Penguin Traits Analysis

Yue Zhu (A1F8C52)
Harry He (7133085)
Tina Zhou (4876165)
Xihua Yao (5927777)
Dongzhen Huangfu (3137544)

Introduction

This study leverages the Palmer Penguins dataset, a widely used ecological dataset that provides morphological measurements and environmental data for three penguin species: Adelie, Chinstrap, and Gentoo. The dataset, originally sourced from Antarctica, includes variables such as body mass, flipper length, culmen length, and culmen depth, along with dietary information represented by nitrogen (Delta 15 N) and carbon isotopes (Delta 13 C). These measurements offer insights into the penguins' physical characteristics and ecological niches.

By analyzing this dataset, the study aims to understand interspecies differences and the factors that influence penguin body mass, contributing to our understanding of biodiversity and adaptation to harsh Antarctic environments.

Research Questions

The analysis focuses on answering two research questions:

1. What is the correlation between penguin traits (body mass, flipper length, culmen length, and culmen depth)?
2. How do external factors (food resource and environment, represented by isotopes) and internal factors (gender) influence body mass, and which has a greater impact?

By addressing these questions, we aim to highlight biodiversity among penguin species, explore factors influencing their adaptation to harsh Antarctic environments, and draw conclusions about ecological interactions.

Hypotheses

To answer the research questions, the analysis tests the following hypotheses:

1. Hypothesis 1:
    - Null Hypothesis (H_0): There are no significant differences in body measurements (body mass, flipper length, culmen length, and culmen depth) among the three penguin species.
    - Alternative Hypothesis (Ha): At least one body measurement differs significantly among the three species.
2. Hypothesis 2:
    - Null Hypothesis (H_0): External and internal factors have no significant impact on body mass.
    - Alternative Hypothesis (Ha): External factors have a greater influence on body mass compared to internal factors.

Dataset Overview

The analysis utilized the penguins_raw dataset, which was cleaned to remove missing or irrelevant data. The cleaned dataset comprises 324 penguins across three species: Adelie, Chinstrap, and Gentoo. It includes key morphological traits such as body mass, flipper length, culmen length, and culmen depth, as well as environmental and dietary data represented by nitrogen (Delta 15 N) and carbon (Delta 13 C) isotopes. Gender information is also provided, with adequate representation of both male and female penguins, ensuring balanced statistical comparisons. The age of each

penguin is not provided, so we assume they are same. This comprehensive dataset enables a thorough examination of interspecies differences and the factors influencing body mass.

Model Selection

Approach to Selecting and Implementing Models

To address the research questions, the following methods were applied:

1. Exploratory Data Analysis (EDA)

   Our Exploratory Data Analysis (EDA) involved several steps to ensure the quality and comprehensiveness of the analysis. First, rows with missing values were removed to enhance data accuracy. Descriptive statistics, including mean, median, minimum, and maximum, were calculated for each species to summarize key traits. Additionally, visualizations such as heatmaps and scatter plots were generated to explore relationships between variables and validate the study's hypotheses.

```
[1] "Summary for Adelie Penguins (Selected Variables):"

 Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm) Body Mass (g)
 Min.   :32.10      Min.   :15.50     Min.   :172.0       Min.   :2850
 1st Qu.:36.70      1st Qu.:17.45     1st Qu.:186.0       1st Qu.:3350
 Median :38.80      Median :18.40     Median :190.0       Median :3700
 Mean   :38.79      Mean   :18.32     Mean   :190.3       Mean   :3703
 3rd Qu.:40.65      3rd Qu.:19.00     3rd Qu.:195.0       3rd Qu.:4000
 Max.   :46.00      Max.   :21.50     Max.   :210.0       Max.   :4775
 Delta 15 N (o/oo) Delta 13 C (o/oo)
 Min.   :7.698     Min.   :-26.79
 1st Qu.:8.567     1st Qu.:-26.24
 Median :8.881     Median :-25.99
 Mean   :8.859     Mean   :-25.81
 3rd Qu.:9.166     3rd Qu.:-25.31

Max.   :9.795     Max.   :-23.90
```

```
[1] "Summary for Gentoo Penguins (Selected Variables):"

 Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm) Body Mass (g)
 Min.   :40.90      Min.   :13.10     Min.   :203.0       Min.   :3950
 1st Qu.:45.33      1st Qu.:14.20     1st Qu.:212.0       1st Qu.:4700
 Median :47.45      Median :15.00     Median :216.0       Median :5050
 Mean   :47.57      Mean   :14.99     Mean   :217.2       Mean   :5091
 3rd Qu.:49.60      3rd Qu.:15.78     3rd Qu.:221.0       3rd Qu.:5500
 Max.   :59.60      Max.   :17.30     Max.   :231.0       Max.   :6300
 Delta 15 N (o/oo) Delta 13 C (o/oo)
 Min.   :7.632     Min.   :-27.02
 1st Qu.:8.106     1st Qu.:-26.69
 Median :8.260     Median :-26.22
 Mean   :8.249     Mean   :-26.18
 3rd Qu.:8.444     3rd Qu.:-25.56
 Max.   :8.834     Max.   :-25.00

[1] "Summary for Chinstrap Penguins (Selected Variables):"

 Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm) Body Mass (g)
 Min.   :40.90      Min.   :16.4      Min.   :178.0       Min.   :2700
 1st Qu.:46.30      1st Qu.:17.5      1st Qu.:191.0       1st Qu.:3475
 Median :49.50      Median :18.4      Median :196.0       Median :3700
 Mean   :48.79      Mean   :18.4      Mean   :195.7       Mean   :3730
 3rd Qu.:50.95      3rd Qu.:19.3      3rd Qu.:200.5       3rd Qu.:3950
 Max.   :58.00      Max.   :20.8      Max.   :212.0       Max.   :4800
 Delta 15 N (o/oo) Delta 13 C (o/oo)
 Min.   : 8.472    Min.   :-25.15
 1st Qu.: 9.104    1st Qu.:-24.69
 Median : 9.374    Median :-24.58
 Mean   : 9.356    Mean   :-24.56
 3rd Qu.: 9.620    3rd Qu.:-24.41
 Max.   :10.025    Max.   :-23.89
```

2. Correlation Analysis

To address the research questions, correlation analysis was conducted to explore the relationships between key traits such as body mass, flipper length, and culmen dimensions. This analysis focused on identifying significant correlations both within individual penguin species and across the three species, providing insights into how these traits are interconnected.

3. Regression Models

Regression models were applied to assess the effects of gender and food resources, represented by Delta 15 N and Delta 13 C isotopes, on body mass. These models operated under key assumptions, including the normal distribution of residuals and the independence of observations, ensuring the validity and

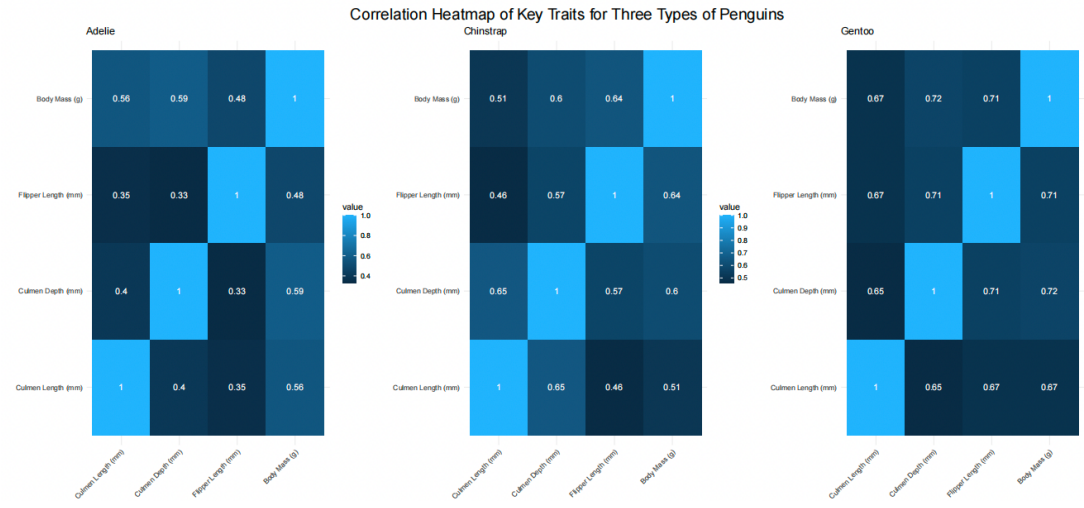
reliability of the results.

