Capstone Project Modeling Report

Introduction to Data Science

Kevin Tajkowski

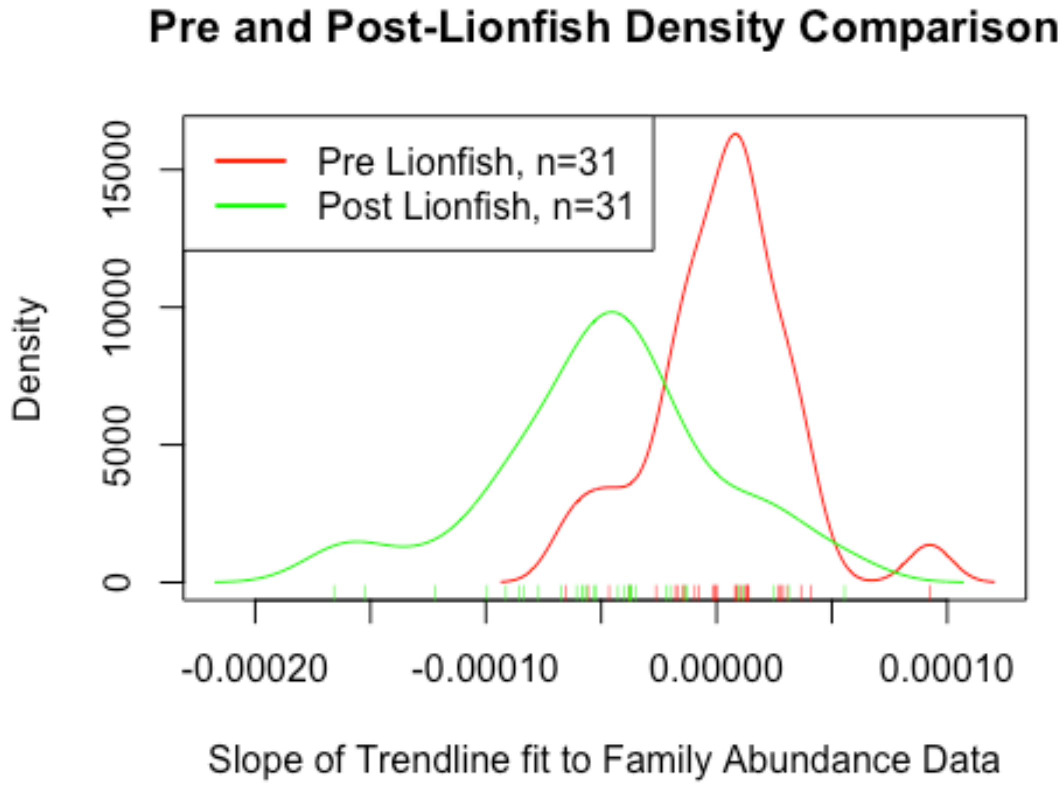
Data Frame Manipulation

The final steps of preparing the data frame consisted of removing various species. First, the red lionfish was removed from the data set leaving four species of scorpionfish. Second, the total number of surveys of all species in each family was determined. Families below a certain threshold were removed from the data set. Two hundred ten (210) was chosen as the threshold.

Density Comparison

A linear model was used to study the abundance of fish families over the twenty-four year period. Each fish family was analyzed individually. Two plots (shown earlier) were generated of abundance versus date. The first plot consisted of survey data before the first lionfish sighting, which occurred on 17 February 2009. The second plot consisted of survey data recorded after this date.

The linear model was used to generate a trendline of the data to show a general increase in fish numbers, a decrease in fish numbers or a steady quantity for each period. The slope of each trendline was calculated and saved to the appropriate vector: pre_slope or post_slope. These two vectors were used to generate a density plot of the slopes of trendlines for the two time periods. The plot below shows the two density curves of this data. As indicated in the plot, the pre-lionfish data shows a concentration of slopes higher than the slopes of the post-lionfish data.

