

Project II: Cuckoo Eggs Dimension

Nangamso Lindeni

Contents

Introduction	1
Importing Data	2
Exploratory Data Analysis	5
Data Visualisation	6
Inferential Statistical Analysis	11
Discussion and Conclusion	13

Introduction

The Common Cuckoo (*Cuculus canorus*) is a bird species known for its distinctive behaviours and characteristics, particularly its unusual reproductive strategy. As brood parasites, cuckoos lay their eggs in the nests of other, often smaller, bird species. The host bird then incubates the cuckoo egg and unknowingly raises the chick as its own (Davies, 2011; Lahti, 2005). According to Kruger and Davies (2003), Davies (2011), and Lahti (2005), the survival of the cuckoo species hinges on their remarkable adaptation strategies, which include evading host defences through deception and matching their egg and chick development with the host's incubation and provisioning patterns.

Research conducted in Russia has provided evidence of evolutionary adaptation in cuckoos, particularly in their ability to vary egg size to match that of their hosts. This adaptation reduces the likelihood of egg rejection, as host birds are more inclined to reject eggs that are significantly different or larger than their own clutch (Meshcheryagina et al., 2018). The study found that while the breadth of cuckoo eggs closely matched that of the host species, the length did not, suggesting a trade-off that limits the cuckoo's ability to maintain larger egg volumes. This alignment in egg breadth is likely an evolved response to avoid host rejection (Meshcheryagina et al., 2018).

This project aims to investigate the egg dimensions—length and breadth—of six different cuckoo bird species, analysing 120 eggs. Various statistical analyses will be performed to determine differences in these dimensions both between and within the species.



Figure 1: Common Cuckoo Bird

Importing Data

For data manipulation, visualisation and analysis and to effectively read and access the relevant cuckoo data, the “tidyverse” was installed and activated. The `read.csv()` function was used to import and read the data that is in CSV format. The data was assigned a new name termed `cuckoo_data`.

```
library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.3.3
```

```
Warning: package 'ggplot2' was built under R version 4.3.3
```

```
Warning: package 'readr' was built under R version 4.3.3
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.2      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.0
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.2      v tidyr      1.3.0
v purrr      1.0.1
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()      masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
cuckoo_data <- read.csv("https://raw.githubusercontent.com/xrander/bootcamp-test/master/data,
```

To view and understand the contents of the data `names()` function was used to return the names of the columns in the `cuckoo_data`, this also ensures that the data has been imported correctly. The names of the `cuckoo_data` are “length” “breadth” “species”

```
names(cuckoo_data)
```

```
[1] "length" "breadth" "species"
```

The names of the variables in the data is given above. The `head()` function was implemented to preview the first 10 rows of the data, while the `tail()` function was run to preview the last 10 rows of the dataset, thus making it easier to observe patterns in the data.

A preview of the data is presented in Table 1.

To further understand and observe the `cuckoo_data` the `car::some()` function was run, first setting a seed, to randomly select and generate a sample of rows from the `cuckoo_data` dataset. The `set.seed()` allows the same results every time the code is run.

```
set.seed(3891)
car::some(cuckoo_data)
```

	length	breadth	species
21	22.0	17.0	Meadow Pipit
31	22.5	17.1	Meadow Pipit
34	22.0	17.2	Meadow Pipit
58	23.2	16.4	Tree Pipit
80	23.0	16.1	Robin
90	23.0	17.0	Robin
99	24.9	16.8	Pied Wagtail
109	20.9	15.7	Wren
112	22.3	16.0	Wren
117	20.0	15.7	Wren

