Jellyfish Experiment Analysis

Jack Cunningham

Goals

The goal of this analysis is to assess how temperature and water treatments impact various measures of Jellyfish quality, specifically their physical characteristics, ability to reproduce or advance to the next life stage and DNA.

Experimental Design

A sample of 50 Polyps are placed into 18 beakers. Those 18 beakers are evenly distributed across 6 treatments. There are two crossed factors, Water Temperature with three levels: (20 degrees 'low', 28 degrees 'ambient', 33 degrees 'high)' and Microbial Condition with two levels: (ASW 'Regular Artificial Salt Water', sterilized Antibiotic treated ASW 'Antibiotic treated Artificial Salt Water').

Measurements with different methods are performed. I have broken them up as follows.

Count Measurements

On a biweekly basis researchers count the number and stage of jellyfish development for each beaker. There are three stages tracked, Buds, Ephyra and Polyps.

Physical Measurements

On a weekly basis researchers measure each Polyp's width and Ephyra's diameter.

DNA Measurements

DNA from sampled Polyps are extracted at the start and end of the experiment. DNA is sampled from each beaker's water at the end of the experiment.

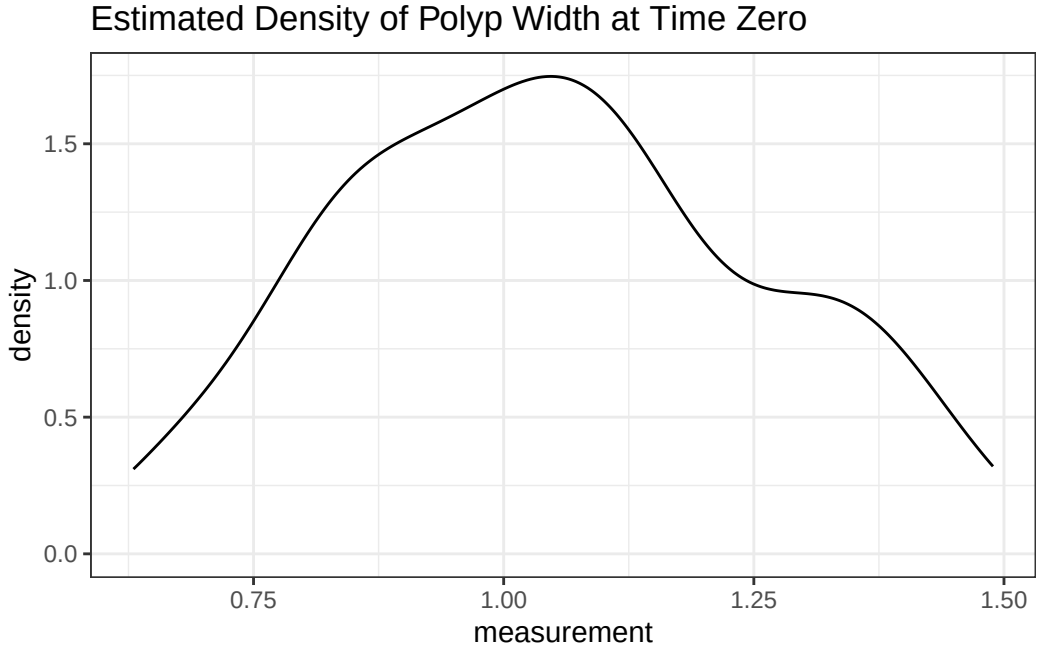
Analysis Results

Polyp Width Data

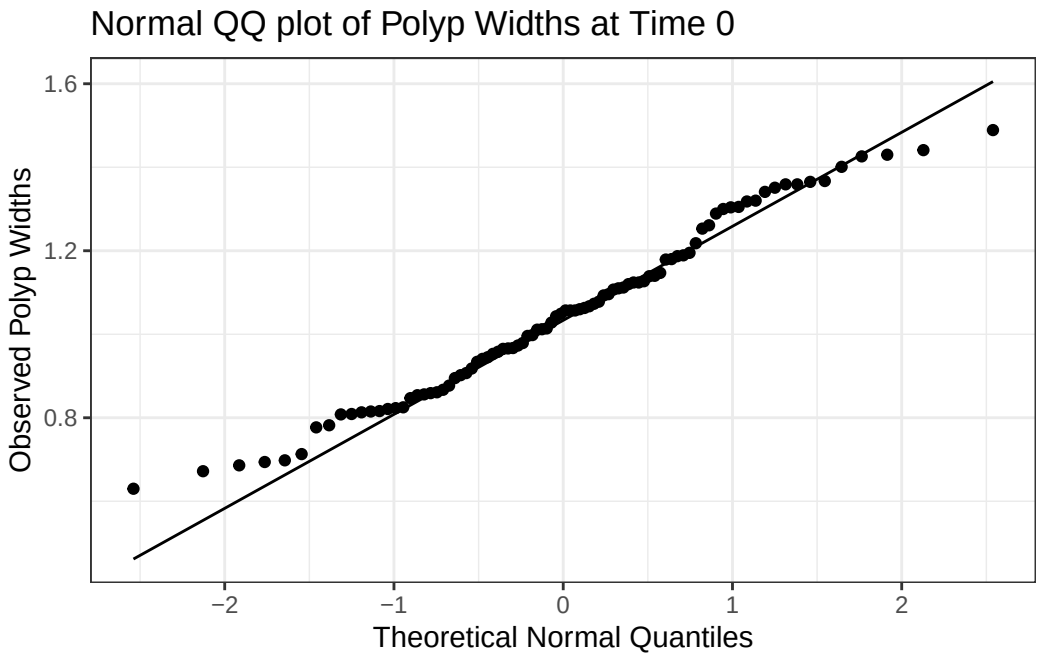
Are all the T0 samples the same (not significantly different)

This question asks whether the polyp samples have widths that are not significantly different than each other.

T0 samples will not have the exact same width, but since they are all subject to the same initial conditions they should be drawn from the same distribution. First we look at the estimated density function of polyp widths at time zero.



This looks close to a normal distribution. I use a normal QQ plot to visually inspect further.



Since the observed data follows the QQ line closely it appears the Polyp width for time 0 samples are from the same normal distribution.

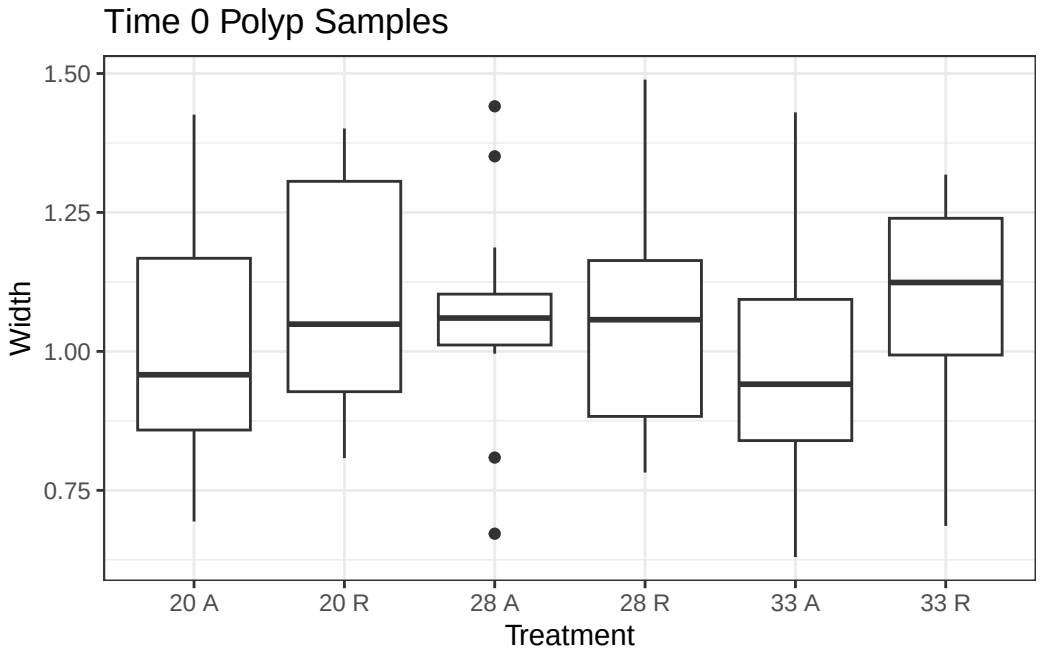
We can formally test whether this sample is from a normal distribution by using a Shapiro-Wilk test. So, we are testing:

H_0 : Sample Obtained from Normal Distribution, H_a : Sample Not Obtained from a Normal Distribution

```
    Shapiro-Wilk normality test

data:  width_time_0
W = 0.9791, p-value = 0.1562
```

With a p-value of .1562 we cannot reject the null hypothesis that this sample is from a normal distribution.

These results have shown that there is no evidence that time 0 polyp samples are drawn from separate distributions. But it would also be reassuring if we could show that the width of time 0 polyp samples are similar across treatments.



The median of polyp samples at time 0 appears consistent across treatments.

To formally test whether sample widths are different depending on treatments we fit the model:

$$y_{ijk} = \mu + w_i + t_j + w_i t_j + e_{ijk}, \quad i = 1, 2, \quad j = 1, 2, 3 \quad k = 1, ..., 15$$

Where y_{ijk} is the width of the polyp sample, w_i is the water treatment i , t_j is temperature treatment j.

We obtain an ANOVA table to test:

$$H_0 : w_i = 0, t_j = 0, w_i t_j = 0 \quad H_a : w_i \neq 0, t_j \neq 0, w_i t_j \neq 0$$

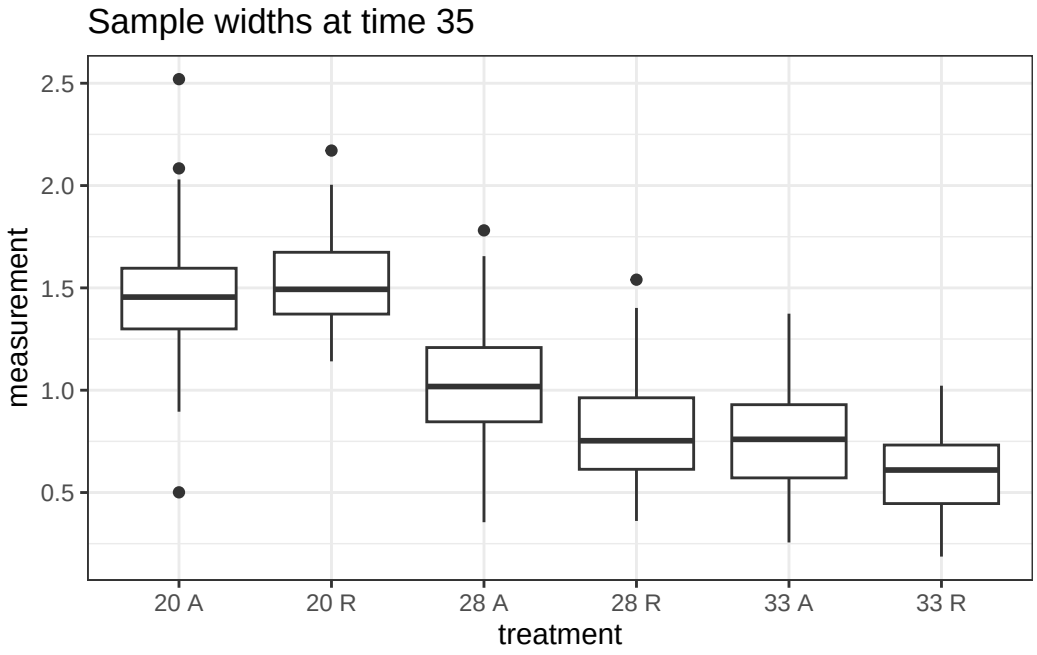
```
Analysis of Variance Table

Response: measurement
                  Df Sum Sq  Mean Sq F value Pr(>F)
water              1 0.0794 0.079388  1.8172 0.1813
temperature        2 0.0109 0.005434  0.1244 0.8832
water:temperature  2 0.0579 0.028967  0.6631 0.5179
Residuals         84 3.6696 0.043686
```

We cannot reject the null hypothesis that there is no difference in sample means across treatments.

