

Brain Weight and Fecundity in *Actinopteri*

Nina Sun, Jeanny Zhang, Xingyi Zhang

November 21, 2022

1 Abstract

According to Fischer and Jungwirth’s research paper in 2022, larger brains are associated with lower fecundity in fish classes *Petromyzonti* and *Elasmobranchii*, which might be due to the evolutionary costs of larger brains (Fischer & Jungwirth, 2022). We built on top of this study and further investigated the relationship between brain size and fecundity in fish class *Actinopteri*. This study used a Bayesian multiple linear regression model and found that brain weight is positively correlated with both fecundity and marine status in *Actinopteri*. This shows us that different fish classes might have distinct evolution strategies. Our analysis highlights the need for more studies in evolutionary biology for fishes in the future.

2 Introduction

In evolution biology, it is believed that brain size evolution comes with behavioral and physiological trade-offs, such as higher energy computation (Martin, 1981). A study finds that larger brain weight is associated with a shorter lifespan and lower fecundity, where fecundity refers to the maximum potential offsprings of an individual over its lifetime (Fischer & Jungwirth, 2022; Bagenal & Braum, 1978). Fischer and Jungwirth (2022) claimed that those are the evolutionary costs of larger brains. However, they only found significant negative correlations in classes *Petromyzonti* and *Elasmobranchii* and didn’t test for a positive correlation. We therefore want to test if there is a positive correlation in the fish class *Actinopteri*. If so, how is brain size associated with fecundity?

This project used Bayesian analysis to examine the relationship between fecundity and brain size among *Actinopteri*. This was implemented through a multiple linear regression model with uninformative priors.

3 Data

The dataset is derived by Dr. Arne Jungwirth from a publicly available resource FishBase (Jungwirth, 2022). It was used for his paper “The Costs and Benefits of Larger Brains in Fishes” (2022). FishBase is a dataset hosted by a nonprofit organization that collects and maintains fish information contributed by many collaborators and scientists globally (*FishBase* 2015). From the original dataset, we dropped the null values and filtered out species of class *Actinopteri* with columns:

- logBrain: the log brain weight in mg of the heaviest brain donor individual for a given species
- logFecundity: the log maximum number of eggs or oocytes recorded for a given species
- Marine: binary variable for whether the species is considered marine (1) or inhabits in freshwater (0)

The resulting dataset has 226 data points, each representing a distinct species under the fish class *Actinopteri*.

4 Methods

4.1 Models

Given the log brain weight, the log fecundity, and the marine indicator variable, we decided to use a Bayesian regression model to investigate the association between the three factors for *Actinopteri*. We first tested both a simple linear regression model and a quadratic regression model with just log brain weight as the explanatory variable, but the models did not appear to be adequately fitting the observed data indicated by the posterior predictive plot. One can find the model fitting results for these two tested models in the code appendix. Therefore, we added marine as an additional explanatory variable, which was also shown to affect the fecundity of a species (Quinn et al., 2004). We eventually used a Bayesian multiple linear regression model. The response variable Y_i is log fecundity, and the explanatory variables are log brain weight and marine.

Given our previous experiences with biology data, we assumed that each Y follows a normal distribution with a unique mean μ_i and a common standard deviation σ . Meanwhile, since each species is evolutionarily distinct, it is reasonable to believe that Y s are independent for the likelihood.

Since we did not have any prior information on *Actinopteri*'s brain weight or fecundity, we placed uninformative priors on the standard deviation, the intercept, and the slope parameters. We used precision ϕ to represent the standard deviation σ where $\phi = \frac{1}{\sigma^2}$. In this case, we used a Gamma(1,1) distribution as the prior for ϕ . Since the data is in log scale, a Normal distribution $\mathcal{N}(0, 10)$ should provide sufficient flexibility and was thus used as the prior distribution of β_0 , β_1 and β_2 . We worked with the centered version of log brain weight $x_{ic,brain} = x_{i,brain} - \bar{x}_{brain}$, so that the intercept would tell us about the expected log fecundity on an average log brain weight (i.e. when $x_{ic} = 0$). Since the marine variable is binomial, there is no tuning process needed for it.