Table 1: Data Preview

```
head(cuckoo_data, n = 10)
```

	length	breadth	species
1	21.7	16.1	Meadow Pipit
2	22.6	17.0	Meadow Pipit
3	20.9	16.2	Meadow Pipit
4	21.6	16.2	Meadow Pipit
5	22.2	16.9	Meadow Pipit
6	22.5	16.9	Meadow Pipit
7	22.2	17.3	Meadow Pipit
8	24.3	16.8	Meadow Pipit
9	22.3	16.8	Meadow Pipit
10	22.6	17.0	Meadow Pipit

```
tail(cuckoo_data, n = 10)
```

	length	breadth	species
111	21.0	15.5	Wren
112	22.3	16.0	Wren
113	21.0	15.9	Wren
114	20.3	15.5	Wren
115	20.9	15.9	Wren
116	22.0	16.0	Wren
117	20.0	15.7	Wren
118	20.8	15.9	Wren
119	21.2	16.0	Wren
120	21.0	16.0	Wren

Exploratory Data Analysis

Various analyses were conducted to investigate the main characteristics and understand the structure of the dataset. The `summary()` function gave statistics of the three variables in the data frame; length, breadth, and species below:

A quick descriptive statistics of the data is given below:

```
summary(cuckoo_data)
```

length		breadth		species
Min.	:19.60	Min.	:15.00	Length:120
1st Qu.	:21.90	1st Qu.	:16.20	Class :character
Median	:22.35	Median	:16.60	Mode :character
Mean	:22.45	Mean	:16.55	
3rd Qu.	:23.23	3rd Qu.	:17.00	
Max.	:25.00	Max.	:17.50	

The `str()` function was run to show the data structure and the type and length of variables in the dataset. For `cuckoo_data` the `str()` function showed that the data has 120 observations and 3 variables, and the variable's length and breadth are numerical, while the species variable is a character, as seen below:

```
str(cuckoo_data)
```

```
'data.frame': 120 obs. of 3 variables:
 $ length : num  21.7 22.6 20.9 21.6 22.2 22.5 22.2 24.3 22.3 22.6 ...
 $ breadth: num  16.1 17 16.2 16.2 16.9 16.9 17.3 16.8 16.8 17 ...
 $ species: chr  "Meadow Pipit" "Meadow Pipit" "Meadow Pipit" "Meadow Pipit" ...
```

To identify distinct species present in the data frame, the `unique()` function was used and it generated six species names in the data: Meadow Pipit, Tree Pipit, Hedge Sparrow, Robin, Pied Wagtail and Wren.

```
unique(cuckoo_data$species)
```

```
[1] "Meadow Pipit" "Tree Pipit"    "Hedge Sparrow" "Robin"
[5] "Pied Wagtail" "Wren"
```

The `skim()` function was run to obtain an extended structure of the data and gave statistical values of the data such as mean, median, and standard deviation along with missing values and histograms of the data.

```
skimr::skim(cuckoo_data)
```

Table 2: Data summary

Name	cuckoo_data
Number of rows	120
Number of columns	3
Column type frequency:	
character	1
numeric	2
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
species	0	1	4	13	0	6	0

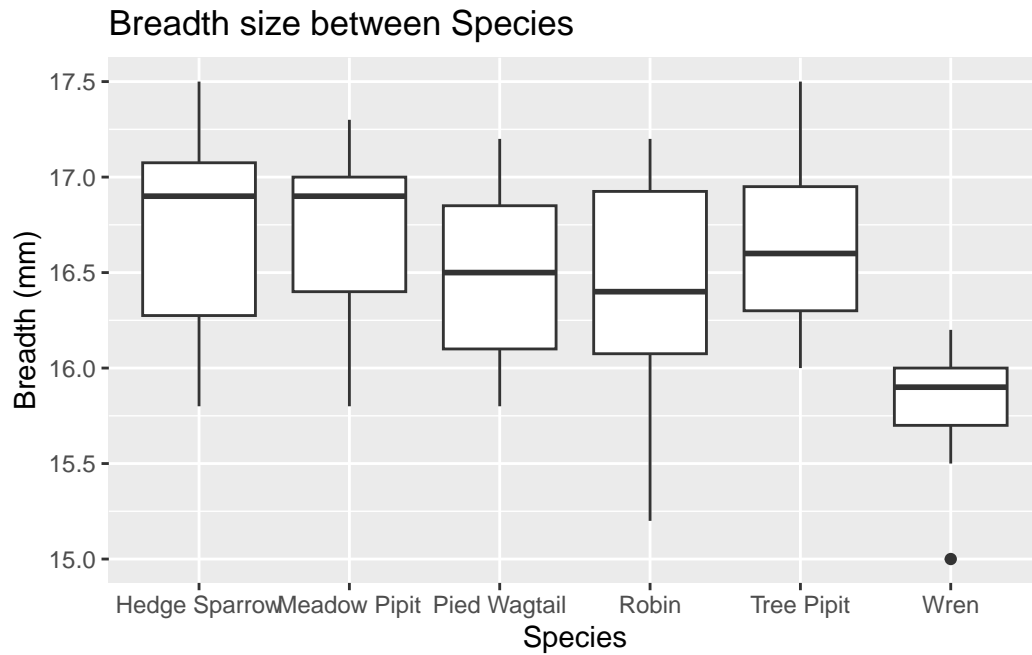
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
length	0	1	22.45	1.07	19.6	21.9	22.35	23.22	25.0	
breadth	0	1	16.55	0.52	15.0	16.2	16.60	17.00	17.5	

