Intersexual differences in body mass across adélie, chinstrap, and gentoo penguins

Omar Morosse, Samyak Shrestha, and Sariyah Sintayehu

Kenyon College
Gambier, OH

February 26, 2024

1. Introduction

Variation in morphology can fundamentally shape how individuals in a population differentially interact with their environment and the other organisms around them (Hutchinson, 1957; Pérez-Barbería, 2006). We wished to explore relationships between the morphometric, taxonomic, and ecological characteristics of adélie (*Pygoscelis adeliae*), chinstrap (*Pygoscelis antarcticus*), and gentoo (*Psygoscelis papua*) penguins that inhabit the Palmer archipelago in Antarctica. Gentoo penguins are the heaviest of the *Pygoscelis* genus, followed by adélie penguins, and then chinstrap penguins. We were specifically interested in testing how culmen length, culmen depth, flipper length, as well as sex predict body masses across the three species.

These data were collected in the austral summers between 2007 and 2010 by Dr. Kristen Gorman at the Palmer station, part of the Antarctica Long Term Ecological Research network (LTER) (Gorman et al., 2014). Gorman et al. sought to understand how differences in sex-specific foraging behavior covaried with environmental variability. They captured each penguin during the nesting season to take blood for sex determination, as well as measure culmen length, culmen depth, flipper length, and body mass.

We used the following variables from these data as organized in the following table.

Variable	**Description**
Body mass (grams)	A quantitative variable that characterizes the weight of the penguin in question
Culmen length (millimeters)	A quantitative variable that characterizes the length of the top ridge of the penguin's bill
Culmen depth (millimeters)	A quantitative variable that characterizes the height of the penguin's bill
Flipper length (millimeters)	A quantitative variable that characterizes the distance between the penguin's sternum and flipper tip
Sex	A binary categorical variable indicating if the penguin is male or female
Species	A categorical variable indicating if the penguin in question is an adélie, chinstrap, or gentoo

2. Methods

After finalizing the variables of interest, we conducted exploratory data analysis to observe any immediate patterns in our data. Culmen length, culmen depth, and flipper length all have a positive linear relationship with body mass. It is important to note that culmen length and flipper length produce a single cluster of points while culmen depth produces two clusters of points (Figure 1). Lastly, independence is satisfied for any model created with these data as each observation denotes an individual penguin's morphometrics, which does not affect any other penguin.

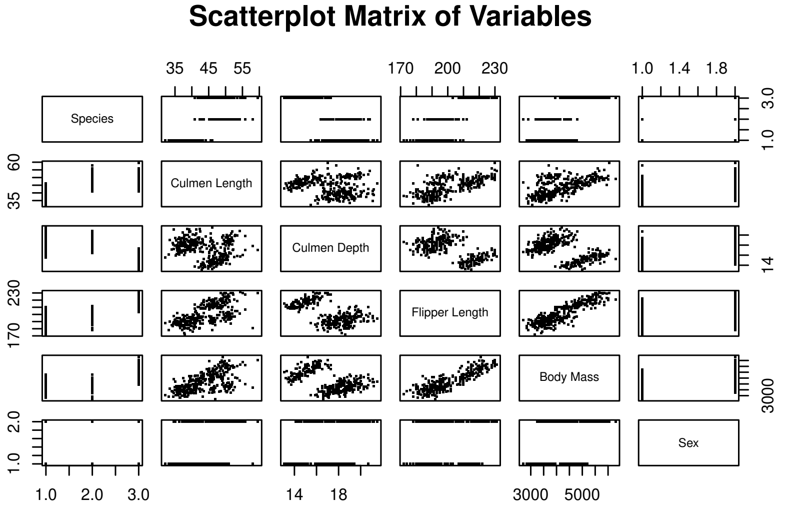
Scatterplot Matrix of Variables

Figure 1. Scatterplot Matrix of the Variables in the Penguins Dataset.

Choosing and Assessing the Model

We removed non-applicable (N/A) values, randomized the indices of our dataset, and split it such that 80% was training data and 20% was testing data. The training data was then used as a subset of the original dataset to build our model. We used the step function in both the forward and backward directions to see what R considered the best model given these data. The full model included culmen length, culmen depth, flipper length, sex, species, and the three-way interaction between sex, species, and flipper length. Excluding missing entries, our dataset has 333 observations.

