Data Analysis in Fisheries Science: Relationship Between Predator and Prey Lengths

Author : Maisie Beckett   **Supervisor** : Gustav Delius

Department of Mathematics

University of York

Heslington, York, UK

March 29, 2021

Abstract

In this project I will investigate the relationship between marine predators and their prey and try to find trends, specifically between their lengths. Using a large data set produced by Barnes et al. [2008], I will use R to identify these relationships using various techniques and plots. Looking at the relationship between predator and prey lengths has important applications in the fishing industry. The research in this paper could give reasons for a minimum length of fish that fisherman are allowed to catch, as the smaller fish could be part of the diet of more predators, and therefore a lack of fish under that minimum length could cause more fish to starve, reducing biodiversity.

Contents

1   Introduction

1.1   The Barnes Data Set

Throughout this project I will be using a data set produced by Barnes et al. [2008] called "Predator and Prey Body Sizes in Marine Food Webs". This data set has 34929 observations of 62 variables, but I will be focussing on just six of these; predator scientific name, predator length, predator length unit, prey scientific name, prey length and prey length unit.

1.2   What other studies have found and how they have influenced me

1.2.1   Costa's Research

Costa found that mean, minimum, maximum and variance of prey size were all positively associated with body size. This suggests that larger predators consume a variety of different sized prey types, increasing both the maximum and minimum size they eat. A similar result can be found in previous studies looking at terrestrial predators conducted by Vézina [1985] and Costa et al. [2008], who found that larger predators tend to consume a wider range of prey sizes. In a similar study which looks at predacious lizards, Costa actually found a negative relationship, concluding that larger predators were actually consuming a narrow diversity of prey categories.

In my project I will investigate this relationship further, looking at the length of each species rather than their mass to see if this follows the same patterns that Costa and Vézina found. Taking inspiration from Costa's calculations and figures, I will calculate the trophic-niche breadth of predators and plot this against their lengths in millimetres.

1.2.2   Scharf's Research

Scharf et al. [2000] used regression analysis to look at ratio-based trophic niche breadths and found that they generally did not expand with predator ontogeny [1] and tended to narrow for the largest predators. They also found that "the range of prey sizes eaten expanded with increasing predator body size for each of the marine predators examined, leading to asymmetric predator size - prey size distributions".

In my project I will investigate whether these results can be replicated using the Barnes data set (Barnes et al. [2008]). I feel this will be important as the Scharf study uses a relatively narrow data set of only 18 species, all off the northeast US coast, so it will be interesting to see whether these results can be found across the 27 locations, 93 predator types and 174 prey types that were observed by Barnes et al. Taking inspiration from figure 4 in the Scharf paper, I will produce a plot showing the ratio of prey length/predator length, the relative frequencies and the cumulative frequencies.

---

[1]"The entire sequence of events involved in the development of an individual organism" - Collins English Dictionary

1.3 Explanation of Mathematics Required

1.3.1 Ratio-Based Trophic-Niche Breadth

Trophic niche breadth B can be calculated using the following formula by Levins [1968]

$$B = \frac{1}{\sum p_i^2}$$

where p_i is the relative frequency of occurrence of the ith prey species; and standardised across prey species:

$$B_A = \frac{B-1}{N-1}$$

ranging from 0 to 1, N is the total number of the resource taxa (Lanszki et al. [2019]).

Scharf et al. [2000] examined trophic-niche breadth on a ratio scale by determining changes in the range of relative prey sizes with increasing predator size. The prey size data can be converted to a ratio scale by dividing each prey size by its corresponding predator size, these can then be plotted against the predator size. The trophic-niche breadth at any given predator size can then be represented by the difference between predicted values of upper and lower bound regressions; in our case we will choose the upper bound to be the 90% regression line and the lower bound to be the 10% regression line. By comparing the gradients of the upper and lower bound regression lines, we can see whether there is an increase or a decrease in ratio-based trophic niche breadth as predator size increases. If there is a significant difference in the gradients of the lines, we can conclude that there is a correlation between the ratio and predator size, whether this is positive or negative. Furthermore, divergent slopes indicate an increase and convergent slopes indicate a decrease in ratio-based niche breadth as predator size increases.

1.3.2 Regression Analysis

Let's say we wish to calculate the standard regression equation for predicting y (the dependent variable) from x (the independent variable), where y is the log of predator length and x is the log of prey length divided by predator length. The equation for this standard regression line will be

$$y = \beta x + \gamma$$

where β is the slope of the line, and γ is the y-intercept (Altman and Gardner [1988]). However, we want to calculate the 10% and 90% regression quantiles. The θth sample quantile, $0 < \theta < 1$, can be defined by any solution of the following minimisation problem

$$\min_{b \in R} \left[\sum_{t \in (t:y_t \geq b)} \theta \left| y_t - b \right| + \sum_{t \in (t:y_t \leq b)} (1-\theta) \left| y_t - b \right| \right]$$

where $y_t : t = 1, \ldots, T$ is a random sample of the random variable Y having distribution function F Koenker and Bassett Jr [1978]. So the 10% quantile would

be

$$\min_{b \in R} \left[\sum_{t \in (t:y_t \geq b)} 0.1 \left| y_t - b \right| + \sum_{t \in (t:y_t \leq b)} 0.9 \left| y_t - b \right| \right]$$

and the 90% quantile would be

$$\min_{b \in R} \left[\sum_{t \in (t:y_t \geq b)} 0.9 \left| y_t - b \right| + \sum_{t \in (t:y_t \leq b)} 0.1 \left| y_t - b \right| \right]$$

