Natural Populations Preliminary Results

Daniel Reyes Corral

7/8/2020

Summary results

Data cleaning

I did a series of PCAs using the variables that were measured continously between 2017 - 2019. I choose these years because 2014 - 2016 were already analysed by Sofia Carvajal. Although we can include these in further analysis (**PTA**) or selection estimates. I removed NAs, check for outliers and corrected some values from the raw data.The clean subset of data was saved as a second table called **TribulusPC** (https://github.com/Winer-DanielR/Tribulus-Natural-population/tree/master/Data/Processed) this new dataset was used for the following analysis.

General PCA

The PCA used the following variables:

- *length*
- *width*
- *depth*
- *longest spine*

I did not include *spine tip distance* since I did not measure that trait in 2019 due to the measurements involving a lot of guessing given that most of the mericarps did not have completed spines.

PCA summary

```
## Importance of components:
##                          PC1    PC2    PC3     PC4
## Standard deviation     1.5897 0.9011 0.6657 0.46669
## Proportion of Variance 0.6318 0.2030 0.1108 0.05445
## Cumulative Proportion  0.6318 0.8347 0.9456 1.00000
```
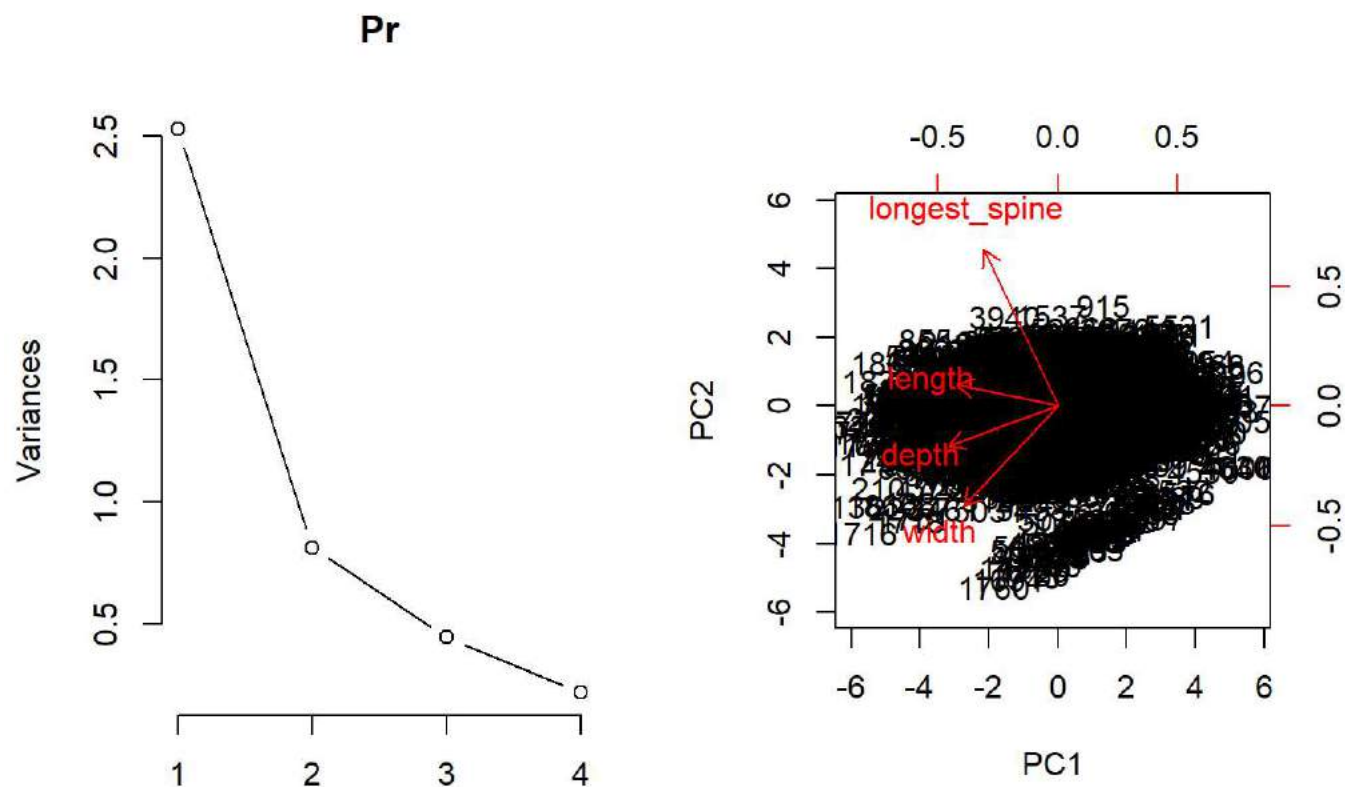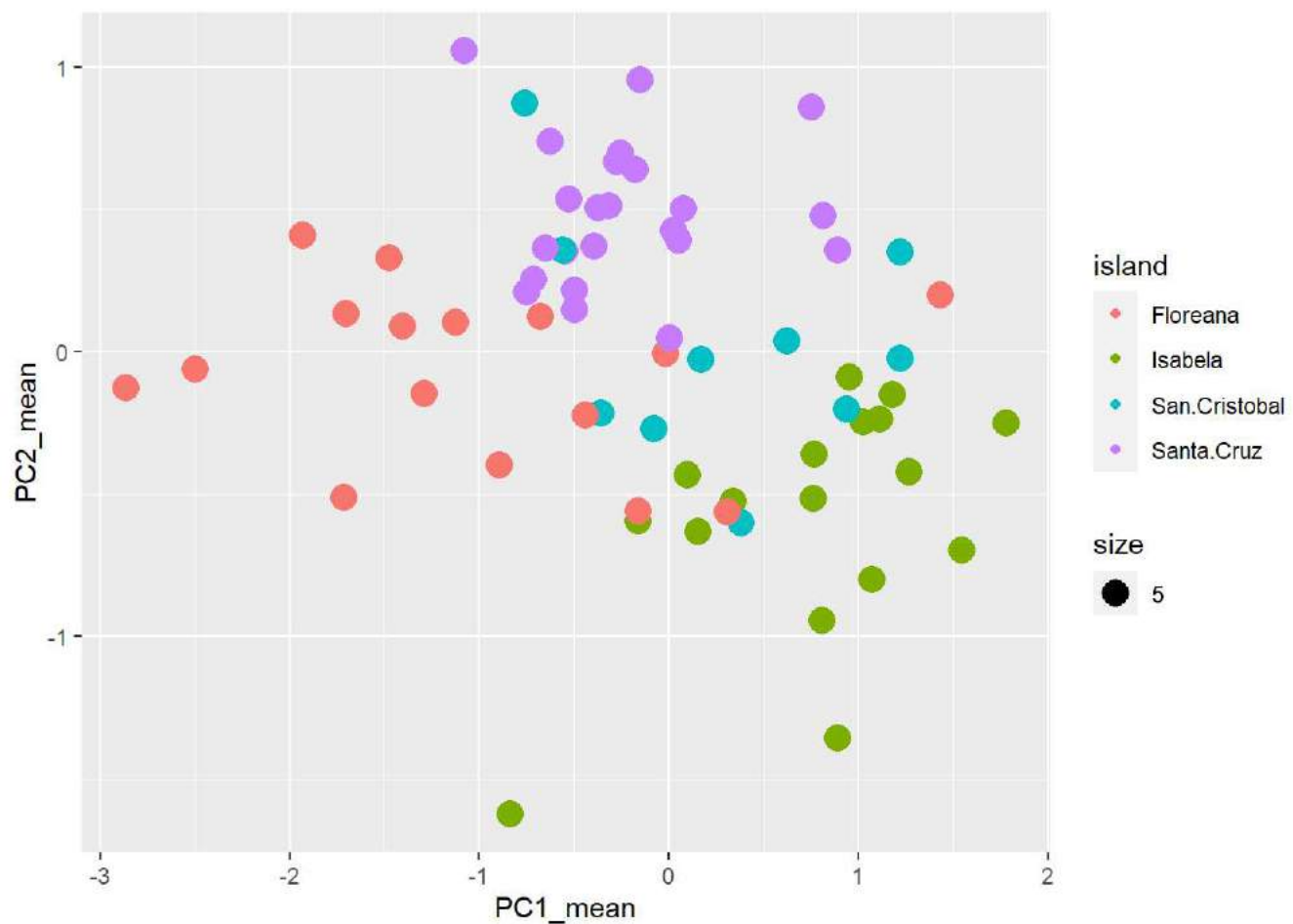
Left: Variance explained per PC. Rigth: Individual PCA as described above.