Model Candidate 1

In the forward selection process, we observed that a majority of the p-values were statistically significant. However, this method produced a model that excluded the interaction terms involving the flipper length variable. This exclusion impacts the model's capacity to accurately predict body mass in relation to flipper length. Despite this conceptual limitation, the model demonstrated strong predictive power, with an adjusted R-squared value of 87.19%. To further assess the model's performance, we conducted a residual analysis. The model satisfied constant variance (even vertical spread throughout the plot), linearity (no trend in the residuals), and zero-mean (residuals evenly spread around the zero line) (Figure 2). The model also satisfied normality as the residuals closely followed the normal line (Figure 3). It should be noted that there is some deviation at the tails which is not necessarily concerning since the plot still passes the pencil test.

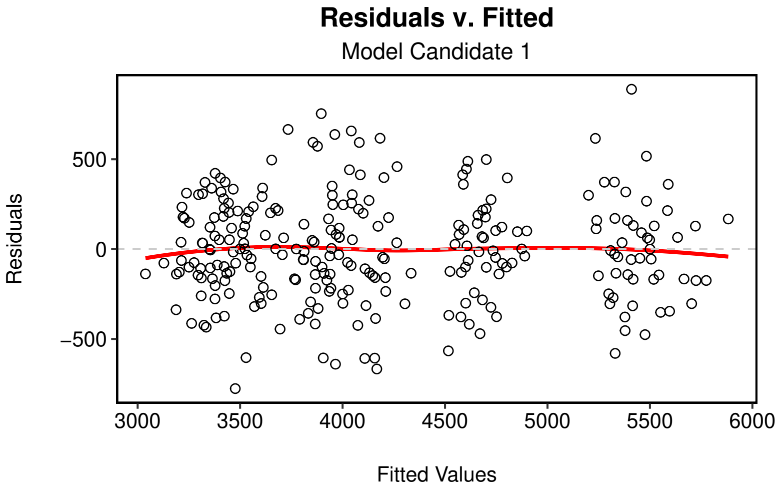

Figure 2. Residuals v. Fitted Plot for Forwards Stepwise Model.

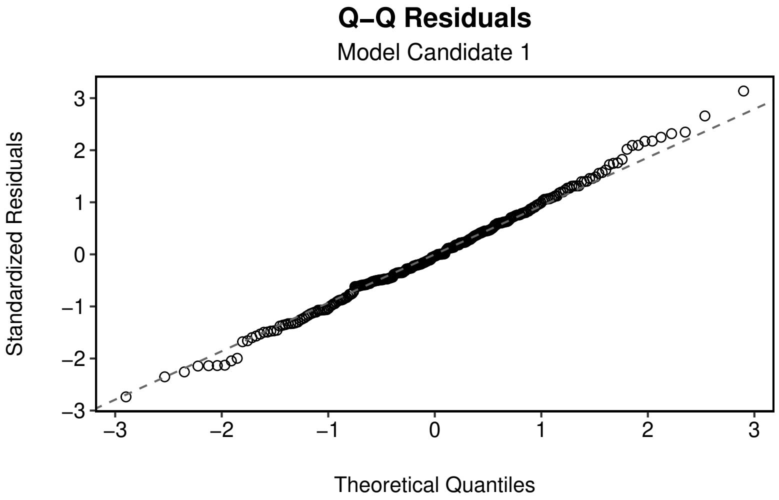

Figure 3. Q-Q Residuals Plot for Forwards Stepwise Model.

Model Candidate 2

The backwards selection process produced a full model which had non-significant p-values for the seven interaction terms except for the quantitative predictor variables of culmen length, culmen depth, and flipper length. These large *p*-values are of little concern since we require the interaction terms for meaningful statistical inference and these data have two categorical variables. This model demonstrated a stronger predictive power than Model Candidate 1, as denoted by an adjusted R-squared value of 87.43%. To further assess the predictive capabilities of this model, we ran a residuals analysis. This model satisfied constant variance (even vertical spread throughout the plot), linearity (no trend in the residuals), and zero-mean (residuals evenly spread around the zero line) (Figure 4). The model also satisfied normality as the residuals closely follow the normal line (Figure 6). Note that there is some deviation at the tails, but this is not concerning since the plot still passes the pencil test. The residuals are closer to the normal line in this plot that in Figure 3, which demonstrates additional benefit in including the flipper length interaction terms.

Adélie and chinstrap residuals clustered on the left side of the plot whereas gentoo residuals clustered at the center and right side of the plot. Gentoos are the heaviest penguins in our dataset, so their distance from our regression line should also be larger than the other species (Figure 4). Furthermore, the right-most clusters of female and male body mass residual points correspond to the gentoo population as shown in Figure 5. Similarly, the left-most clusters of female and male body mass residuals correspond to adélie and chinstrap populations, though they are interspersed with one another. We would expect male penguins to weigh more than their female counterparts, which an intersexual residual analysis confirms (Figure 5). Please reference Figure 6 for the normality of Model Candidate 2.