1.3.3   Hypothesis Testing

In order to determine whether the ratio regression lines I will later fit to the data are accurate, I will conduct a hypothesis test, more specifically, I will use a t-test. I must first make the assumption that the lengths of the predators in the data set are independent and identically distributed from a normal distribution with mean μ and variance σ^2 - i.e. $X_1, ..., X_n$ are iid from $N(\mu, \sigma^2)$. The null hypothesis, H_0, for this test will be that the 10% and 90% regression lines should actually be parallel (the slopes of the lines are equal) as there is no correlation between the ratio of prey size over predator size and the predator size. The alternative hypothesis, H_1, is that there is a significant difference in the gradients of the regression lines (the slopes of the lines are **not** equal), so there is a correlation between ratio and predator size. Say the slope of the 10% regression line is β_{10} and the slope of the 10% regression line is β_{90}, then are hypotheses can be written as

$$H_0 : \beta_{10} - \beta_{90} = 0$$

and

$$H_1 : \beta_{10} - \beta_{90} \neq 0$$

I will use a linear regression t-test to determine whether the difference between the slopes of the regression lines differs significantly from 0, using a significance level of 0.05 (Berman [2021]). For simple linear regression, the degrees of freedom (DF) is equal to:

$$DF = n - 2$$

We have 34929 observations in our sample, so we choose 34927 for our degrees of freedom. The test statistic is defined by

$$T = \left| \frac{\beta_{10} - \beta_{90}}{SE} \right|$$

where SE is the standard error of the slope, defined by

$$SE = \frac{\frac{\sqrt{\left| RSS_{10}^2 - RSS_{90}^2 \right|}}{\sqrt{n-2}}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

where RSS_{10} is the residual sum of squares for the 10% quantile, RSS_{90} is the residual sum of squares for the 90% quantile, x_i is the ith observed ratio of prey

length over predator length, $\bar{x}$ is the mean ratio, and n is the number of observations (34929). The residual sum of squares is defined by the following formula

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

where n is the number of observations, y_i is the ith observed predator length and $f(x_i)$ is the predator length predicted by the regression line.

Our rejection rule will be to reject the null hypothesis H_0 if the test statistic t is less than $t_{\frac{\alpha}{2}}$ (Hao et al. [2007]); the critical value for the two-sided student's t-distribution with confidence level $100(1-\alpha)\%$ that can be found by looking it up in a t-value table (Heagerty [2004]). The P-value is the probability, given H_0 is true, that we would observe a test statistic at least as extreme as the one we actually did observe; this will be calculated later.

2 Preparing the Data

In order to use this data, I first had to ensure it was usable and suitable for my area of research. On initial inspection of the data set, by plotting the log of the predator lengths against the log of the prey lengths, I could see anomalies where the smaller predators were consuming much larger prey.
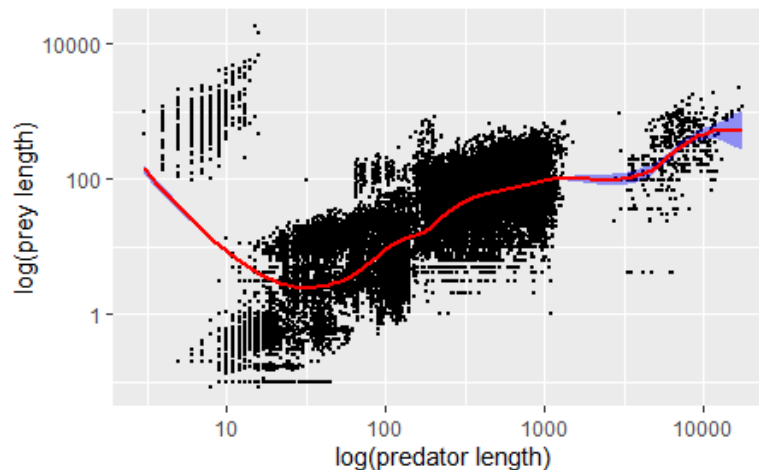

Figure 1: Graph showing the relationship between the log of predator length and the log of the length of their prey