Results and Interpretation

The analysis revealed key findings regarding the relationships between penguin traits and the factors influencing body mass.

Since we have 324 sample, which is large enough to assume that the sample distribution is normal. Hence, we use Pearson coefficient to create correlation heatmaps. It demonstrated that body mass exhibits a strong positive relationship with flipper length across all species, with Gentoo penguins showing the highest correlation coefficient ($r > 0.7$). This suggests that flipper length is a critical factor in determining body mass, likely due to its role in swimming efficiency and environmental adaptation. Culmen dimensions (length and depth) showed moderate correlations with body mass, indicating species-specific influences. These correlations are visually summarized in the heatmaps:

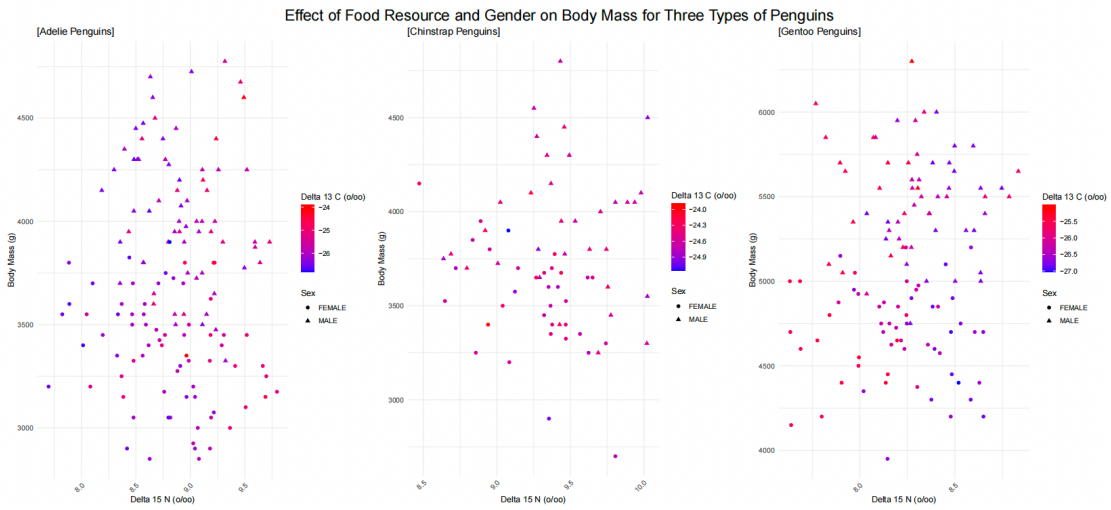
1.6 Data Visualization



Regression analysis provided deeper insights into the role of gender and food resources on body mass. Male penguins were consistently heavier than females across

all species, confirming the presence of sexual dimorphism. This pattern aligns with biological differences in energy requirements and behavior, such as reproduction or territorial defense.

The effects of external factors, represented by dietary isotopes (Delta 15 N and Delta 13 C), were also significant. Higher nitrogen levels (Delta 15 N), indicative of higher-trophic-level prey consumption, were associated with heavier body mass. Scatter plots revealed a clear positive trend between Delta 15 N and body mass across all species, with some variation in slope and distribution, reflecting species-specific dietary preferences. Carbon isotopes (Delta 13 C), representing foraging locations, showed a influence on body mass. These relationships are illustrated in scatter plots of body mass vs. isotopes, separated by species and gender.



