Estimation of North Atlantic Right Whales' Population Size

Chongkyung Jason Kim '22

worked with Horace Shew '22, advised by Professor Steve Wang

1. Introduction

Context

North right whales are one of the most endangered large species in the world, mainly caused by human interventions such as commercial hunting, entanglement with fishing gear and vessel strikes, which includes collision with ships or anything that humans ride on the sea.

Objective

The goal of our project is to accurately assess and estimate the population size of North Atlantic Right whales.

2. Dataset

North Atlantic Right Whale Consortium (NARWC) dataset of 70,000+ sighting entries. North Right Atlantic whales are individually identifiable by their callosities, irregular patches of tissues on their backs that are like human finger prints: no two whales have the same callosity. Two things we need to keep in mind: 1) some sighted whales that are recorded in the dataset are now dead 2) some living whales may have never been sighted.
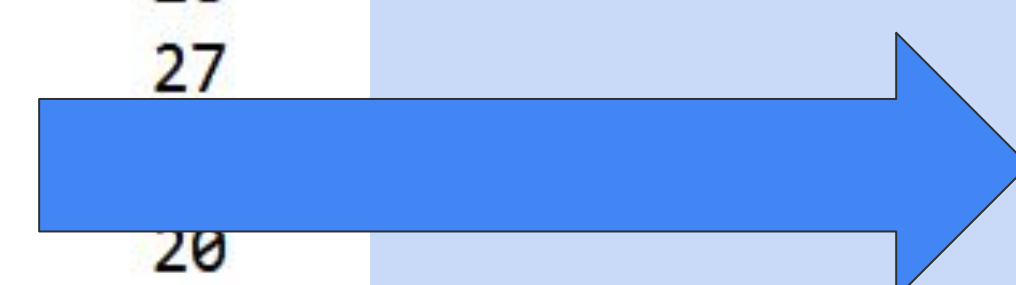
Transformation

Interested in each individual whale's probability of being currently alive, we parsed and transformed the original sighting-level dataset to a whale-level dataset of 724 whale individuals

New Vessel Strike Data

We introduced a new dataset that recorded whales that suffered from vessel strikes and its severity, classified by one of "SUPERFICIAL", "SHALLOW", and "DEEP". Eventually, we added two new indicator variables "strikeEver" and "deepStrike" to our logistic regression model.

SightingEGNo	GenderCode	Age		AgeClassCode	
1001	F	U	U	1978	3	31
1001	F	U	U	1978	10	20
1001	F	U	A	1979	2	28
1001	F	U	A	1983	4	27
1001	F	U	A	1983	5	
1001	F	U	A	1983	5	20
1001	F	U	A	1983	5	22
1001	F	U	A	1983	6	17
1001	F	U	A	1983	6	30
1001	F	U	A	1983	8	4
1001	F	U	A	1983	8	10
1001	F	U	A	1983	8	11
1001	F	U	A	1983	8	11
1001	F	U	A	1983	8	14

NARWC Dataset: each entry is a sighting of a whale

	id	sex	age	firstyear	lastyear	sightings
1	1001	1	1	1978	1992	77
2	1004	1	1	1975	2004	95
3	1005	2	1	1982	1997	23
4	1006	1	3	1979	1986	22
5	1007	1	1	1977	1986	21
6	1012	1	1	1978	2017	97
7	1013	1	1	1980	2007	39
8	1014	1	1	1974	1999	78

Transformed Dataset: each entry pertains to a unique whale individual and its relevant info

3. Methodology

1. Logistic Regression Model

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.83334    0.40514   6.993 2.68e-12 ***
adult          0.30128    0.26187   1.150  0.24994
female         0.12015    0.32654   0.368  0.71290
feedEver       0.64407    0.38173   1.687  0.09155 .
momwcalfEver   0.42251    0.22102   1.912  0.05592 .
momwcalfNow    1.43550    0.66327   2.164  0.03044 *
entglEver     -0.58816    0.61050  -0.963  0.33534
entglNow      -1.00288    0.53078  -1.889  0.05883 .
disentglEver   0.70618    0.55472   1.273  0.20301
strikeEver    -0.06488    0.64126  -0.101  0.91941
deepStrike    -2.08322    0.76071  -2.739  0.00617 **
```

We first estimate the probability that an individual whale is alive given its sighting record, using the method of Solow (1993). Then, we use our logistic regression model to adjust those initial probabilities with our covariates, thereby estimating the effect (the odds ratio) of these variables on the probability of each whale being alive.

Here, odds ratio is equal to e raised to the power of the corresponding coefficient. For example, being female multiplies the odds of being alive by $e^{0.120} = 1.127$; on the other hand, being "DEEP" struck by a vessel multiples a whale's odds by $e^{(-2.08322)} = 0.1245$.

2. Parallelism

Once we adjust the initial Solow probabilities with our logistic regression model, we simulate parallel, random sampling 1000 times by using the foreach package instead of a normal for loop to get our final estimation for the population size of North Atlantic Right whales. With the foreach package, we execute the simulations on multiple cores/processors on the computer, hence a parallel execution. Once we arrive at the sample of 1000 values of population sizes, we take the mean of those figures to conclude with our final estimation.

4. Results/Future

After we execute the parallelized simulation based on the adjusted probabilities, we arrive at an estimated population size of 387. We are still working on better ways to generate more accurate initial probabilities for whales with too little sightings to begin with. I would like to thank Professor Steve Wang for hsi guidance and Horace Shew '22 for being an amazing partner.

Estimated population size