PCA means

Then I estimated the means of the PC scores obtained above and plot them in different ways to visualize the patters of variation.

Here I estimated the means per population and years for each island:

Here I estimated the means per year and sorted by island. -You can see there is a trend per year that goes in different directions-:
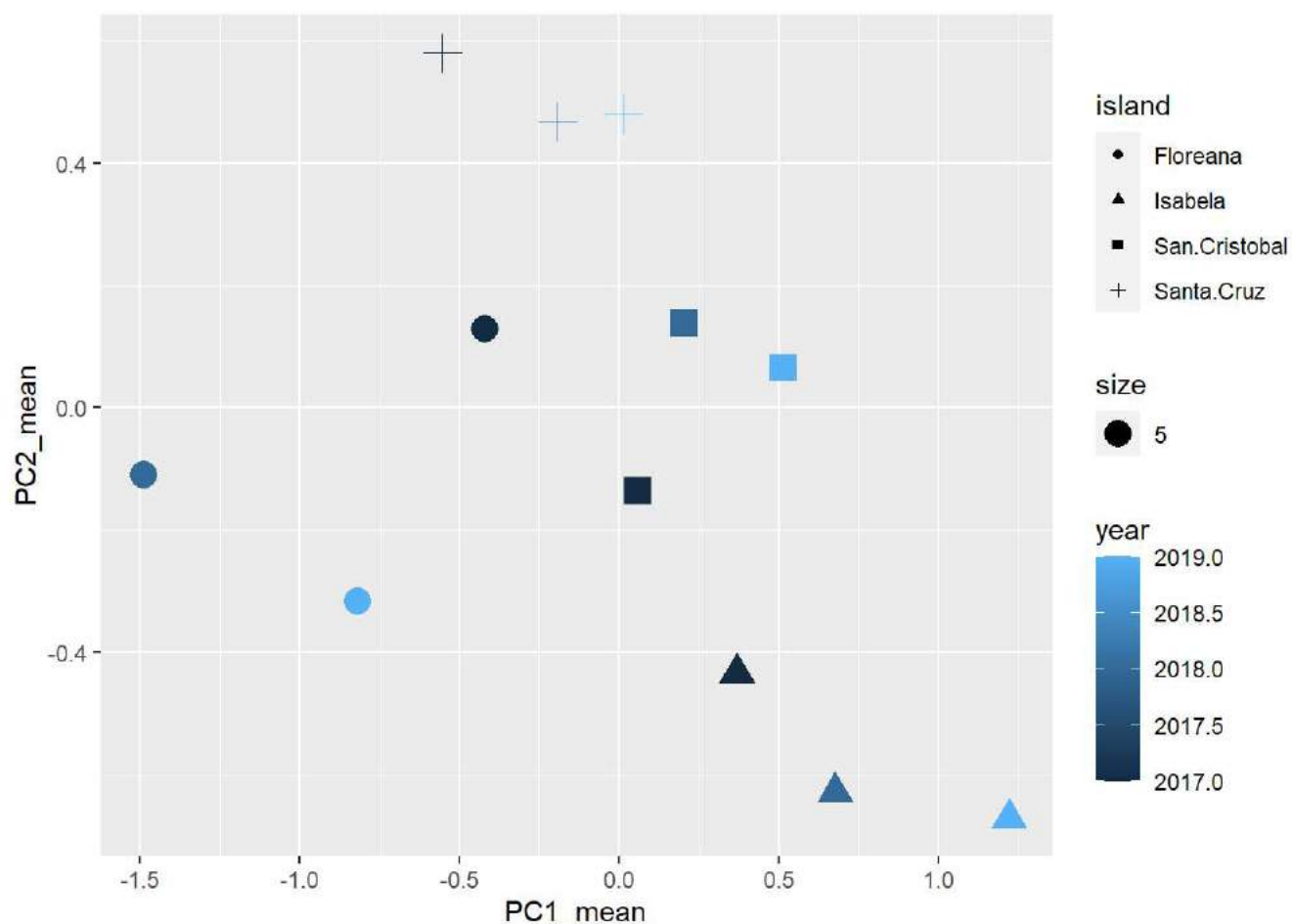
Variance Partition Analysis

I reviewed Sofia's code and use it as a reference to do the partition analysis. I ran a series of individual models per trait. I made sure that population was nested within islands and included island and year as random models. I edited the dataset to make all the other variables factors.

Descriptive data:

```
##                        means        sd  min    max
## length             6.308864  1.359785 2.23  11.89
## width               3.22321 0.6389286 1.03   7.34
## depth              4.782531 0.8408234 2.01   8.54
## longest_spine      5.372706  1.334417    0  10.18
## spine_tip_distance 10.70068  2.451824 2.17     NA
```
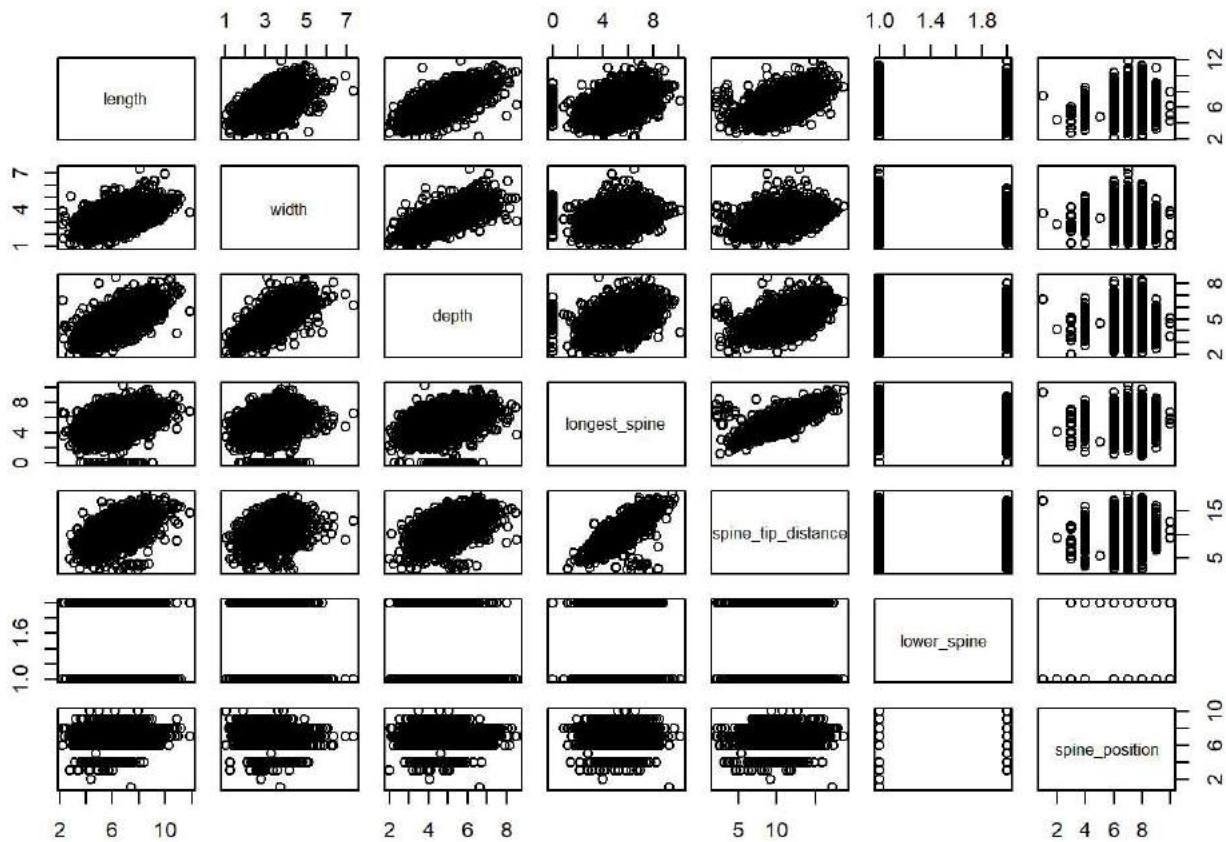
Generalized linear model:

lmer(*trait* ~ 1 + (1|year) + (1|island) + (1|island:population), data = tribulusPC, REML = T)

Where *trait* was each of the measured traits:

- *length*
- *width*
- *depth*
- *longest spine*
- *spine tip distance*

For *lower spine* I used a different model since is a binomial trait:

glmer(lower_spine ~ 1 + (1|year) + (1|island) + (1|island:population), data = tribulusPC, na.action = na.omit, family = binomial)

I did not use *spine position* but I can include that as well I can do a similar analysis per island

For each trait I used those models I checked assumptions (as Sofia did) and looked at the variance explained by each trait and looked at the significance effect of each random factor: year, island and population within island.

