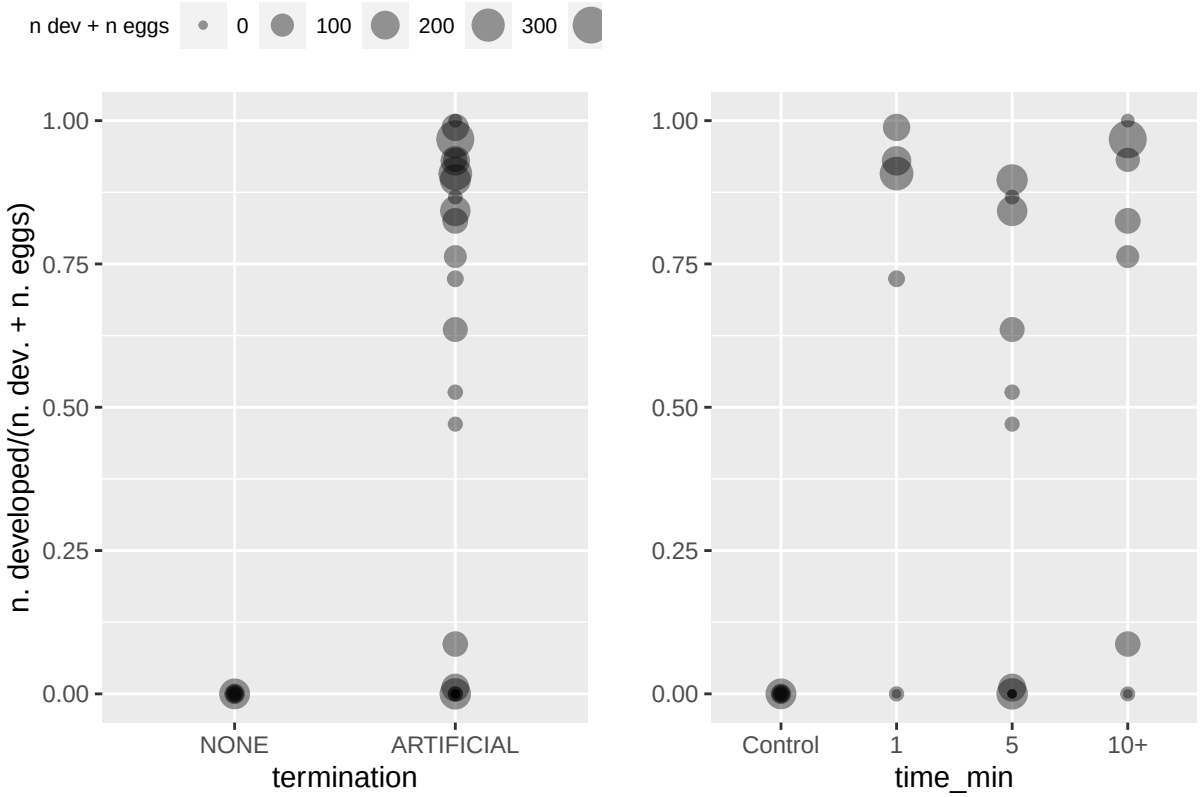
# Fertilisation project - GLM

SP

12/11/2020

The plot on the left is to illustrate the difference in the proportion of developed embryos between control (NONE) and all treatments combined (ARTIFICIAL). The right plot shows the same y variable but divided into the three different treatments. In both figures, the size of the points are proportional to the total number of eggs and embryos in each female (n dev + n eggs).




There were some questions about the above figures.