Data Visualisation

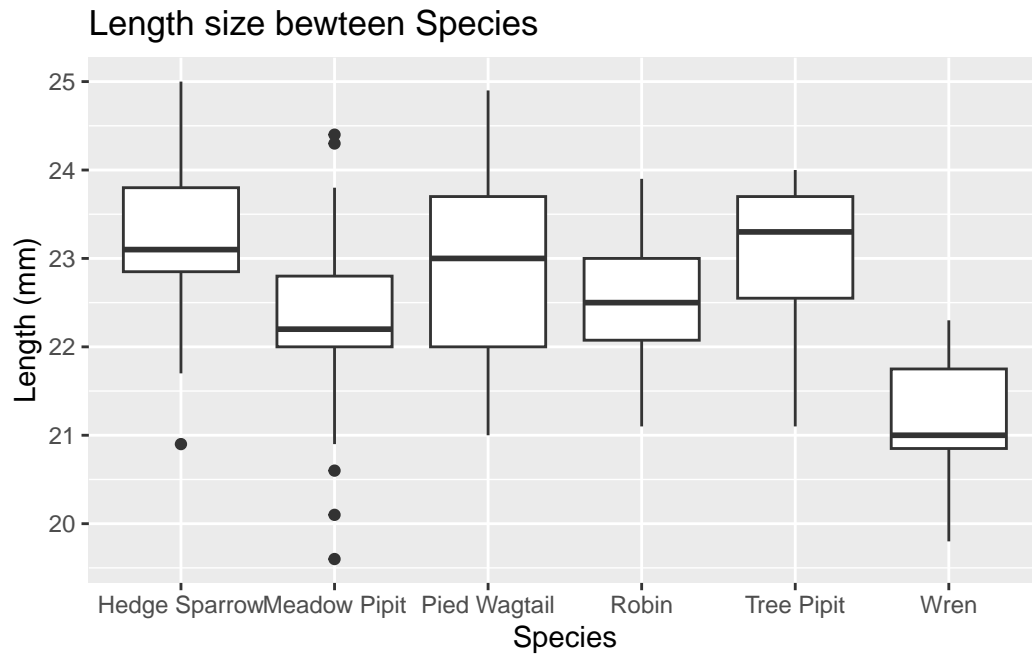
The box plot provides a visual summary of key statistics for the breadth of eggs within each species

```
library(ggplot2)
ggplot(cuckoo_data, aes(species, breadth)) +
  geom_boxplot() +
  labs(x = "Species",
       y = "Breadth (mm)",
       title = "Breadth size between Species")
```