Summary Results Length:

```
summary(lm.length)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: length ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##     Data: tribulusPC
##
## REML criterion at convergence: 16457.2
##
## Scaled residuals:
##      Min      1Q  Median      3Q     Max
## -4.4644 -0.6554 -0.0379  0.6406  5.2236
##
## Random effects:
##  Groups            Name          Variance Std.Dev.
##  island:population (Intercept) 0.40583  0.6371
##  island            (Intercept) 0.44624  0.6680
##  year              (Intercept) 0.02885  0.1698
##  Residual                      1.10964  1.0534
## Number of obs: 5552, groups:  island:population, 25; island, 4; year, 3
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   6.3427     0.3723 3.4968   17.04 0.000172 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing significance of each random effects:

```
ranova(lm.length)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## length ~ (1 | year) + (1 | island) + (1 | island:population)
##                          npar  logLik   AIC     LRT Df Pr(>Chisq)
## <none>                      5 -8228.6 16467
## (1 | year)                  4 -8274.1 16556   90.92  1  < 2.2e-16 ***
## (1 | island)                4 -8233.0 16474    8.84  1    0.002954 **
## (1 | island:population)     4 -8766.3 17541 1075.31  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Checking assumptions *length*:

```
plot(lm.length)
```
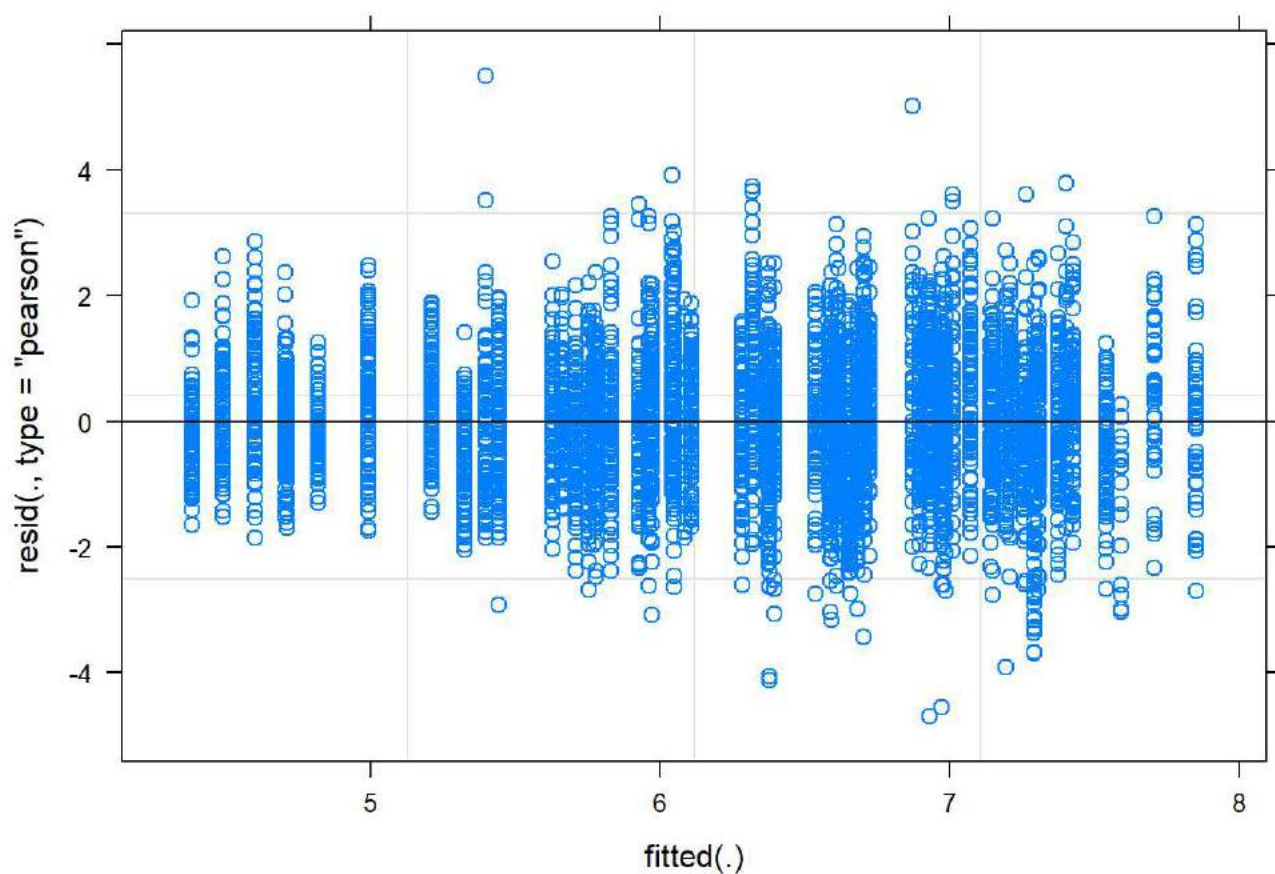
```
opar <- par(mfrow=c(2,2))
hist(tribulusPC$length)
plot(fitted(lm.length),resid(lm.length))
abline(h=0,lty=2,col="red")
qqnorm(resid(lm.length))
qqline(resid(lm.length), col="red")
hist(resid(lm.length))
```

Histogram of tribulusPC$length
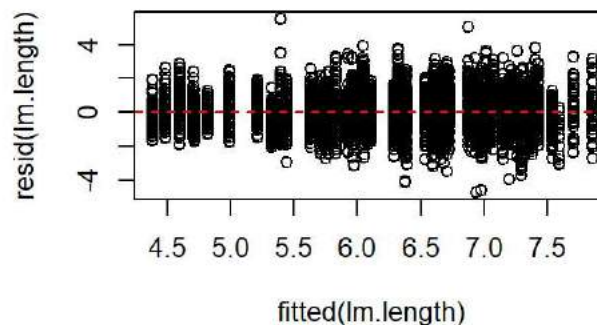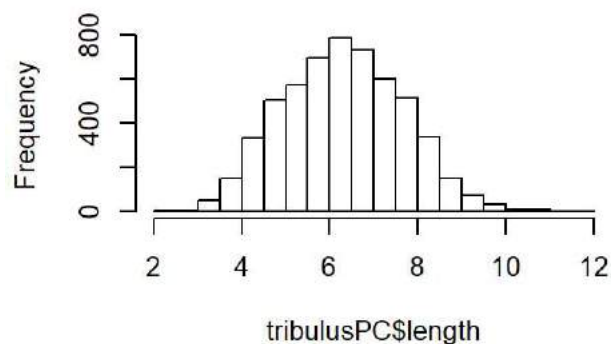
Normal Q-Q Plot
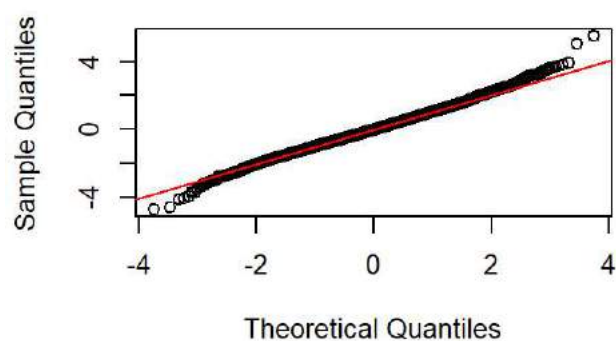
Histogram of resid(lm.length)


```
par(opar)
```
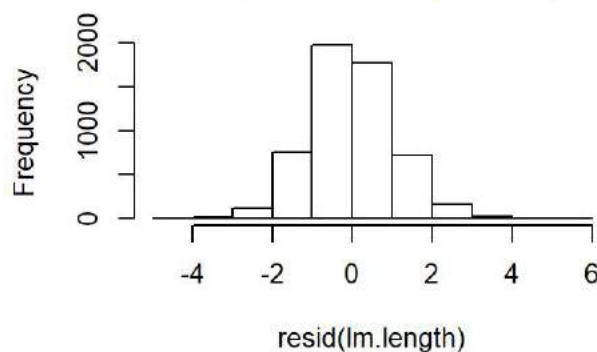
Top

