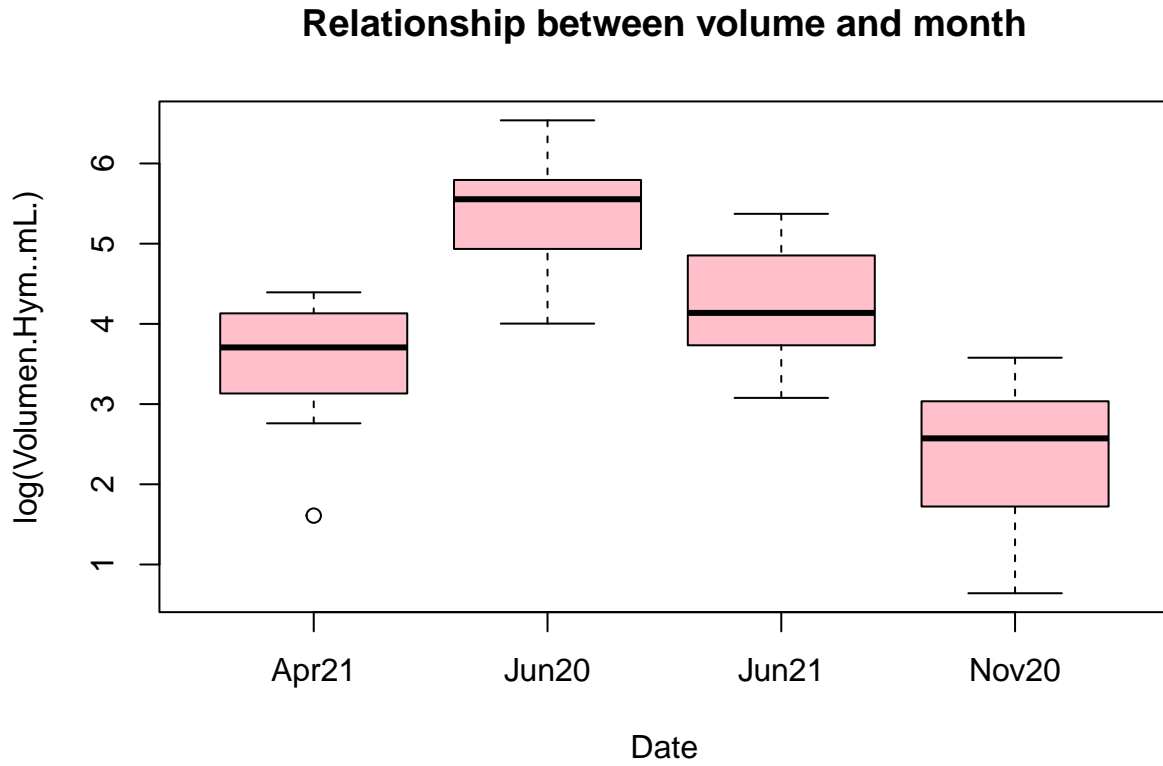


Figure 4: Boxplot of the relationship between month/date and the log of the volume (mL) for *H.perlevis*.(x axis): The month the volume was measured and (y axis): the log of the volume.

