Bird Extinction

Alexander Thoren, Lars Larsen, Malthe Nielsen & Sander Thilo

March 8, 2023

Abstract

Introduction

Over the last several decades, data has been gathered from numerous islands in Britain documenting the extinction of bird species. This data measures Species, average inverse extinction time in years, average number of nesting pairs, species size and species migratory status. The trend of animal species going extinct is a worrying one, and in this report we attempt to analyze the data collected to find any correlation between the size or migratory status of a bird species with its extinction time.

Data Analysis

Preparing data

Before we could analyze the data, it was important to recognize that a great number of nesting pairs would of course positively influence extinction time, as species with more members can survive for longer. To conduct our analysis, we therefore performed a multivariate regression model to control for nesting pairs. To conduct our analysis, we used the popular statistical computing programming language R. The first step was to transform the Size and Status columns into dummy variables, as Size and Status are dichotomous predictors. This is done by replacing S(mall) with 0 and L(arge) with 1 in the Size column, and replacing R(esident) with 0 and M(igrant) with 1 in the Status column.

Species	Time	Pairs	Size	Status
Sparrowhawk	3.03	1.00	1	0
Buzzard	5.46	2.00	1	0
Quail	1.49	1.00	1	1
Wheat-ear	2.61	4.83	0	1
Great_tit	6.06	2.50	0	0

Fitting a full model

This allows any slope coefficients used on these variables to represent, for example, the difference in extinction time for Small birds versus Large birds. We then model a full model:

$$\text{ExtinctionTime} = \beta_0 + \beta_1 \text{Pairs} + \beta_2 \text{Size} + \beta_3 \text{Status}+$$
$$\beta_4 (\text{Pairs} \times \text{Size}) + \beta_5 (\text{Pairs} \times \text{Status}) + \beta_6 (\text{Size} \times \text{Status})+ \tag{1}$$
$$\beta_7 (\text{Pairs} \times \text{Size} \times \text{Status})$$

Residuals:

Min	1Q	Median	3Q	Max
-14.769	-3.644	-0.652	1.128	49.831

Coefficients:

	Estimate Std.	Error	t value	p-value
(Intercept)	-4.3990	3.3227	-1.324	0.191096
Pairs	2.8703	0.7096	4.045	0.000168
Size	9.2254	5.3216	1.734	0.088697
Status	5.3524	5.9619	0.898	0.373298
Pairs×Size	-1.0989	1.2446	-0.883	0.381205
Pairs×Status	-2.5363	1.8422	-1.377	0.174252
Size×Status	-9.4013	9.8736	-0.952	0.345255
Pairs×Size×Status	2.0047	2.7218	0.737	0.464608

Table A.1: Residuals and estimated coefficients of full model

The full model includes all possible interactions between pairs, size, and migratory status. The intercept represents the expected extinction time for a bird with zero pairs, small size, and non-migratory status. The coefficient for Pairs indicates the expected change in extinction time for each additional pair, holding all other variables constant. The coefficient for Size indicates the expected difference in extinction time between birds of small and large sizes, holding all other variables constant. The coefficient for Status indicates the expected difference in extinction time between non-migratory and migratory birds, holding all other variables constant.

The interactions between Pairs and Size, Pairs and Status, Size and Status, and all three variables together are also included in the model. These interactions allow the effect of Pairs on extinction time to vary depending on the values of Size and Status, and vice versa.

Model analysis

The overall fit of the model is significant, as indicated by the F-statistic (3.97) and associated p-value (0.001446). However, not all coefficients are significant at the 0.05 level. Specifically, Size and Status are not significant, and neither are the interactions between Pairs and Size, and Pairs and Status. (Appendix A.1).