Summary Results Width

```
summary(lm.width)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: width ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##    Data: tribulusPC
##
## REML criterion at convergence: 9820.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7678 -0.6146 -0.0292  0.6219  5.8806
##
## Random effects:
##  Groups            Name        Variance Std.Dev.
##  island:population (Intercept) 0.085086 0.29169
##  island            (Intercept) 0.014726 0.12135
##  year              (Intercept) 0.007487 0.08653
##  Residual                      0.336490 0.58008
## Number of obs: 5552, groups:  island:population, 25; island, 4; year, 3
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  3.24292    0.09881 4.73038   32.82 9.17e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing significance of each random effects:

```
ranova(lm.width)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## width ~ (1 | year) + (1 | island) + (1 | island:population)
##                         npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>                     5 -4910.3  9830.7
## (1 | year)                 4 -4947.8  9903.5  74.84  1    <2e-16 ***
## (1 | island)               4 -4910.8  9829.6   0.92  1    0.3367
## (1 | island:population)    4 -5236.4 10480.9 652.18  1    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

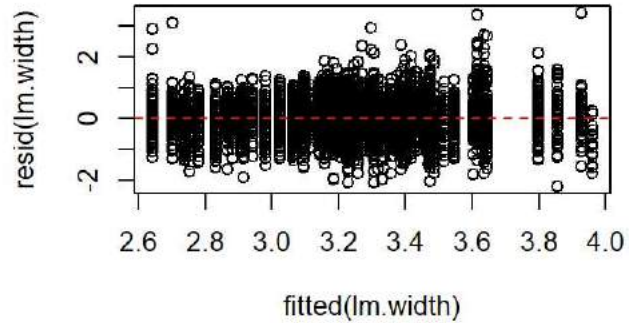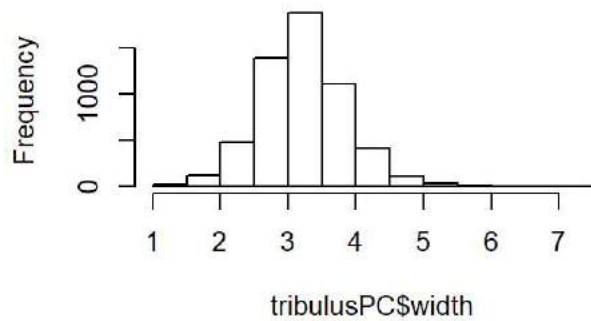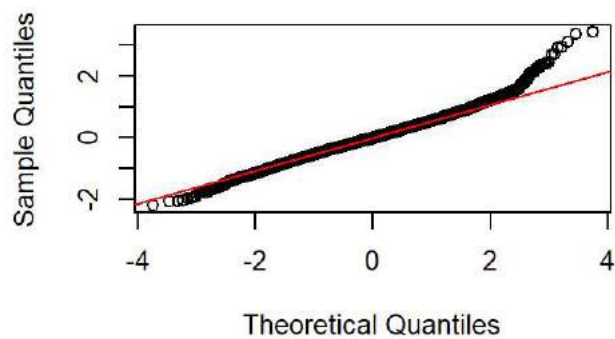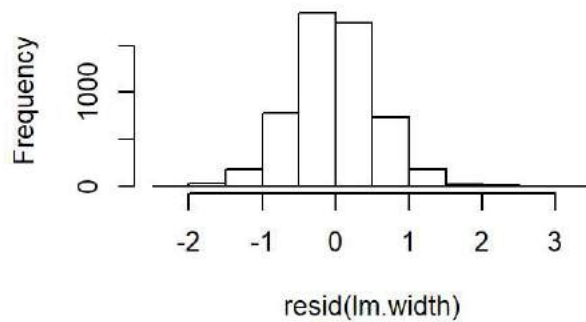## Checking assumptions *width*:

```
plot(lm.width)
```
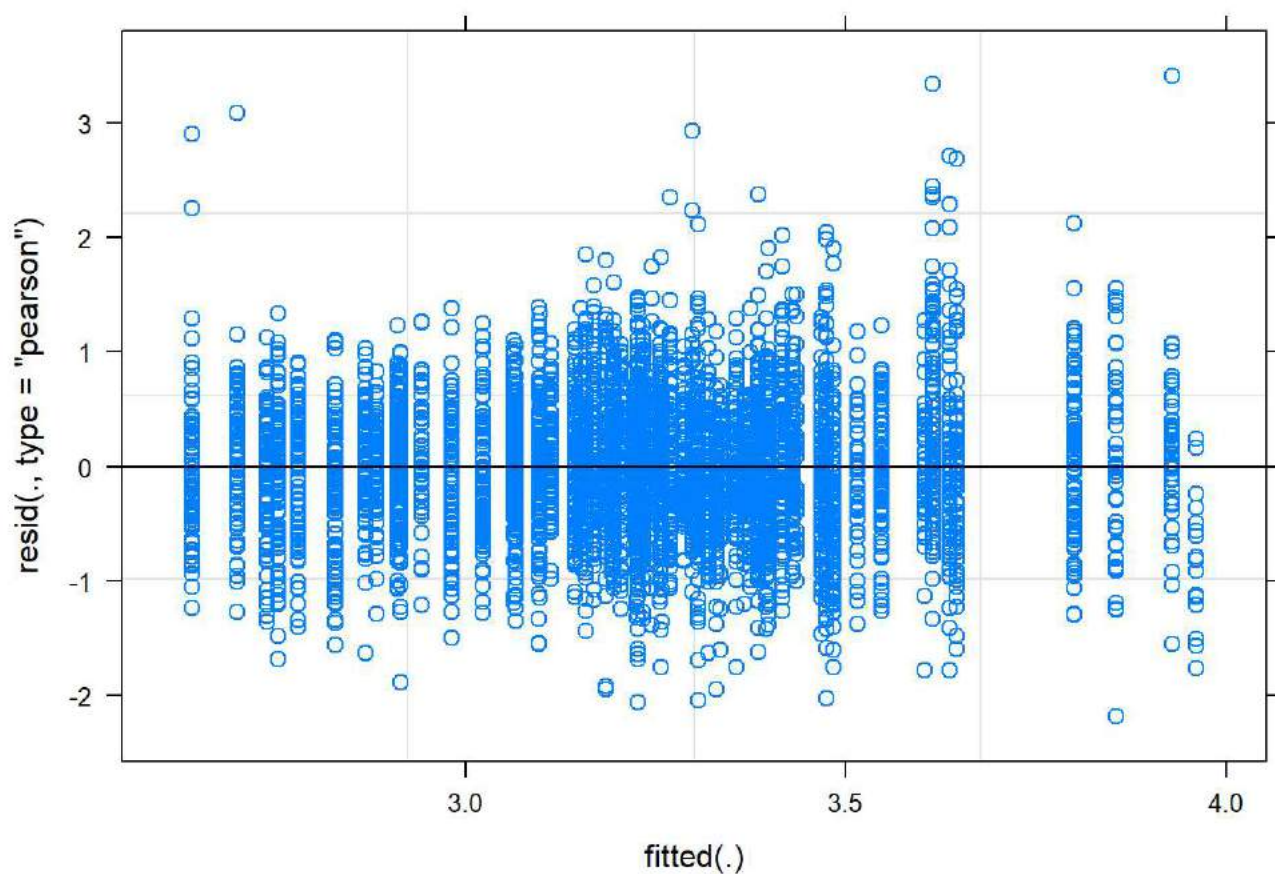
```
opar <- par(mfrow=c(2,2))
hist(tribulusPC$width)
plot(fitted(lm.width),resid(lm.width))
abline(h=0,lty=2,col="red")
qqnorm(resid(lm.width))
qqline(resid(lm.width), col="red")
hist(resid(lm.width))
```