Comparing external and internal factors revealed that gender has a much larger influence on body mass than other factors. Following are three regression models for the three species, using Delta 15 N, Delta 13 C and gender to study body mass:

Adelie:

```
Call:
lm(formula = `Body Mass (g)` ~ `Delta 15 N (o/oo)` + `Delta 13 C (o/oo)` +
    Sex, data = adelie_2)

Residuals:
    Min      1Q  Median      3Q     Max
-627.12 -203.77   -6.72  195.07  771.32

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          7176.74    1723.73   4.163 5.55e-05 ***
`Delta 15 N (o/oo)`  -205.02      70.70  -2.900  0.00436 **
`Delta 13 C (o/oo)`    77.83      50.94   1.528  0.12894
SexMALE               717.92      51.86  13.842  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 299 on 135 degrees of freedom
Multiple R-squared:  0.5869,    Adjusted R-squared:  0.5777
F-statistic: 63.94 on 3 and 135 DF,  p-value: < 2.2e-16
```

Chinstrap:

```
Call:
lm(formula = `Body Mass (g)` ~ `Delta 15 N (o/oo)` + `Delta 13 C (o/oo)` +
    Sex, data = chinstrap_2)

Residuals:
    Min      1Q  Median      3Q     Max
-715.85 -188.77   16.76  203.41  846.08

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          4084.84    4444.47   0.919   0.3616
`Delta 15 N (o/oo)`  -219.21     114.80  -1.909   0.0608 .
`Delta 13 C (o/oo)`   -59.85     182.95  -0.327   0.7446
SexMALE               459.16      83.17   5.521  6.8e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 324.4 on 63 degrees of freedom
Multiple R-squared:  0.3268,    Adjusted R-squared:  0.2947
F-statistic: 10.19 on 3 and 63 DF,  p-value: 1.459e-05
```

Gentoo:

```
Call:
lm(formula = `Body Mass (g)` ~ `Delta 15 N (o/oo)` + `Delta 13 C (o/oo)` +
    Sex, data = gentoo_2)

Residuals:
    Min      1Q  Median      3Q     Max
-699.02 -188.57   14.13  204.89  724.95

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         6434.92    1337.29   4.812 4.62e-06 ***
`Delta 15 N (o/oo)`   23.76     132.85   0.179    0.858
`Delta 13 C (o/oo)`   74.43      63.97   1.163    0.247
SexMALE              804.42      57.24  14.053  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 298.9 on 114 degrees of freedom
Multiple R-squared:  0.6564,    Adjusted R-squared:  0.6474
F-statistic: 72.59 on 3 and 114 DF,  p-value: < 2.2e-16
```

For Adelie, we can see the figure for Delta 15 N and gender have significant relationship with body mass. For Chinstrap and Gentoo, we can see only gender has

significant relationship with body mass.

Conclusion and Recommendations

This analysis highlights significant interspecies differences in penguin traits and emphasizes the dominant role of internal factor (gender) in shaping body mass for Chinstrap and Gentoo, but for Adelie, although gender still plays an important role, Delta 15 N is also important. Delta 13 N has nothing to do with body mass for all three species. However, the study's robustness may be limited by sample size variations, missing data, and in balance between number of samples. Future research could address these limitations by incorporating multivariate models to better capture interactions between factors. Additionally, extending the analysis to other environmental variables, such as temperature or habitat and examining temporal trends in body mass through longitudinal data would provide deeper insights into penguin adaptations over time.

Appendix

Code:
https://www.dropbox.com/scl/fi/2iureqe044pb6mu2adfln/output.pdf?rlkey=qpes0gck141zdmbn6m2vpr6w6&st=bnhyda7j&dl=0