All evidence points in the direction of time 0 polyp samples being the same (not significantly different).

Are the T35 samples significantly different from one another across treatments?

After treatments are applied, samples are drawn again.

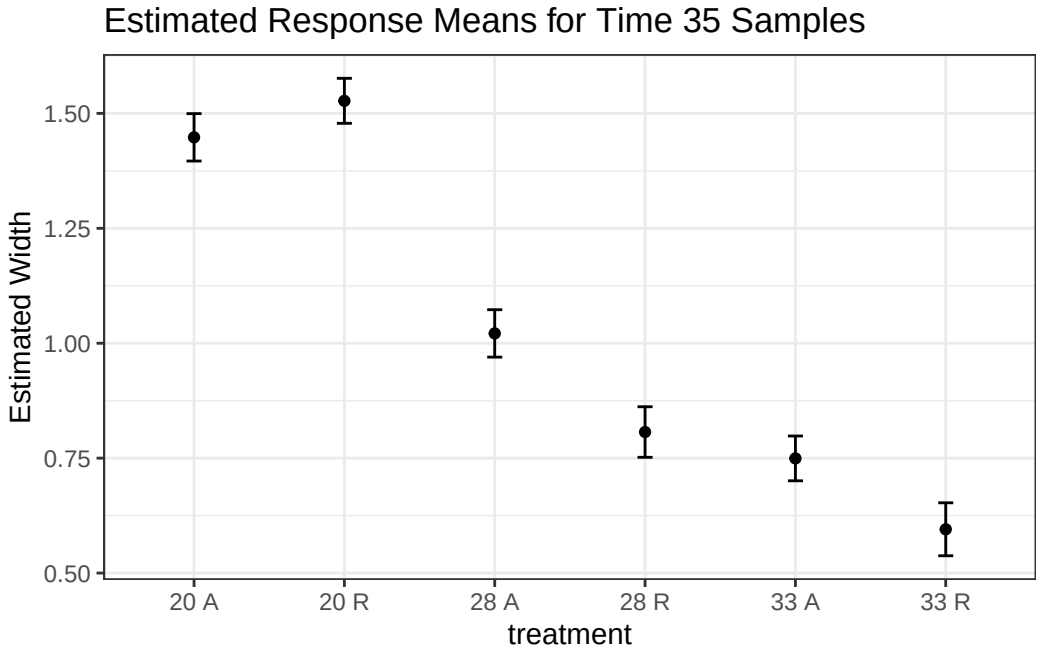


A box-plot indicates that there does appear to be treatment effects.

We formally test this using the same model as before, but on the time 35 samples.

```
Analysis of Variance Table

Response: measurement
                   Df Sum Sq Mean Sq  F value    Pr(>F)
water               1  0.104   0.104   1.6592    0.1983
temperature         2 64.222  32.111 511.7473 < 2.2e-16 ***
water:temperature   2  2.209   1.105  17.6053 3.949e-08 ***
Residuals         532 33.382   0.063
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We reject the null hypothesis of $w_i t_j = 0$ at any reasonable choice of a p-value. There is a treatment effect and an interaction effect between water and temperature.



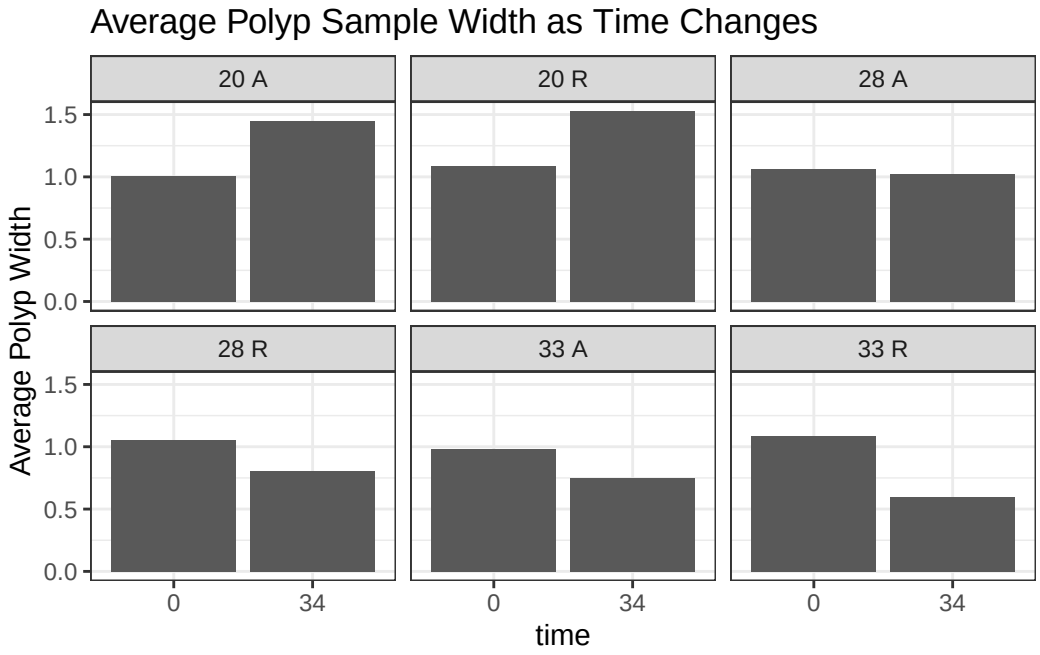
A plot of the estimated response means shows that, as reported in our anova table, the interaction between temperature and water is significant. For temperature 20 regular water and artificial water have similar widths, in the other two temperatures artificial water leads to higher widths.

The biggest takeaway is the impact of temperature, as temperature gets higher the widths of polyps get smaller.

T35 samples are significantly different from one another across treatments.

Are there differences between T0 and T35 between temperatures and microbe treatments

```
`summarise()` has grouped output by 'time'. You can override using the
`.groups` argument.
```



There is initial evidence pointing towards differences between treatments, for low temperatures widths seem to increase for high temperatures widths seem to decrease. The impact of water treatment seems to depend on the temperature of the water.

With these findings, we fit the model:

$$y_{ijkl} = \mu + w_i + t_j + w_i t_j + s_k + s_k w_i + s_k t_j + e_{ijkl}, \quad i = 1, 2, \quad j = 1, 2, 3 \quad k = 1, 2, \quad l = 1, ..., n_{ijk}$$

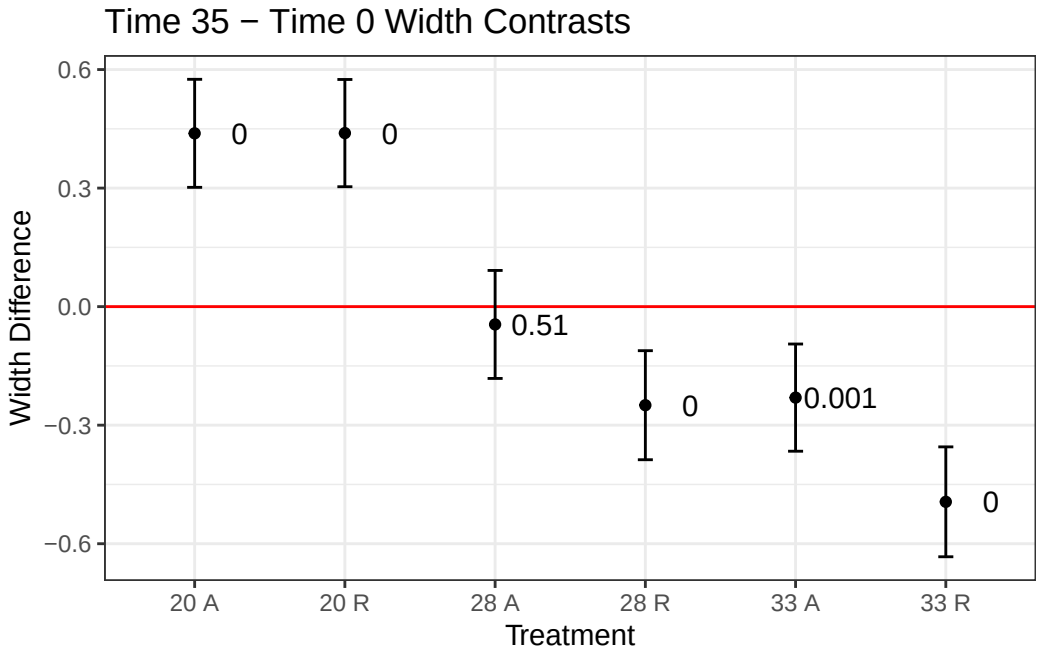
Where y_{ijkl} is the width of the polyp sample, w_i is the water treatment i , t_j is temperature treatment j, s_k is an indicator variable for whether the sample was performed at the start (time 0) or end (time 35) of the treatment. Of note n_{ijk}, the number of observations in each treatment, time group is uneven as there as more samples are conducted at the end of the experiment.

We will first test the significance of $s_k, s_k w_i, s_k t_j$ and then conduct contrasts between each treatment at time 0 and time 35.

```
Anova Table (Type III tests)

Response: measurement
                     Sum Sq  Df  F value    Pr(>F)
(Intercept)          17.761   1 295.2798 < 2.2e-16 ***
temperature           0.011   2   0.0873    0.9164
```

```
water                    0.047   1   0.7743    0.3792
time                     2.519   1  41.8837 1.978e-10 ***
temperature:water        0.058   2   0.4816    0.6180
temperature:time         6.009   2  49.9502 < 2.2e-16 ***
water:time               0.000   1   0.0000    0.9950
temperature:water:time   0.248   2   2.0611    0.1282
Residuals               37.051 616
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that the two-way interaction between the sample time s_k and temperature t_j is significant. This indicates that the effect depends on temperature. The two-way interaction between the sample time s_k and water treatment w_i is not significant.



We have plotted the contrasts of time 35 - time 0 polyp widths.

We test:

$$H_0 : T35 - T0 = 0; T_{35} - T0 \neq 0$$

To the right of each confidence interval we have the tukey adjusted p-value of this test.