Histogram of tribulusPC$width

Normal Q-Q Plot

Histogram of resid(lm.width)


```
par(opar)
```

Top

Summary Results Depth

```
summary(lm.depth)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: depth ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##     Data: tribulusPC
##
## REML criterion at convergence: 12043.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1191 -0.6260  0.0180  0.6526  4.9630
##
## Random effects:
##  Groups            Name        Variance Std.Dev.
##  island:population (Intercept) 0.17119  0.4137
##  island            (Intercept) 0.14782  0.3845
##  year              (Intercept) 0.01652  0.1285
##  Residual                      0.50120  0.7080
## Number of obs: 5552, groups:  island:population, 25; island, 4; year, 3
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   4.8287     0.2232 3.7845   21.64 4.17e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing significance of each random effects:

```
ranova(lm.depth)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## depth ~ (1 | year) + (1 | island) + (1 | island:population)
##                          npar  logLik   AIC    LRT Df Pr(>Chisq)
## <none>                      5 -6021.8 12054
## (1 | year)                  4 -6078.3 12164 113.02  1  < 2.2e-16 ***
## (1 | island)                4 -6025.3 12059   7.06  1   0.007879 **
## (1 | island:population)     4 -6396.2 12800 748.78  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
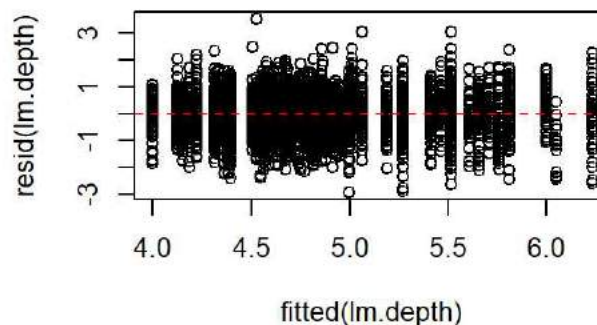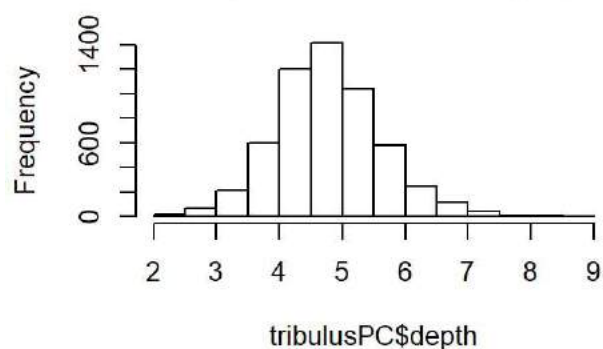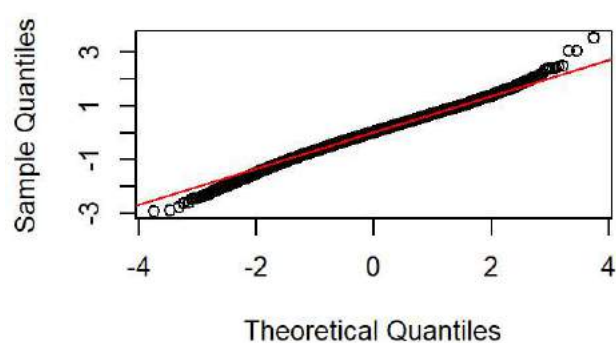
Checking assumptions *depth*:

```
plot(lm.depth)
```
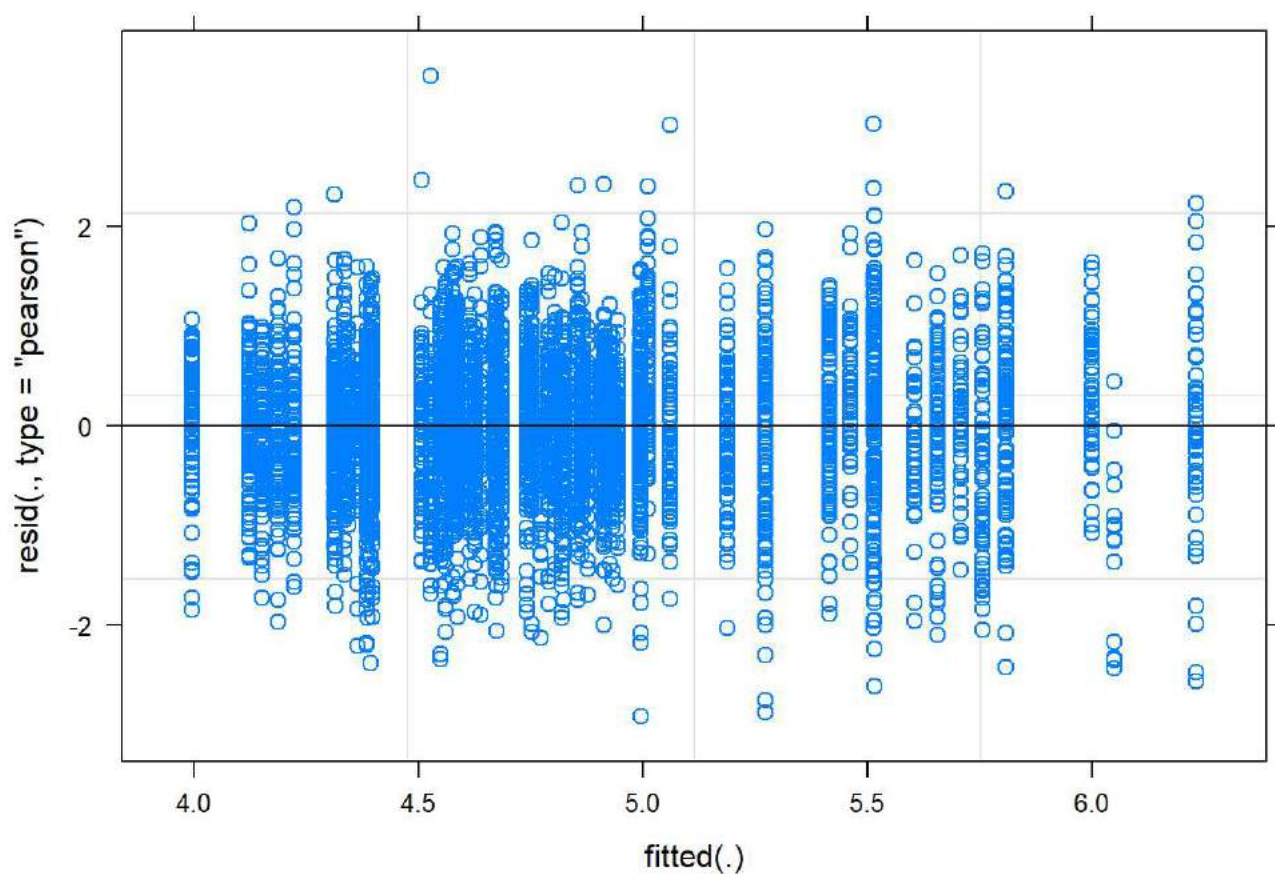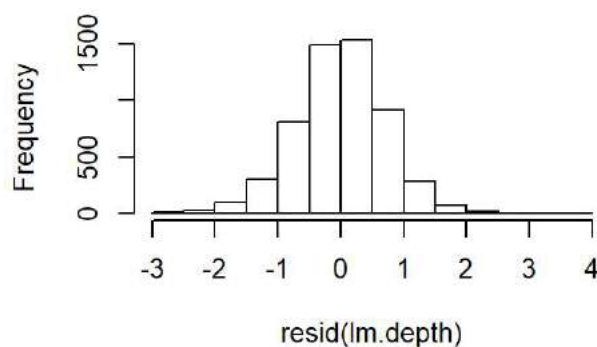
```
opar <- par(mfrow=c(2,2))
hist(tribulusPC$depth)
plot(fitted(lm.depth),resid(lm.depth))
abline(h=0,lty=2,col="red")
qqnorm(resid(lm.depth))
qqline(resid(lm.depth), col="red")
hist(resid(lm.depth))
```

