To determine the relationship between survival and the other variables of sparrows in the Bumpus data

Vinayak Kumar Gupta (24066272)

September 6th ,2023

Executive Summary

The analysis is based on an incident which happened on 1, February 1898 after an uncommonly severe storm with snow, rain and sleet and of long duration, 136 samples of moribund house sparrows were brought to the biological laboratories of Hermon Bumpus at Brown University, Providence, Rhode Island. Almost half these birds died as a result of stress from the storm. Bumpus made specimens of all the birds, recording gross body weight at the time, and later took measurements of different variables[1]. The variables of birds are Sex (male, female), Total Length of the bird(mm), Survival (T = Survived, F = Died), Alar Extent - measured from tip to tip of the extended wings (mm), Weight(g), Beak Head: length of beak and head(mm), Humerus Bone(in), Femur Bone(in), Tibiotarsus Bone(in), Skull Width(in), Sternum Bone(in). Our aim is to find the relationship between Survival of the birds and the other Variables. A Generalized Linear Model with only few significant variables is used. Those variables are Sex, Total Length, Weight and Humerus. The reason is that rest of the variables did not have much effect on the survival of the birds, alone and with the interactions as well. The model became too complicated to understand. Although from statistical point of view interactions were better but nothing was significant. The final model showed that male birds have higher chances of Survival than females, shorter birds in length and smaller birds in weight are more likely to survive, longer the humerus bone higher the chances of Survival of Birds.

Introduction

Due to an unfortunate event in 1898 these sparrows became a part of Hermon Bumpus Project. Bumpus saw an opportunity to see whether he could find any support for Charles Darwin's theory of natural selection. 125 years ago in the age of no Computers, multivariate statistical methods had hardly begun. The correlation coefficient as a measure of the relationship between two variables was devised by Francis Galton in 1877. As a fact Bumpus did not even calculate standard deviations. It was another 56 years before Harold Hotelling described a practical method for carrying out a principal components analysis. Nevertheless, his methods of analysis were sensible. Many authors have re-analysed his data and have confirmed his conclusions [2].

Richard F Johnson, David M. Niles and Sievert A. Rohwer did published a paper [1] in 1971 in which they used principal component analysis for analysis and summarized their findings and also wrote about the adequacy of variables in Bumpus data.

Insights into Adequacy of Variables

According to them "The measurements of total body length and of alar spread in birds are subject to sufficient observational error to be unreliable for the study of morphologic variation. Today no one uses these variables, but they are precisely the two external size characters that Bumpus chose to measure [1]. Bumpus tells us virtually nothing about how he treated the birds on receipt in his laboratory, but we assume that had surviving birds been fed after he obtained them. These assumptions are sufficient to allow an explanation of observed differences between the weights of Survivors and Non-Survivors that has nothing to do with presumed selection by the storm. Also, having survived cold stress longer than some of the Non survivors, survivors would have been characterized by a continuous expenditure of energy ultimately to be reflected in a loss of body weight [1]. They assumed that the skeletal characters are adequate for testing the likelihood that what Bumpus hypothesized was derivable from the data, but that body length and wing spread are less adequate, and body weight unacceptable, as useful variables [1]."

Because of these reasons, two sets of 9 and 6 variables were used for analysis [1]. Each set had different observations. I will focus only on 9 variable sets just to get a broader view of the analysis and its significance(table given in the paper). [1] Principal component analysis was employed.

Analysis showed that for males the disposition of individuals along component 1 generate the hypothesis that males were subject to directional selection in which large individuals had a higher probability of surviving than small ones, but Fisher's exact probability ($p = .274$) computed for a median test (Siegel, 1956) did not show this ordering to be significant [1]. An additional point of interest is that most survivors had relatively long keels and skulls, giving them relatively large negative component II scores. For females each of the nine variables also loads significantly on component 1, which accounts for 64% of the variation; component II represents an inverse relationship of total length and body weight against femur and tibiotarsus lengths, account for 11% of variation. [1] Observations of some analysts are collected in the table 1 below [3] -