Table 11: Levenes test for homogeneity of variance

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  0.4505 0.7178
##      67
```

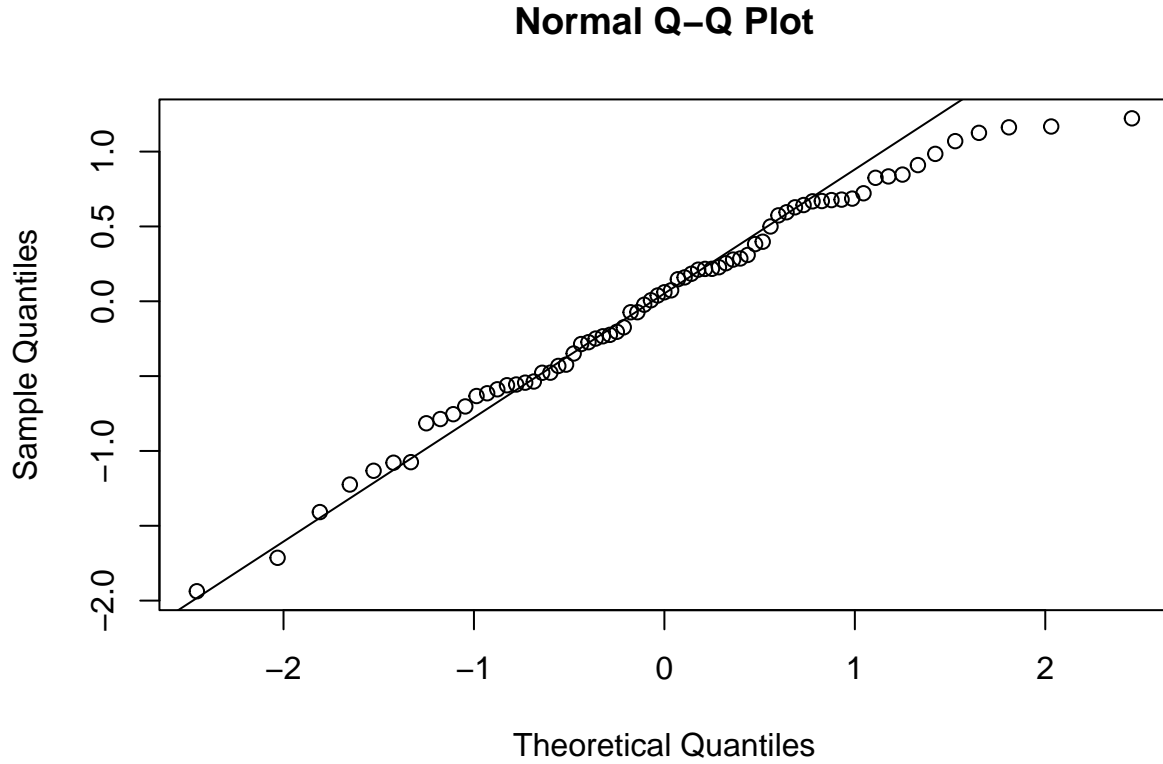


Figure 5: QQ plot for ANOVA model

Results

The ANOVA analysis indicated a significant effect of the variable ‘date’ on the logarithmically transformed variable ‘volume of H.perlevis’ ($F_{3,67} = 51.380$, $p = 2.35 \times 10^{-17}$ (<0.05)). A post hoc Tukey test showed that the volume of H.perlevis among the months differed significantly at $p < .05$ and confidence level of 95% [April 21- Nov 20 ($p = 3.2 \times 10^{-4}$), Jun 21 – Nov 20 ($p = 0$), Jun 20- Nov 20 ($p = 0$), Jun 21 – Apr 21 ($p = 4.4 \times 10^{-2}$), Jun 20- Apr 21 ($p = 0$), Jun 20 – Jun 21 ($p = 5.7 \times 10^{-6}$)]. The results of the ANOVA test support the alternative and reject the null hypothesis. A boxplot was used to visualise the relationship between date and logarithmically transformed volume (refer to figure 4). Levene’s test was conducted to examine the homogeneity of variables across the levels for the independent variable ‘date’. A non-significant result was observed ($F_{3,67} = 0.4505$, $p = 0.7178$). Levene’s test supports the null hypothesis of homogeneity of variance using pearson distribution ($(F_{3,67} = 0.4505, p = 0.7178)$)

Assumptions of Analyses

Independence of observations: No two observations in the dataset are related to each other or affect each other in any way.

Normal distribution: The residuals are normally distributed (refer to figure 5)

Homogeneity of variance: The variance of the volume of H.perlevis is equal across all groups for the months (refer to table 11).

Hypothesis 4: Is there a relationship between average volume of H.perlevis per month and pH, salinity and temperature averages for the months

Metadata from: Nudibranch predation boosts sponge silicon cycling (2023) Temp, Salinity and pH values obtained from Somlit.

IV= average Temp (T), Salinity (S) and PH values for the month the volume was measured

D= average sponge volume per month (mL sponge m-2) given by ((ponge volume mL./abundance) per quadrant, averages for the months)

Null Hypothesis (H_0): There is no significant relationship between the average volume of *H.perlevis* per month (D) and the average temperature (T), salinity (S), and pH values for the corresponding months.

Alternative Hypothesis (H_A): There is a significant relationship between the average volume of *H.perlevis* per month (D) and the average temperature (T), salinity (S), and pH values for the corresponding months.

Analyses: pearsons correlation

Table 12: Summary statistics for temperature, salinity and pH data from SOMLIT

