**Fish Population Dynamics**

**Laboratory Exercise #2**

**Fitting and Comparing Nonlinear Models in Excel**

**Marcel Gietzmann-Sanders**

Data for this laboratory are from two Florida Spotted Seatrout *Cynoscion nebulosus* populations (Indian River and Charlotte Harbor). The dataset contains only females from each population. Fish were collected from 1986 to 1988 by Murphy and Taylor (1994, *Transactions of the American Fisheries Society* 123:482-497).

The objectives of this laboratory activity are to:

**1)** Write your own likelihood function for the von Bertalanffy growth equation and estimate , , , and ;

**2)** Estimate the mean length at age for each population;

**3)** Use an AIC model selection method to compare growth curves and parameters between populations.

**Please conduct the following analyses in *Microsoft Excel* and answer the questions. Each question or bullet point should be responded to either with text, a table, or a plot.**

1. For each population make a plot of TL as a function of YRSOLD.

A graph of a graph with a red line and a blue line

Description automatically generated with medium confidence

* Do the relationships appear linear?
  + Definitely not, both clouds of data clearly asymptote.

2. Add the following columns defined by equations for: (this question can be answered by attaching your Excel spreadsheet to your submission)

* Model predictions of length at age using the von Bertalanffy growth equation.
* Residuals of the observed length at age and predicted length at age .
* Log likelihood function, , using a normal distribution to determine the probability of the data, , given the model parameters, ; i.e. .

3. For the Charlotte Harbor population:

* Maximize the log likelihood to estimate parameters of the von Bertalanffy growth equation.
* Does this model appear to fit the data?

A graph with a red line and blue dots

Description automatically generated

It seems to fit the data reasonably well.

* Plot the residuals as a function of age.

A graph with blue dots

Description automatically generated

* Based on looking at the plot, does this appear to be a reasonable model for the data?

It seems we’ve got a lot more variance in the central ages. The variance tapers off as you approach 0 or get above ~5. We also see that the model is slightly overpredicting the very high values (thus the negative residuals). Finally the model is definitely be driven largely by the cluster of data in ages 1 – 2. There’s a lot more “power” from those ages than the higher ones. It’s not a terrible fit but definitely could be improved.

* Are there ages where the residuals do not appear to be normally distributed?

Yes especially in the 1 – 2 year range it looks like we’ve got a skew in the observations toward higher lengths away from the estimated length at age. This makes sense as we’ve got the whole gear selectivity issue where smaller fish are just getting through the nets. Then for the higher ages (6-7) there are just so few points that it’s impossible to tell what kind of distribution is present at those ages.

4. Conduct the same analysis for the Indian River population (question 3) and address the same bullet points.

A graph with blue and red dots

Description automatically generated

A graph with blue dots

Description automatically generated

By and large the same conclusions apply here as before. The fit is reasonable however when looking at the residuals we still see that positive skew in the observations, the centralization of the data at lower ages, a tad of overprediction at high ages, and the changes in variance at low and high ages. What is interesting is that the variance and asymptotic length are both higher for this population.

5. Using the model parameters for each population:

* Calculate the mean length at each age, for age 0 to age 8, for each population. A table of numbers and a number

  Description automatically generated with medium confidence
* How do the lengths at age differ between the populations?
  + They both start at the same length at age 0 but Indian River fish grow faster (according to the model) and end up much larger by age 8 (~130mm or 20% longer)

6. Set up an AIC table:

* Determine the log likelihood, the number of parameters, and the AIC for:
* A model with different parameters (, , , and ) for each population.
* A model with shared parameters for both populations. A close-up of a table

  Description automatically generated
* Which hypothesis had the most support based on AIC?
  + The model with different parameters for each population did the best.

7. Using what you know about the biological meaning of the von Bertalanffy growth parameters, choose a parameter (, , , or ) that you think should be shared between the populations:

* Justify your parameter choice.
  + is the variance in length v age. It makes sense that given different environmental conditions the variance could differ greatly. So we’ll leave this parameter separate for each population.
  + is supposed to be, on average, the largest a fish of this species could ever get. While at first one would assume that this is based on the specific genetics and biology of the fish itself rather than something dependent on its environment the fact is that one environment may simply not be able to support as large fish as the other. So we’ll let this parameter be free.
  + this parameter is giving us an indication of how quickly the fish are growing. This is definitely influenced dramatically by the environment so we’ll leave this parameter free.
  + is supposed to be the age at length zero. It’s been noted in lectures how this is pretty much a fake degree of freedom so let’s fix it between the two populations so I’m interested to see what happens if we actually just force this to be 0 (and thus effectively saying that at age 0 the fish are smaller than the measurement error bounds).
* Determine the log likelihood, the number of parameters, and the AIC for:
* A model with a different value for your chosen parameter for each population.
* A model with a shared value for your chosen parameter for both populations.

A group of graphs showing different types of sea levels

Description automatically generated

* Which hypothesis had the most support according to AIC?A close-up of a table

  Description automatically generated

The completely free model definitely does better (which is expected as it has those two additional free degrees of freedom).

* Based on which parameter you chose and the subsequent AIC results, what is your conclusion regarding the biology of the populations?

Clearly the growth rates and “maximum” size in these two populations are different. In general the Indian River population is getting to larger sizes and faster. It’s hard to tell if the fact that model fits better with being free is really saying anything about the differences in populations or is just pointing out that our data is skewed by the size selectivity and/or the growth at small ages isn’t really going to fit this particular model. Clearly it is having trouble hitting the higher sizes which just might indicate that growth at early ages is quite a bit faster than this model would predict and so the whole curve could end up getting shifted up toward higher lengths.

**Now go to** [***https://zsiders.shinyapps.io/lab\_2\_part\_1/***](https://zsiders.shinyapps.io/lab_2_part_1/) **and answer this last question.**

8. By changing which von Bertalanffy growth parameters are shared between the populations:

* Determine the model for the two populations that has the lowest AIC.A table with numbers and letters

  Description automatically generated

Looks like holding as shared is the best model which is pretty interesting as I’d expect the growth to vary based on the environmental conditions.

* Determine the top models for the two populations according to AIC.

Sharing just , both and , or just all worked pretty well too. What’s interesting is holding both and does pretty poorly which just goes to show that the two are working in tandem to represent growth (i.e. what one gets set to affects the other, likely because the largest fish don’t really have much representation in this data set). I ended up refitting the model where was shared rather than fixed as I did in my hypothesis and it spat out . So the model is just using to flatten out the curve, not to give us anything biologically meaningful. Very interesting.

Graduate students only

9. What might cause the residuals to be non-normally distributed at a particular age? Think carefully regarding the mechanisms that might be the same or different at different ages e.g. young versus old fish.

To me at least it seems clear this is size selectivity. The residuals max out at ~150mm which given the >200mm means we aren’t hitting into some kind of a-fish-can’t-be-less-than-0-mm issues. Besides there’s a clear cutoff around 300mm length. At the extremely large lengths you’re just ending up with a data sparsity issue which could either be that larger fish are rare or that something about how you’re trying to catch them is once again causing a skew of some kind.

10. When might it make sense to fix to zero? Are there fish species where it might not make sense to fix to zero?

When fish are born at lengths that are under the measurement error should really be quite close to zero. However in cases where fish are born with some significant length (viviparous or organisms with larger eggs) it could make sense to have a negative as the fish are never actually size zero.