Upon further inspection of the data set itself, I found that there were three different measurement units used for the lengths of the marine animals; these were micrometres (Âμm), millimetres (mm) and centimetres (cm). With a view to create accurate plots from this data set, I converted all length data into millimetres by multiplying the centimetre data by ten, and dividing the micrometre data by a thousand. The length data was then ready to be presented. As can be seen below, the anomalies are no longer present and a positive correlation between predator and prey lengths can be seen.
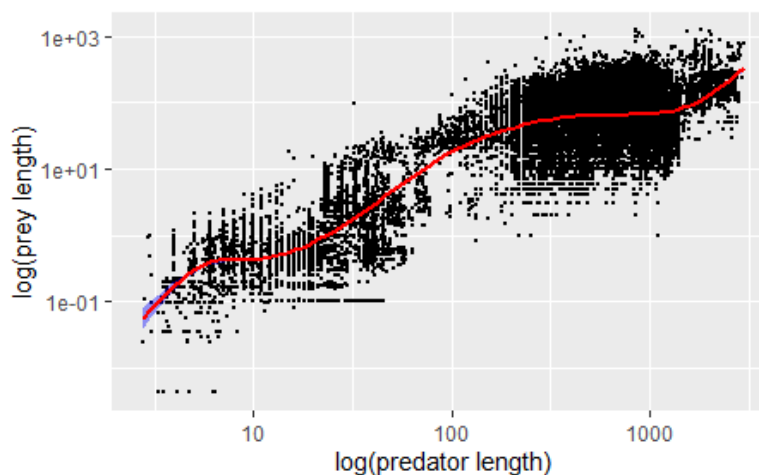

Figure 2: Corrected graph showing the relationship between the log of predator length and the log of the length of their prey

3   Results

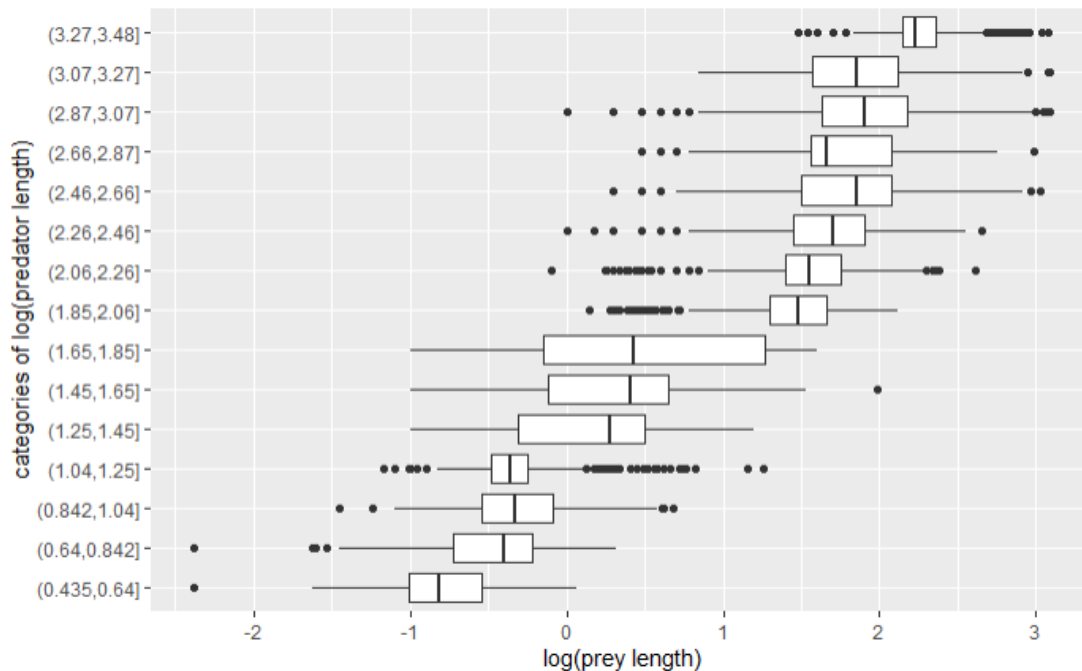## 3.1   Predator Length - Prey Length Patterns

Figure 3: Box plot, using a log scale, showing the range of prey lengths eaten by different categories of predator lengths

First, looking at the medians in the box plot above, we can see that, in general, the length prey increases with the length of predators. However, there are some cases where the median prey length actually decreases as the predator length increases, the clearest one of these is change in median from the $(2.46, 2.66]$ to the $(2.66, 2.87]$ predator length category.

The length of the box for the midsize predators appear to be the largest, with the longest being just under 2, which means that predators in the $(1.65, 1.85]$ category consume the largest variety in prey sizes.

There appear to be the most outliers for the largest and smallest predators; the outliers for the smaller predators could be due to difficulties measuring the length of the smallest prey.

The range of prey lengths consumed by each predator, including the outliers, appear to be very similar for each predator length category, with a value of around 2.5. This range suggests that predators consume prey with a difference of up to around $316mm$ between the largest and smallest species. On the other hand, if the outliers were omitted from the plots, the ranges for most of the predator categories decrease drastically, with the biggest decrease being for predators in the $(1.04, 1.25]$ category, who would consume prey with a difference of only $10mm$ between the longest and shortest prey species they consume.

The box plot for the $(2.66, 2.87]$ predator length category shows a right skew: the right whisker is longer than the left. For most of the categories, there doesn't appear to be a significant skew in either direction. Be that as it may, the drastic

outliers would undoubtedly have an effect on calculations of the skewness, mean and standard deviation, so these results should be taken reluctantly.