Publisher	Tool	Observations
Bumpus [2]	Multivariate Assessment	He observed that the birds which perished, perished not through accident, but because they were physically disqualified, and that the birds which survived, survived because they possessed certain physical characters. Specifically, he found that the survivors "are shorter and weigh less, have longer wing bones, longer legs, longer sternums, and greater brain capacity than the non-survivors.[2]
RDS [1]	Principal Component Analysis	Large males have a selective advantage over small ones under conditions of severe winter cold stress. Females of intermediate size have a selective advantage over larger and smaller individuals under conditions of severe winter cold stress. Small subadult (=first year) males are at a selective disadvantage relative to small adult males under severe winter cold stress.
Harris_1911 [3]	Mean/SD/Standard Error	Among females, adult males, and juvenile males, shorter, lighter birds survived better. Survivors in all three groups had longer humeri and femurs but showed no consistent differences in wing, sternum, or tibiotarsus length or head size. No evidence of stabilizing selection.
Calhoun-1947[3]	Chi-square Tests	Treated only humerus and femur length; no evidence of stabilizing selection in either sex in these traits. Some evidence of directional selection for longer humeri in males ($.05 < P < .1$).

In my model I have used Total Length, Weight, Sex and Humerus of the Bird because they had a major impact on survival of the birds. In the next section I will describe statistical method which was used for making the model followed by the results and their interpretation.

Methodology

I first started out at looking at the data. It seamed to be fine and generally did not had any observable outliers. Then comparison of Survival was done with other 10 variables (only Body parts + Sex) to get a rough idea of how they are related to each other. To compare, box plots were made and it showed high variations in lines of variables with respect to Survival. I found out that Survival showed strong variation with Length, Weight, Humerus and Sex. A logistic regression model was used. With reference level set to Survival = 'TRUE' and Sex = 'M', a model was made. I first made a model with all the interaction terms with 30 + variables, but this made my model more complicated and was statistically more insignificant even after doing Step AIC. So I sticked to only single variables which was easy to interpret and those variables were also statistically significant. My statistically significance value (p-value) was taken as $\alpha = 0.05$. I chose the final model by Drop-in Deviance Test Method. It is more accurate than Wald's Test and is used to compare two nested models based on an estimator of the 'fit'. One by one I removed the less significant variable and compared the two models with ANOVA function in R and their results, plots and interpretation are discussed below.

Results

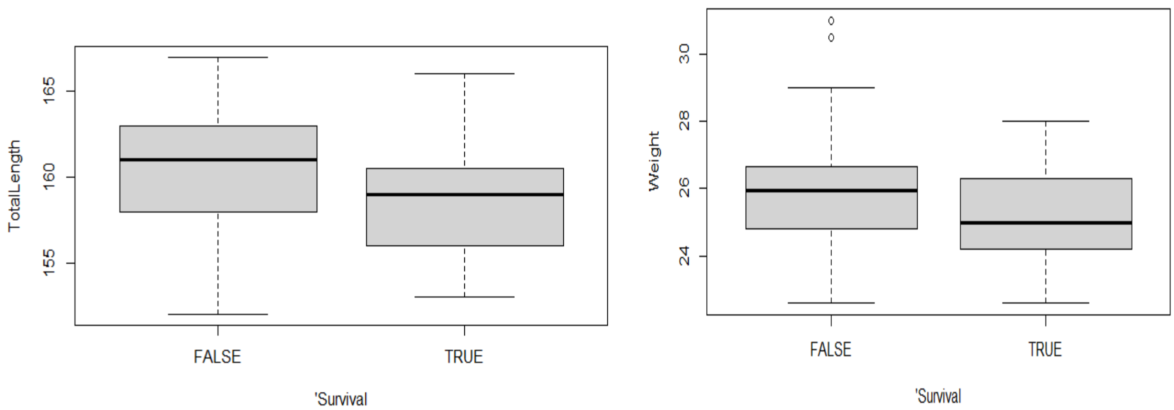
Variables	*Description*	*Summary*	*Confidence Intervals*		
Survival	*T = Survived, F = Died*	*T = 72, F = 64*	*p-value*	*2.5 %*	*97.5 %*
Sex	*Female, Male*	*F =49, M = 87*	*0.0006(F)*	*0.7486*	*2.7063*
Total Length	*measured from tip of the beak to the tip of the tail (mm)*	*Min = 152.0 Median = 160.0 Max = 167.0*	*0.000156*	*0.1685*	*0.5146*
Weight	*weight of the bird (g)*	*Min = 22.60 Median = 25.55 Max = 31.00*	*0.002836*	*0.2408*	*1.0760*
Humerus	*length of humerus (inches)-the forelimb's bone that connects with the pectoral girdle*	*Min = 0.6590 Median = 0.7330 Max = 0.7880*	*4.8e-06*	*-93.4798*	*-38.4054*
DOF = 131, Residual Deviance = 139.02			*AIC = 149.02*		