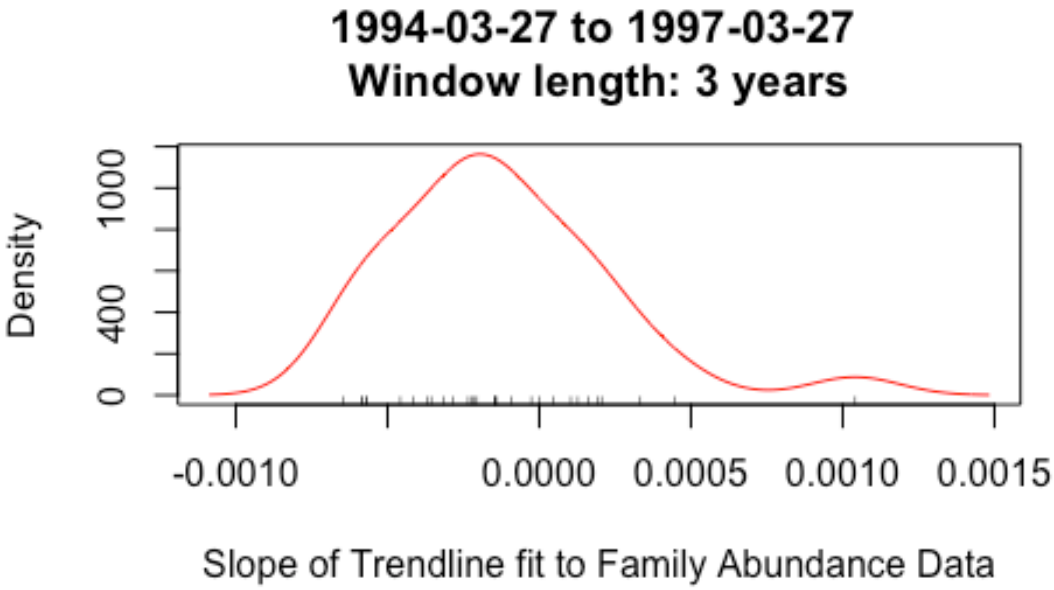


Moving Window Average Slope

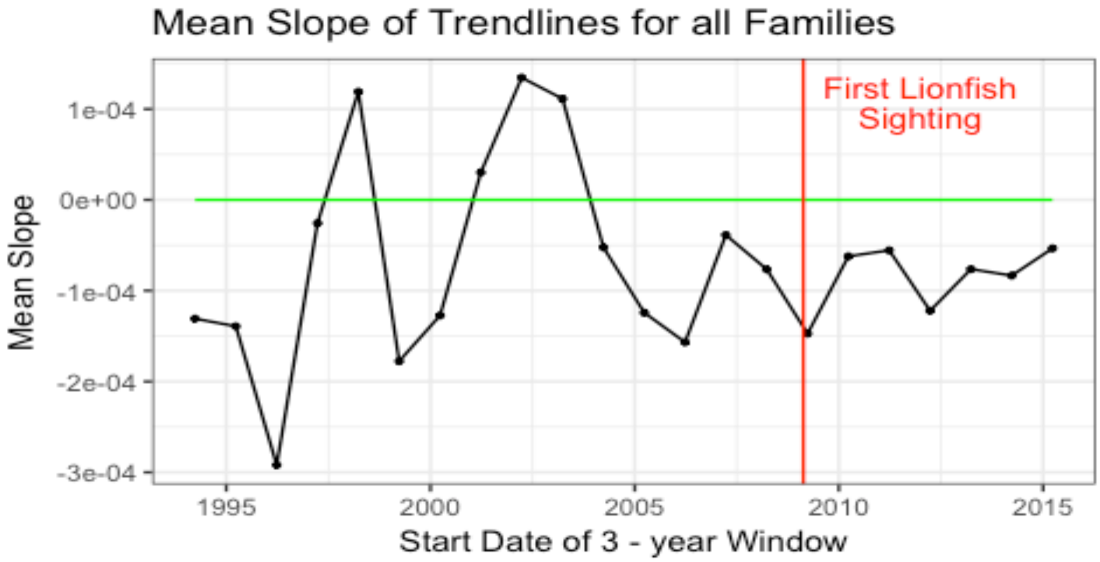
The next step of the analysis consisted of using a moving window to study the survey data of each family. The date of the first survey recorded is 27 March 1994 and the date of the last survey available is 17 April 2018. Survey windows of 3, 4, 5 and 6 years were studied. The start date of each window was chosen as March 27th and the survey window was shifted by one year for each consecutive window. Partial windows at the end of the 24-year period were not used.

The linear model was again used to identify trendlines in the abundance versus date data and to calculate the slope of each trendline. Density plots of slopes were generated for each window, in the same manor as the pre-lionfish and post-lionfish data above. The first 3-year window density plot is shown below.

Finally, the average slope for each density plot was determined and these average values were plotted over the coarse of the 24 years (see below for the 3-year window). The average slope remains negative following the introduction of the lionfish suggesting that there are fewer fish families with a positive slope to the trendline of abundance versus date.



The slopes of the trendlines of abundance versus date are extremely small (on the order of 1e-04) and cannot be used as an indication of significant decline in fish numbers. Hunting efforts of the invasive red lionfish near Little Cayman may be successful in preventing a serious decline in native species.


Mean Slope of Trendlines for all Families
First Lionfish Sighting
Mean Slope
1e-04
0e+00
-1e-04
-2e-04
-3e-04
Start Date of 3 - year Window