Bumpus Data Exploration

Dharun Somalingam

Executive Summary

The given report provides a detailed analysis of the relationship between the survival and morphological characteristics of the sparrow. The Dataset consists of 136 randomly selected samples of House sparrows(English sparrows) that were available, including the Sex of the bird, Total length of the bird, AlarExtent(wings measurement), Weight of the bird, BeakHead(Length of beak and head), Humerus, Femur, Tibiotarsus, SkullWidth(width of the skull bone), Sternum(Length of knee sternum) and the Survival status of the sparrow.
We employed a Binary Logistic Regression Model to examine the Bumpus data, as it offers a more meaningful interpretation for a dichotomous outcome, namely survival (either alive or perished). Our analysis indicates that the survival of English Sparrows is significantly influenced by factors such as Total Length, Weight, Sex, and other size-related variables, which are likely to play a crucial role in determining the survival rates of Bumpus Sparrows.

Introduction

Background

house sparrow, (Passer domesticus), also called the English sparrow, is one of the world's best-known and most abundant small birds, sometimes classified in the family Passeridae (order Passeriformes). It lives in towns and on farms, worldwide, having accompanied Europeans from its original home—most of Eurasia and northern Africa. It was introduced into North America at Brooklyn, N.Y., in 1852 and within a century had spread across the continent. It is a 14-cm (5.5-inch) buffy-brown bird with a black bib (male only) [1]. House sparrows breed nearly year-round in warm regions. The nest, containing four to nine eggs, is an untidy bundle of straw and feathers—usually quite dirty—placed in house eaves. Both birds of the pair take part in building the nest. Formerly, large sparrow populations were supported by waste grain from the feed of horses, and the number of sparrows in urban areas declined as horses were replaced by automobiles.

Research Related

Hermon Bumpus, a professor of Comparative Zoology at Brown University, conducted a well-known study involving domestic sparrows (House Sparrows, Passer domesticus) that were subjected to an ice and snowstorm in 1898 [2]. The study aimed to test Charles Darwin's Theory of Natural Selection[3]. During the storm, Bumpus collected a total of 136 sparrows that had been affected by the adverse weather conditions.Out of the 136 sparrows collected, 72 survived, while the remaining 64 perished. Bumpus then examined and compared various physical characteristics and traits between the surviving sparrows and those that did not survive

Researchers have reanalysed Bumpus data and compared the different ways of measuring selection based on their findings. Based on Bumpus data Lande and Arnold introduced the concept of "the selection gradient" and "the selection differential" to quantify and understand the strength and direction of selection acting on traits[5]. The study by Crespi and Bookstein in 1989 likely examined the phenomenon of coefficient fluctuation in multiple

regression analysis. Multiple regression is a statistical technique used to analyze the relationship between a dependent variable and two or more independent variables.

Source of DataSet

This dataset comes from a research project conducted by Hermon Bumpus in 1898. Bumpus's study is remarkable because it contains precise measurements for 136 birds, all of which are discussed in his published paper[4]. The dataset includes information about the birds' sex (a categorical variable), their survival status (indicated as either surviving or not, a binary variable), and various measurements related to the birds' physical characteristics(continuous variables).

About Analysis and Data

The analysis aims to predict the survival of the sparrows based on their morphological estimations and the Sex of the bird (10 different variables). So, the classification of Data into Categorical and Numerical Variables. Categorical: At first Sex, sparrows have two different sexes. Those are male and female. Numerical Variables: The TotalLength and AlarExtent, which are the measures from the tip of the beak to the tip of the tail and tip to tip of the extended wings. The weight of the bird, BeakHead(length of the beak and head). Lengths of Humerus, Femur, Tibiotarsus and SkullWidth (width of the skull). Sternum is the length of the keel of the sternum and Survival is a Binary Variable. In a further section, we will describe the statistical methodology used in our analysis.

Methodology

Data processing

Check the data clearly, find any mismatchings or null values in the given data before performing the analysis and interpret the Summary.The data contains the survival status of the sparrows along with their identifier and sex and several morphological measurements made on the birds including TotalLength, AlarExtent, Weight, BeakHead, Humerus, Femur, Tibiotarsus, SkullWidth and Sternum. Variations in characteristics were analysed to determine how they affected the chances of survival.

Generalized Linear Model (Binary logistical regression)

We will first explore the data by numerical summaries. Following this a Generalised Linear Model will be fitted to the data with Survival as a response while factoring out the 'Sex' variable to identify which were Male and Female. Interaction terms will also be investigated.