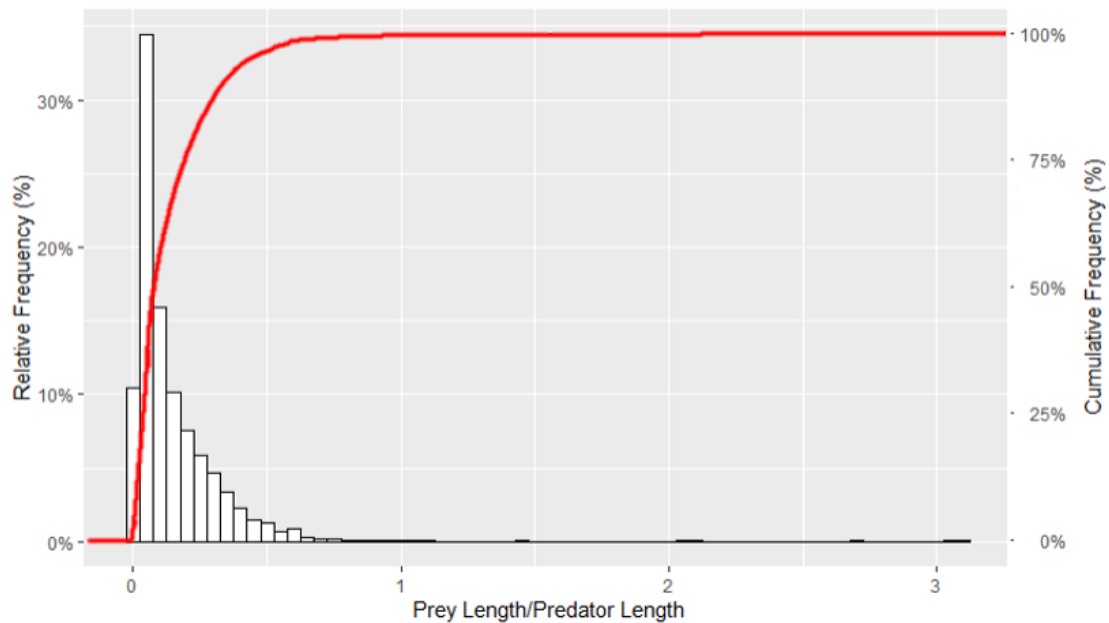## 3.2 Ratio-Based Trophic-Niche Breadth

Figure 4: Histogram showing relative frequency distributions of prey size/predator size ratios consumed by marine fish predators and empirical cumulative distribution function shown in red

This shows that predators generally prefer to consume prey that are around $\frac{1}{10}$th of their own size, with around 35% of their diets consisting of this prey size, and around 75% of their diets consisting of prey between $\frac{1}{20}$th and $\frac{1}{4}$th of their own size.

There appear to be a few instances where predators eat prey that are actually larger than or the same size as them (this is where Prey Length/Predator Length is greater than 1); however this is just a very small percentage of predators and could just be due to an error in the measurement of very small prey species.
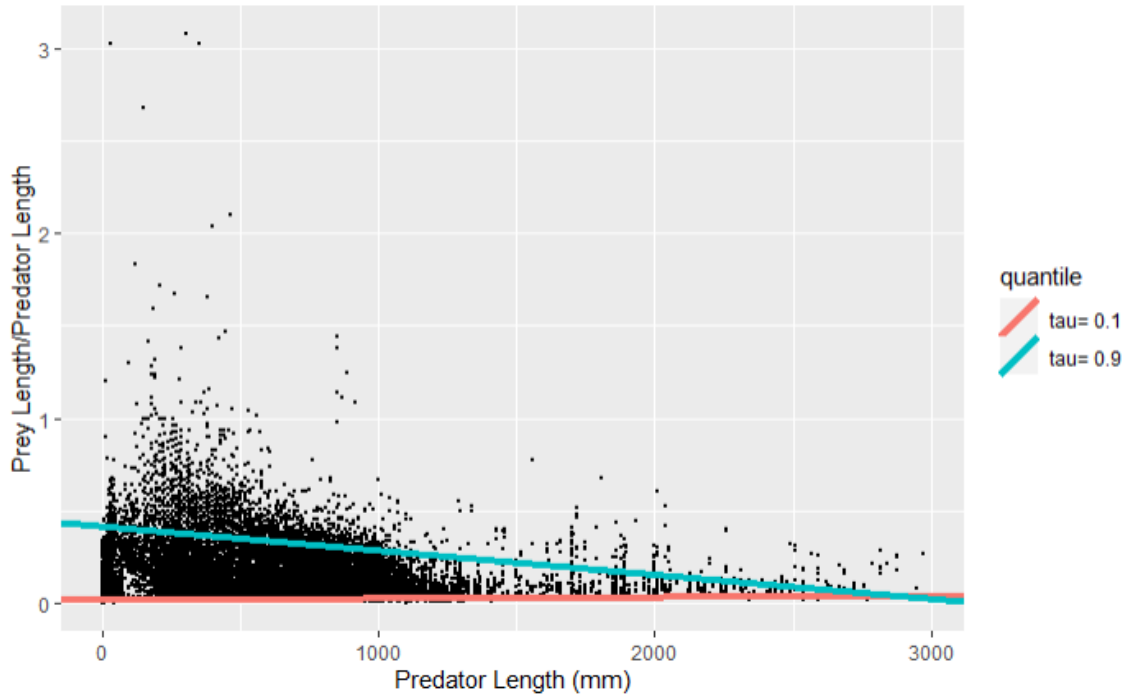
Figure 5: Scatter Diagram showing predator size as a function of prey size/predator size ratios and regression slopes for the 90th (in blue) and 10th (in red) quantiles.