##	ID_SITE	DATE	T	S	PH
##	Min. :5	Length:14	Min. :10.97	Min. :33.62	Min. :7.791
##	1st Qu.:5	Class :character	1st Qu.:12.90	1st Qu.:34.34	1st Qu.:7.885
##	Median :5	Mode :character	Median :14.24	Median :34.59	Median :7.920
##	Mean :5		Mean :14.03	Mean :34.48	Mean :7.921
##	3rd Qu.:5		3rd Qu.:15.62	3rd Qu.:34.73	3rd Qu.:7.950
##	Max. :5		Max. :16.82	Max. :35.01	Max. :8.138

[1] "Standard Deviation (T): 2.02518274307706 Standard Deviation (pH): 0.0953984414828807 Standard D

Table 13: Averages data for Temp, pH, salinity and volume per month brackets

##	Month	Year	T	S	PH
##	1	6 2020	16.04400	34.68967	8.016667
##	2	11 2020	13.13850	34.46150	7.859750
##	3	4 2021	11.09100	33.85867	7.913333
##	4	6 2021	15.61075	34.81675	7.915000

##	Date	Volumen.Hym..mL..N.H..perlevis
##	1 Apr21	3.8781924
##	2 Jun20	12.6688840
##	3 Jun21	3.0448443
##	4 Nov20	0.8122513

Table 14: Correlation analysis for volume of *H.perlevis* vs T,S and pH

##	Volumen.Hym..mL..N.H..perlevis	T
##	Volumen.Hym..mL..N.H..perlevis	1.00000000 -0.2758412
##	T	-0.27584123 1.00000000
##	S	-0.08480915 0.9491472
##	PH	-0.52623615 0.5763816
##	S	PH
##	Volumen.Hym..mL..N.H..perlevis	-0.08480915 -0.5262361
##	T	0.94914725 0.5763816
##	S	1.00000000 0.2922510
##	PH	0.29225097 1.0000000

Results

A correlation test was used to show the relationship between volume of *H.perlevis* for the different months and T, S and Ph for the corresponding month. Volume had a weak negative correlation with temperature

($r = -0.276$) and a weak negative correlation with salinity ($r = -0.085$). It has a moderate negative correlation with the variable pH ($r = -0.526$).

Assumption of Analyses

Independence of observations: No two observations in the dataset are related to each other or affect each other in any way.

Linearity: The relationship between Date and the volume of *H.perlevis* is linear

Homogeneity of variance: The variance of the volume of *H.perlevis* is equal across all groups for the months.

Hypothesis 5: What is the relationship between the abundance of *H.perlevis* and biomass per habitat (2022 study)

metadata: Sponge contribution to the silicon cycle of a diatom-rich shallow bay (2022) IV= habitat& abundance (N sponge m-2) D= biomass (mL sponge m-2) Habitats= Rocky intertidal,Rocky subtidal,Maerl,Shallow muds,Heterog seds,Circ.coarse sed

Null Hypothesis (H_0): There is no significant relationship between biomass and the interaction between Habitat and abundance

Alternative Hypothesis (H_A): There is a significant relationship between biomass and the interaction between Habitat and abundance

Analyses: scatterplot, linear regression analysis, anova (m10b and m10), regression analysis

Table 15: Summary statistics for abundance (N sponge m-2) and biomass (mL sponge m-2) of *H.perlevis* sponges from 2022 study

