Modeling Population Decline Report

Peter Boyd

11/27/2018

Abstract

As species around the globe face declining populations, it is pertinent to model these endangered species. Through regression and simulation techniques, we may consider when these species may become extinct at the current rate of change with no preventative measures taken. In this report, I will discuss framework and data used, function creation and output, an example of my function at work, and possible future work.

Framework and Data

My goal for this poject is to create a function that selects the best fitting model a particular data set and uses this fit to simulate data and predict an eventual point of extinction for the species.

All of the data used in this project was downloaded from ______, a website that contains information for many species counts at various locations, some of which are population totals, which will be the data chosen for my analysis.

Function Creation and Output

The function will have quite a few sections in order to fully implement all of the desired operations.

First, the the function fits several models to the data. Due to their frequent use in discrete glm scenarios, the function will fit generalized linear models with specified distributions Gaussian, Negative Binomial, and Poisson. That is, the models are called as

```r
glm(Population ~ Year, data = df)
glm.nb(Population ~ Year, data = df)
glm(Population ~ Year, data = df, family = "poisson")
```

Once these models have been fit, the Akaike Information Criterion is assessed for each model, and the best fitting model is selected as the model producing the lowest AIC value.

Once the best model has been selected, the function will simiulate data based using the fitted values (and other parameters if the distribution requires them) from the best model. Once many simulations have been created, they are used to fit new models. Finally, after fitting many new models, the model parameters are averged to produce an overall simulated model.

The predicted extinction date can be found by solving the general regression equation:

$$\hat{y} = \beta_0 + \beta_1 x$$

for $\hat{y} = 0$; so, the predicted date at which the population is extinct can be approximated by

$$\text{Predicted Extinction Date} = \frac{-\beta_0}{\beta_1}$$

.

Finally, the function will return each fitted model's AIC, the predicted extinction date from the best fitting model, a plot with each model depcited, and a residual plot for the best model.

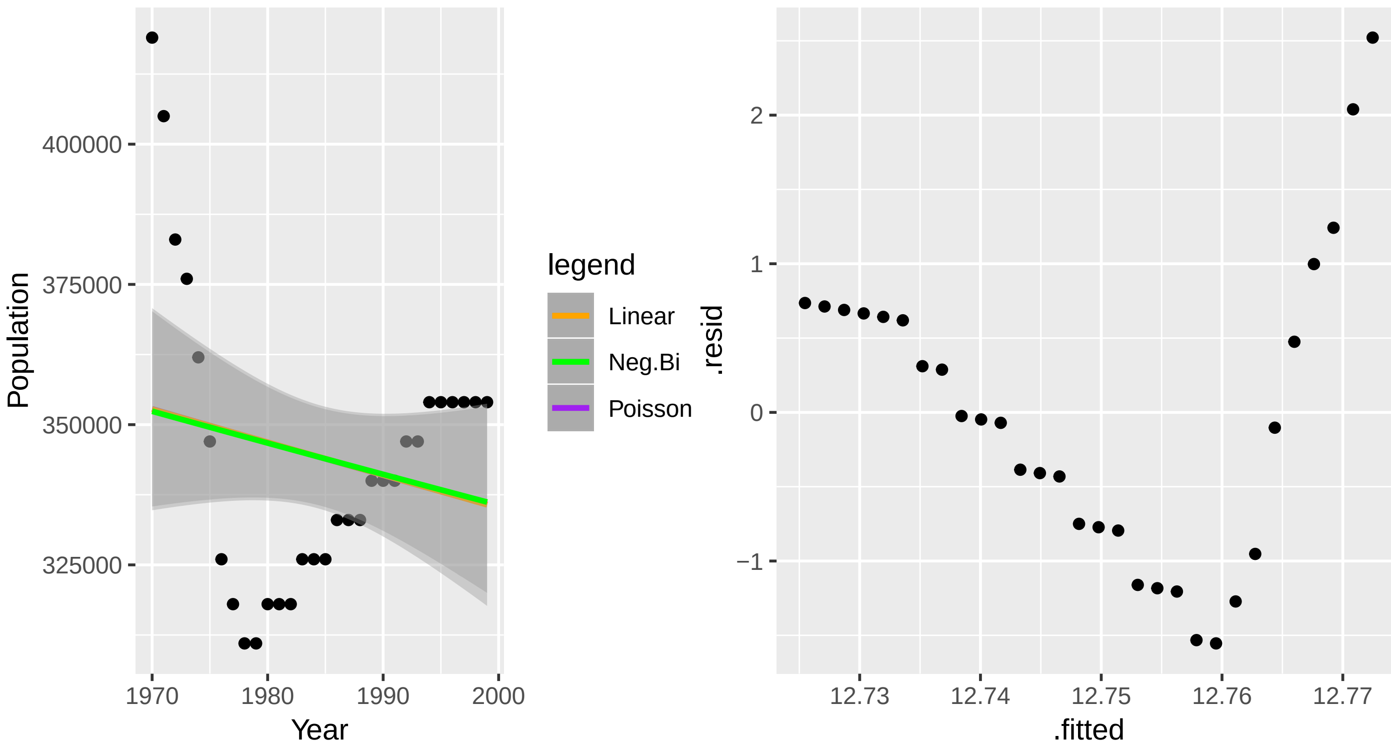
Example Using Whale Data Set

In order to demonstrate the usefulness of my function, I will produce an example of the function implemented for a data set. I will first use the function for annual estimates of the global population of sperm whales. This data set contains estimates between the years 1970 and 1999. After importing the file, we must also define `ntimes`, or the number of simulations we will produce. In this example, we will call the function as