The model will be reduced to significant terms only. The final model will be interpreted to explain the dependence of Survival on morphological characteristics. All statistical analysis will be conducted in the R statistical environment.

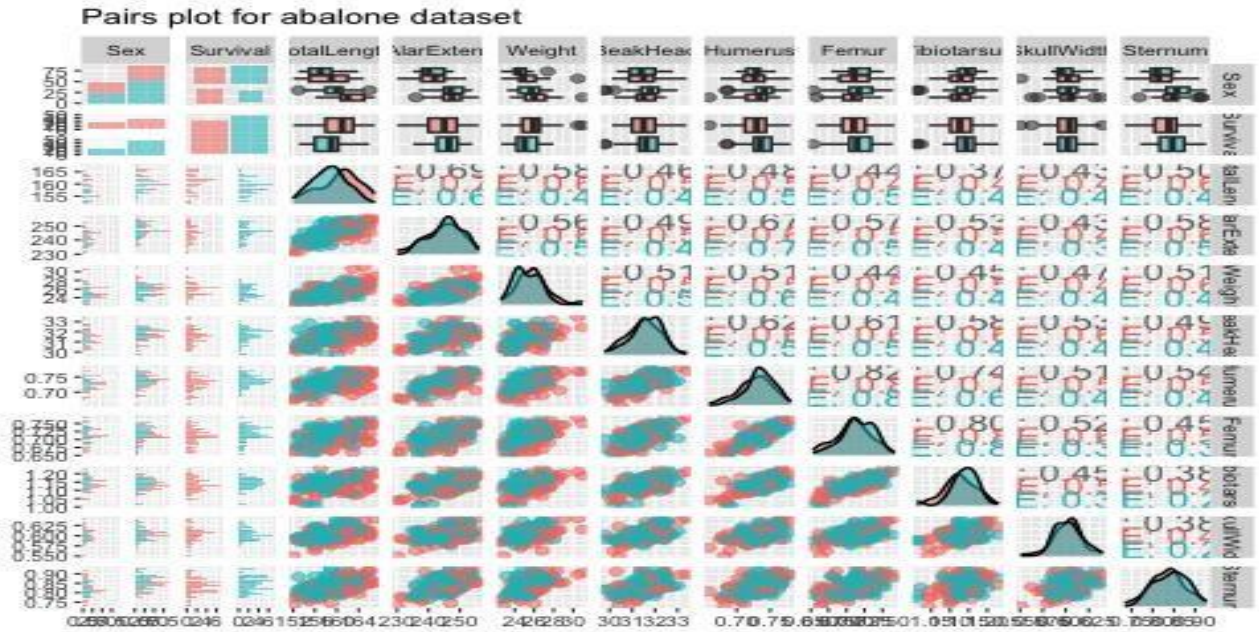
Result

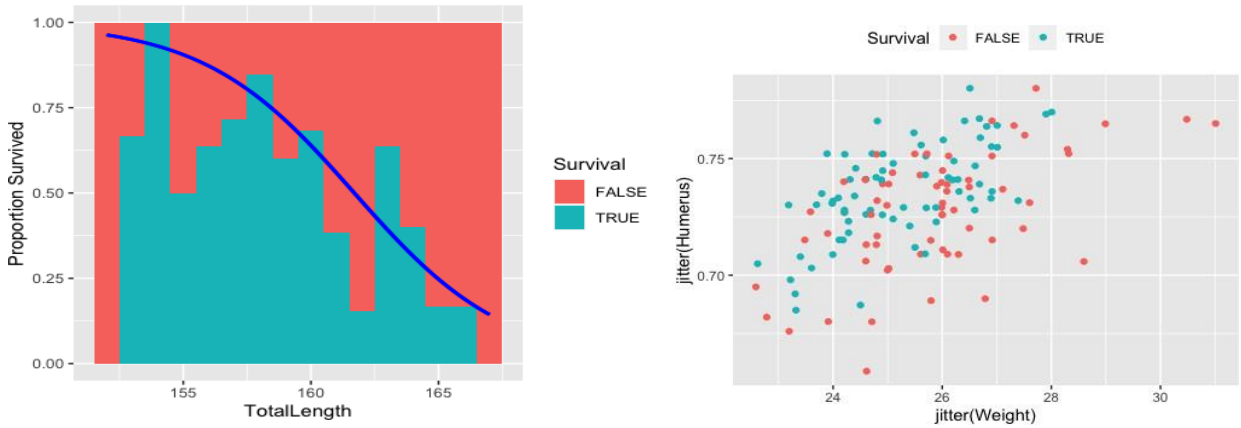
In 1898, Hermon Bumpus, an American biologist working at Brown University, collected data on one of the first examples of natural selection directly observed in nature. Immediately following a bad winter storm, he collected 136 English house sparrows, Passer domesticus, and brought them indoors. Of these birds, 64 had died during the storm, but 72 recovered and survived. By comparing measurements of physical traits, Bumpus demonstrated physical differences between the dead and living birds.All variables were included in a Generalised Linear Model, giving the reader a good interpretation of the variables and the survival status. The linear predictor for survival is:

Variable	Description	Summary
ID	an identifier of bird	Min. 1st Qu. Median Mean 3rd Qu. Max. 1.00 34.75 68.50 68.50 102.25 136.00
Sex	(m = Male, f = Female)	Length:136 , f : 49, m: 87
Survival	survival status (T = Survived, F = Died)	FALSE:64 TRUE :72
TotalLength	measured from tip of the beak to the tip of the tail (mm)	Min. 1st Qu. Median Mean 3rd Qu. Max. 152.0 157.0 160.0 159.5 162.0 167.0
AlarExtent	measured from tip to tip of the extended wings (mm)	Min. 1st Qu. Median Mean 3rd Qu. Max. 230.0 242.0 246.0 245.2 249.0 256.0
Weight	weight of the bird (g)	Min. 1st Qu. Median Mean 3rd Qu. Max. 22.60 24.57 25.55 25.52 26.50 31.00
BeakHead	length of beak and head, measured from tip of the beak to the occiput (mm)	Min. 1st Qu. Median Mean 3rd Qu. Max. 29.80 31.10 31.60 31.57 32.02 33.40
Humerus	length of humerus (inches)	Min. 1st Qu. Median Mean 3rd Qu. Max. 0.6590 0.7177 0.7330 0.7319 0.7482 0.7800
Femur	length of femur (inches)	Min. 1st Qu. Median Mean 3rd Qu. Max. 0.6530 0.7017 0.7130 0.7130 0.7312 0.7670
Tibiotarsus	length of tibiotarsus (inches)	Min. 1st Qu. Median Mean 3rd Qu. Max. 1.011 1.112 1.133 1.134 1.162 1.230
SkullWidth	width of skull measured from the postorbital bone of one side to the postorbital bone of the other (inches)	Min. 1st Qu. Median Mean 3rd Qu. Max. 0.5510 0.5920 0.6020 0.6025 0.6110 0.6400
Sternum	length of keel of sternum (inches)	Min. 1st Qu. Median Mean 3rd Qu. Max. 0.7340 0.8090 0.8410 0.8399 0.8652 0.9270

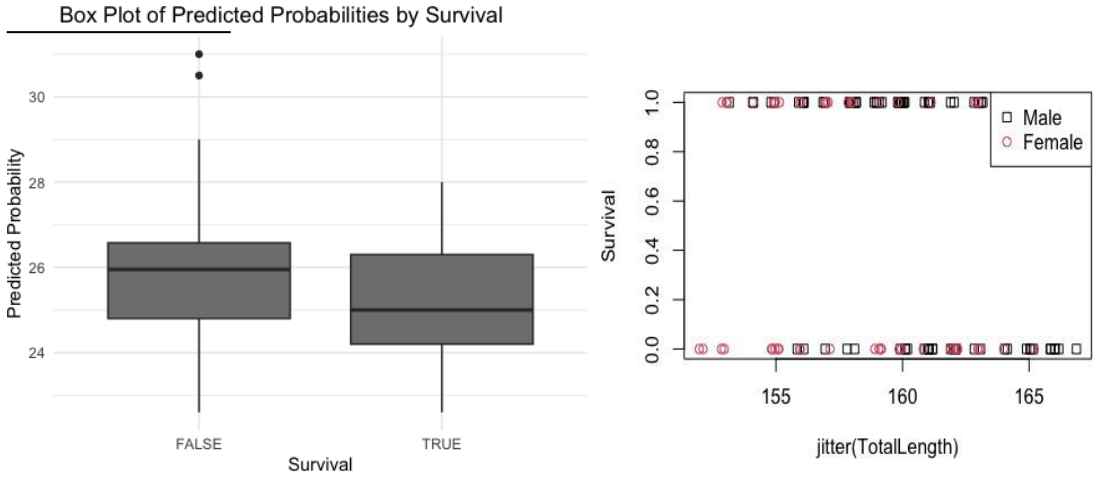
β0 + β1Sex + β2TotalLength + β3AlarExtent + β4Weight + β5BeakHead + β6Humerus + β7Femur + β8Tibiotarsus + β9SkullWidth + β10Sternum. A Generalised Linear Model for all variables was fitted to the equation against survival, to determine if any significance could be drawn.