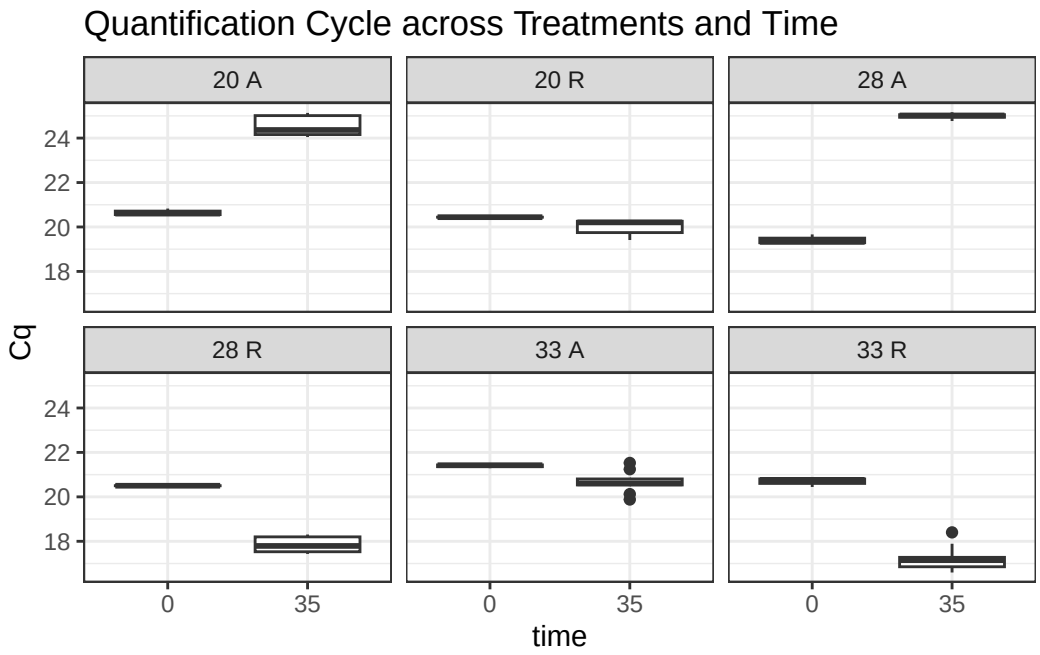
The only treatment that does not differ over time is the treatment where temperature is 28 and water is antibiotic treated. All other treatments have differences between time 0 and time 35 between treatments.

For most treatments there is a difference between time 0 and time 35, but the nature of that difference depends primarily on the temperature of the water. For low temperatures the width of polyps are larger after the treatments are applied, for high temperatures the widths are smaller.

DNA Data

Are there differences between samples significantly different (across temperatures and microbe) significantly different from one another?

DNA samples were done on polyps and water. We start with the polyp samples.



The response variable C_q is the Quantification Cycle, the cycle number where we can detect bacterial load. A low C_q indicates a high presence of bacterial load.

It appears that antibiotic treated water leads to lower C_q compared to regular water across all treatments.

Time 0 samples are all quite similar to each other, so we focus our analysis to time 35 samples.

So we fit the model:

$$y_{ijk} = \mu + w_i + t_j + w_i t_j + e_{ijk}, \quad i = 1, 2, \quad j = 1, 2, 3 \quad k = 1, .., 9$$

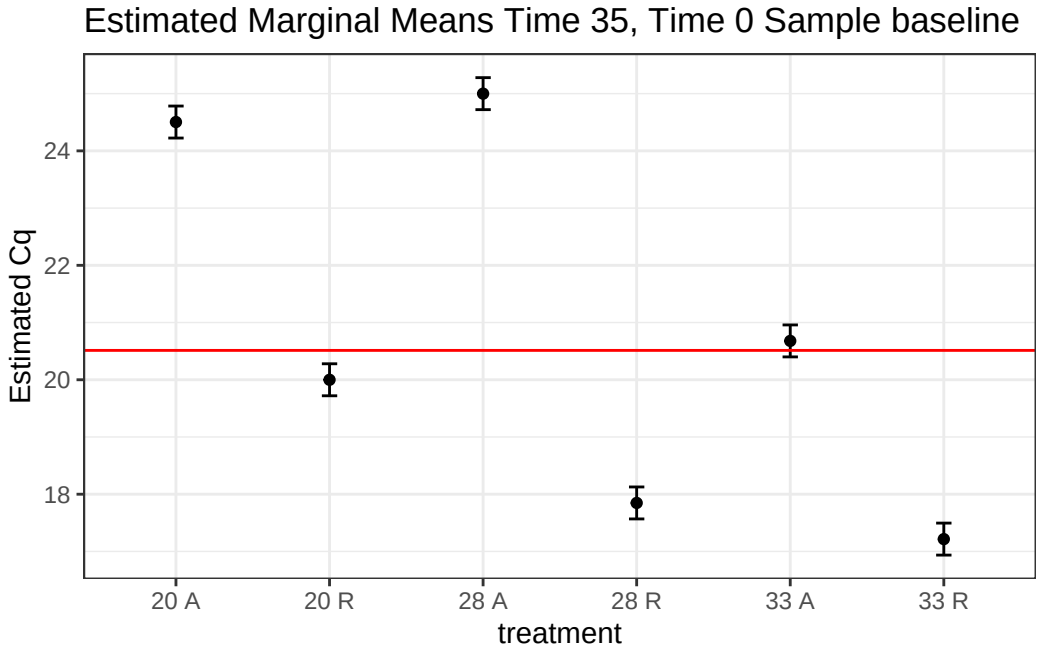
Where y_{ijkl} is the C_q measure of the sample, w_i is the water treatment i , and t_j is temperature treatment j.

We use the ANOVA table to look for the significance of each effect.

Then look at the contrasts across main effects.

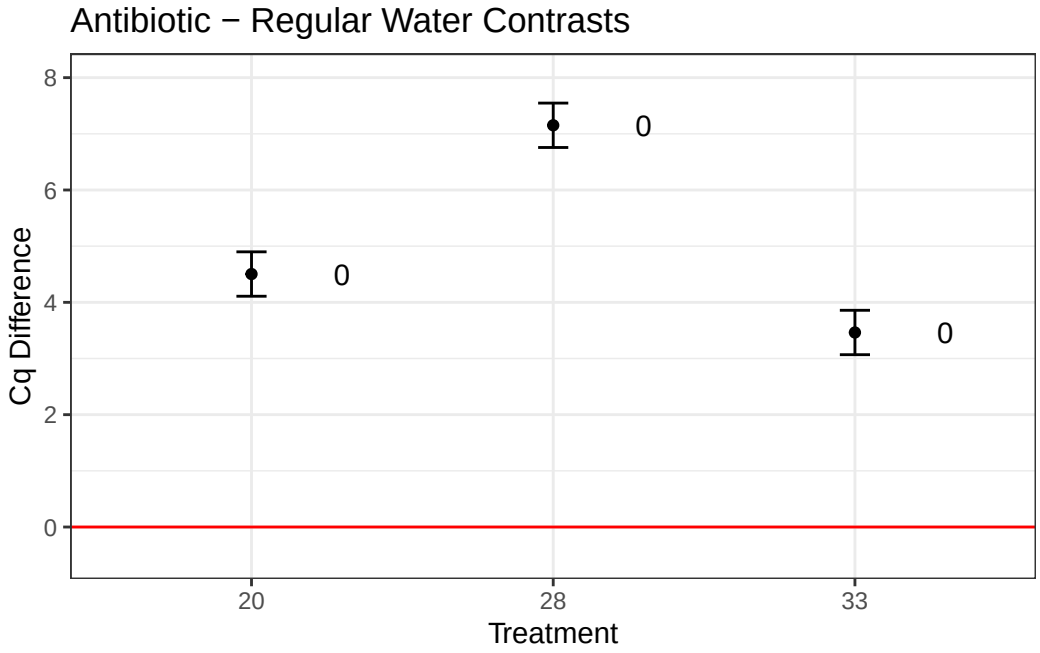
```
Analysis of Variance Table

Response: Cq
                  Df Sum Sq Mean Sq  F value    Pr(>F)
water              1 342.88  342.88 1951.148 < 2.2e-16 ***
temperature        2 106.41   53.20  302.751 < 2.2e-16 ***
water:temperature  2  32.56   16.28   92.639 < 2.2e-16 ***
Residuals         48   8.44    0.18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimated means are plotted:

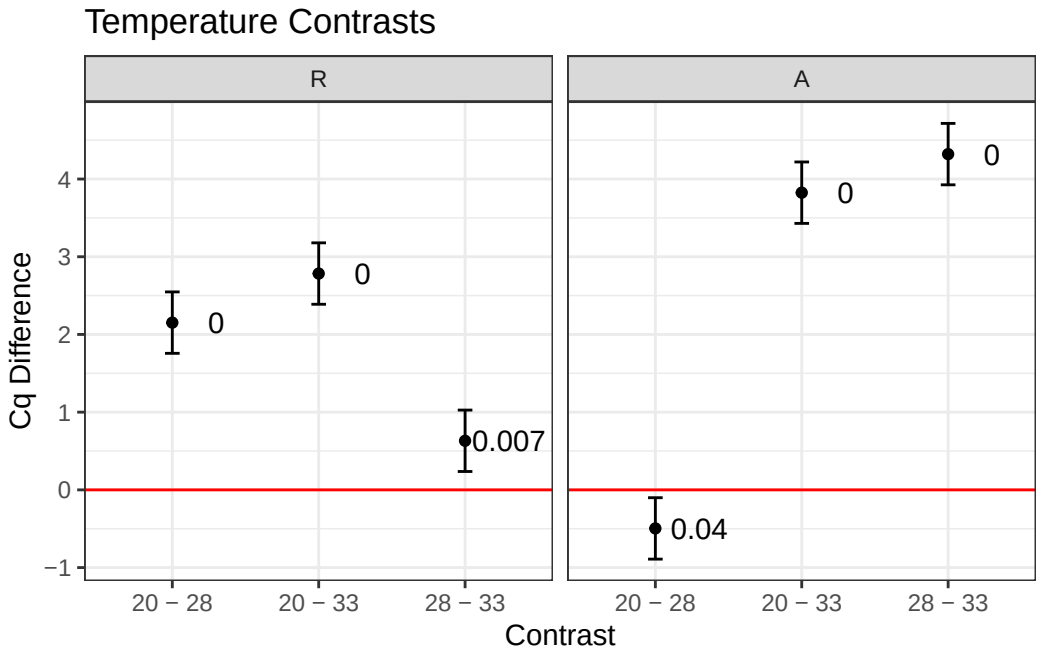


We can see that the antibiotic treated water always has a C_q larger than regular water in the same temperature. As temperature increases C_q generally drops.

The water, temperature effect is significant. So we will look at the main effect of water at each level of temperature. And the look at the main effect of temperature at each level of water.



We can see that antibiotic water has much higher C_q compared to regular water at each temperature. All tests of $H_0 : C_i = 0$ are rejected at any reasonable p-value. The effect of water treatment is great.



Given untreated water, C_q is significantly different across all temperature combinations at a p-value of 0.05. As differences in temperature get larger, the differences in C_q get larger as well. Colder temperatures have higher C_q.

Given treated water, C_q is significantly different across all treatment combinations at a p-value 0.05. The nature of that comparison depends on the contrast. The 20 - 28 contrast is negative, in this case the colder temperature has a lower C_q. The two other contrasts continue the general pattern we have seen, as temperature gets colder C_q increases.

If we were seeking to minimize bacterial load, and thus maximize C_q we should choose the temperature of 33 and treat the water with antibiotics.

The general trend holds for water samples, so I have left out that analysis.

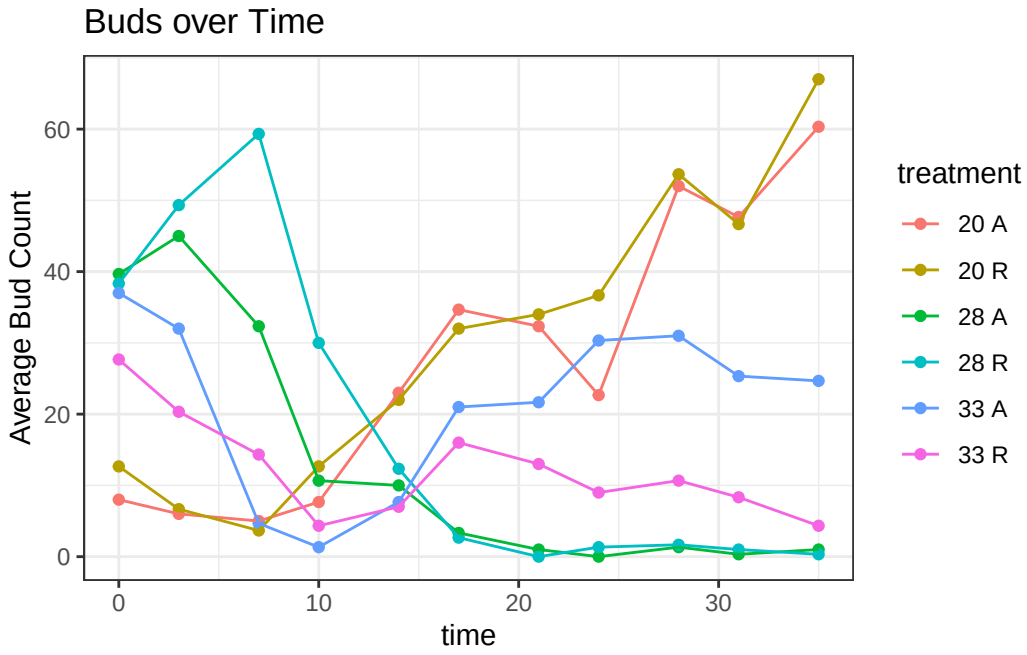
I would also recommend using C_q as the measurement of bacterial load, as its close to normal distribution allows for easy comparisons.

Count Data

Are trends across temperatures significant?

This part of the analysis is focused mainly on buds. A complication in the analysis is that as jellyfish develop through life stages they lose the ability to asexually reproduce. So assessment of which treatment to choose depends on whether the goal is to encourage reproduction, or encourage development.