```
##           N           Habitat      Abundance.sponges Sponge.biomass
## Min.      : 1.0    Length:142    Min.      : 1.00    Min.      :   3.75
## 1st Qu.: 8.0    Class :character 1st Qu.: 3.00    1st Qu.:   35.21
## Median :15.5    Mode  :character Median : 7.00    Median :  131.58
## Mean   :16.5                      Mean   :12.77    Mean   :  985.79
## 3rd Qu.:24.0                      3rd Qu.:15.00    3rd Qu.:  388.68
## Max.   :40.0                      Max.   :92.00    Max.   :47446.35
##
##      row_id
## 1         : 1
## 2         : 1
## 3         : 1
## 4         : 1
## 5         : 1
## 6         : 1
## (Other):136
```

```
## [1] "Standard Deviation (Abundance): 17.1338764821148 Standard Deviation (Biomass): 4739.46115648504"
```

Table 16: ANCOVA analysis for homogeneity of slopes in model 10 (m10). With log biomass as the dependent variable and habitat,abundance (log.n) and their interaction as independent variables.

```
## Analysis of Variance Table
##
## Response: log.biomass
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Habitat      5 184.086   36.817  25.4346 < 2.2e-16 ***
```

```
## log.n          1  49.178  49.178 33.9737  4.17e-08 ***
## Habitat:log.n   5   8.061   1.612  1.1137   0.3563
## Residuals      130 188.178   1.448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 17: ANOVA test for model (M10b) without interaction between log biomass and abundance per habitat. Log biomass is the dependent variable, and habitat and abundance (log.n) are the separate independent variables.

```
## Analysis of Variance Table
##
## Response: log.biomass
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Habitat     5 184.086  36.817  25.328 < 2.2e-16 ***
## log.n        1  49.178  49.178  33.831 4.146e-08 ***
## Residuals  135 196.239   1.454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 18: ANOVA Summary of linear model chosen (m10b)

```
##
## Call:
## lm(formula = log.biomass ~ Habitat + log.n, data = data2022)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3431 -0.6598 -0.2098  0.5976  3.9337
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.0795     0.2935   7.085 6.96e-11 ***
## HabitatHeterog seds      1.7064     0.3752   4.548 1.19e-05 ***
## HabitatMaerl        1.2676     0.3837   3.304 0.001221 **
## HabitatRocky intertidal  1.5486     0.3742   4.138 6.12e-05 ***
## HabitatRocky subtidal    2.7984     0.3944   7.094 6.64e-11 ***
## HabitatShallow muds      2.3146     0.6098   3.796 0.000222 ***
## log.n              0.6314     0.1085   5.816 4.15e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.206 on 135 degrees of freedom
## Multiple R-squared:  0.5431, Adjusted R-squared:  0.5228
## F-statistic: 26.75 on 6 and 135 DF, p-value: < 2.2e-16
```

Table 19: Posthoc test on ANOVA

```
## Warning in replications(paste("-", xx), data = mf): non-factors ignored: log.n