The adjusted R-squared value is 0.2542, which indicates that the model explains about a quarter of the variation in extinction times. Overall, the full model suggests that the effect of Pairs on extinction time varies depending on the values of Size and Status, but additional analysis is needed to determine the best model for the data. (Appendix A.1).
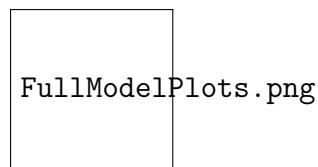

Figure A.2: Model Plots

The Q-Q plot suggests that the standardised residuals are approximately normally distributed, although there are some larger positive standardised residuals than expected. This positive skew indicates that the full model does not do a great job at predicting very high values of inverse extinction time.

When observing the residuals, there is an indication of mild heteroscedasticity: the spread of values is largest for predicted values between 10 to 20, and narrows towards the edge of the plot. This suggests tendency for the residual variance to get larger for birds with high inverse extinction time, as observed in the Q-Q plot. Furthermore, the linearity of this model is questionable.

The species rock dove, raven and skylark show the largest three positive residuals as also illustrated in the graphs in appendix A.2 (Appendix A.3). In addition, the starling species also exhibits substantial influence on our data set as illustrated by the large Cook's

distance in the Residual vs Leverage plot.

When removing the largest three influential data points (rock dove, raven and starling) and comparing it to the previous model, we see that the coefficients and p-values have largely remained unchanged. Specifically, the coefficient for the Pairs predictor has decreased (by 0.9811 points), and the p-value has increased (from 0.000168 to 0.00029), indicating that removing the three species slightly increases the probability that the relationship between Time and Pairs is random. Adjusted R-squared value is largely unchanged (0.2583). (Appendix B.1).

Overall, upon deleting these outliers from the data set and re-estimating the model, the coefficient estimates are almost the same as before. So there is no reason to think that the conclusions are strongly influenced by including these bird species, and therefore there is no reason to exclude them.

Data transformations

When trying three different transformations of the time variable ($\log_2(\text{Time})$, $\sqrt{\text{Time}}$ and 1/Time), there is a systematic curvature in the residual plot of the inverse time transformed model (Appendix C.3). Models with square root and log transformations are more acceptable for linearity. (Appendix C.1 & C.2).

In order to access whether there are linear relationships between $\log_2(\text{Time})$ and Pairs in the four combinations of Size and Status, we performed an adjusted R^2 analysis on the transformed model. From this, it appears that transforming the "Time" variable to logarithmic scale does improve the model fit. The adjusted R-squared value of the log-transformed model (0.5739) is significantly higher than that of the full model (0.2542), indicating that the log-transformed model explains a larger proportion of the variability in the data. Additionally, the p-value for the Pairs coefficient in the log-transformed model is even more significant (1.74e-07) than in the full model (0.000168), indicating a stronger linear relationship between $\log_2(\text{Time})$ and Pairs. However, there is no evidence that the relationship between $\log_2(\text{Time})$ and Pairs varies across the four combinations of "Size" and "Status", so there is no need to include all interaction terms in the log-transformed model. (Appendix C.4)

To assess whether the slopes for all four combinations of Size and Migratory Status are equal, we fitted a linear model with an interaction term between Pairs and the combination of Size and Migratory Status. The interaction term allows us to test whether the effect of Pairs on Time differs across the four combinations. Based on the ANOVA output for the regression of $\log_2$(Time) on pairs, size, and migratory status, we can assess whether the slopes for all four combinations of size and migratory status are equal.

Analysis of Variance Table
Response: $\log_2$(**Time**)

	Df	Sum Sq	Mean Sq	F-value	p-value
Pairs	1	55.499	55.499	62.7593	1.318e-10
Size	1	14.301	14.301	16.1719	0.000181
Status	1	6.749	6.749	7.6315	0.007825
Pairs:Size	1	1.469	1.469	1.6608	0.202995
Pairs:Status	1	0.003	0.003	0.0034	0.953872
Size:Status	1	0.378	0.378	0.4270	0.516245
Pairs:Size:Status	1	1.770	1.770	0.0015	0.162886
Residuals	54	47.753	0.884		

Table D: Interaction Term Model, ANOVA

Looking at the interaction terms, we can see that none of them are statistically significant at the 0.05 level. This suggests that there is no evidence of a significant interaction effect between Pairs and Size, Pairs and Status, Size and Status, or Pairs, Size, and Status combined. Therefore, the slopes for all four combinations of size and migratory status are equal, as there is no evidence to suggest otherwise. (Appendix D).

