Lens Regeneration Data Analysis

I. Introduction

This data comes from the Katia Del Rio-Tsonis lab at Miami University, from a study with the preprint title of "*Regenerative hallmarks of aging: Insights through the lens of Pleurodeles waltl*". This study investigated the differences in the kinetics of lens regeneration across three developmental stages of newts: pre-metamorphic larvae, post-metamorphic juvenile, and adult. The response and predictor variables were as follows:

$\mathbf{Y_1}$ = EdU+, a count of cells that entered the G1 phase of the cell cycle.

$\mathbf{X_1}$ = Day, a numeric value of the day at which cells were counted. Although it is treated as a continuous variable, in the experiment it only took on 4 values: 1, 4, 10, and 15.

$\mathbf{X_2}$ = Age, a categorical variable indicating whether an observation is from a larvae, juvenile or adult.

The response variable also had to take into account the total number of cells present in an eye section. The variable hoechst, a measure of the total number of cells present in an observation, was used as an offset term to standardize the EdU+ count.

The hypothesis is that there is a decline in regenerative ability for older newts. In other words, larvae are expected to have a high proportion of cells reentering the cell cycle (high standardized EdU+) a short time interval after lens injury. Juvenile lenses are expected to start regenerating after larvae, and adult lenses after juveniles. More specifically, the authors originally speculated that the iris cells of larvae re-enter around 1 day post injury, 4 days for juveniles, and 10 days for adults. After a standardized EdU+ count peaks, there is an anticipated return to baseline as cells either transdifferentiate into lens cells or withdraw from the cell cycle.

The goal of my analysis is to construct a confirmatory regression model to check when the standardized EdU+ count peaks for each of the three newt life stages.

II. Exploratory Data Analysis (EDA)

To visualize the response variable, I plotted the log of the ratio of cells that entered the cell cycle to the total number of cells (figure 1). It is immediately evident that the cells of larvae reenter the cell cycle much faster than for juveniles and adults. For juveniles, there are more observations of cells reentering on day 4 than for adults, but the difference in peak of reentry appears smaller than between larvae and juveniles. For all three groups, there is a characteristic parabolic shape. The lower end represents the time immediately after lens injury, when cells have not yet responded to the stimulus. Gradually, the number of cells that enter the G1 phase of the cell cycle climbs, until a peak is reached. After the peak, there is a descent back to baseline levels.

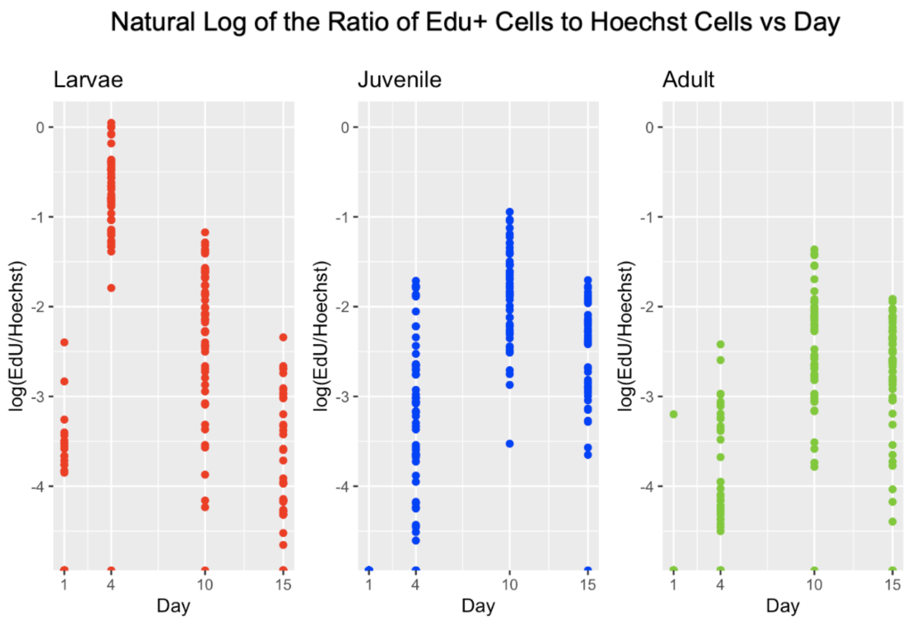

Figure 1: The response variable is plotted as a log of a ratio, since that is the form it appears in a later regression model.

III. Models and Analysis

The mean of the count response variable EdU+ is 4.37, and the variance 22.65. Because the variance is much larger than the mean, it is not appropriate to use a poisson distribution to model EdU. The overdispersion may produce biased or inefficient estimates, leading to incorrect conclusions. I will instead represent EdU+ as a negative binomial random variable, which has a distribution that allows the variance to be greater than the mean.