Here is a summary of notation used in this report:

- the centered log brain weight for species i : $x_{ic,brain}$
- whether species i is marine: $x_{i,marine}$
- the average log fecundity for species i : $\mu_i = \beta_0 + \beta_1 \cdot x_{ic,brain} + \beta_2 \cdot x_{i,marine}$
- an observation specific mean u_i for the i -th species' log fecundity Y_i : $Y_i | \mu_i, \sigma \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_i, \sigma), i = 1, \dots, n$
- the standard deviation within species: σ

The above assumptions and information lead to the following Bayesian model:

$$\begin{aligned}
 Y_i | \beta_0, \beta_1, \beta_2, \sigma &\stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_i, \sigma) \\
 \mu_i &= \beta_0 + \beta_1 \cdot x_{ic,brain} + \beta_2 \cdot x_{i,marine} \\
 \beta_0 &\sim \mathcal{N}(0, 10) \\
 \beta_1 &\sim \mathcal{N}(0, 10) \\
 \beta_2 &\sim \mathcal{N}(0, 10) \\
 \phi &\sim \text{Gamma}(1, 1)
 \end{aligned}$$

4.2 Computation

The above-mentioned model was implemented through the R package JAGS (Just Another Gibbs Sampler) which uses the MCMC sampling algorithm. Three separate chains with randomly selected initial values were simulated. Each chain has an adaption period of 1000 draws, a burn-in period of 5000 draws, and a sample of 8000 draws. The thinning rate is set to 5. The parameters monitored are $\beta_0, \beta_1, \beta_2$, and σ . The three chains have the following starting values and random seeds:

1. $\beta_0 = 0, \beta_1 = 1, \beta_2 = -1, \beta_3 = 1$, random seed = 95063432.
2. $\beta_0 = -1, \beta_1 = 0, \beta_2 = 1, \beta_3 = 1$, random seed = 95073452.
3. $\beta_0 = 1, \beta_1 = -1, \beta_2 = 0, \beta_3 = 1$, random seed = 95086345.

The convergence of the chains was assessed through various methods. The trace plot, the density plot, and the ACF plot for each chain were checked to make sure the simulations converge and the correlation between the samples is low. Geweke diagnostics and Gelman-Rubin diagnostics were used to determine if the chains converge. Effective sample sizes for all parameters from each chain were checked to make sure that we ran the sampler long enough after convergence.

Besides checking the diagnostics for MCMC chains, we also checked if the model violates any regression assumptions. We used the residual plots to diagnose linearity and constant variance assumptions. A residual normal Q-Q plot was generated to check the skewness and the normality assumption for the model. Posterior prediction plot and predictive residual plot were used to diagnose if the model adequately describes the actual data. We also compared the empirical density and the CDF distribution of the data to the distributions of the simulated data from the posterior predictive distribution. Moreover, we investigated if the model captures the appropriate level of skewness and median by plotting the predictions against the observed response.

5 Results

By using the methods mentioned above, we generated the estimations for the parameters after running and combining the three MCMC simulations. The estimated median for β_0 is 7.997 with a 95% credible interval of 7.425 and 8.574. The estimated median for β_1 is 1.442 with a 95% credible interval of 1.260 and 1.626. The estimated median for β_2 is 3.139 with a 95% credible interval of 2.462 and 3.801. The estimated median for σ is 2.175 with a 95% credible interval of 1.971 and 2.376.

Param	Median	Mean	SD	Lower95	Upper95
beta0	7.997	7.998	0.294	7.425	8.574
beta1	1.442	1.442	0.093	1.260	1.626
beta2	3.139	3.136	0.341	2.462	3.801
sigma	2.175	2.179	0.103	1.971	2.376

Table 1. Summary of MCMC Chains

The trace plots for all parameters in each chain appear to be stationary. The ACF plots for all parameters in each chain decay quickly within 5 lags. The Geweke diagnostics for each chain are all below |2|. The Gelman-Rubin diagnostics show that the R_j values for all three chains equal 1. These two diagnostics indicate that the chains converge well. The effective sample sizes for all parameters in each chain are at least a few thousand, showing that the model is well mixed. The residual plots for both explanatory variables show that the residuals are evenly distributed around 0, indicating that the linearity and constant variance assumptions are not violated. The normal Q-Q plot is slightly S-shaped. This means that even though the distribution is symmetric, the tails are relatively too long compared to those of a normal distribution. However, we are not using this model for prediction. Therefore, the assumption of normality doesn't affect our statistical inference for the research question. The posterior prediction plot (Figure 1) shows the predicted y with prediction intervals versus observed y . Most observed y values with prediction intervals are close to the slope = 1 line, which means the predicted y are fairly close to the observed y , except for several points on the larger end of observed y .