logit(Survival)=10.83 + 1.64Sexm -0.43TotalLength +0.02AlarExtent -0.86Weight +0.48BeakHead +40.85Humerus -2.72Femur +6.89Tibiotarsus +13.74344SkullWidth +16.25Sternum





A logistic linear regression model was constructed to predict survival using various independent variables. To refine the model, stepAIC was employed for variable selection. The final model retained the significant predictors: Sex, TotalLength, Weight, Humerus, and Sternum, each associated with a coefficient representing its impact on survival prediction. This model highlights the importance of gender (Sex) and various measurements (TotalLength, Weight, Humerus, and Sternum) in determining the likelihood of survival.



Predicted

Actual	FALSE	TRUE
FALSE	46	18
TRUE	14	58

logit(Survival)=22.92+ 1.45Sexm -0.38TotalLength -0.75Weight +46.35Humerus+17.96Sternum

Variable	Odds_Ratio
(Intercept)	1.069826e+10
Sexm	4.690000e+00
TotalLength	6.800000e-01
Weight	4.700000e-01
Humerus	1.309506e+20
Sternum	6.324761e+07

In the logistic regression model for Survival, significant findings emerged. Notably, males had substantially higher odds of survival, approximately 4.69 that of females, indicating a gender-related disparity in survival rates. Additionally, Total Length and Weight were inversely associated with survival odds, with each unit increase in these variables leading to reduced odds of survival, at 0.68 and 0.47, respectively. In contrast, Humerus and Sternum exhibited positive effects on survival odds, with one-unit increases resulting in odds ratios of 1.3 respectively. These insights, derived from the odds ratios within the logistic regression model, shed light on how each variable influences survival odds while accounting for the influence of other variables. From the prediction table

we can explicitly seen that 18 birds were wrongly prediction as dead and 14 birds were incorrectly prediction as alive out of 58.

Discussion

In this report, we conducted an analysis of the Bumpus sparrow dataset using a logistic regression model to predict sparrow survival, considering factors such as Sex and various morphological variables. The model, described by equation (1), highlights the significant negative impact of TotalLength, Weight, and Femur on survival. However, a more nuanced understanding of these variables emerges when interaction terms are introduced. Notably, a model without interaction terms appears to provide a better fit compared to models with interactions.

The final model's predictions suggest that smaller male sparrows with reduced Total Length and lower Weight tend to have a higher likelihood of survival. It's important to note that several researchers have explored this dataset using diverse analytical techniques. For instance, Butterman utilized the Differential method and found that smaller female sparrows with larger sizes exhibit greater consistency [7]. In contrast, Johnston et al. employed the principal component method for a detailed analysis, excluding certain variables like Total Length, Wingspan, and Weight [6].

However, according to our model, lower values of Humerus and Sternum are associated with an increased risk of perishing among the sparrows.

Reference

[1].Britannica, The Editors of Encyclopaedia. “house sparrow”. Encyclopedia Britannica, 21 Sep. 2023, https://www.britannica.com/animal/house-sparrow. Accessed 25 September 2023.

[2].Bumpus, H. C. (1899). “The Elimination of the Unfit as Illustrated by the Introduced Sparrows (Passer domesticus and Passer ake)” in the journal Biological Lectures from the Marine Biological Laboratory, Woods Hole, Volume 6, pp. 209-226.

[3].Darwin, Charles (1859). “On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.”

[4].Bumpus, H. C. 1898. Eleventh lecture. The elimination of the unfit as illustrated by the introduced sparrow, Passer domesticus. (A fourth contribution to the study of variation.). Biological Lectures: Woods Hole Marine Biological Laboratory 209–225.

[5].Lande, R., & Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. Evolution, 37(6), 1210. https://doi.org/10.2307/2408842

[6] SJohnston, Richard F., David M. Niles and Sievert Rohwer. “HERMON BUMPUS AND NATURAL SELECTION IN THE HOUSE SPARROW PASSER DOMESTICUS.” Evolution 26 (1972): n. pag.

[7] Buttemer, William A. “Differential Overnight Survival by Bumpus’ House Sparrows: An Alternate Interpretation.” *The Condor* 94, no. 4 (1992): 944–54. https://doi.org/10.2307/1369291.