In general, the ratio of prey length over predator length remains between 0 and 1, with some outliers for smaller predators showing them consuming prey up to 3 times the length of them, this does not seem possible and could be a result of difficulties in measuring the microscopic species to the same accuracy. As predators get larger, the spread of the points decreases, as predators only consume prey less than half the length of themselves. From the points alone, we can deduce that as predator length increases, the ratio-based trophic niche breadth decreases.

Furthermore, the slopes of the regression lines in figure 5 converge to the same point, so, by the mathematics explained in section 1.3.1, the convergent slopes indicate a decrease in ratio-based niche breadth as predator size increases. Several previous studies have not found a change in ratio-based trophic niche breadth with predator size in general, and only found a "general trend to indicate a decrease in the breadth of relative prey sizes eaten with ontogeny" Scharf et al. [2000]. In order to check that the regression lines in figure 5 are accurate, I will conduct a hypothesis test as described in section 1.3.3. Using R to calculate the RSS and denominator values, I was able to calculate the standard error to be the following

$$
\begin{aligned}
SE &= \frac{\frac{\sqrt{\left|RSS_{10}^2 - RSS_{90}^2\right|}}{\sqrt{n-2}}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \\
&= \frac{\frac{\sqrt{|1702.604 - 2334.381|}}{\sqrt{34929-2}}}{\sqrt{\sum_{i=1}^{n} (x_i - 327.3272))^2}} = \frac{\frac{\sqrt{|1702.604 - 2334.381|}}{\sqrt{34913-2}}}{\sqrt{16753917841}} \\
&= 0.000001039065853
\end{aligned}
\tag{1}
$$

From the above, I was able to calculate the test statistic

$$
\begin{aligned}
T &= \left| \frac{\beta_{10} - \beta_{90}}{SE} \right| \\
&= \left| \frac{0.000008163265 - (-0.0001316345)}{0.000001039065853} \right| \qquad (2) \\
&= 134.5417758
\end{aligned}
$$

Using a confidence level of 99.75%, we find that $t_{\frac{\alpha}{2}} = 127.32$. Our test statistic is larger than this t-value ($134.5417758 > 127.32$), so we can reject our null hypothesis H_0 and conclude that the 10% and 90% regression lines should not in fact be parallel, so there is a relationship between the ratio of prey length over predator length, and predator length. In particular, this suggests that there is a decrease in ratio-based trophic niche breadth as predator length increases.

Due to my findings from figure 5, I would like to look into species specific ratio-based trophic niche breadth, focussing on just 6 of the species in the data set of varying sizes. In order to choose predators of various sizes, I will look at a summary of the standard predator lengths, and choose those predators that have sizes corresponding to the minimum, 1st quantile, median, 3rd quantile and maximum.