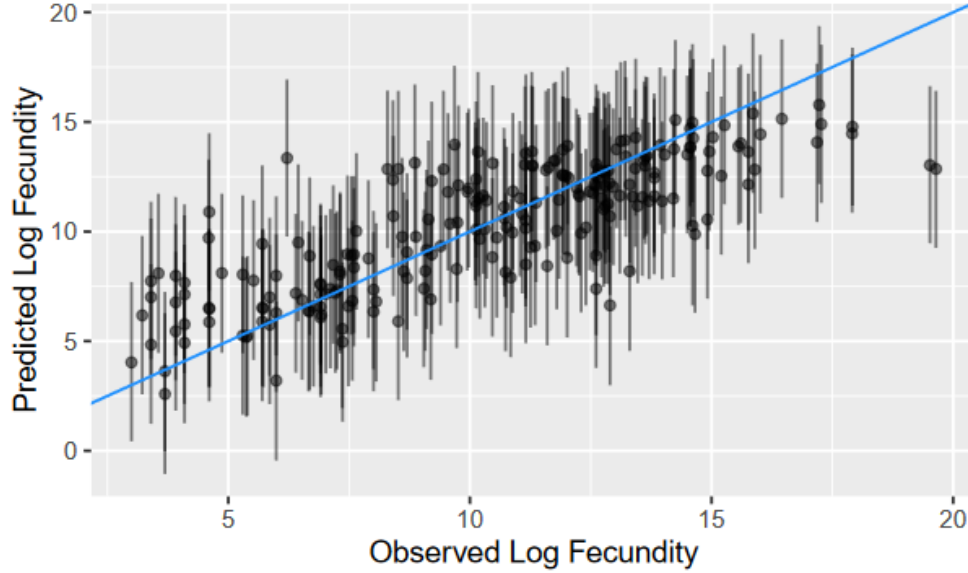


Figure 1. Posterior Prediction Plot for Log Fecundity

The predictive residual plots for both explanatory variables show that the residuals are evenly distributed around 0. The overlayed density curve plot shows that the model adequately describes the observed data, with a little under-fitting at the left end of the distribution. The overlayed CDF density curve demonstrates that the simulated values perfectly include the observed values. The skewness and median are also captured well by the model.

6 Discussion and Conclusions

Given the data and the model, we were able to convert the logged parameter values back to their original values and interpret them. A fish under class *Actinopteri* with an average brain weight living in freshwater is expected to have a median fecundity value of 2970.644. An m -fold change in brain weight is associated with a $m^{1.442}$ -factor change in the median fecundity value after accounting for the marine category. If the species is considered marine, there is a 23.089 factor change in the median fecundity value after accounting for brain weight. After calculating the estimated mean fecundity for a certain species, the standard deviation for that species' mean fecundity is 8.799.

There are some limitations in this Bayesian model. Firstly, we do not have the tool to determine the collinearity between the two explanatory variables. This has exceeded our current knowledge in the area. Although we plotted a boxplot of log brain size vs. marine and there does not seem to be a relationship between the two factors, we cannot be certain about this deduction. Secondly, there are probably some outliers in our data. However, we do not have the tools to identify them in a Bayesian setting. We suspect that the two data points on the furthest right in Figure 1 are potential outliers. Thirdly, the inadequacy in our model is probably still due to the lack of features. The performance of our model might improve if we are able to include more relevant features into the model.

These results add an alternative perspective regarding the trade-offs of brain weight. While Fischer and Jungwirth (2022) claimed that larger brains come with an evolutionary cost of lower fecundity, we see that this is not necessarily true in all fish classes. In *Actinopteri*, larger brain weight is associated with higher fecundity. Additionally, compared to freshwater, marine fish is associated with higher fecundity.

This project only focused on one fish class and accounted for two factors. There are other factors such as migration behavior, mating system, or social behavior. We couldn't determine any casual relationship between these factors and brain size, and mediating factors might act on the evolution

and different traits. However, this methodology could be generalized to other species and additional factors could be added to the model to achieve a more accurate representation. This study also shows how Bayesian inference about species statistics might provide insights into evolution biology.

In conclusion, brain weight and fecundity are positively correlated in *Actinopteri*. Each fish class might have its own evolution strategy, and increased brain weight doesn't always come with a cost of decreased fecundity.

7 References

- Bagenal, T. B., & Braum, E. (1978). Eggs and early life history. Methods for assessment of fish production in freshwaters. *IBP Handbook*, (3), 165-201.
- Fischer, S., & Jungwirth, A. (2022). The costs and benefits of larger brains in fishes. *Journal of Evolutionary Biology*.
- FishBase*. Search FishBase. (2015, October). Retrieved November 19, 2022, from <https://www.fishbase.se/search.php>
- Jungwirth, A. (2022, May 13). *Data for: The costs and benefits of larger brains in fishes*. Zenodo. Retrieved November 20, 2022, from <https://zenodo.org/record/6546756#.Y3ATxHbMJPY>
- Martin, R. D. (1981). Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature*, 293(5827), 57-60.
- Quinn, T. P., Vøllestad, L. A., Peterson, J., & Gallucci, V. (2004). Influences of freshwater and marine growth on the egg size–egg number tradeoff in coho and Chinook salmon. *Transactions of the American Fisheries Society*, 133(1), 55-65.