Nested model comparisons

We then made nested models to asses whether variables and interaction of variables made contributions. To compare the models, we used the adjusted R^2 statistics to see whether the addition of variables or interaction terms significantly improves the model fit.

Nested models:

Model 1: $\log_2$(Time) $= \beta_0 + \beta_1$Pairs $+ \beta_2$Size $+ \beta_3$Status

Model 2: $\log_2$(Time) $= \beta_0 + \beta_1$Pairs $+ \beta_2$Size $+ \beta_3$Status $+ \beta_4$Pairs $\times$ Size

Model 3: $\log_2$(Time) $= \beta_0 + \beta_1$Pairs $+ \beta_2$Size $+ \beta_3$Status $+ \beta_4$Pairs $\times$ Status

Model 4: $\log_2$(Time) $= \beta_0 + \beta_1$Pairs $+ \beta_2$Size $+ \beta_3$Status $+ \beta_4$Size $\times$ Status

Model 5: Full model (From section **Fitting a full model**)

Based on the adjusted R^2 statistics, Model 2 has the highest value (Appendix E.1). This suggests that the addition of the interaction term between Pairs and Size in Model 2 is an improvement and is the best overall model fit.

To further compare the models, we can use the ANOVA test to see whether the addition of variables or interaction terms significantly improves the model fit. Comparing Model 1 to Model 2, we see that the addition of the interaction term between Pairs and Size results in a significant decrease in RSS (1.4686), with an associated F-statistic of 1.6608 and a p-value of 0.2030. This suggests that the addition of the interaction term improves the model fit, but the improvement is not statistically significant at the 0.05 level. (Appendix E.2)

Comparing Model 1 to Model 5, we see that the addition of all possible interaction terms results in a smaller decrease in RSS (0.9339) compared to the addition of only the Pairs:Size interaction in Model 2, with an associated F-statistic of 1.0994 and a p-value of 0.3632. This suggests that the improvement in model fit from adding all possible interaction terms is not statistically significant at the 0.05 level, and Model 2 may be the preferred model since it has fewer parameters and is easier to interpret. (Appendix E.2, E.3, & E.4)

Overall, the ANOVA results suggest that Model 2, which includes the Pairs×Size interaction term, is the best model among the five nested models, as it provides an improvement in model fit compared to Model 1 (not statistically significant), while also being simpler than Model 5.

Thus, Model 2 also acts as our preferred reduced model, which is a linear regression model with $\log_2(\text{Time})$ as the response variable and Pairs, Size, Status, and Pairs×Size as predictor variables.

Results

The coefficients table shows the estimated regression coefficients for the model. The intercept is the estimated value of the response variable when all predictor variables are equal to zero. Pairs, Size, and Status are the estimated effects of the respective

predictor variables on the response variable, holding all other predictor variables constant. Pairs:Size is the estimated interaction effect between Pairs and Size. (Appendix F).

Residuals				
Min	1Q	Median	3Q	Max
-2.3073	-0.5359	-0.0962	0.3315	3.5352

Coefficients				
	Estimate	Std. Error	t-value	p-value
(Intercept)	0.22890	0.29750	0.769	0.4448
Pairs	0.43225	0.06529	6.620	1.36e-08
Size	1.43373	0.44919	3.192	0.0023
Status	-0.69336	0.26323	-2.634	0.0108
Pairs:Size	-0.14271	0.11019	-1.295	0.2005

Multiple R^2: 0.6099, Adjusted R^2: 0.5825
F-statistic: 22.28 on 4 and 57 DF, p-value: 4.105e-11

Table F: Reduced Model Summary

The t-values and p-values indicate the significance of each predictor variable. Pairs, Size, and Status are statistically significant predictors with p-values less than 0.05, meaning there is strong evidence that their effects on the response variable are different from zero. Pairs:Size, however, is not a statistically significant predictor with a p-value of 0.2005. (Appendix F).