## Tukey multiple comparisons of means
## 95% family-wise confidence level
```

```
##
## Fit: aov(formula = log.biomass ~ Habitat + log.n, data = data2022)
##
## $Habitat
##
```

	diff	lwr	upr	p adj
Heterog sed	2.0590719	0.9886933	3.1294505	0.0000020
Maerl	2.3458599	1.3746685	3.3170512	0.0000000
Rocky intertidal	1.6536569	0.5730455	2.7342683	0.0002815
Rocky subtidal	3.8419796	2.8264024	4.8575569	0.0000000
Shallow muds	1.9187712	0.1667901	3.6707524	0.0229877
Maerl-Heterog sed	0.2867879	-0.6132068	1.1867827	0.9404453
Rocky intertidal-Heterog sed	-0.4054150	-1.4225182	0.6116881	0.8582929
Rocky subtidal-Heterog sed	1.7829077	0.8351867	2.7306286	0.0000036
Shallow muds-Heterog sed	-0.1403007	-1.8538397	1.5732383	0.9998964
Rocky intertidal-Maerl	-0.6922029	-1.6043439	0.2199380	0.2474236
Rocky subtidal-Maerl	1.4961198	0.6620470	2.3301925	0.0000113
Shallow muds-Maerl	-0.4270886	-2.0804846	1.2263073	0.9755802
Rocky subtidal-Rocky intertidal	2.1883227	1.2290596	3.1475858	0.0000000
Shallow muds-Rocky intertidal	0.2651143	-1.4548352	1.9850638	0.9977419
Shallow muds-Rocky subtidal	-1.9232084	-3.6030603	-0.2433564	0.0148821

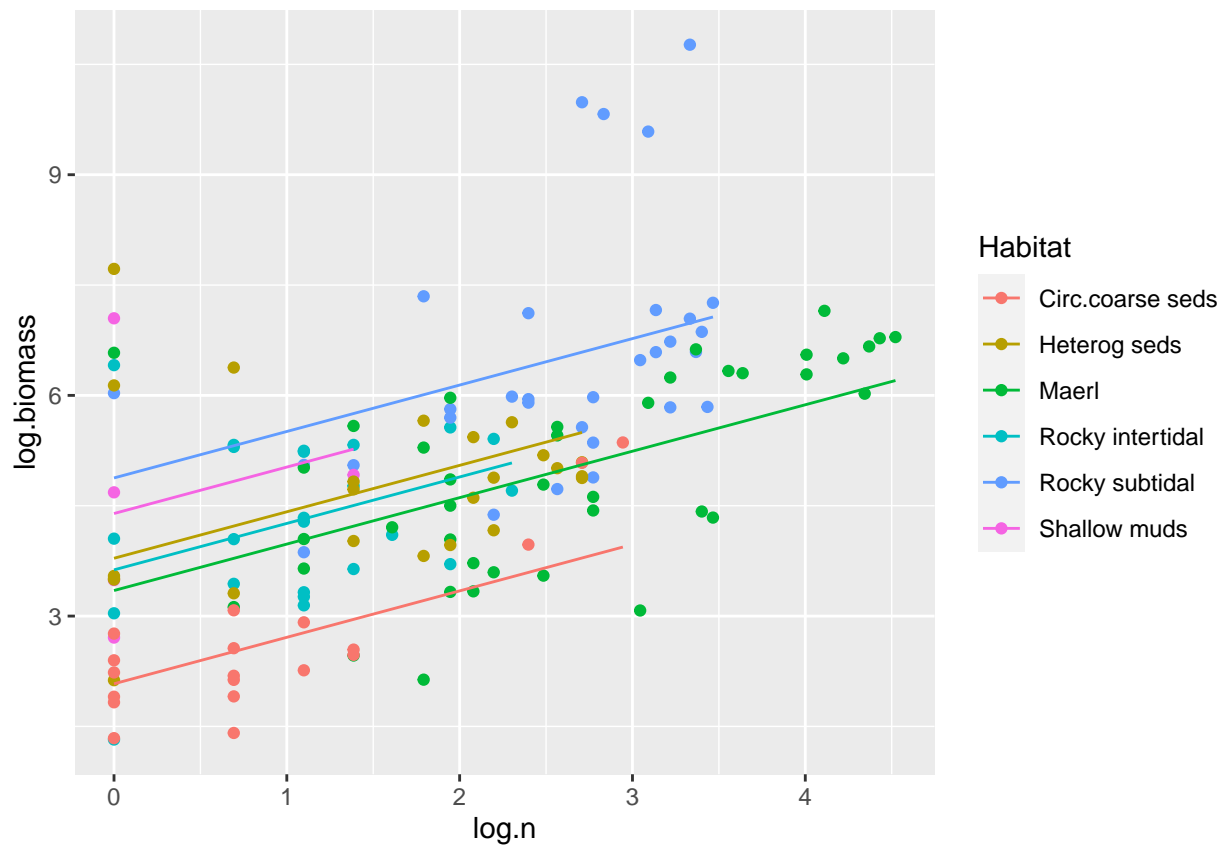


Figure 6: Scatterplots of the relationship between (x): log abundance (N sponge m^{-2}) and (y): log biomass (mL sponges m^{-2}), per habitat.

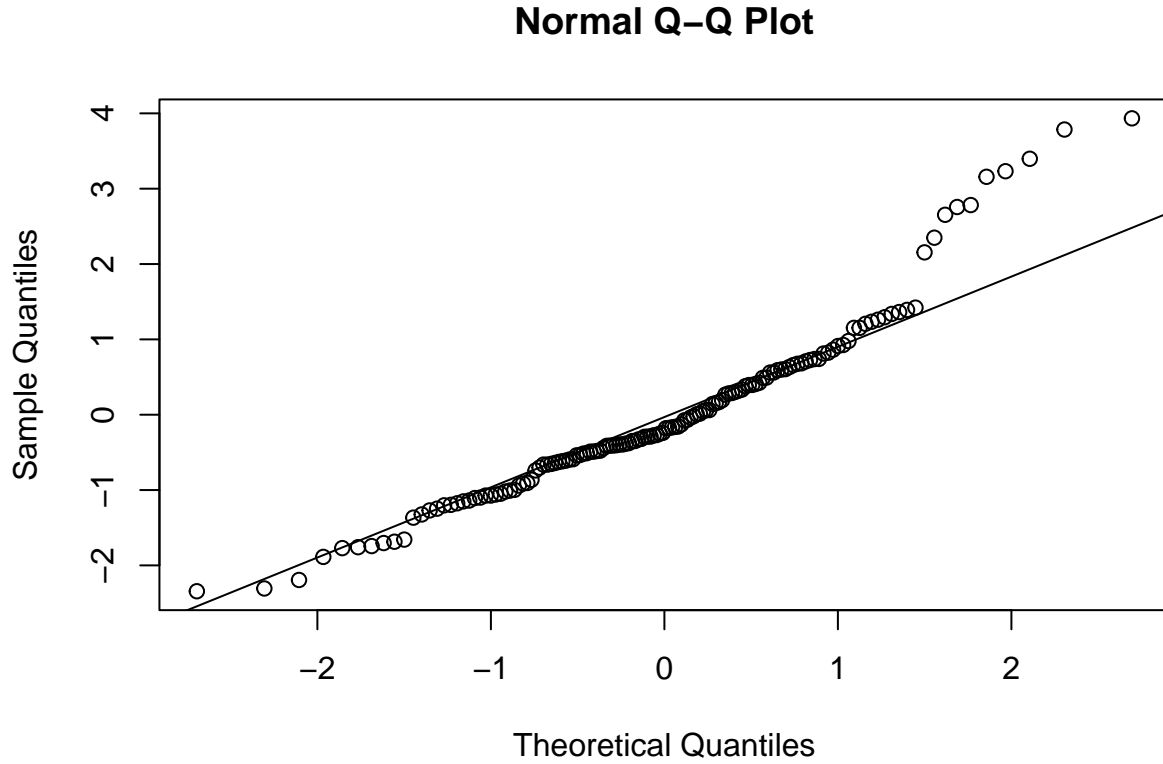


Figure 7 : QQ plot for linear model 10b

Results