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
2.74  280.00  630.00  653.65  910.00 2970.00
```

Figure 6: Summary of Standard Predator Lengths from R

I will therefore choose to focus on Lampanyctus crocodilus (89 entries), Merluccius bilinearis (1389 entries), Thunnus alalunga (3581 entries), Gadus morhua (2518 entries) and Thunnus thynnus (1909 entries). Looking up these predators in the table, we find that Lampanyctus crocodilus is a Myctophidae fish larva, Merluccius bilinearis is an Silver hake, Thunnus alalunga is an Albacore, Gadus morhua is an Atlantic cod and Thunnus thynnus is an Atlantic bluefin tuna.
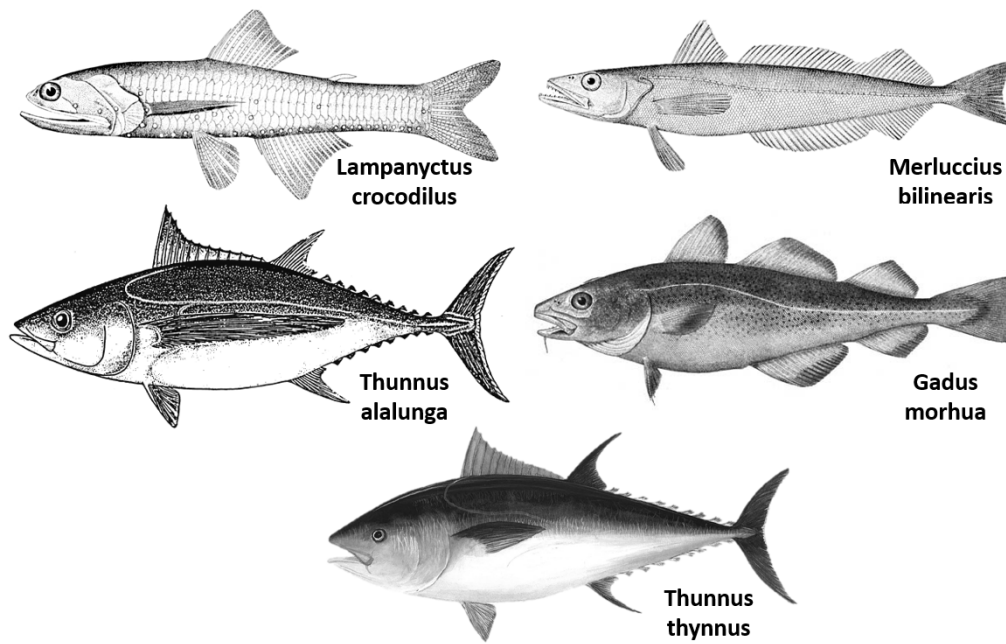
Figure 7: Image of the predators I have chosen to focus on.

Below is a table showing the slopes of the regression lines for my 6 species, influenced by table 4 in the Scharf et al. [2000] paper.

Predator Species	Predator length vs prey length/predator length ratio		Change in trophic-niche breadth
	Lower bound (10th quantile) slope	Upper bound (90th quantile) slope	
Lampanyctus crocodilus	0.006867515	0.013084018	Increase
Merluccius bilinearis	0.0003200000	0.0002645503	Very slight decrease
Thunnus alalunga	0.0001103343	0.0003577488	Increase
Gadus morhua	0.00003267916	-0.00005580357	Decrease
Thunnus thynnus	-0.00001796194	0.000005937222	Increase

Figure 8: Table showing the slopes of the upper and lower bound regression lines, and what this means for the ratio-based trophic niche breadth.

For the smallest predator I looked at, Lampanyctus crocodilus, the regression lines corresponding to the 10th and 90th quantiles appear to diverge, suggesting that, for the smallest predators, there is a positive correlation between ratio of prey size over predator size and predator size. The next predator I looked at was Merluccius bilinearis, the regression lines in figure 7 for these predators appear to be parallel, but by looking at the slopes we can see that they in fact converge,

suggesting a very slight increase in ratio-based trophic niche breadth as predator size increases. Next, I looked at Thunnus alalunga, whose regression lines diverge, which suggests that as the size of the Thunnus alalunga increases, so does the ratio of prey size over predator size. Penultimately, I looked at Gadus morhua and found that, due to the regression lines converging, the ratio-based trophic niche breadth decreases as predator size increases; this follows the relationship found previously within the entire data set. The final, and largest predator I looked at was Thunnus thynnus, the regression lines for which appear to diverge, suggesting a positive correlation between ratio-based trophic niche breadth and predator size, which is not what I expected to find for larger predators.
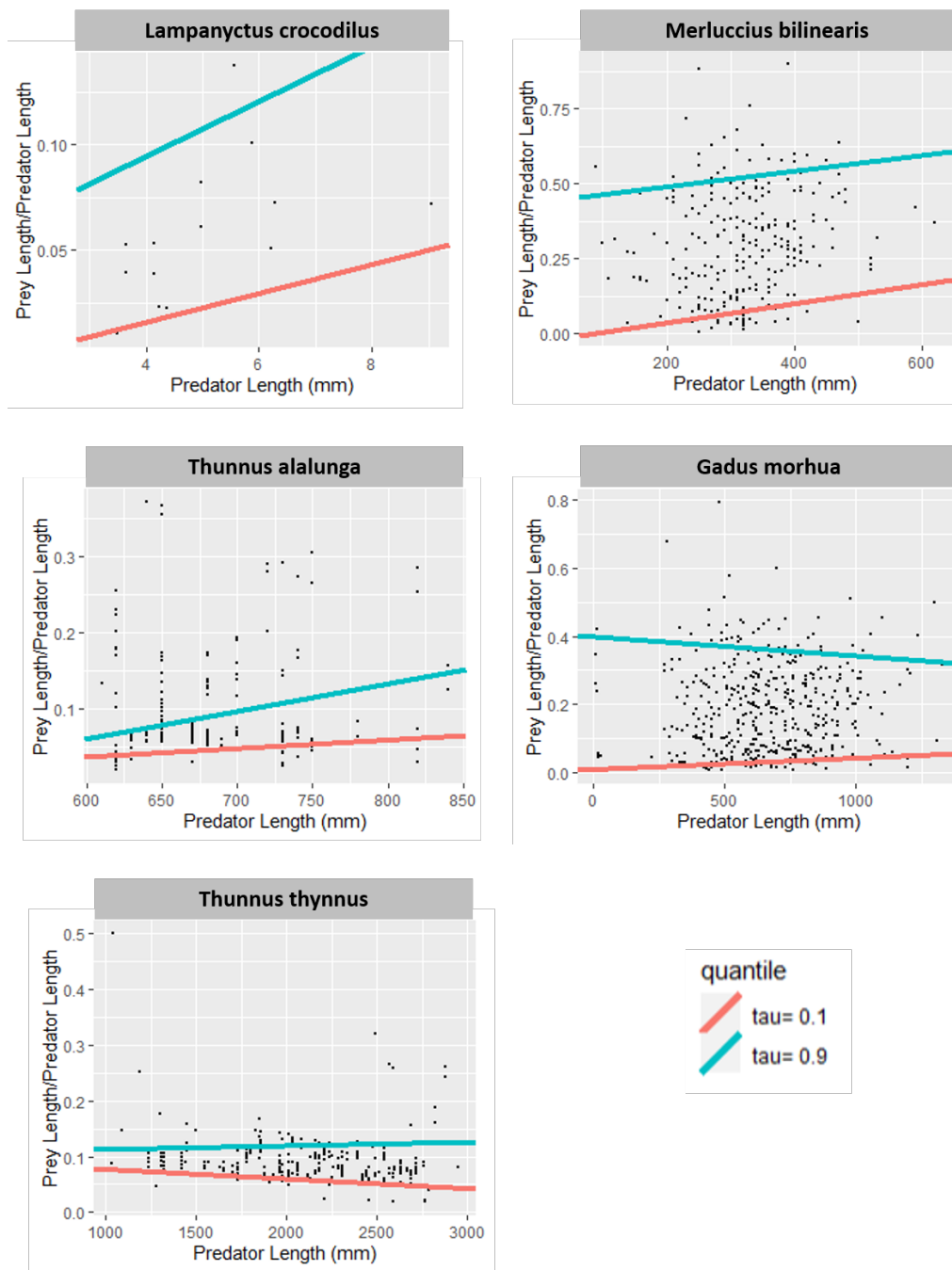
Figure 9: Scatter Diagram showing prey size/predator size ratios of 6 different predators as a function of predator size and regression slopes for the 90th and 10th quantiles.

According to the results of Scharf et al. [2000] "the general trend was for smaller average-sized predators to show no significant change, while the largest predators tended to show a decrease in ratio-based trophic-niche breadth with increasing body size", which is supported by my findings for Gadus morhua. This is also supported by the smallest species, Lampanyctus crocodilus, however, the sample size for this species is very small (there are only 18 points on the plot), so this finding could be due to overfitting, reducing the accuracy of the results.