Histogram of tribulusPC$depth

Normal Q-Q Plot

Histogram of resid(lm.depth)

```
par(opar)
```

Top

Summary Results Longest Spine

```
summary(lm.longspine)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## longest_spine ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##    Data: tribulusPC
##
## REML criterion at convergence: 16980.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.5275 -0.5493  0.0718  0.6419  3.9052
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  island:population (Intercept) 0.30468  0.5520
##  island            (Intercept) 0.34562  0.5879
##  year              (Intercept) 0.08652  0.2941
##  Residual                      1.22096  1.1050
## Number of obs: 5552, groups:  island:population, 25; island, 4; year, 3
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   5.3717     0.3584 4.5178   14.99 5.03e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing significance of each random effects:

```
ranova(lm.longspine)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## longest_spine ~ (1 | year) + (1 | island) + (1 | island:population)
##                       npar  logLik   AIC     LRT Df Pr(>Chisq)
## <none>                   5 -8490.5 16991
## (1 | year)               4 -8604.5 17217 228.03  1  < 2.2e-16 ***
## (1 | island)             4 -8495.1 16998   9.29  1   0.002309 **
## (1 | island:population)  4 -8843.1 17694 705.24  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

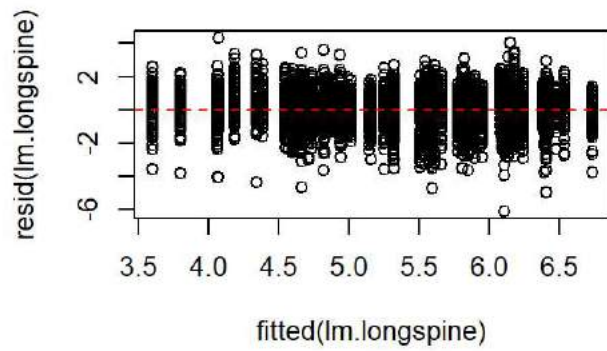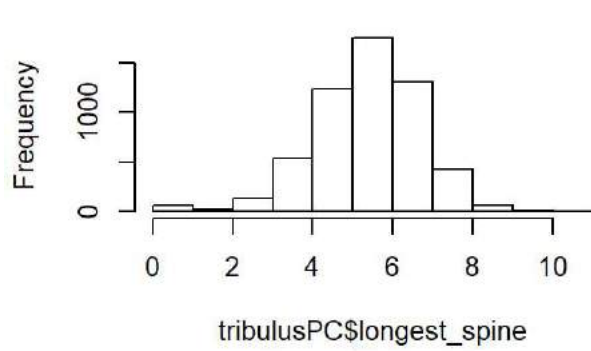## Checking assumptions *longest spine*:

```
plot(lm.longspine)
```
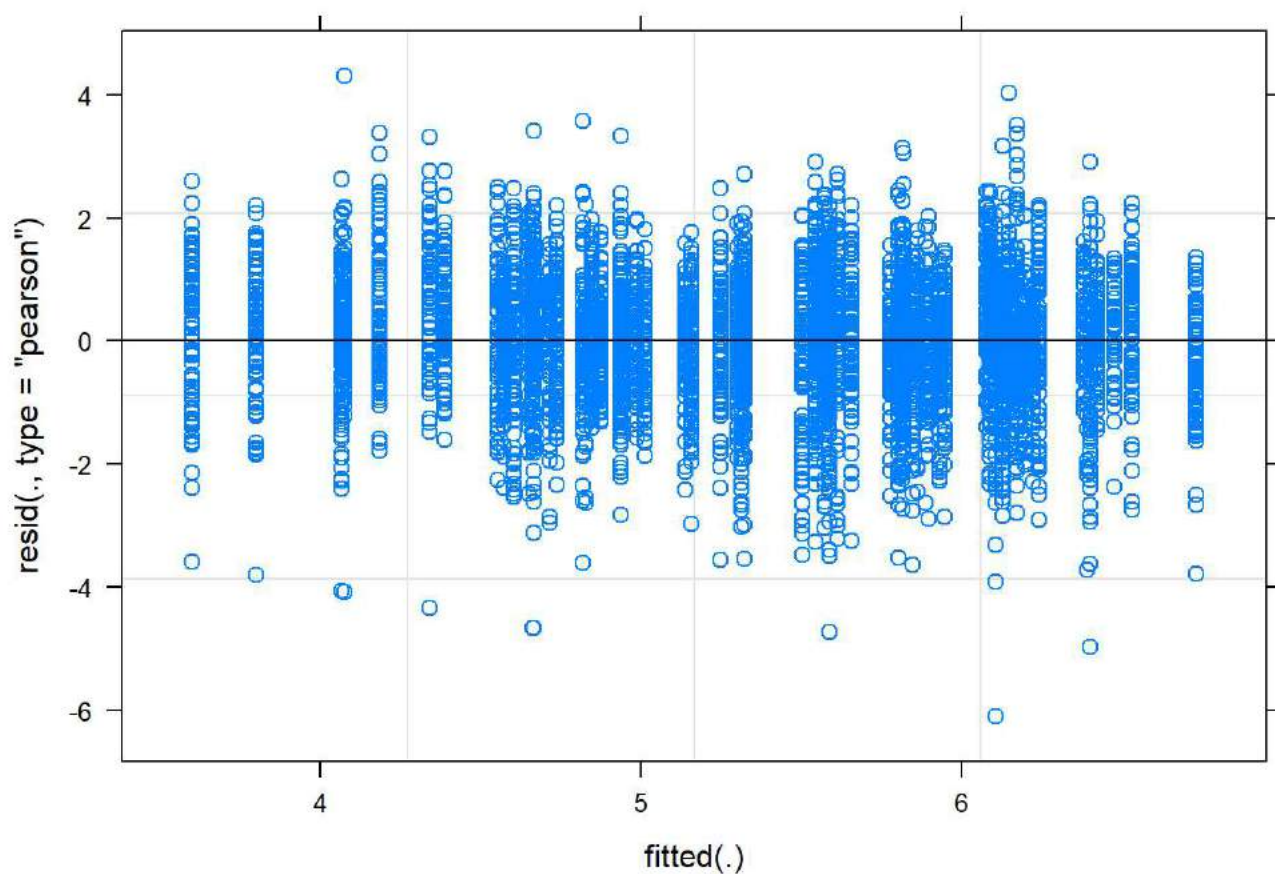
```
opar <- par(mfrow=c(2,2))
hist(tribulusPC$longest_spine)
plot(fitted(lm.longspine),resid(lm.longspine))
abline(h=0,lty=2,col="red")
qqnorm(resid(lm.longspine))
qqline(resid(lm.longspine), col="red")
hist(resid(lm.longspine))
```

Histogram of tribulusPC$longest_spine

Normal Q-Q Plot
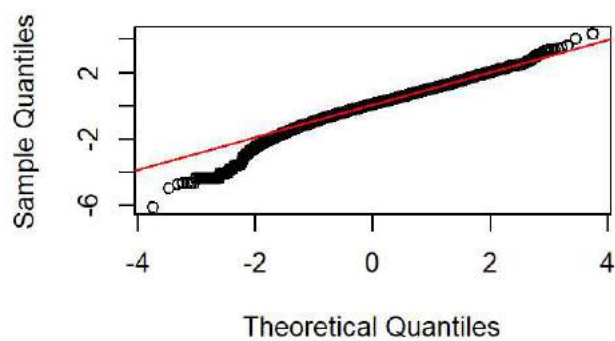
Histogram of resid(lm.longspine)


```
par(opar)
```
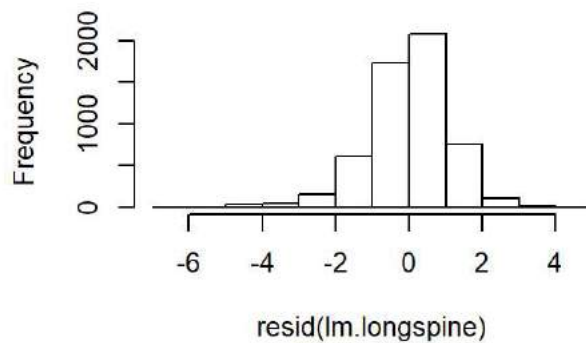
Top