```
`summarise()` has grouped output by 'treatment'. You can override using the
`.groups` argument.
```



An initial analysis shows that temperatures appear to have different trends over time. For example, in the temperature of 20 we don't see any development into Ephyra, so we have more polyps to reproduce. This leads to an increasing trend in buds.

Generally the difference between regular and antibiotic treated water is not large.

A model to demonstrate the trends across temperatures are significant is:

$$y_{ijkl} = \text{temp}_i + \text{time}_j + \text{temp}_i\text{time}_j + \text{beaker}(\text{time}_j)_k, \quad i = 1,2,3 \quad j = 1,3,7,10,...35 \quad k = 1,2,3$$

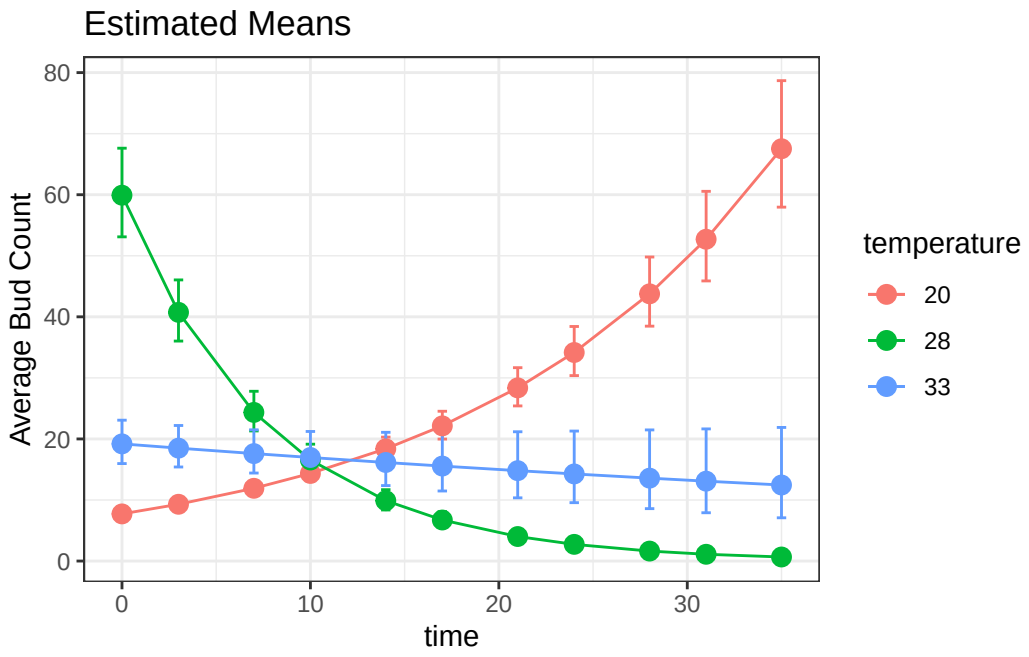
The beaker, which is nested inside time, is a random effect. We use poisson regression to adjust for the fact that the response is a count variable. We drop the water effect, because if we use it we run against issues with over populating the model.

```
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: bud
                Chisq Df Pr(>Chisq)
temperature    67.737  2  1.955e-15 ***
```

```
time                17.256  1  3.267e-05 ***
temperature:time 403.243  2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the interaction between temperature and time is significant, we can state that trend differences across temperatures are significant.

The estimated means from the model are below:



At temperature 20 we quadratic growth, at temperature 33 we see a negative linear relationship with time. And at temperature 28 we see a concave upward decreasing curve.

Its important to note that this is highly related to whether polyps are developing into a new stage, or staying static and continuing to reproduce.

Are the differences within a temperature between microbe treatments significant?

As previously mentioned, we could not add the water treatment to the temporal model. So we attempt to answer this question by collapsing the data across time and modeling in a different way.

Since we are averaging over 11 observations of time, the bud variable approaches normality. So we can fit a weighted linear model, adjusting for $Var(y_{ijk}) = \sigma^2/11$.

Since we saw some evidence that the relationship between water and temperature exists, particularly for temperature 33, we include the interaction term.

So our model is:

$$y_{ijk} = \mu + w_i + t_j + w_i t_j + e_{ijk}, \quad i = 1, 2, 3 \quad j = 1, 2 \quad k = 1, 2, 3$$

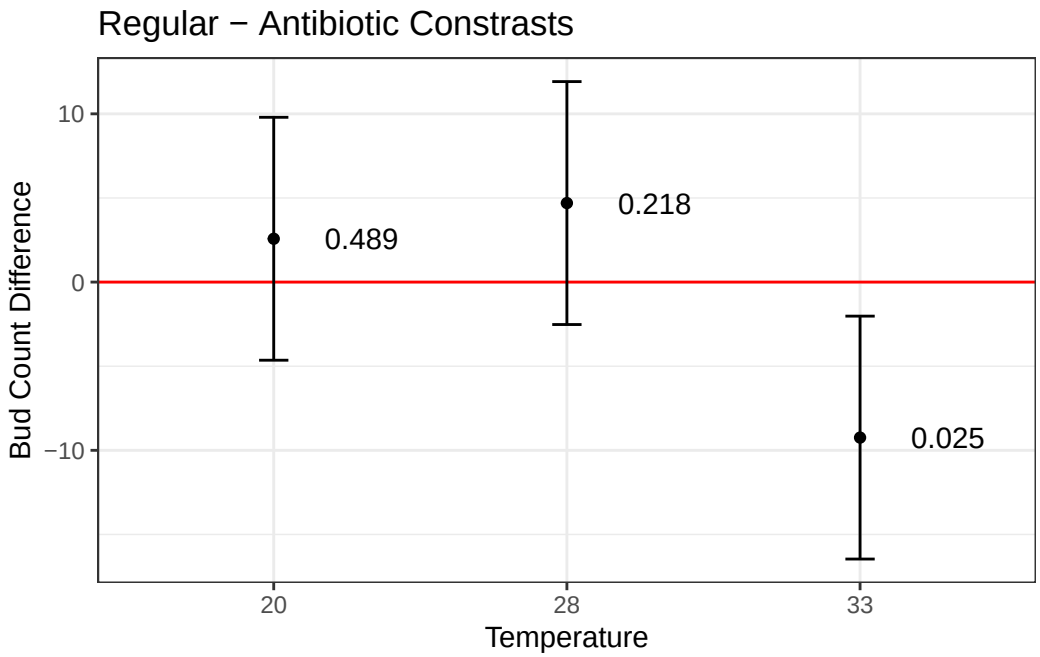
Where y_{ijk} is the average bud count for beaker k with treatment of water w_i and temperature t_j and $Var(e_{ijk}) = \sigma^2/11$.

First we fit the model and analyze the significance of terms, then we will look at contrasts between water treatments.

```
`summarise()` has grouped output by 'temperature', 'water'. You can override
using the `.groups` argument.
```

```
Analysis of Variance Table

Response: avg_bud
                  Df Sum Sq Mean Sq F value    Pr(>F)
water              1  0.176  0.1764  0.0993 0.7581417
temperature        2 55.572 27.7858 15.6380 0.0004546 ***
water:temperature  2 15.385  7.6926  4.3294 0.0384102 *
Residuals         12 21.322  1.7768
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the water, temperature interaction term is significant we need to analyze the contrasts between water at each temperature.



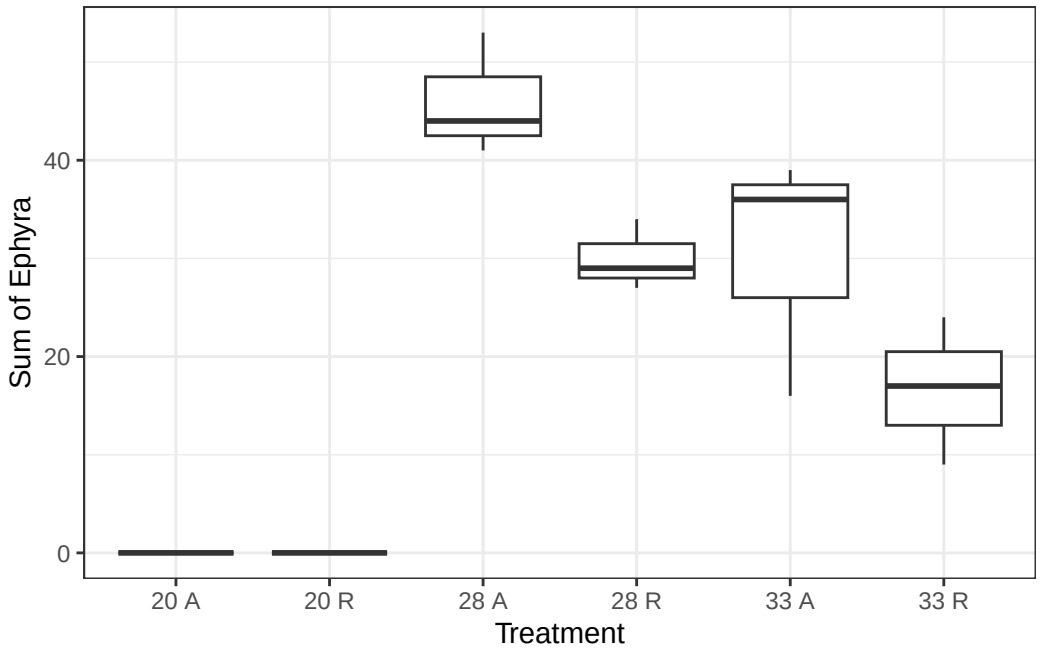
Temperature 33 is the only time the water treatment effect contrast is significantly different than 0. That contrast found regular water to be associated with less buds than antibiotic water.

Ephyra Data

What factors are significant for determining the number of Ephyra produced?

Unfortunately in the data provided we are unable to obtain an accurate sum of ephyra in each beaker since the count data will include duplicate counts. However, we can model the total sum as an estimate.