I will now conduct a hypothesis test as before, but for the individual species. I will display my findings in a table for clarity.

Predator Species	Standard Error (SE)	Test Statistic (T)	Can we reject H_0 with confidence level 99.75?	Can we reject H_0 at a lower 95% confidence level?
Lampanyctus crocodilus	0.002499509616	2.487089052	No	No
Merluccius bilinearis	0.000009637433582	5.75357532	No	No
Thunnus alalunga	0.0000009926846672	249.2377572	Yes	Yes
Gadus morhua	0.000009324124494	9.489655577	No	Yes
Thunnus thynnus	0.000004637961019	51.52945853	No	Yes

Figure 10: Table showing the values used in a hypothesis test, similar to the one done previously, but for the individual species, rather than the entire population.

The values in the table show that we cannot reject the null hypothesis that the regression lines should be parallel at the same 99.75% confidence level for all but the Thunnus alalunga species, so we can't conclude that there is a relationship between predator length and ratio of prey length over predator for each individual species in the same way we could for the entire population.
We cannot reject H_0 at either confidence level for the Lampanyctus crocodilus and Merluccius bilinearis species; by never rejecting this H_0 we are completely ruling out the possibility of a type I error (rejecting H_0 when H_0 is actually true), meaning the size of the test is 0.
The 5th column in the table in figure 10 shows that if we had chosen a 95% confidence level rather than 99.75%, it would be possible to reject H_0 and conclude (with less confidence) that there is in fact a correlation between predator length and ratio, for 3 out of the 5 species. The findings for the 95% confidence level support the findings by Scharf et al. [2000], as for the 2 smaller species we cannot reject the null hypothesis that the regression lines are parallel, showing that there is

no significant change in ratio-based trophic niche with increasing predator length. Our findings for the Gadus morhua species also supports their findings, as the regression lines converge, so the ratio-based trophic niche breadth decreases with increasing predator length. However, the regression lines for the Thunnus alalunga and Thunnus thynnus species diverge, implying the ratio-based trophic niche breadth actually increases with predator size; this contradicts what Scharf et al. [2000] had said.
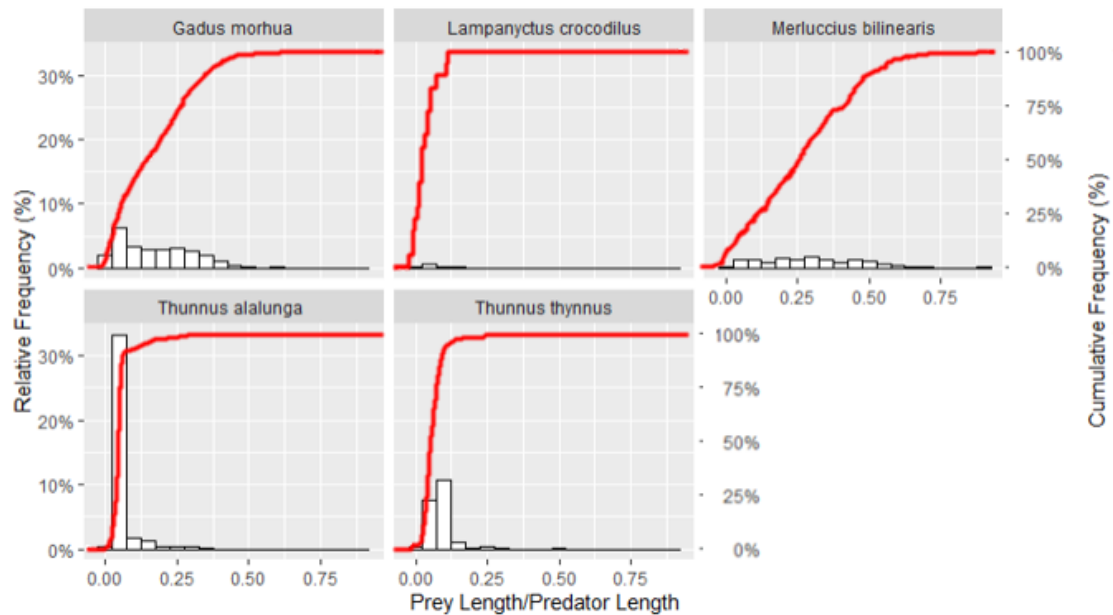
Figure 11: Histogram showing relative frequency distributions of prey size/predator size ratios consumed by 6 chosen predators and empirical cumulative distribution function shown in red

Although all predators only consume prey smaller than them, there appears to be some variation between the different predators in the relative frequency distributions of prey size/predator size ratios.