Summary Results Spine Tip Distance

Spine tip distance model removed NAs only uses data from 2017-2018

```
summary(lm.tipdist)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## spine_tip_distance ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##    Data: tribulusPC
##
## REML criterion at convergence: 13918
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.2632 -0.5685  0.0254  0.6195  3.5055
##
## Random effects:
##  Groups            Name        Variance Std.Dev.
##  island:population (Intercept) 1.077    1.038
##  island            (Intercept) 1.690    1.300
##  year              (Intercept) 0.000    0.000
##  Residual                      3.697    1.923
## Number of obs: 3334, groups:  island:population, 25; island, 4; year, 3
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)  10.6863     0.6854 3.0548   15.59 0.000518 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Testing significance of each random effects:

```
ranova(lm.tipdist)
```

```
## boundary (singular) fit: see ?isSingular
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## spine_tip_distance ~ (1 | year) + (1 | island) + (1 | island:population)
##                        npar  logLik   AIC    LRT Df Pr(>Chisq)
## <none>                    5 -6959.0 13928
## (1 | year)                4 -6959.0 13926   0.00  1  1.0000000
## (1 | island)              4 -6965.1 13938  12.22  1  0.0004735 ***
## (1 | island:population)   4 -7171.6 14351 425.09  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

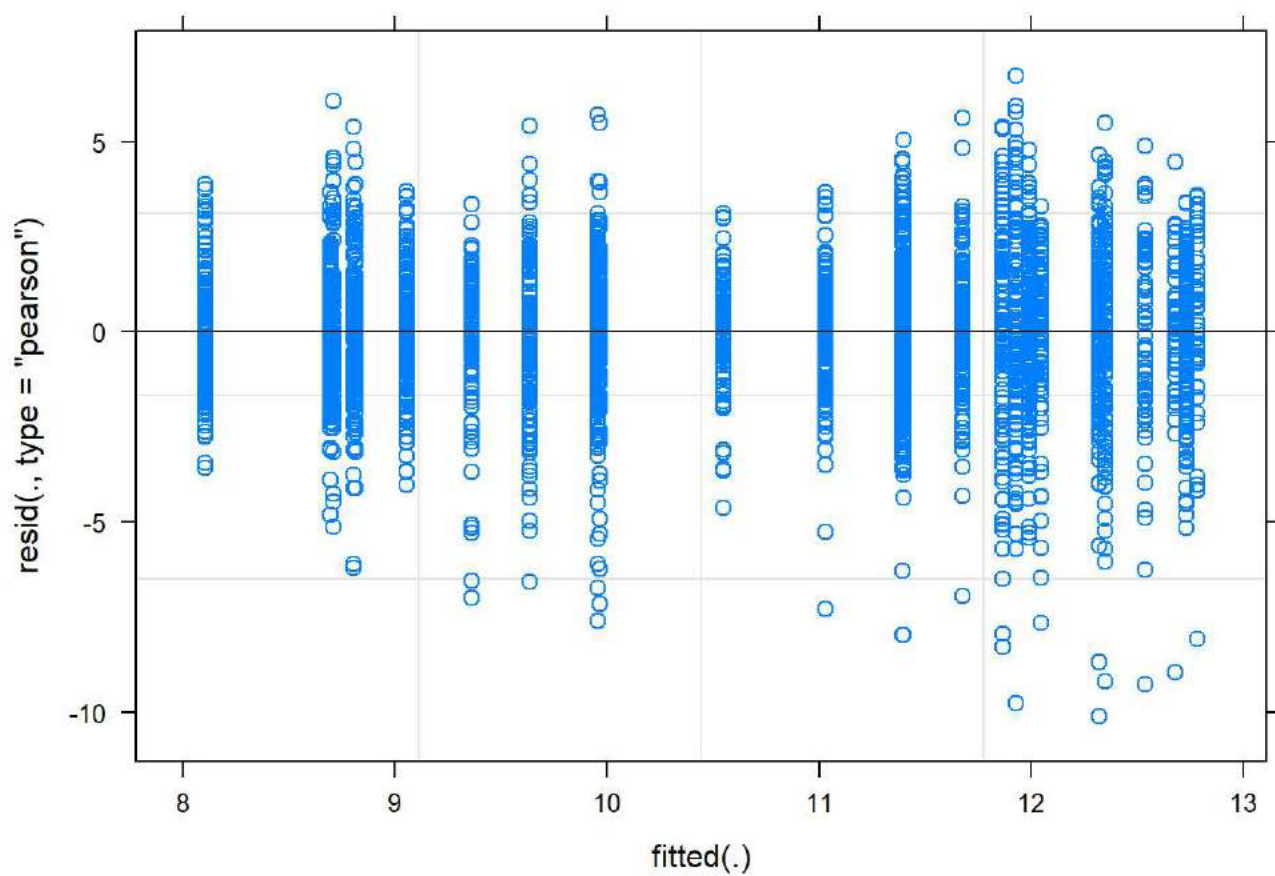## Checking assumptions *spine tip distance*:
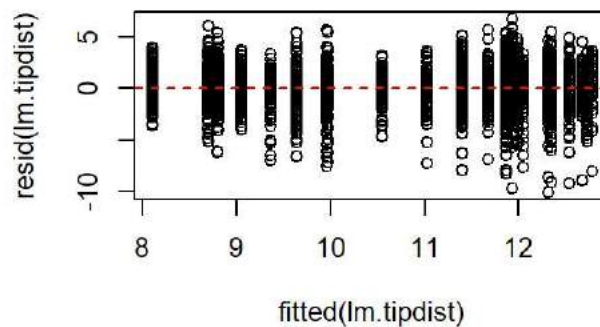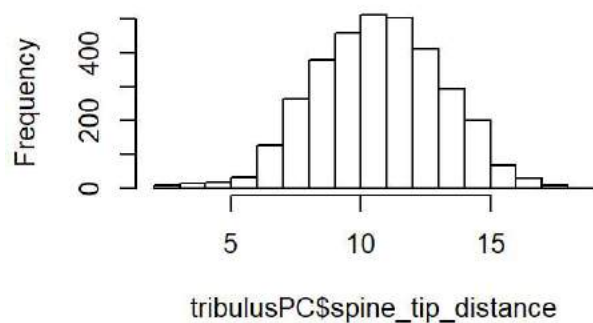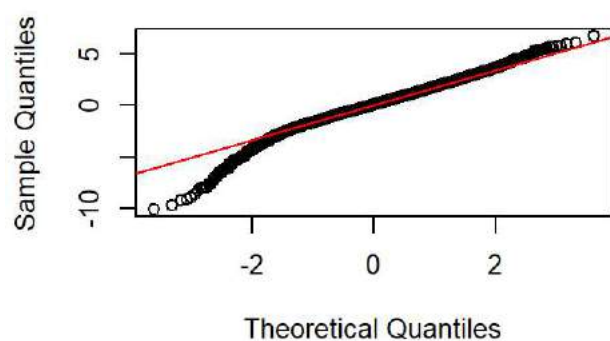
```
plot(lm.tipdist)
```

```
opar <- par(mfrow=c(2,2))
hist(tribulusPC$spine_tip_distance)
plot(fitted(lm.tipdist),resid(lm.tipdist))
abline(h=0,lty=2,col="red")
qqnorm(resid(lm.tipdist))
qqline(resid(lm.tipdist), col="red")
hist(resid(lm.tipdist))
```
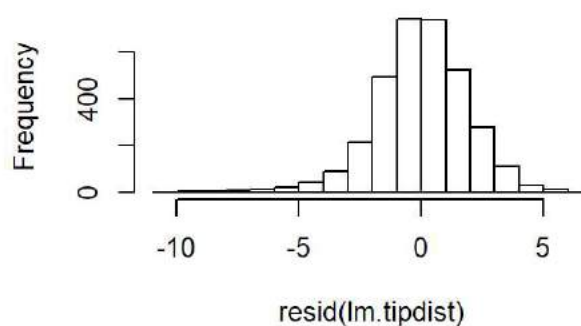
## Histogram of tribulusPC$spine_tip_distan