1. Do females with few of no developing embryos have also very low number of eggs?

The size of the dots in the figure is proportional to the sum of eggs and embryos of each female and it looks like three females with low or no developing embryos do not have low number of eggs. Here is a table with the number of eggs for such females and other information.

| snail_ID | termination | size_mm | sex | ecotype | time_min | egg | dev | vip | notes | stage | p_dev | proj | tot | p_egg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | ARTIFICIAL | 9.3 | female | wave | 5 | 258 | 0 | S | pilot | U | 0 | pilot | 258 | 1 |
| H05 | ARTIFICIAL | 13.0 | female | crab | 5 | 0 | 0 | S | in tank 20200525 | U | 0 | followup | 0 | NaN |
| H11 | ARTIFICIAL | 12.0 | female | crab | 1 | 0 | 0 | S | in tank 20200528 | U | 0 | followup | 0 | NaN |
| H12 | ARTIFICIAL | 12.5 | female | crab | 10+ | 16 | 0 | S | in tank 20200528 | U | 0 | followup | 16 | 1 |
| H15 | ARTIFICIAL | 10.0 | female | crab | 10+ | 0 | 0 | S | in tank 20200601 | U | 0 | followup | 0 | NaN |
| H19 | ARTIFICIAL | 9.5 | female | crab | 5 | 0 | 0 | S | in tank 20200604 | U | 0 | followup | 0 | NaN |
| H21 | ARTIFICIAL | 10.0 | female | crab | 1 | 18 | 0 | S | in tank 20200602 | U | 0 | followup | 18 | 1 |
| H22 | ARTIFICIAL | 9.0 | female | crab | 5 | 0 | 0 | S | in tank 20200602 | U | 0 | followup | 0 | NaN |

2. Can the few and no sperms group be separated?

From the table above, females cannot be separated by any of the columns/variables. These females are small and big, wave and crab, etc.

3. Assuming that fertilization did not occur for these females (e.g., either sperms were not transferred because of the male, or the male-female combination was incompatible and embryos do simply not develop properly) what happens if these "failing females" are removed before analysing the rest?

   a) Analysis without "failing females"

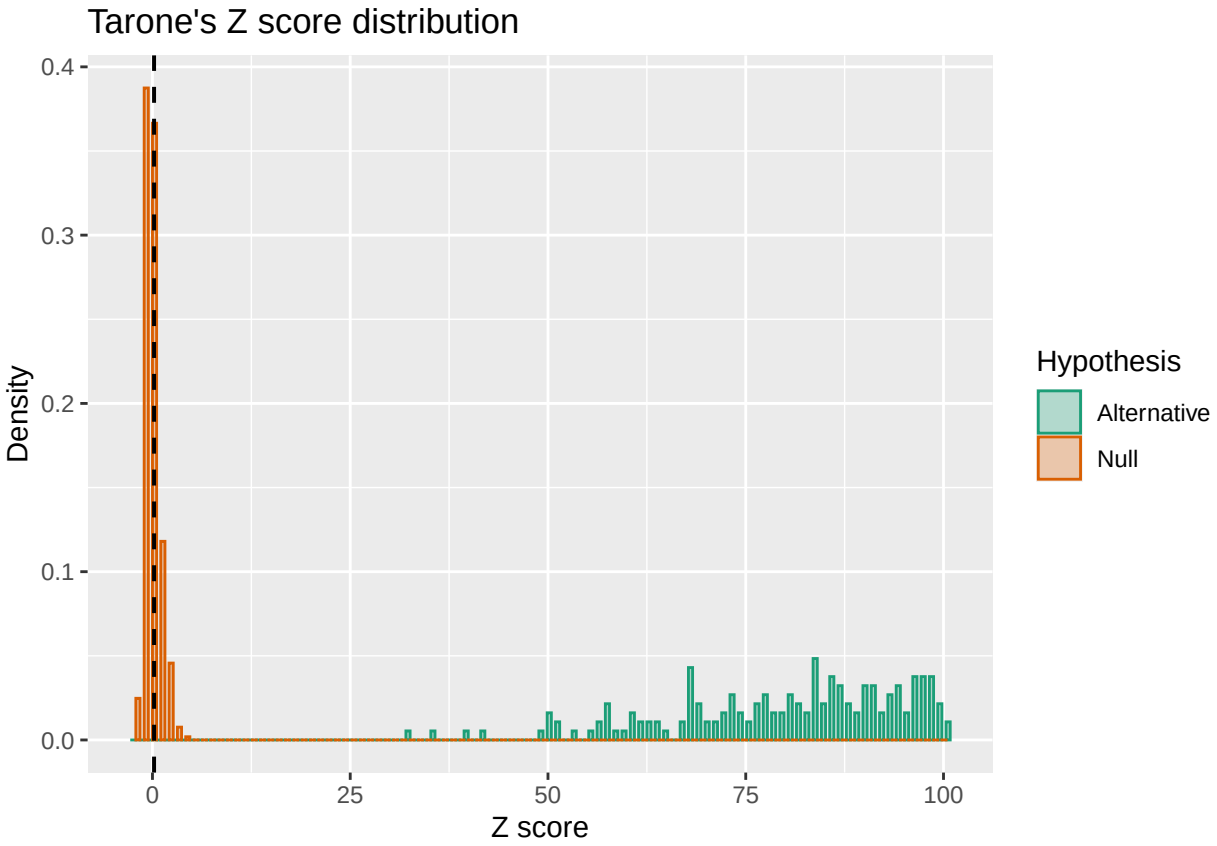
   The beta-binomial (BB) model returned much lower SEs of the parameter estimates than the binomial (B) and quasi-binomial (QB) model (table below onlt for BB model). However, I was not able to compare the three models using the likelihood ratio test because fitting the BB model using `glm()` did not return the log-likelihood nor the deviance.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -5.378 | 2.843 | -1.891 | 0.072 |
| terminationARTIFICIAL | 6.851 | 2.861 | 2.395 | 0.026 |