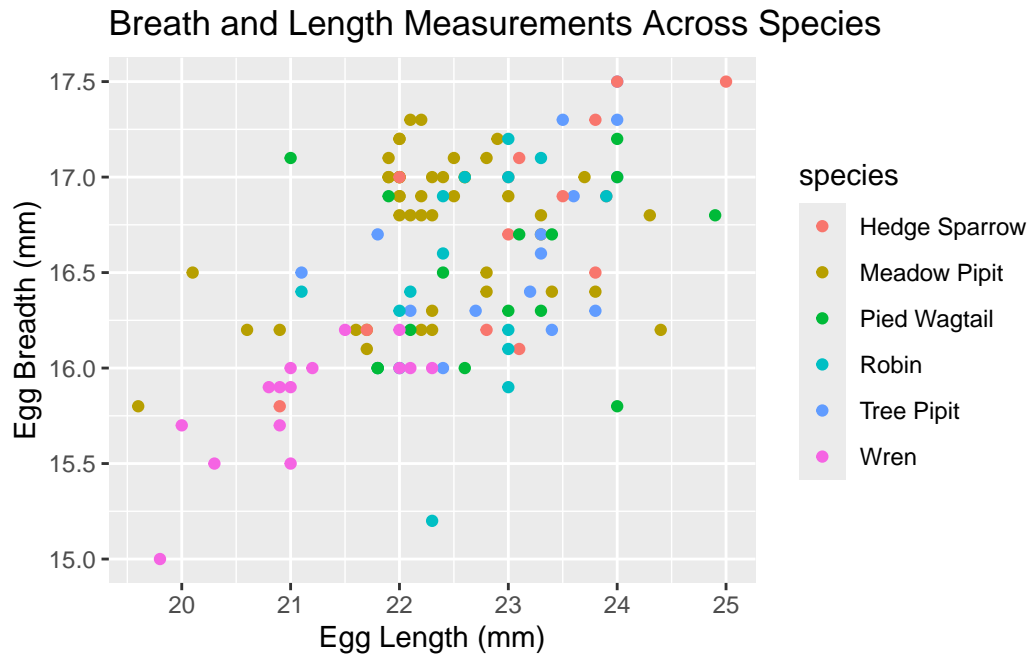
The box plot provides a visual summary of key statistics for the length of eggs within each species

```
ggplot(cuckoo_data, aes(species, length)) +
  geom_boxplot() +
  labs(x = "Species",
       y = "Length (mm)",
       title = "Length size bewteen Species")
```



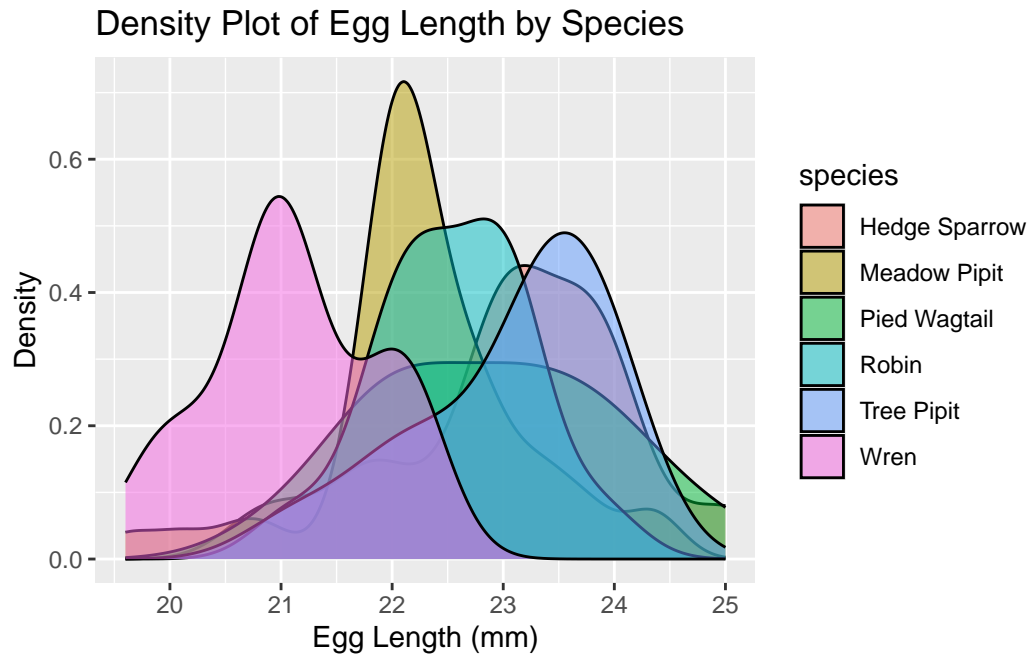
Scatter plot illustrating the relationship between breadth and length measurements within cuckoo bird species.

```
ggplot(cuckoo_data, aes(x = length, y = breadth, color = species)) +
  geom_point() +
  labs(x = "Egg Length (mm)",
       y = "Egg Breadth (mm)",
       title = "Breath and Length Measurements Across Species")
```