Summary Results Lower Spine

```
summary(lm.lowspines)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: lower_spine ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##    Data: tribulusPC
##
##      AIC      BIC   logLik deviance df.resid
##   4798.8   4825.2  -2395.4   4790.8     5548
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.1378 -0.3373  0.3023  0.5127 14.1088
##
## Random effects:
##  Groups            Name        Variance Std.Dev.
##  island:population (Intercept) 4.64714  2.1557
##  island            (Intercept) 2.35354  1.5341
##  year              (Intercept) 0.07092  0.2663
## Number of obs: 5552, groups:  island:population, 25; island, 4; year, 3
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1810     0.8992  -0.201     0.84
```

Lower Spine plot

```
plot(lm.lowspines)
```
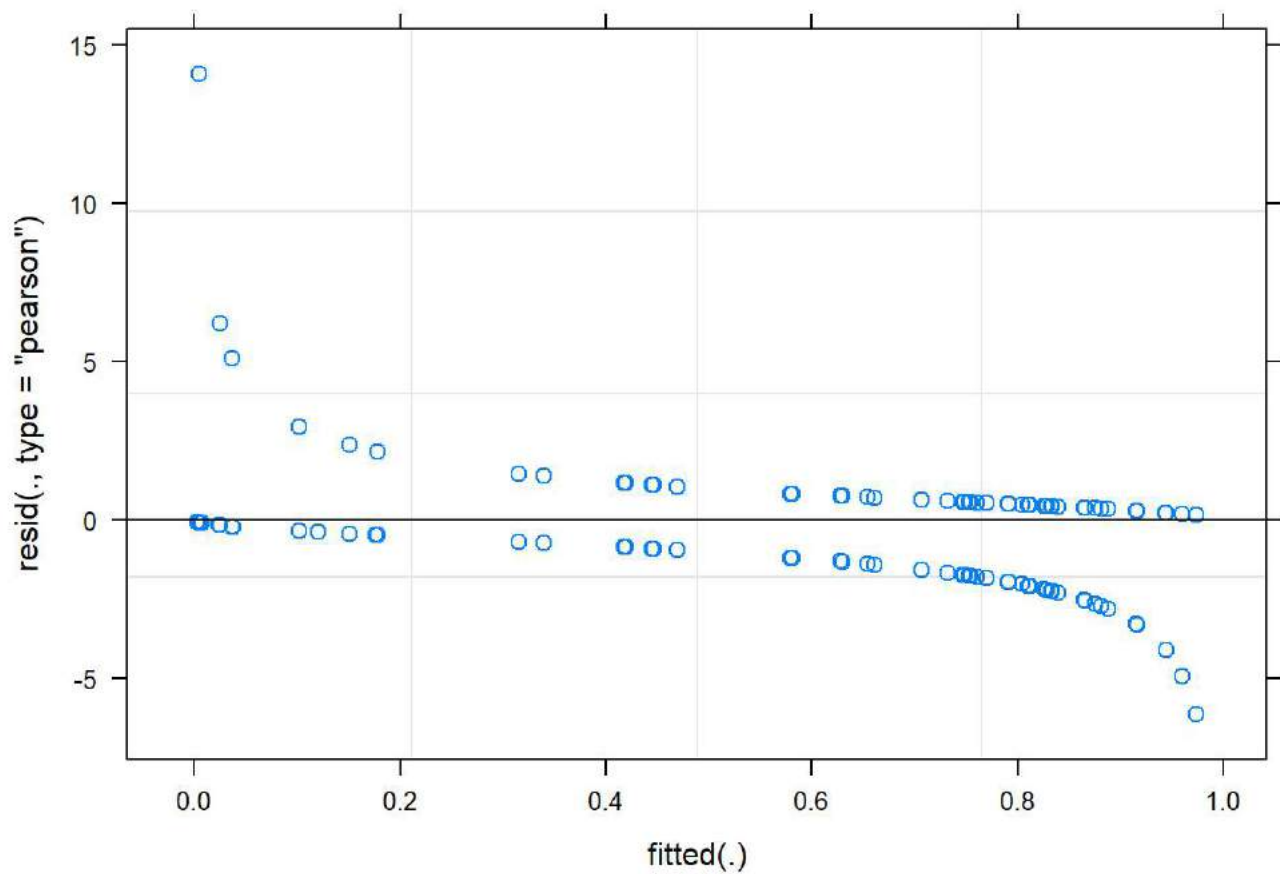
Testing significance of each random effects:

This was done manually given that lower spine is a binomial trait. Similar to Sofia's approach I created a model with all the random effects: year, island, island:population and then compared with other models removing each effect and test for significance.

Island:Population removed

```
summary(mred)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: lower_spine ~ 1 + (1 | year) + (1 | island)
##     Data: tribulusPC
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
##
##      AIC      BIC   logLik deviance df.resid
##   6060.1   6079.9  -3027.0   6054.1     5549
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0760 -0.4049  0.4969  0.6871  2.8873
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  island (Intercept) 1.54519  1.2431
##  year   (Intercept) 0.02984  0.1727
## Number of obs: 5552, groups:  island, 4; year, 3
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.1356     0.6290   0.216    0.829
```

```
anova(mfull, mred)
```

```
## Data: tribulusPC
## Models:
## mred: lower_spine ~ 1 + (1 | year) + (1 | island)
## mfull: lower_spine ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mred     3 6060.1 6079.9 -3027.0   6054.1
## mfull    4 4798.8 4825.2 -2395.4   4790.8 1263.3  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Island removed

```
summary(mred2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: lower_spine ~ 1 + (1 | year) + (1 | island:population)
##    Data: tribulusPC
##
##      AIC      BIC   logLik deviance df.resid
##   4801.9   4821.8  -2397.9   4795.9     5549
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.1372 -0.3396  0.3029  0.5140 13.9500
##
## Random effects:
##  Groups            Name        Variance Std.Dev.
##  island:population (Intercept) 7.48875  2.7366
##  year              (Intercept) 0.06921  0.2631
## Number of obs: 5552, groups:  island:population, 25; year, 3
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1649     0.5757  -0.286    0.775
```

```
anova(mfull, mred2)
```

```
## Data: tribulusPC
## Models:
## mred2: lower_spine ~ 1 + (1 | year) + (1 | island:population)
## mfull: lower_spine ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mred2     3 4801.9 4821.8 -2397.9   4795.9
## mfull     4 4798.8 4825.2 -2395.4   4790.8 5.1273  1    0.02355 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Population removed

```
summary(mred3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: lower_spine ~ 1 + (1 | island) + (1 | island:population)
##    Data: tribulusPC
##
##      AIC      BIC   logLik deviance df.resid
##   4824.6   4844.4  -2409.3   4818.6     5549
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.4260 -0.3808  0.3277  0.5179 11.7428
##
## Random effects:
##  Groups            Name        Variance Std.Dev.
##  island:population (Intercept) 4.625    2.151
##  island            (Intercept) 2.311    1.520
## Number of obs: 5552, groups:  island:population, 25; island, 4
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1790     0.8844  -0.202     0.84
```

```
anova(mfull, mred3)
```

```
## Data: tribulusPC
## Models:
## mred3: lower_spine ~ 1 + (1 | island) + (1 | island:population)
## mfull: lower_spine ~ 1 + (1 | year) + (1 | island) + (1 | island:population)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mred3    3 4824.6 4844.4 -2409.3   4818.6
## mfull    4 4798.8 4825.2 -2395.4   4790.8 27.814  1  1.335e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Top