The relative frequency bars for the diets of the Lampanyctus crocodilus species are extremely short, so it is difficult to deduce the size of prey they eat from this graph, this predicament could be a result of problems with measuring the microscopic-sized dietary choices found in the stomachs of the equally microscopic predator species, the smallest of the 5 species we have analysed in further detail. However from the cumulative frequency line of Lampanyctus crocodilus, we can see that 100% of the prey they eat are less than $\frac{1}{8}$th of the predator's size. The size of the relative frequency bars for the Merluccius bilinearis species propose the same problem as that of the Lampanyctus crocodilus, but we can't learn much more from the cumulative frequency line either. The Thunnus alalunga have a clear preference for prey $\frac{1}{10}$th of their size, with this constituting over 30% of their diets, and 100% of the prey they consume are less than $\frac{3}{8}$ths of their size. The diet of the Gadus morhua is pretty equally spread between prey up to half the length of the predator. For the largest species, Thunnus thynnus, we can see that about 10% of their diets comprises prey $\frac{3}{20}ths$ of their size, and almost 100% of the prey they eat are less than half their own length, a figure we also saw in the slightly smaller Gadus morhua species.

4   Conclusions

List of Figures

References

Carolyn Barnes, DM Bethea, RD Brodeur, J Spitz, V Ridoux, C Pusineri, BC Chase, ME Hunsicker, F Juanes, A Kellermann, et al. Predator and prey body sizes in marine food webs: Ecological archives e089-051. *Ecology*, 89 (3):881–881, 2008.

Alain F Vézina. Empirical relationships between predator and prey size among terrestrial vertebrate predators. *Oecologia*, 67(4):555–565, 1985.

Gabriel C Costa, Laurie J Vitt, Eric R Pianka, Daniel O Mesquita, and Guarino R Colli. Optimal foraging constrains macroecological patterns: body size and dietary niche breadth in lizards. *Global Ecology and Biogeography*, 17(5):670–677, 2008.

Frederick S Scharf, Francis Juanes, and Rodney A Rountree. Predator size-prey size relationships of marine fish predators: interspecific variation and effects of ontogeny and body size on trophic-niche breadth. *Marine Ecology Progress Series*, 208:229–248, 2000.

Collins English Dictionary. Ontogeny definition and meaning: Collins english dictionary. URL https://www.collinsdictionary.com/dictionary/english/ontogeny.

Richard Levins. *Evolution in changing environments: some theoretical explorations*. Princeton University Press, 1968.

József Lanszki, Miklós Heltai, György Kövér, and Andrzej Zalewski. Non-linear relationship between body size of terrestrial carnivores and their trophic niche breadth and overlap. *Basic and Applied Ecology*, 38:36–46, 2019. ISSN 1439-1791. doi: https://doi.org/10.1016/j.baae.2019.06.004. URL https://www.sciencedirect.com/science/article/pii/S1439179118302354.

Douglas G Altman and Martin J Gardner. Statistics in medicine: calculating confidence intervals for regression and correlation. *British medical journal (Clinical research ed.)*, 296(6631):1238, 1988.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Harvey Berman. Stat trek, 2021. URL https://stattrek.com/regression/slope-test.aspx.

Lingxin Hao, Daniel Q Naiman, and Daniel Q Naiman. *Quantile regression*. Number 149. Sage, 2007.

Patrick Heagerty. t-tables, 2004. URL https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/t-Tables/.