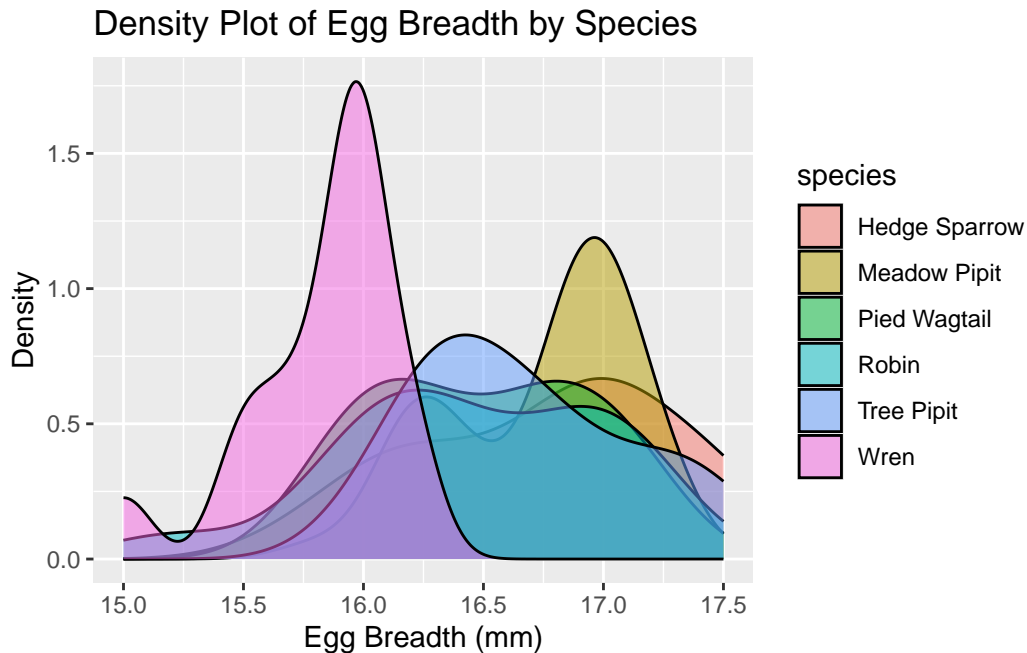
The distribution of length measurements by species is shown below using a density plot.

```
ggplot(cuckoo_data, aes(x = length, fill = species)) +
  geom_density(alpha = 0.5) +
  labs(x = "Egg Length (mm)", y = "Density",
       title = "Density Plot of Egg Length by Species")
```



The distribution of breadth measurements by species is shown below.

```
ggplot(cuckoo_data, aes(x = breadth, fill = species)) +  
  geom_density(alpha = 0.5) +  
  labs(x = "Egg Breadth (mm)", y = "Density",  
        title = "Density Plot of Egg Breadth by Species")
```



Inferential Statistical Analysis

The results ANOVA test were conducted to observe if there is a statically significant difference in the breadth of eggs between species is below:

```
breadth_anova <- aov(breadth ~ species, data = cuckoo_data) |>
  anova()
breadth_anova
```