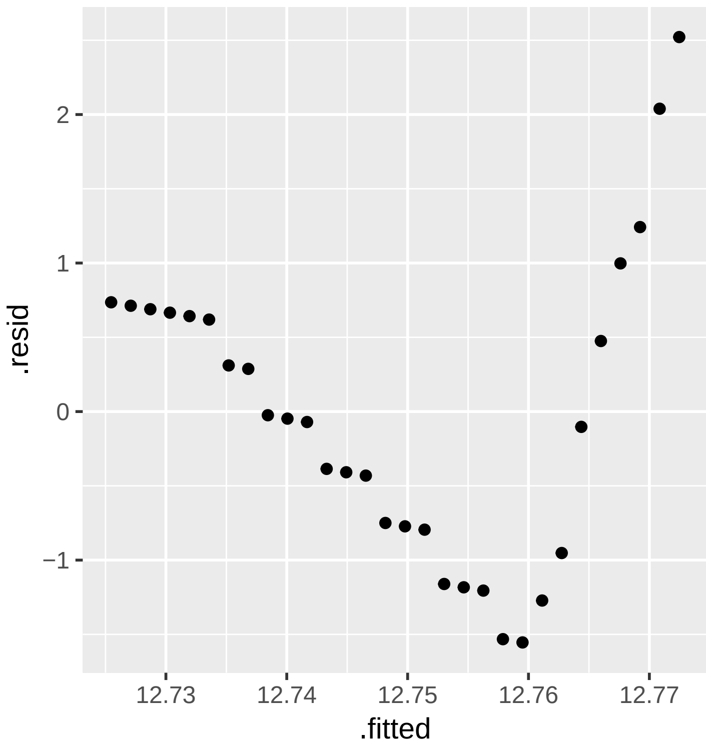
```
pop.decline(df = whale, ntimes = 100)
```

As outlined above, the function will return a list of each model's AIC, predicted extinction, and 2 plots:

```
## AIC for Each Model
##  Linear Model :  698.5794
##  Poisson :  53225.4
##  Negative Binomial :  697.0227
## Predicted Extinction:  2111.485
```




```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
```