The estimated coefficient for Pairs is 0.43225, which means that for every one-unit increase in Pairs, the log of extinction time increases by 0.43225 units, holding other variables constant. This coefficient is statistically significant at the 0.001 level. (Appendix F).

The estimated coefficient for Size is 1.43373, which means that on average, large bird species have a log of extinction time that is 1.43373 units greater than small bird species, holding other variables constant. This coefficient is statistically significant at the 0.05 level.

The estimated coefficient for Status is -0.69336, which means that on average, migratory bird species have a log of extinction time that is 0.69336 units less than residential bird species, holding other variables constant. This coefficient is statistically significant at the 0.05 level.

The estimated coefficient for the interaction between Pairs and Size is -0.14271, which means that the effect of Pairs on the log of extinction time depends on the size of the bird

species. However, this interaction term is not statistically significant at the 0.05 level, meaning that we cannot reject the null hypothesis that the coefficient is equal to zero.

The adjusted R-squared value of 0.5825 indicates that the model explains 58.25% of the variance in the log of extinction time. The F-statistic of 22.28 with a p-value of 4.105e-11 indicates that at least one of the independent variables is significantly associated with the log of extinction time.

Conclusion

In conclusion, the number of pairs is predictive of extinction time. Both large Size and Migratory Status have positive effect on extinction time. However, compared to Status, Size has a greater effect. We set up equations to help interpret the coefficients:

$$\log_2\left(\text{ExtTime}\right) = 0.2289 + 0.43225\text{Pairs} + 1.43373\text{Size} - 0.69336\text{Status} - 0.14271\text{Pairs} \times \text{Size} \Leftrightarrow$$

$$2^{\log_2(\text{ExtTime})} = 2^{0.2289 + 0.43225\text{Pairs} + 1.43373\text{Size} - 0.69336\text{Status} - 0.14271\text{Pairs} \times \text{Size}} \Leftrightarrow$$

$$\text{ExtTime} = 2^{0.2289} \cdot 2^{0.43225\text{Pairs}} \cdot 2^{1.43373\text{Size}} \cdot 2^{-0.69336\text{Status}} \cdot 2^{-0.14271\text{Pairs} \times \text{Size}} \Leftrightarrow$$

$$\text{ExtTime} = 1.17194 \cdot 1.34933^{\text{Pairs}} \cdot 2.70144^{\text{Size}} \cdot \frac{1}{1.61704^{\text{Status}}} \cdot \frac{1}{1.10397^{\text{Pairs} \times \text{Size}}}$$

For each new pair, the time to extinction increases by around 35% when other variables are accounted for. A variable with a massive impact is species size. Large bird species seem to 170% larger time to extinction than small bird species. The last two coefficients can be interpreted by removing the exponent:

$$\frac{1}{1.61704} = 0.618414 \text{ and } \frac{1}{1.10397} = 0.905822$$

https://www.overleaf.com/project/640853f72e8ddf5505a1177b This means that migrant bird species have a time to extinction around 38% smaller than native species, and that large migrating species together have a time to extinction around 9% smaller than small resident species. This also tells us that despite large species increasing the time to extinction by such a large amount, it is still not enough to offset the negative change that comes with migrant species.
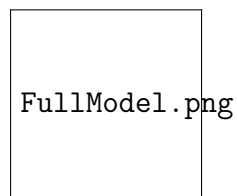
We can conclude that larger birds seem to survive for more than twice as long, and that migratory birds seem to die out quicker. The size of birds doesn't seem to be enough to offset the negative impact from being a migratory bird however.
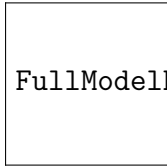
Appendix

Appendix A: Full Model

Appendix A.1: Model Summary

Coefficients:

	Estimate Std.	Error	t value	p-value
(Intercept)	-4.3990	3.3227	-1.324	0.191096
Pairs	2.8703	0.7096	4.045	0.000168
Size	9.2254	5.3216	1.734	0.088697
Status	5.3524	5.9619	0.898	0.373298
Pairs×Size	-1.0989	1.2446	-0.883	0.381205
Pairs×Status	-2.5363	1.8422	-1.377	0.174252
Size×Status	-9.4013	9.8736	-0.952	0.345255
Pairs×Size×Status	2.0047	2.7218	0.737	0.464608

FullModel.png

Appendix A.2: Model Plots

FullModelPlots.png

Appendix A.3: Largest three positive residuals
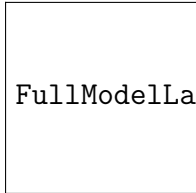
FullModelLargestResiduals.png

Appendix B: Model without Three Largest Outliers

Appendix B.1: Model Summary

CheatModel.png
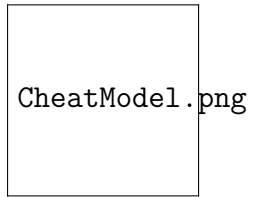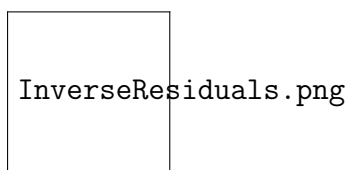
Appendix C: Models with Transformed Time Variables
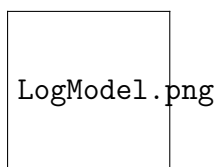
Appendix C.1: Residual Plot, Log$_2$ Time Transformation

LogResiduals.png
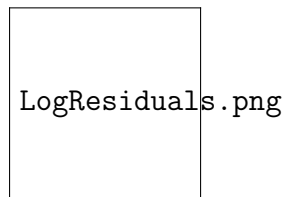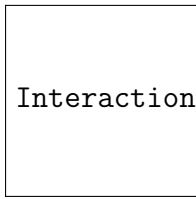
Appendix C.2: Residual Plot, Square Root Time Transformation

SqrtResiduals.png
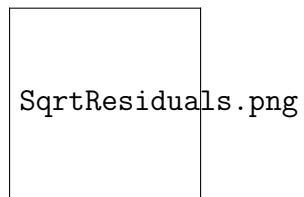
Appendix C.3: Residual Plot, Inverse Time Transformation

InverseResiduals.png

Appendix C.4: Model Summary, Log$_2$ Time Transformation

LogModel.png

Appendix D: Interaction Term Model, ANOVA

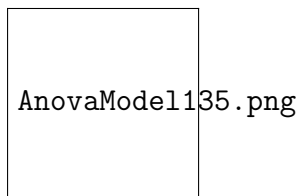InteractionTermModel.png

Appendix E: Nested Models

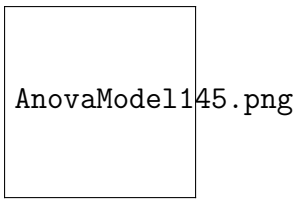Appendix E.1: Adjusted R^2 statistics for each model

NestedModelsR2.png
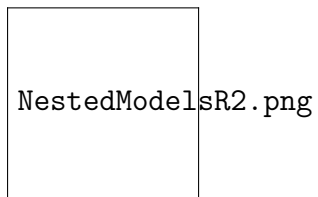
Appendix E.2: ANOVA for model 1, 2, 5

AnovaModel125.png
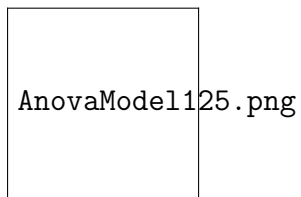
Appendix E.3: ANOVA for model 1, 3, 5

AnovaModel135.png

Appendix E.4: ANOVA for model 1, 4, 5

```
AnovaModel145.png
```

Appendix F: Reduced Model Summary

```
ReducedModel.png
```