Analysis of Variance Table

Response: breadth

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
species	5	10.373	2.07457	10.885	1.412e-08 ***
Residuals	114	21.727	0.19059		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The ANOVA results to see if there is a statically significant difference between the length of the eggs and species has a P-value of

```
length_anova <- aov(length ~ species, data = cuckoo_data) |>
  anova()

length_anova
```

Analysis of Variance Table

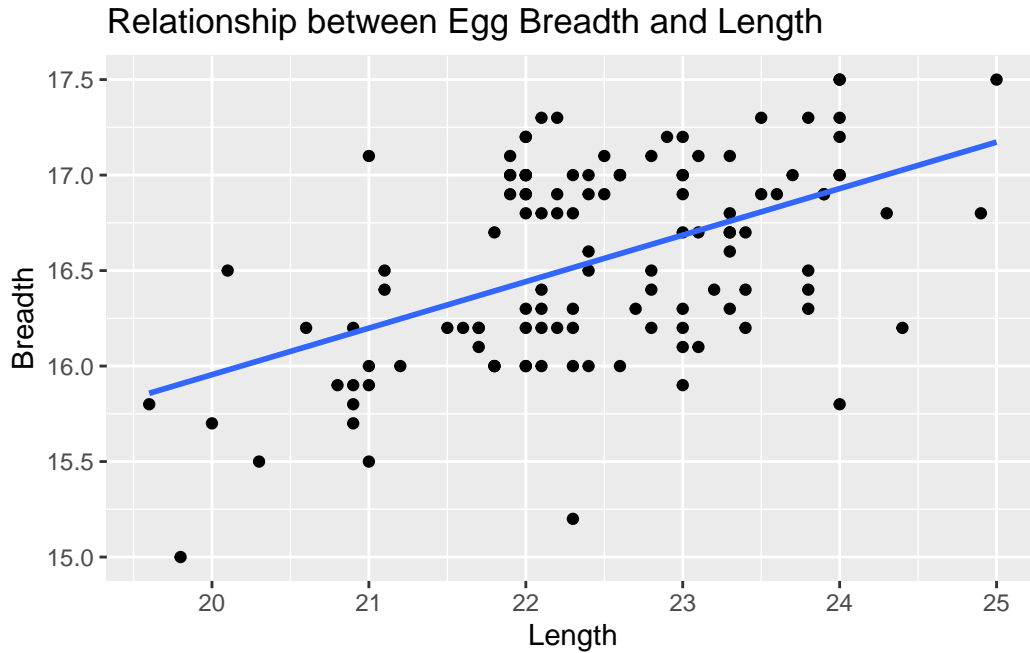
```
Response: length
      Df Sum Sq Mean Sq F value    Pr(>F)
species    5  42.81   8.5620   10.449 2.852e-08 ***
Residuals 114   93.41   0.8194
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Linear regression model accessing a relationship between lengths and breadths of eggs shows a positive relationship.

```
library(ggplot2)
lm_model <- lm(breadth ~ length + species, data = cuckoo_data) |>
  summary()

ggplot(cuckoo_data, aes(x = length, y = breadth)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(x = "Length", y = "Breadth") +
  ggtitle("Relationship between Egg Breadth and Length") +
  theme_gray()
```

```
`geom_smooth()` using formula = 'y ~ x'
```



Discussion and Conclusion

The ANOVA test assessing the influence of species on the breadth of cuckoo eggs reveals a statistically significant difference in egg breadth among the different cuckoo species. This is proved by the small p-value ($1.412\text{e-}08$), the highly significant code (***), and a high F value (10.885), which indicate that the variation between species is greater than the variation within species. Therefore, species have a significant effect on the breadth of eggs.

Similarly, the ANOVA test evaluating the influence of species on the length of cuckoo eggs also shows a statistically significant difference in egg length across species. The p-value of $2.852\text{e-}08$ and the associated significance code (***) provide strong evidence to reject the null hypothesis, concluding that species significantly affect egg length.

Additionally, the linear regression model demonstrates a positive relationship between the breadth and length of the eggs, suggesting that an increase in breadth is associated with an increase in length, and vice versa.