```
`summarise()` has grouped output by 'temperature', 'water'. You can override
using the `.groups` argument.
```



On an initial pass it appears both water and temperature are strong effects in determining Ephyra count. As previously noted, there is no Ephyra production at temperature 20. There does not appear to be an interaction between water and temperature, the drop off in count seems consistent under both temperatures.

For the sake of modeling, we leave out the temperature of 20 data since we are sure to run into a problem with uneven variance:

We fit the model:

$$y_{ijk} = \mu + w_i + t_j + w_i t_j + e_{ijk}, \quad i = 1, 2 \quad j = 1, 2, 3 \quad k = 1, 2, 3$$

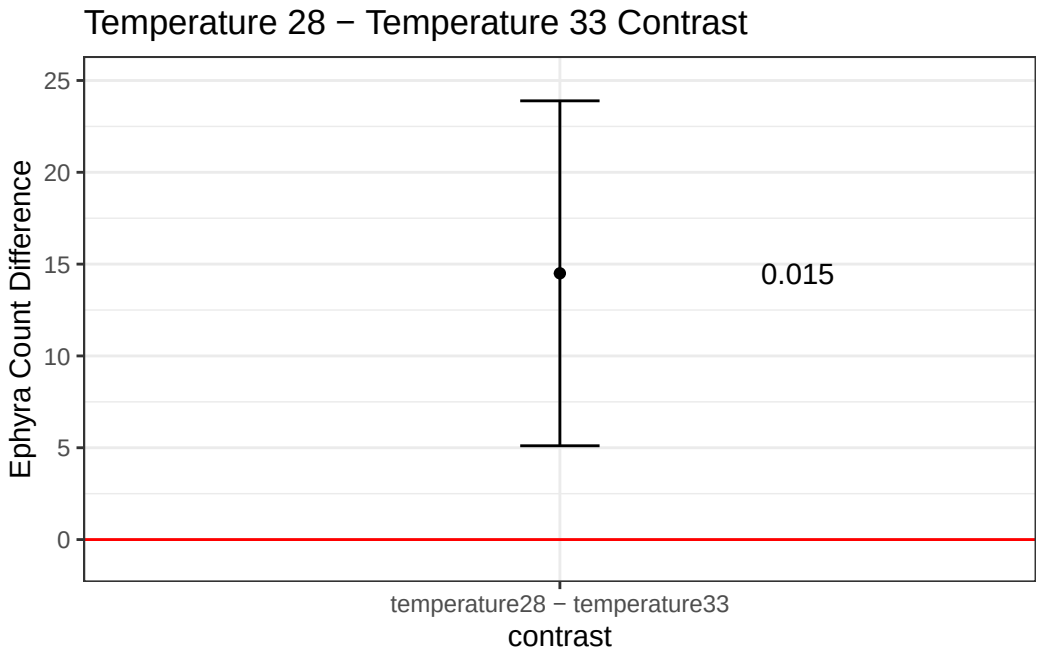
First, we use an ANOVA table to see what effects are significant:

```
Analysis of Variance Table

Response: ephyra_sum
                 Df Sum Sq Mean Sq F value  Pr(>F)
water             1 660.08  660.08  9.9761 0.01343 *
temperature       1 630.75  630.75  9.5327 0.01495 *
water:temperature 1   4.08    4.08  0.0617 0.81007
Residuals         8 529.33   66.17
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As figured from the visual inspection, the interaction term is not significant. Both the water and temperature effects are significant.

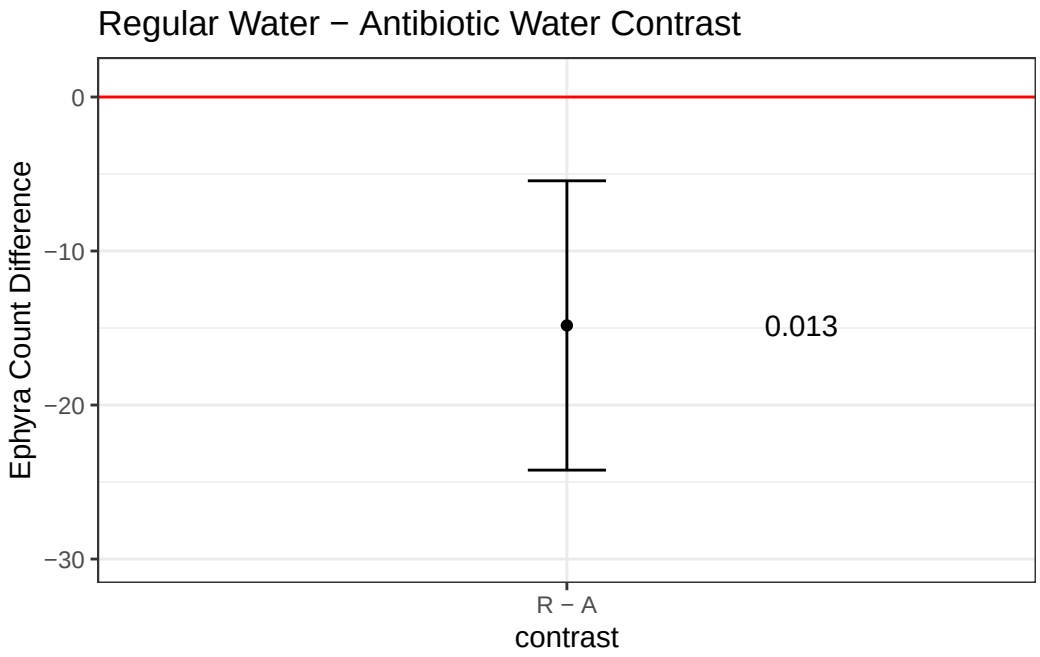
We look at the contrasts of both:

```
NOTE: Results may be misleading due to involvement in interactions
```



The contrast of the Ephyra count between temperature 28 and temperature 33 is significantly different from 0 at a p-value of 0.05. Temperature 28 is associated with 15 more buds than temperature 33.

```
NOTE: Results may be misleading due to involvement in interactions
```

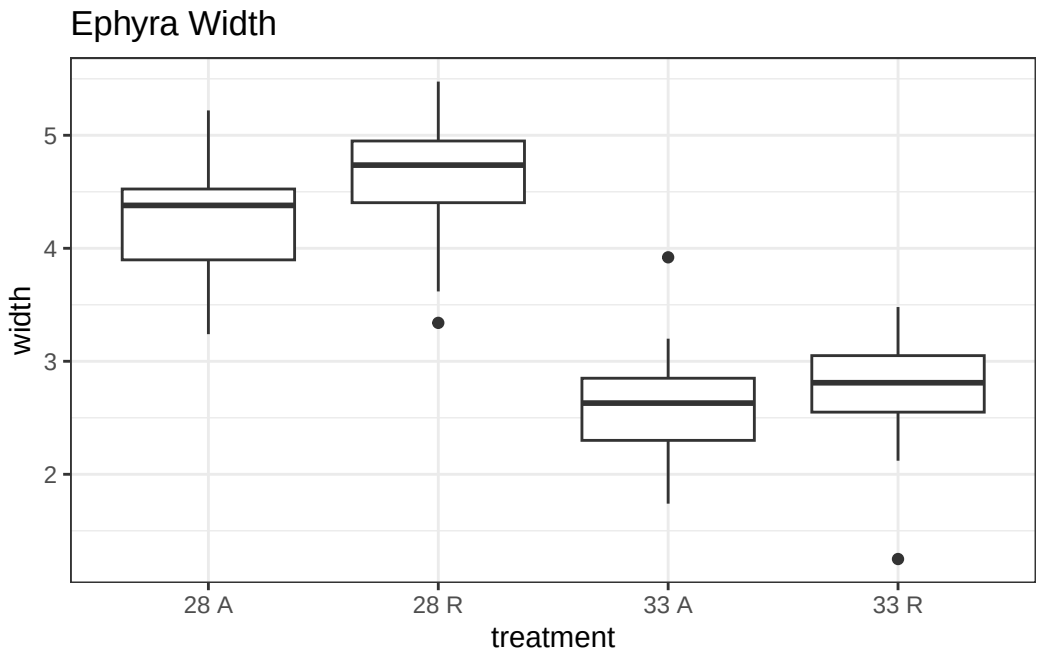


The contrast of the Ephyra count between regular water and antibiotic water is significantly different from 0 at a p-value of 0.05. Regular water is associated with about 15 less Ephyra when compared to antibiotic water.

Both water and temperature are meaningful factors for determining how many Ephyra will be produced, however these results are only an estimate because we do not have access to a unique count of Ephyra produced in each beaker.

What factors are significant for determining Ephyra Widths?

Measurements of Ephyra width were only performed for the first replicate of each treatment's beaker.



A visual inspection suggests that temperature has a meaningful effect while water has a smaller effect. There does not appear to be an interaction.

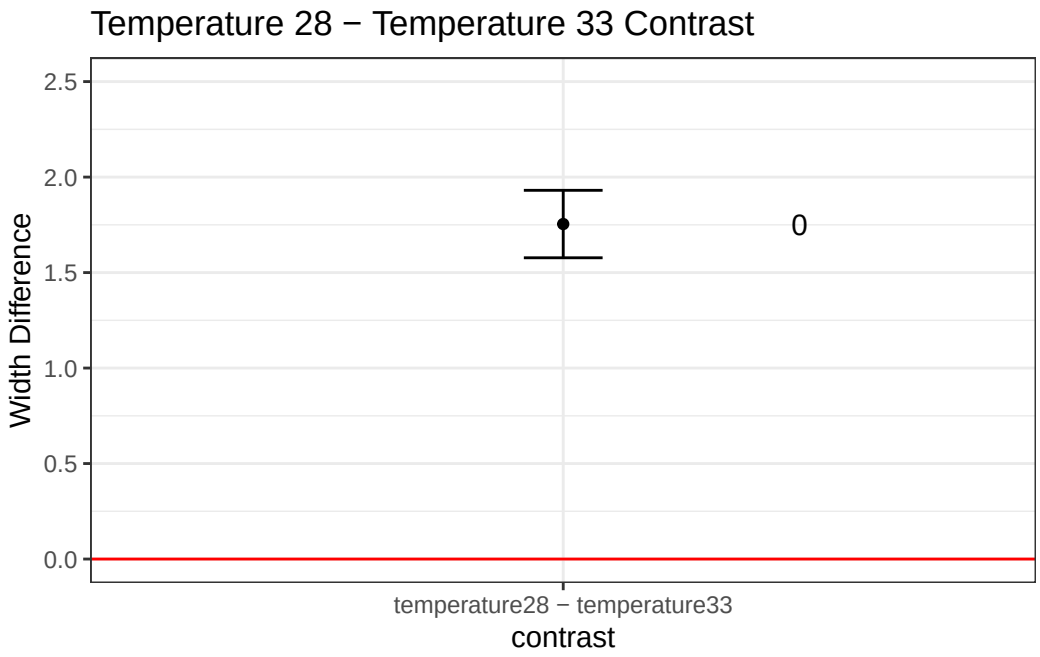
We fit a similar model to before:

$$y_{ijk} = \mu + w_i + t_j + w_i t_j + e_{ijk}, \quad i = 1, 2 \quad j = 1, 2, 3 \quad k = 1, 2, ..n_{ij}$$

```
Anova Table (Type III tests)

Response: width
                  Sum Sq  Df   F value    Pr(>F)
(Intercept)       489.67   1 2209.3201 < 2.2e-16 ***
water               1.66   1    7.5052  0.007116 **
temperature        39.11   1  176.4493 < 2.2e-16 ***
water:temperature   0.23   1    1.0375  0.310496
Residuals          25.93 117
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

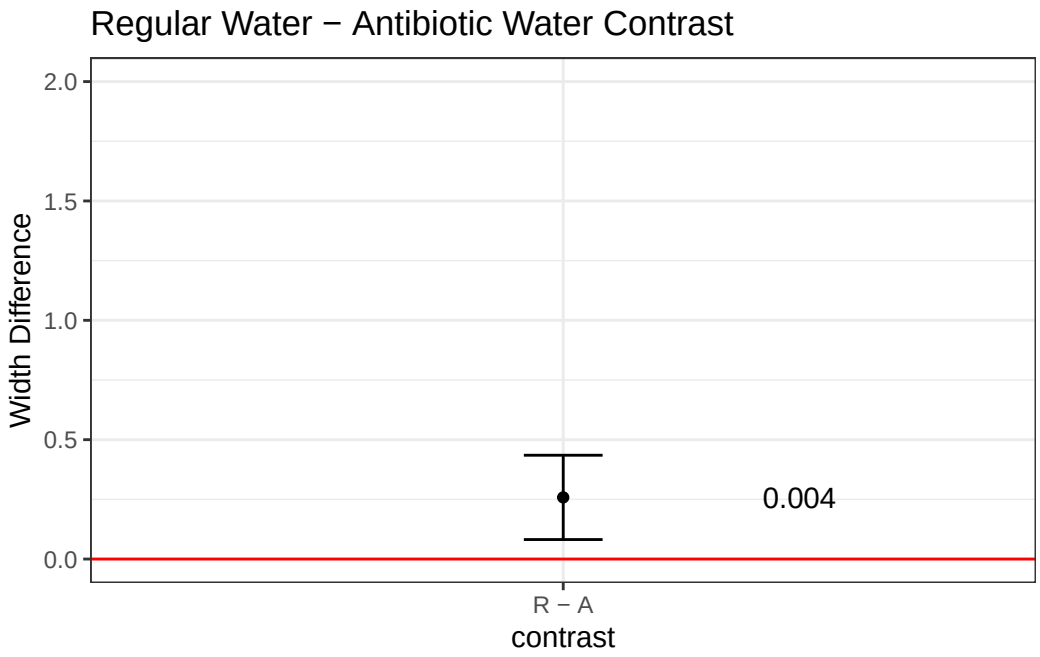
The interaction term is not significant, but the main effects of water and temperature are. So we compute their contrasts.

```
NOTE: Results may be misleading due to involvement in interactions
```



The contrast of the Ephyra width between temperature 28 and temperature 33 is significantly different from 0 at any reasonable p-value. A temperature of 28 is associated with a 1.75 cm (unit not provided by researcher) increase in Ephyra width when compared to a temperature of 33.

