Manuscript- Data Analysis Project

Quinn Moon

3/31/23

# 1. Summary/Abstract

We used a built in data set that includes several measurements of three species of penguins to analyze the relationships between sex, species, mass, and flipper length. The project involved an initial data cleaning step follow by analyses. Our results indicate a significant relationship between both sex and species with mass. For research highlighted the statistically significant positive correlation between mass and flipper length.

# 2. Introduction

## 2.1 General Background Information

Penguins is a general term used to desribe a collection of cold climate dwelling bird species.

## 2.2 Description of data and data source

Here we describe an analysis of the Palmer Penguins dataset [@Horst:2020], which contains data on three species of penguins from the Palmer Archipelago, Antarctica. Data were collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER, a member of the Long Term Ecological Research Network [@Gorman:2014].

## 2.3 Questions/Hypotheses to be addressed

Our research questions are:

1. Do sexes differ in size?
2. Does species have an affect on size?
3. Does mass have an affect on flipper length?

# 3. Methods

We used the Palmer Penguins dataset to do an initial data cleaning step. From the cleaned data, we created numerous plots and tested for statistical significance using ANOVA tests.

## 3.1 Data import and cleaning

For this project, species names were corrected and variables were plotted to identify typos in the data. Incorrect data was either removed or replaced with the correct value. More details on the data cleaning step, including .R and .qmd files, can be found online in a github repo within the Code/Processing\_code folder https://github.com/quinnmoon11/QUINN-Rclass-project2

## 3.2 Statistical analysis

We used a combination of scatter and density plots followed by ANOVA tests to test for statistical significance.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

Our first result serves to be an introductory summary of the data.

[Table 1](#tbl-summary_table) shows a summary of the data.

Table 1: Data summary table.

| Variable | N | Mean | SE | Counts |
| --- | --- | --- | --- | --- |
| Individual ID | 339 |  |  |  |
| Clutch Completion | 339 |  |  |  |
| Date Egg | 339 |  |  |  |
| Species | 339 |  |  | Ade: 151, Gen: 120, Chi: 68 |
| Island | 339 |  |  | Bis: 164, Dre: 124, Tor: 51 |
| Sex | 330 |  |  | MAL: 168, FEM: 162 |
| Culmen Length (mm) | 339 | 43.90 | 0.30 |  |
| Culmen Depth (mm) | 339 | 17.17 | 0.11 |  |
| Flipper Length (mm) | 339 | 200.79 | 0.76 |  |
| Body Mass (g) | 339 | 4196.17 | 43.62 |  |
| Delta 15 N (o/oo) | 327 | 8.74 | 0.03 |  |
| Delta 13 C (o/oo) | 328 | -25.68 | 0.04 |  |

## 4.2 Size

We found that male penguins are substantially larger than female penguins ([Figure 1](#fig-mass_sex_bars), ANOVA P-value < 0.001 ).

|  |
| --- |
| Figure 1: Body mass distributions of the two sexes of penguin. |

Table 2: Size differences between sexes.

|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| --- | --- | --- | --- | --- | --- |
| Sex | 1 | 40572797 | 40572796.7 | 76.72 | 0 |
| Residuals | 328 | 173452840 | 528819.6 |  |  |

Differences in mass across sexes is also related to the species ([Figure 2](#fig-mass_species_sex_density)).

|  |
| --- |
| Figure 2: Body mass distributions of both sexes for the three species of penguin. |

Table 3: Size differences between sexes by species.

|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| --- | --- | --- | --- | --- | --- |
| Species | 2 | 144209469 | 72104734.5 | 714.23 | 0 |
| Sex | 1 | 36905104 | 36905104.1 | 365.56 | 0 |
| Residuals | 326 | 32911063 | 100954.2 |  |  |

## 4.3 Mass and Flipper Length

Mass and flipper length appear to have a significantly correlated linear relationship.

|  |
| --- |
| Figure 3: Body mass distribution compared to flipper length. |

Table 4: Size differences by flipper length.

|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| --- | --- | --- | --- | --- | --- |
| Flipper\_Length | 1 | 165443187 | 165443186.7 | 1059.76 | 0 |
| Residuals | 337 | 52610578 | 156114.5 |  |  |

# 5. Discussion

## 5.1 Summary and Interpretation

By cleaning an availabe penguin data set, we were able to explore the relationships between sex and mass, species and mass, and mass and flipper length. We also ran ANOVA tests to asses the statistical significance of our findings.

## 5.2 Strengths and Limitations

A strength of our study is that we utilized public data and had a low budget. A short coming of the study is that we only studied 3 species of penguins. Similar questions could be asked about all species of penguins to test how the relationships between variables changes in each species.

## 5.3 Conclusions

The most significant take away message is that writing R scripts and embedding them into .qmd files and then rendering them to html is actually fun to work on! I appreciaite how reproducible the analyses are when they follow the above format.

In regards to the data, our results demonstrate that sex and species both significantly influence the mass of penguins. And, penguin flipper length is highly correlated to mass.