What I have estimated instead was the Tarone's Z statistic which can be used as goodness of fit test of the Binomial distribution against the BB distribution. I have followed this tutorial link. First, I have inferred two parameters, the probability of observing developed embryos $\mu$ and the dispersion $\rho$. The probability $\mu$ was calculated for the whole dataset of no "failing females" and without using "termination" as the fixed effect. This is certainly not appropriate for what we ultimately want to test (i.e., difference between control and treatments) but it can be still fine if we simply want to see whether the a model with dispersion (BB model) fits the data better than a model without (B model). The analysis of the Tarone's Z statistic should come earlier than the summary of the BB model (table above) and it is here now because I wanted to show first that I could not use the likelihood ratio test to compare B, BB and BQ models. Actually, I did compare B and BQ and the difference was not significant which is strange given that BQ and BB are both models with a dispersion parameter.

```
## [1] "mu_hat: 0.439     rho_hat: 0.611"
```

To assess which of the two models (B and BB) fitted the data best, I have simulated 1000 data points given a beta-binomial distribution with $\mu = 0.4$ and $\rho = 0.6$ and also 1000 data points given a beta-binomial distribution with $\mu = 0.4$ and $\rho = 0$ (i.e., a binomial distribution with $\mu = 0.4$). A dispersion of 0.6 is high and makes the plot look squeezed so I have used $\rho = 0.3$ to show the difference between the two simulated datasets (figure below).




The light orange histogram is the empirical distribution under the null/binomial model and the green coloured histogram is the distribution of the Z-scores under a Beta Binomial distribution. If what I have done makes sense, then it should be clear from the figure above that the BB model is a better fit to our data.

I have performed the same steps for the observed dataset with "failing females" and the difference is that $\mu$ is lower and $\rho$ is higher which is what we expected if we included treated females with no developed embryos.

b) Analysis with "failing females"

```
## [1] "mu_hat: 0.376     rho_hat: 0.652"
```

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -5.378 | 2.684 | -2.004 | 0.056 |
| terminationARTIFICIAL | 6.806 | 2.700 | 2.521 | 0.019 |

With or without "failing females", the proportions of developed embryos is significantly higher than that of the control. The next step is to check for differences between treatments but I am going to wait for your feedbacks on what I have tried so far.