```
NOTE: Results may be misleading due to involvement in interactions
```



The contrast of the Ephyra width between regular water and antibiotic treated water is significantly different from 0 at any reasonable p-value. However the size of effect is quite small, regular water is associated with a .025 cm increase in Ephyra Width over the antibiotic treated water. This might not be a meaningful difference.

Conclusion

After investigating various measures, the choice of treatment depends on priorities.

If the goal is to encourage reproduction and maintain the polyp strength and polyp population, you would want to choose lower temperatures. These lower temperatures prevent the jellyfish from developing to new life stages,

If the goal is for jellyfish to continue to develop, choose a temperature of 28. This temperature seems to be a happy medium for development, the most amount of Ephyra were produced and their widths were longer than those at temperature 33.

Temperature 33 seems to be a weak option, the widths of both polyp and ephyra samples were smaller and it does not seem to have a clear strength. It produces some Ephyra, but not as many as temperature 28 and maintains some bud population, but not as much as temperature 22.

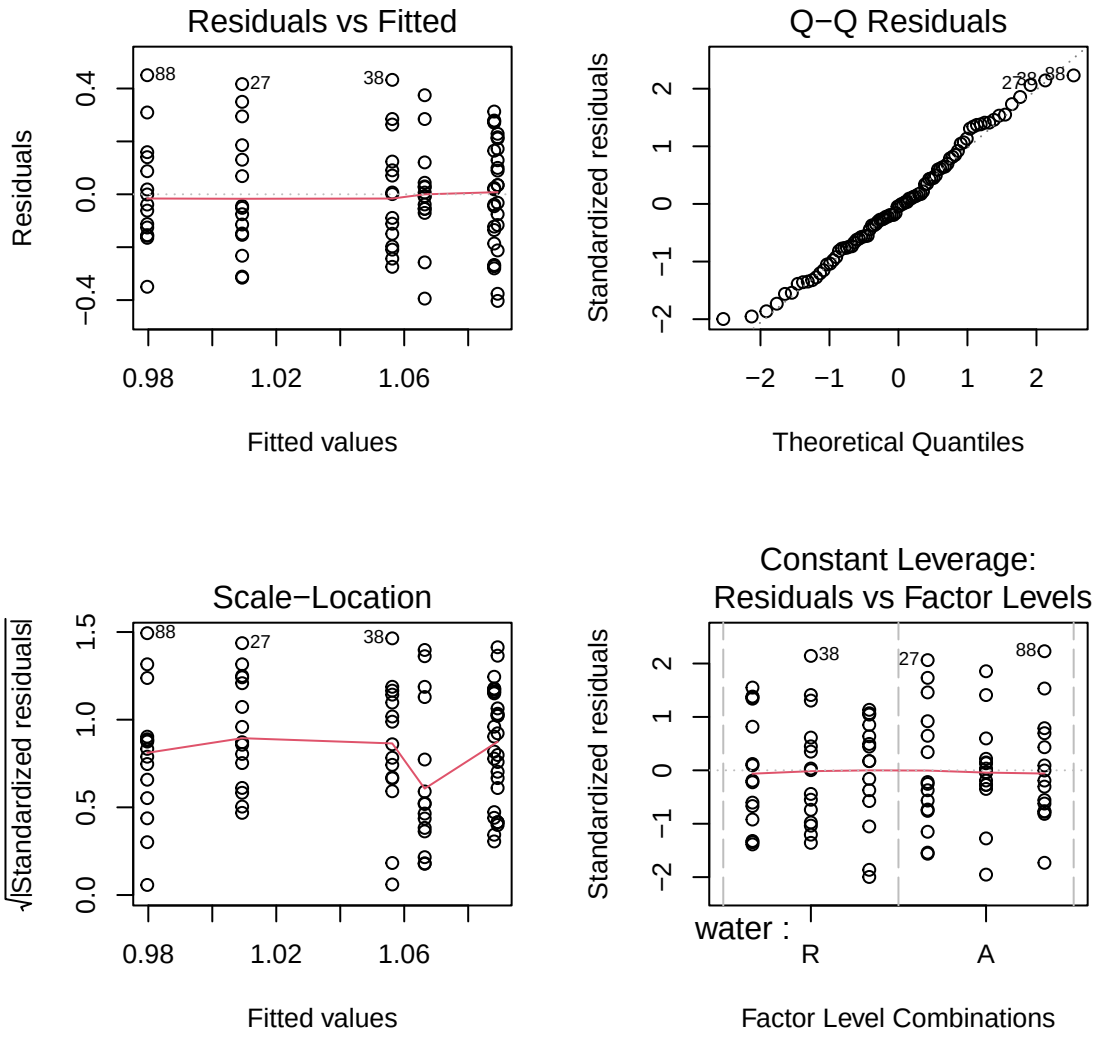