The parabolic shape of the data necessitates use of the squared predictor variable day. Because we are interested in the relationship between day and cell cycle reentry for different life stages of newts, I will fit the full linear model with interactions:

$$\begin{aligned} ln(EdU) = \beta_0 + \beta_1(Day) + \beta_2(Day^2) + \beta_3(Juvenile) + \beta_4(Adult) + \beta_5(Day * Juvenile) + \beta_6(Day * Adult) \\ + \beta_7(Day^2 * Juvenile) + \beta_8(Day^2 * Adult) + offset(log(Hoechst)) \end{aligned}$$

The natural log of EdU represents the link function between the count response variable and the predictor variables. The offset term adds the log of total cell counts into the model equation and does not have a parameter estimate of its own. When a negative binomial glm is fit by the specifications above, the model parameters estimates are in terms of the natural log of the ratio between EdU and Hoechst.

The fitted model appears in the appendix. All the parameter estimates are significant except for the interaction between Day and the age indicator variables. This means that the parameter estimate of the linear relationship between day and EdU is not significantly different for juvenile or adult newts versus larvae. The interaction terms of day^2 and the age indicator variables, however, are significant.

Figure 2 shows the fitted regression model plotted by the age categorical variable.

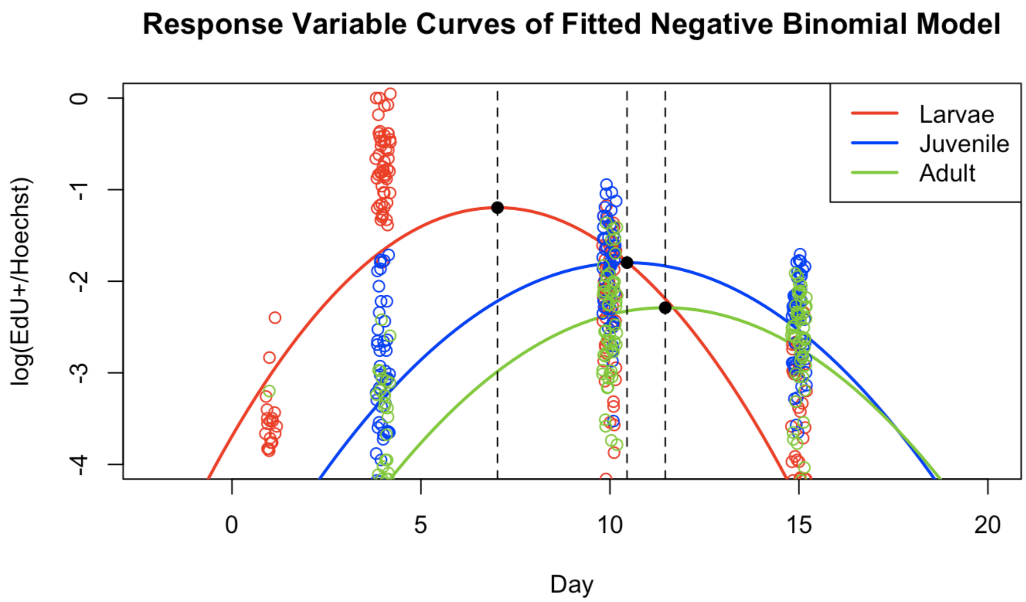

Figure 2: Data points were jittered for visibility. The maximum for the larvae curve occurs at day 7.02. For juveniles, on day 10.45. For adults, on day 11.46.

IV. Conclusions

Figure 2 indicates clear differences in the peaks of the response variable curves for different newt life stages. Cells from larvae peak ("maximally" reenter the cell cycle) much later than the initial hypothesis of day 1. Juveniles re enter later than day 4. Adults renter slightly later than the hypothesized day 15. Overall, there is a trend of decreased regenerative speed across newt life stages.

Between the curves of EdU versus day across newt life stages, there is a significant difference in the coefficient of the squared term (day^2 and age interaction). Depending on how the linear and constant terms of the curve change, this could indicate significant differences between the maxima. This is one crude method of performing statistical inference on the difference between the curve peaks. Another method is to perform bootstrap sampling to find confidence intervals for the maxima, and is the future direction of this analysis.

Appendix

```
Call:
glm.nb(formula = EdU ~ Day + I(Day^2) + Juvenile + Adult + Day *
    Juvenile + Day * Adult + I(Day^2) * Juvenile + I(Day^2) *
    Adult + offset(log(DAPI)), data = data2, init.theta = 3.996114573,
    link = log)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       -3.691082   0.141250 -26.132  < 2e-16 ***
Day                0.710309   0.044531  15.951  < 2e-16 ***
I(Day^2)          -0.050556   0.002898 -17.445  < 2e-16 ***
Juvenile          -2.004527   0.268132  -7.476 7.67e-14 ***
Adult             -3.240144   0.335496  -9.658  < 2e-16 ***
Day:Juvenile       0.035717   0.071052   0.503 0.615188
Day:Adult          0.099725   0.080923   1.232 0.217821
I(Day^2):Juvenile  0.014868   0.004132   3.599 0.000320 ***
I(Day^2):Adult     0.015223   0.004501   3.382 0.000718 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(3.9961) family taken to be 1)

    Null deviance: 1896.52  on 646  degrees of freedom
Residual deviance:  679.49  on 638  degrees of freedom
```