Table – Summary of Data

The first draft of the model included all the nine variables (except ID) against survival. Only Total Length and Weight was highly significant with AIC = 152.21. To simplify the model I used Drop-In-Deviance Test. To simplify I removed each Variable one by one and checked the p-value and deviance. I assumed null hypotheses H0 = The simpler model. Then after comparing two models I found out that higher the p -value of the model less significant the model with additional parameters. The p-value of the last model was 0.5238 so it was concluded that we do not have the strong evidence to reject null hypothesis and it suggests that there is a relatively high probability of observing the data if the null hypothesis were true. Also simpler model was sufficient to explain observed variation in data. The final equation of the model is –

$$(\text{logit})\pi = \frac{\exp(-23.037 + 1.682Sex(f) + 0.332(TotalLength) + 0.632(Weight) - 63.830(Humerus))}{1+ \exp(-23.037+1.682Sex(f) + 0.332(TotalLength) + 0.632(Weight) - 63.830(Humerus))} \quad (1)$$

Coloured Scatter Plot of Length and Weight of the Bird with Survival /Sex parameter. Fig[1]



Box Plots showing corelation of Survival of Birds with their Length, Weight Bone. Fig[2]

Discussion

The model has taken Survival = 'T' and Sex = 'M' as a reference level hence it estimates the change in log-odds of 'Survival' for a bird who is a male as compared to a female bird. This is for categorical variable. For continuous it means how survival of bird depends upon its length and weight. From the model equation (1), all the variables are significant predictors of Survival.

The effects of variables on survival of birds is as follows –

1. For the binary variable Sex, we see that exp(1.682) = 5.372 means that males are 5.375 times likely to survive than females (Scatter Plot). Hence, males survival is more.
2. For continuous variable Total Length, we observed that for one unit increase in length (keeping other variables constant) we could expect $\frac{\exp(\,0.332)}{1+\exp(0.332)} = 0.582$. Means as the length increases chances of survival gets reduced by almost 58%.
3. For one unit increase in weight $\frac{\exp(0.632)}{1+\exp(0.632)} = 0.652.$ Chances reduce by 65.2 %.

4. For one unit increase in humerus $\frac{\exp(-63.830)}{1+\exp(-63.830)} = 4.379e - 28.$ This is an extremely small value, it means there is not much dependency of humerus on Survival of the birds.

As compared with other summaries in table 1 [3] , my model resembles most closest with Calhoun – 1947, Harris -1911 and Bumpus only in terms of Sex, Weight and Length. There is no dependency of Humerus in Calhoun-1947. In layman terms our model is simple and easy to Interpret. Variations maybe due to inadequacy of variables or more effective statistical methods as discussed in Introduction section.

References

[1] *Academic.oup.com*. Available at: https://academic.oup.com/evolut/article-abstract/26/1/20/6867368 (Accessed: 08 October 2023).

Johnston, R. F., Niles, D. M., & Rohwer, S. A. (1972). Hermon Bumpus and Natural Selection in the House Sparrow Passer domesticus. *Evolution*, *26*(1), 20–31. https://doi.org/10.2307/2406980

[2] Manly, B. F., & Manly, B. F. J. (1994). Multivariate statistical methods.(Accessed: 08 October 2023)

[3] *Biology of the ubiquitous house sparrow: From genes to populations*. Available at: https://api.pageplace.de/preview/DT0400.9780198041351_A23606110/preview-9780198041351_A23606110.pdf (Accessed: 08 October 2023).