It appears that applying an antibiotic water treatment is a good idea. It is strongly associated with a higher quantification cycle, which indicates that bacterial load is lower than regular water. It also some association with better polyp widths, an indication of stronger jellyfish.

Next steps would be to re-analyze Ephyra data if a unique count is available, and extend research to the Medusa life stage if it is of interest.

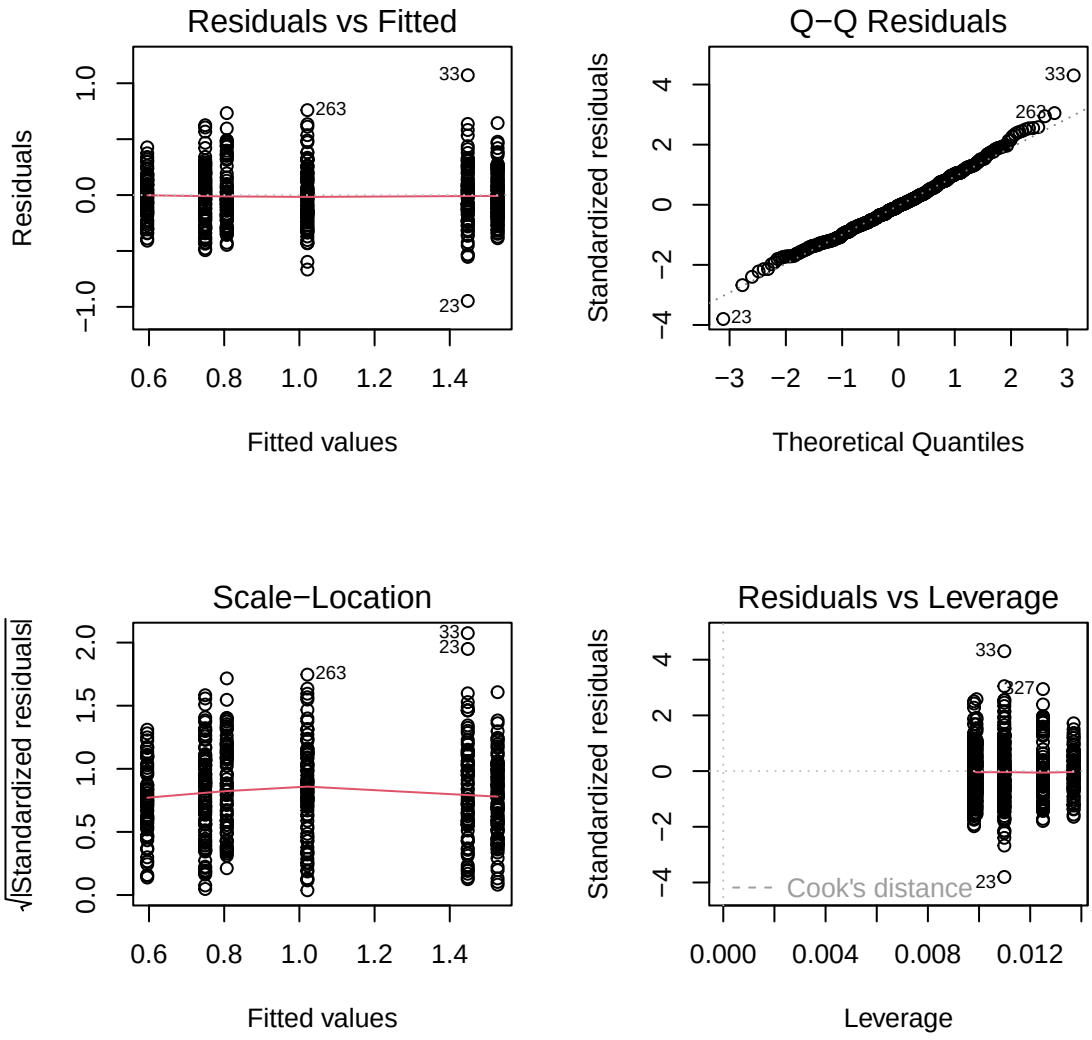
Appendix

Residual plots for models used:

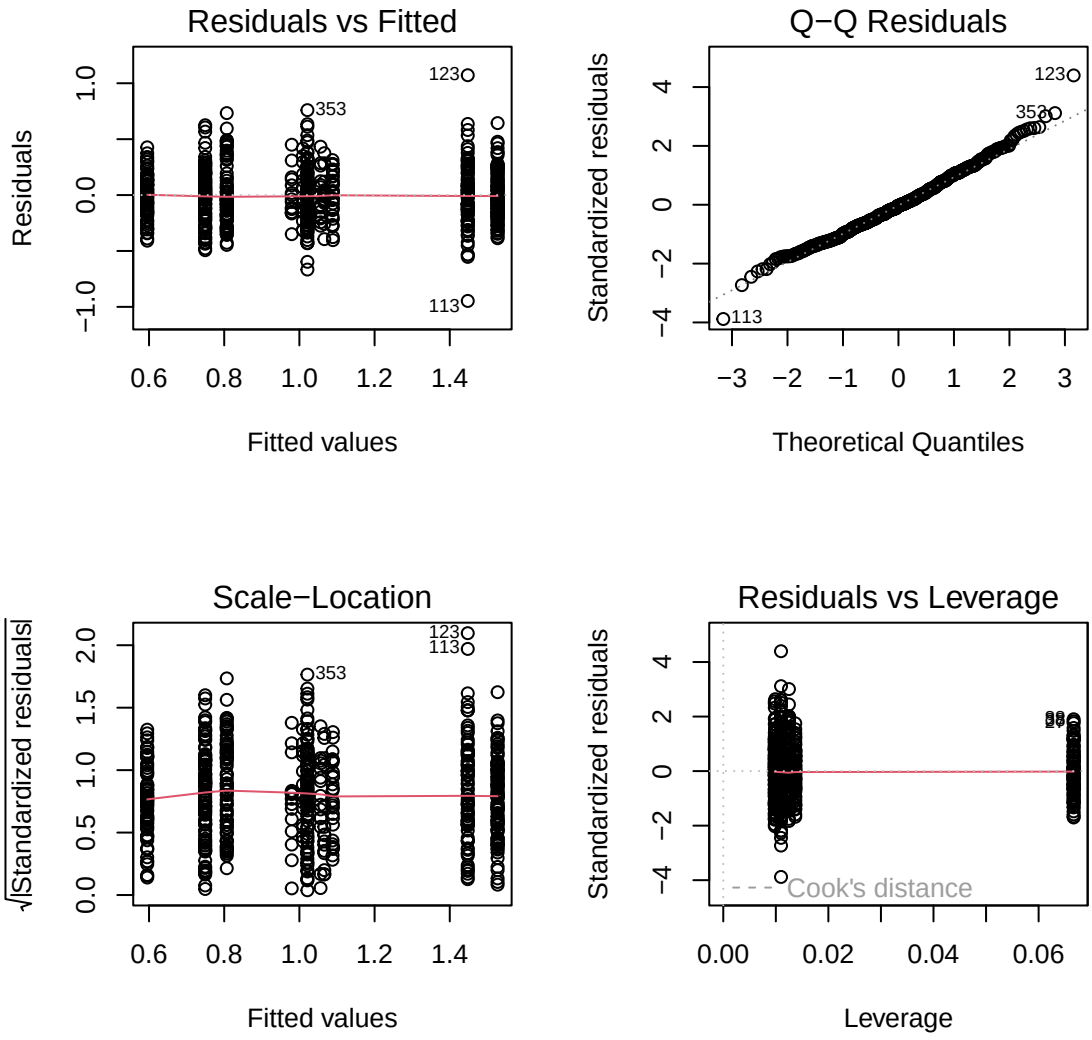
time 0 sample model:



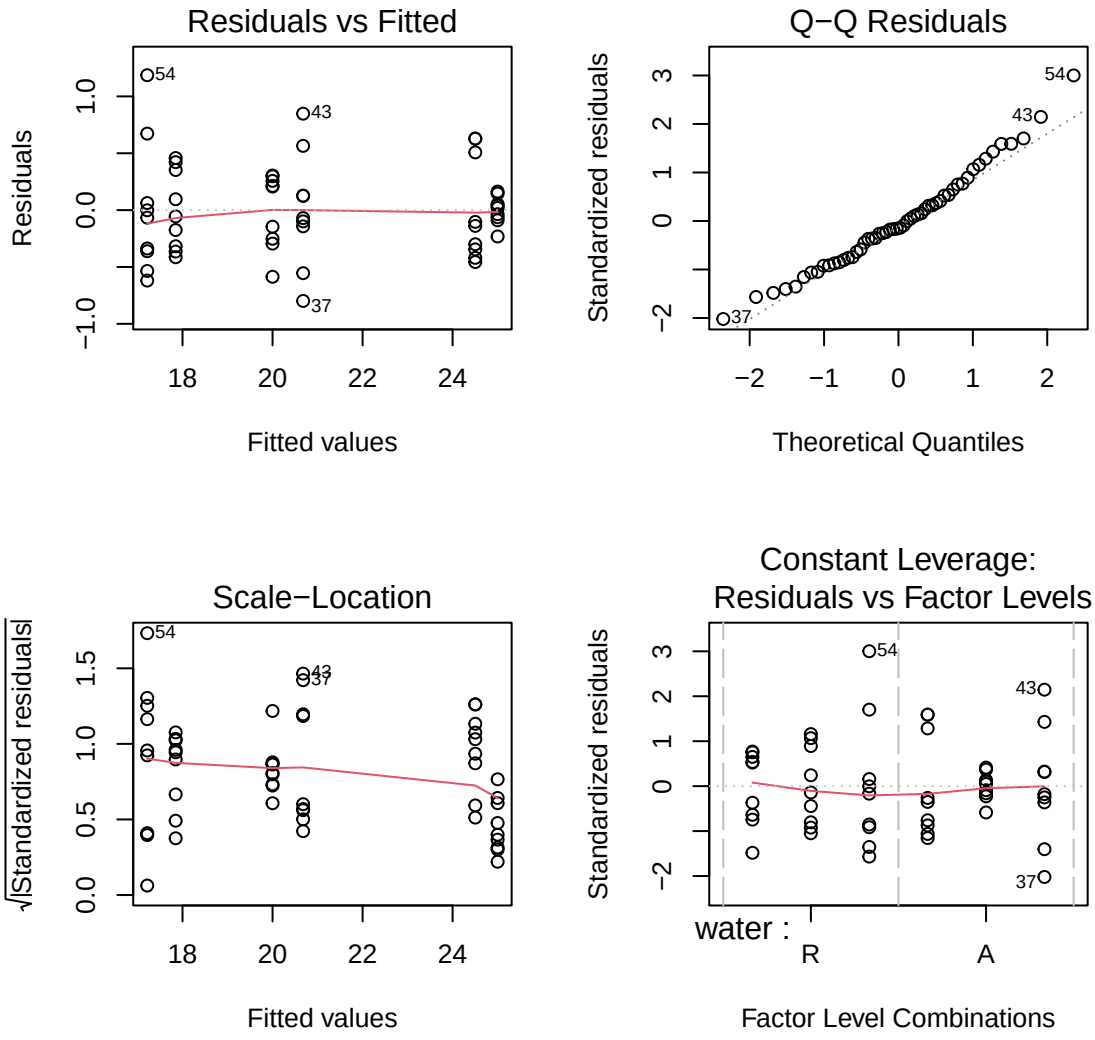
time 35 sample model:



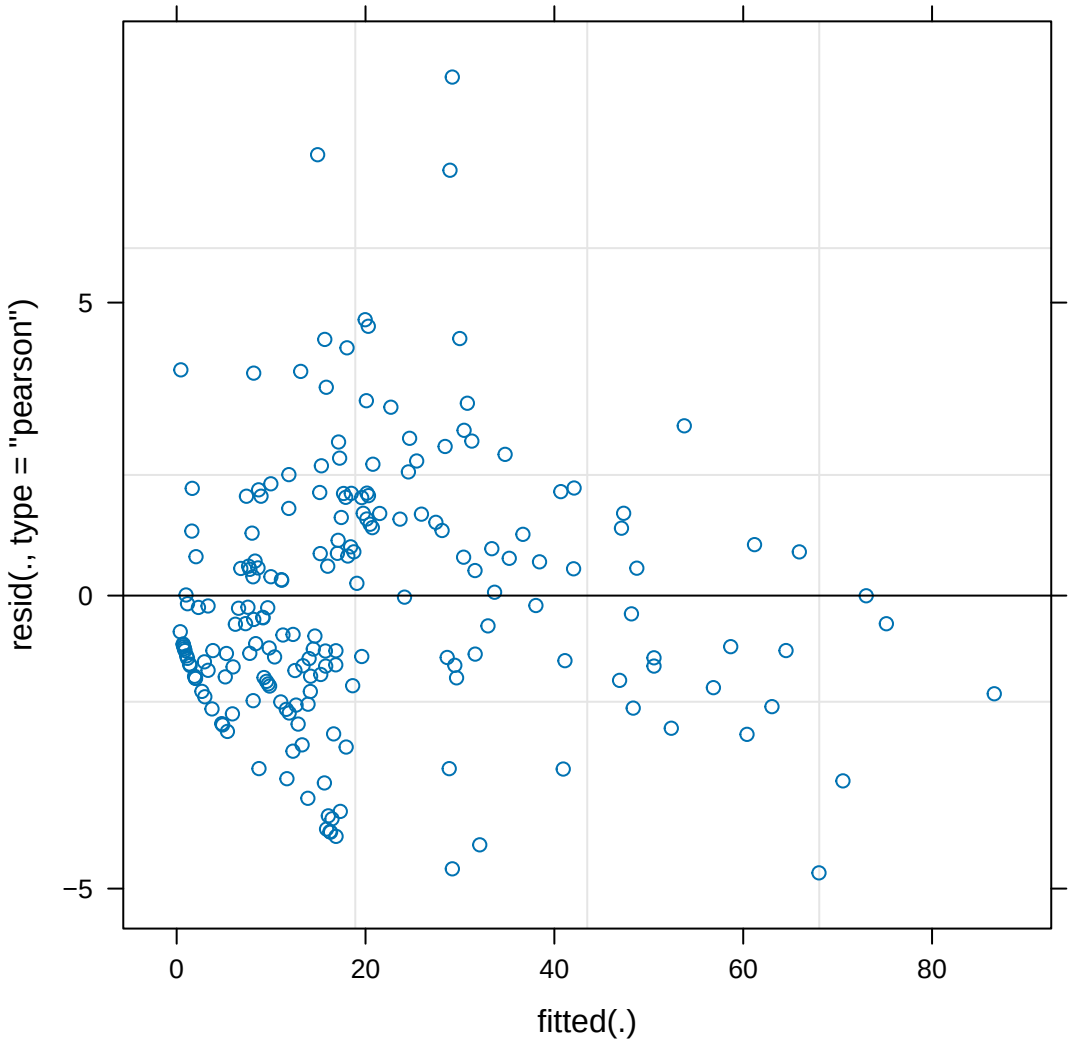
time 0, time 35 sample model:



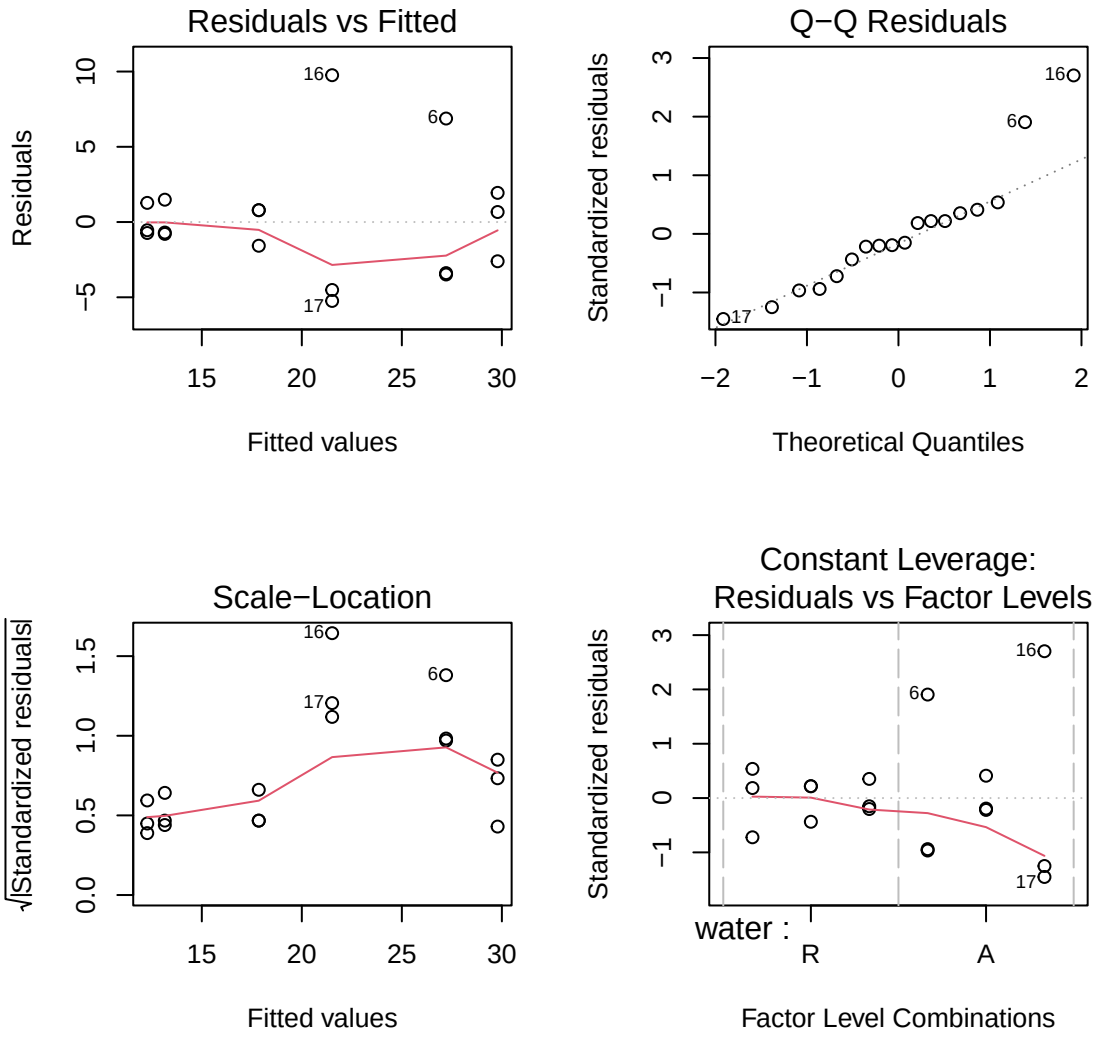
polyp width model:



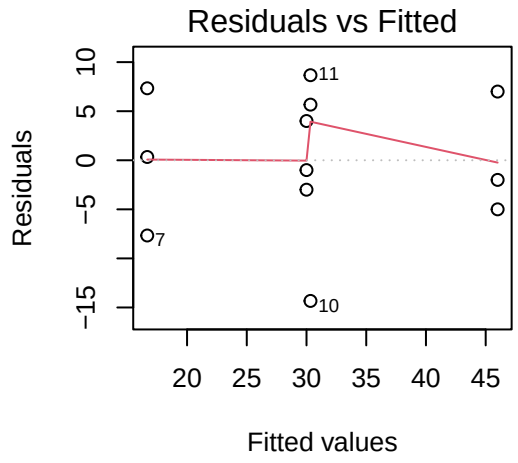
bud count model:

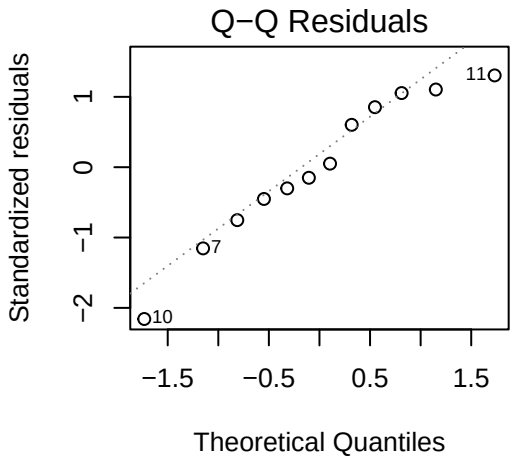


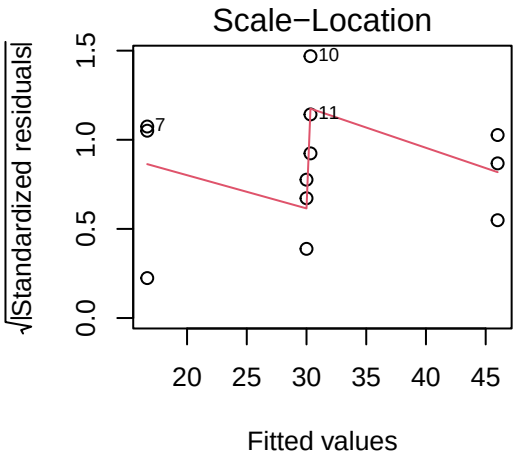
bud count model collapsed over time:

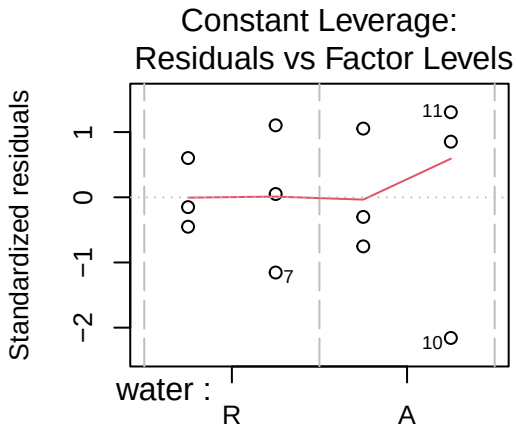


ephyra count model:

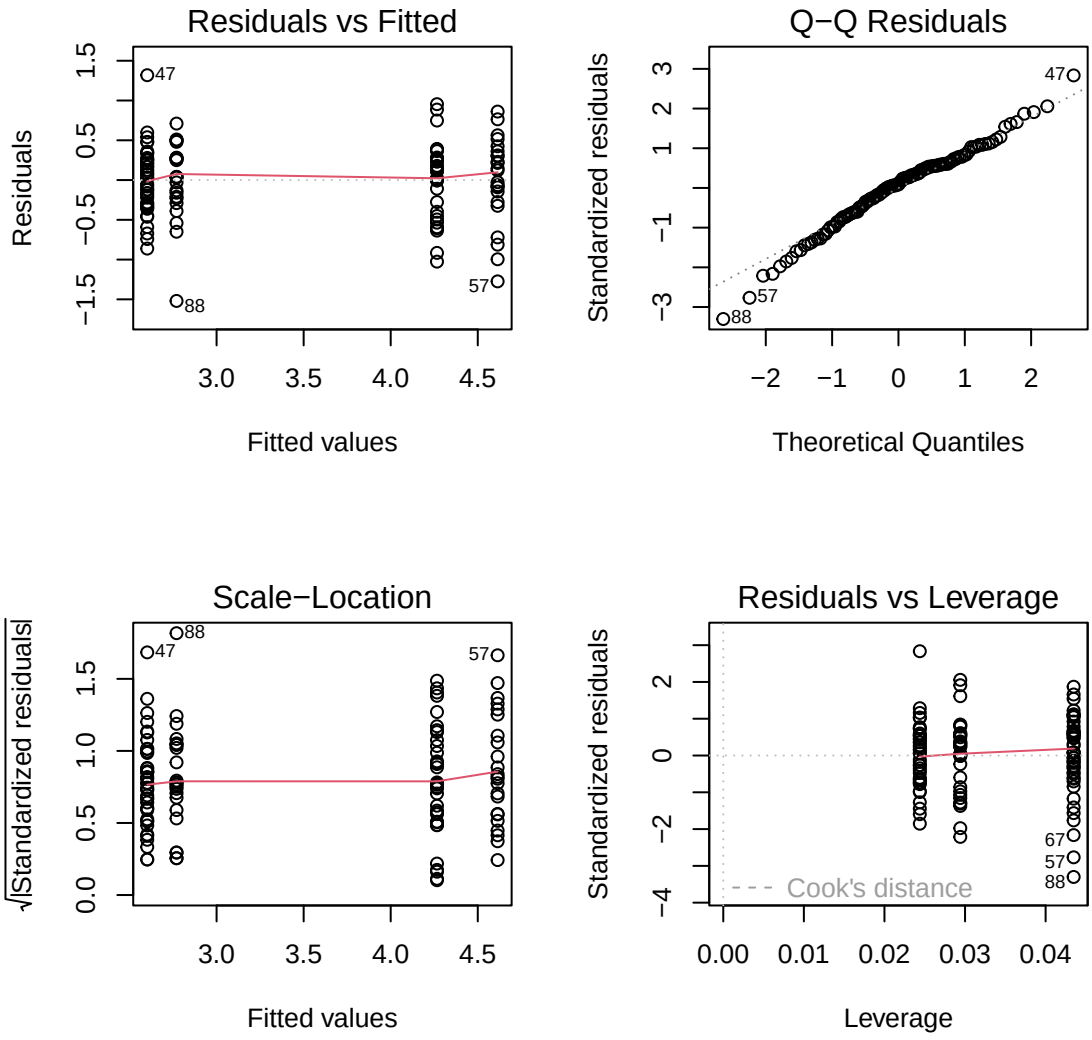








ephyra width model:


Residuals vs Fitted
Residuals
Fitted values
47
88
57
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
47
57
88
Scale−Location
√|Standardized residuals|
Fitted values
88
47
57
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
67
57
88