

IT UNIVERSITY OF CPH

Factors Affecting Extinction

March 14, 2023

Rinalds Lipenitis Jakub Mráz Adam Rosenørn Costel Gutu Richard Kentoš
March 2023

1 Exercise 1

In exercise 1 we fitted

2 Exercise 2

3 Exercise 3

4 Exercise 4

Results can be significantly impacted by the outlier in each end of the scale. Outliers can also affect the measures of the central tendency, more specifically mean, median and mode. Moreover, they can also affect standard deviation and range - the spread of the data.

We believe that it is vital for us to keep the outliers in the dataset for various reasons:

1. Outliers provide important information about the data distribution. To be more specific, they can indicate the presence of extreme or unusual values, which might be beneficial for understanding the range and variability of our data. For instance, the outlier for the Peregrine species in this dataset indicates that it has a much lower average time until it becomes extinct than the other species. This could also be because of the specific environmental or ecological factors.

2. Preventing loss of information and biased analysis. Outliers may be representing rare but important occurrences that we should not leave out. Removing them might result in a distorted view of the true data distribution and lead to incorrect conclusions. For instance, if we decided to remove the outlier for the Raven species in this dataset, we would miss the fact that it has much longer average extinction time than the other species.

3. Allowing more robust statistical analysis. Methods such as robust regression or non-parametric tests are able to handle outliers and provide more accurate results.

We therefore think that it is inevitable to keep outliers in our dataset and carefully consider how they impact the results. Rather than removing them, it may be better to investigate why they exist in the first place and how they may affect the analysis.

5 Exercise 5

Firstly we need to understand the motivation behind transforming a variable. Transformation of a variable is often done to linearize a relationship between two variables which is not linear in its original form. However, in this question we are specifically asked to assess whether there are linear relationships between

$\log(\text{'time'})$ and 'pairs' in all possible combinations of 'size' and $\text{'migratory status'}$.

6 Exercise 6

7 Exercise 7

8 Exercise 8

$$\hat{y} = \alpha + \beta X_{HDI} \tag{1}$$

$$\hat{y} = \alpha + \beta X_{HDI}$$

The three types of males:

1. \exists
2. \forall
3. Σ

$$\sum_{i=1}^{\infty} 2^i$$

$$Heightofchildren = \alpha + \beta * heightofparent + \gamma * diet + \delta pollution + \epsilon$$