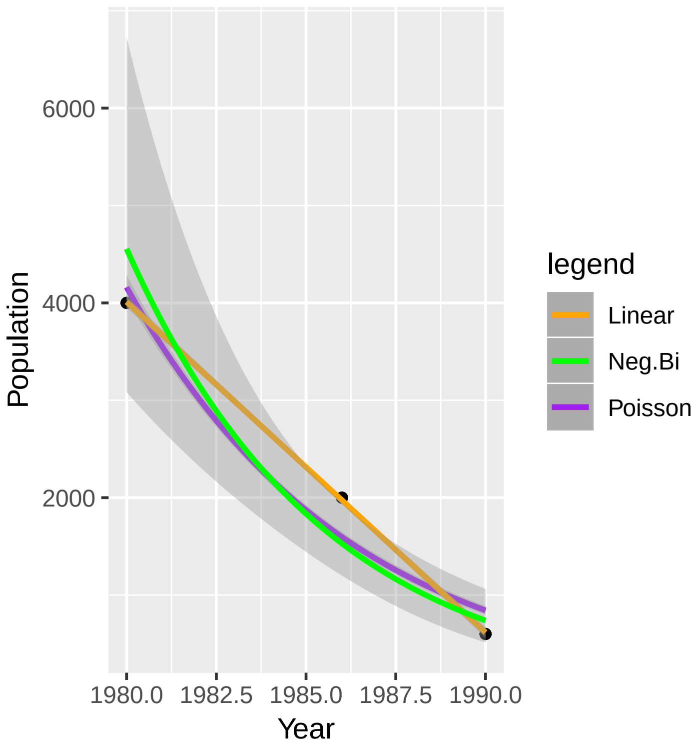


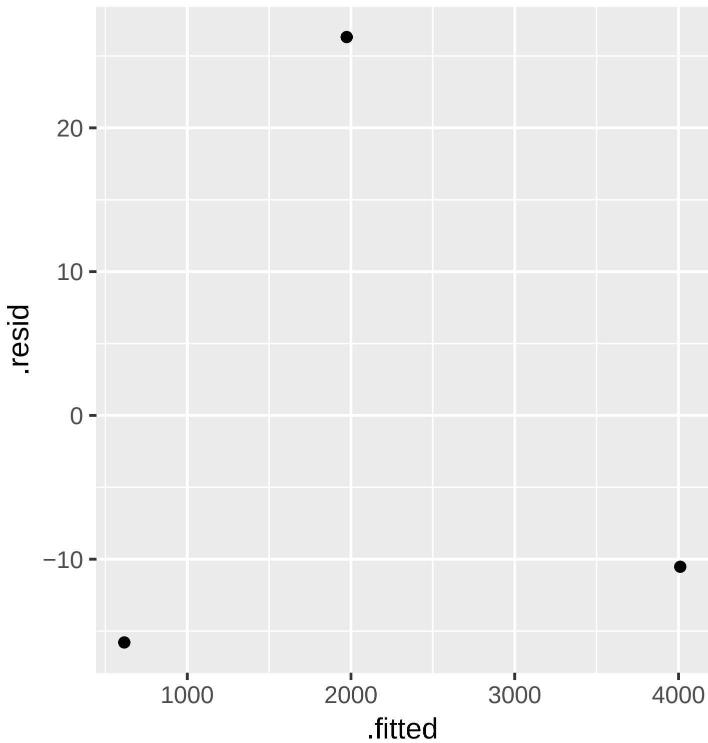
So, for the sperm whale, simulation is done via negative binomial glm regression, and predicted extincion is around the year 2114. This prediction is many years into the future, so we are extrapolating. However, it is still worth investigating the notion of a predicted extinction.

Example Using Addax Data

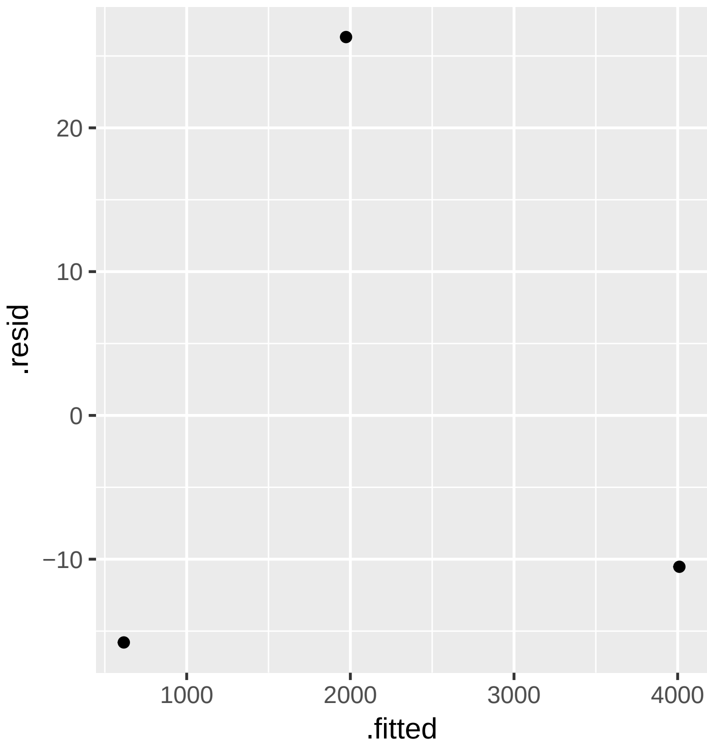
As a supplemental example, we repeat the above process for a data set containing population values for addax.

```
## AIC for Each Model
##  Linear Model :  32.09494
##  Poisson :  210.3932
##  Negative Binomial :  49.82695
## Predicted Extinction:  1990.351
```






```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
```



Future Work

Although my function runs quickly and produces seemingly reasonable estimates, there is always room for improvement. In this section, I will outline several ideas that I had attempted to implement but abandonded after spending a disproportionate amount of time that seemed to be beyond the scope of this course.

In simulating new values based on the best model, I utilized the `apply()` function. I initially had employed for loops, but ran timing experiments (found in Week 10 Tasks.Rmd) and found that the apply function was considerably faster. I had hoped to use `map()`, but found that a by-column iteration of a function over columns of a matrix may not be within the scope of the purrr library (https://github.com/tidyverse/purrr/issues/341).

I had attempted a more rigorous simulation methodology, but eventually abandoned the attempt after spending many hours that did not yield an accurate solution. In this approach, I tried to follow the following steps:

- use original data to fit a model
- simiulate data from fitted values
- fit many models for the simulated values and predict the next value for each model
- find the average of these predicted values and assume that this average will be the next "observed" value
- repeat this process as long as the predicted population is above 1

Unfortunately, this process proved to be too taxing with the other components of my already existing work.

Conclusions

In this project, I model population decline by creating a function that selects the best fitting model, simulates new data, and predicts when the population will become extinct. While there are possible extenstions and improvements, I feel that the function as it currently stands does a good job of creating a plausible model of population decline and extinction.