ANCOVA was used to assess relationships between biomass and abundance and whether those relationships varied among habitats. A homogeneity of slopes test indicated that the slopes of the log-log relationship between biomass and abundance were similar for the six habitats ($F_{5,130} = 1.1137$, $P = 0.3563$). The scatterplot (figure 6) was used to visualise the relationship between abundance and biomass for each habitat. Assuming a common slope of 0.6314 for all sites, the intercept differed significantly among habitats ($F_{5,135} = 25.328$, $p = 2.2 \times 10^{-16}$) and abundance ($F_{135,1} = 33.831$, $p = 4.146 \times 10^{-08}$). Tukey's posthoc test of the model (Residual standard error: 1.206 on 135 degrees of freedom) showed each habitat level shows a significant effects on log.biomass ($F_{135} = 26.75$, $p = 2.2 \times 10^{-16}$). The model explains approx. 54.31% of the variance in the biomass variable (multiple R-squared). There is strong evidence to support the alternative hypothesis and reject the null.

Assumption of Analyses

Independence of observations: No two observations in the dataset are related to each other or affect each other in any way.

Normal distribution: The residuals of the data are normally distributed (represented by figure 7).

Homogeneity of variance: The variance of the volume and abundance of *H. perlevis* is equal across all groups for the habitats.