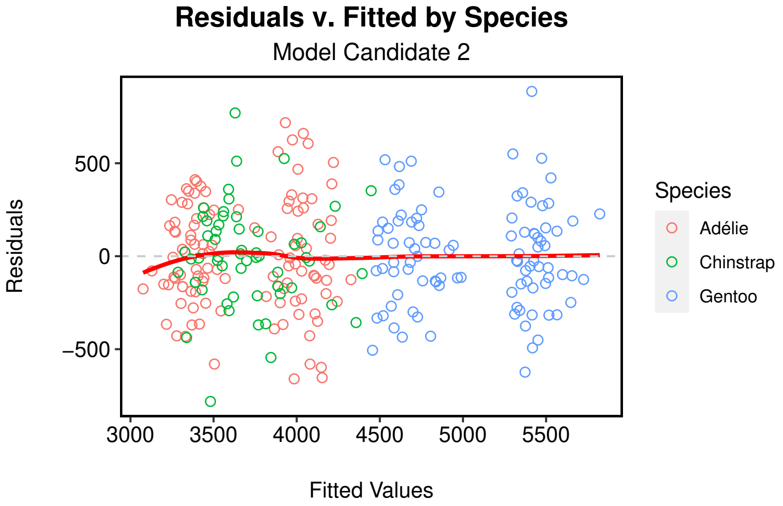

Figure 4. Residuals v. Fitted Plot by Species.

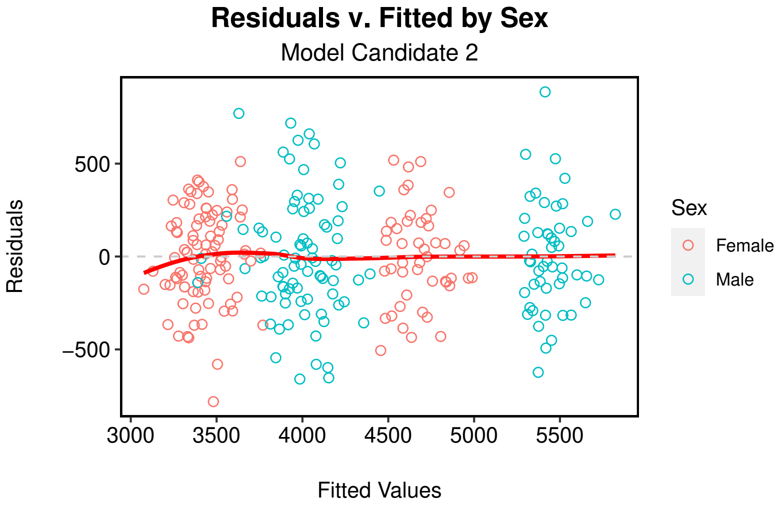

Figure 5. Residuals v. Fitted Plot by Sex.

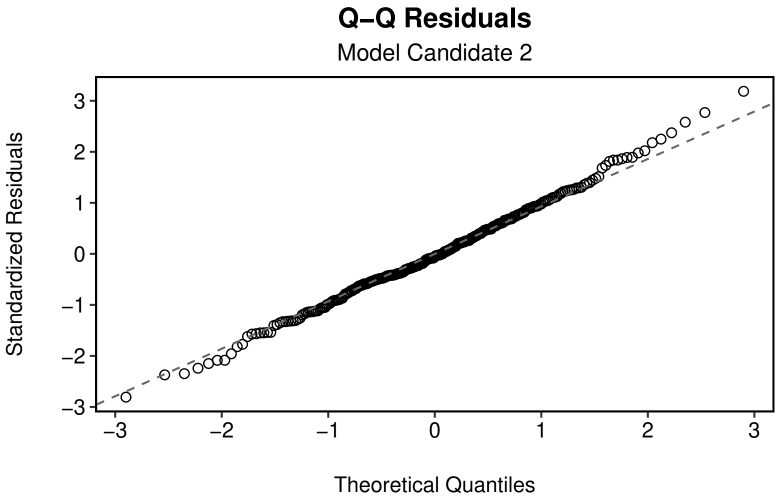

Figure 6. Q-Q Residuals Plot for Full/Backwards Stepwise Model.

Our residual analyses confirmed that both model candidates met the necessary conditions for linear regression. However, we adopted Model Candidate 2 because of its higher adjusted R-squared value and its inclusion of the flipper length interaction terms. These terms are crucial for capturing the dynamics of flipper length across the *Psygoscelis* penguins, which allow us to holistically address our research questions. Our general and fitted model equations are listed below.

$$
\begin{aligned}
\text{BodyMass} = {} & \beta_0 + \beta_1(\text{CulmenLength}) + \beta_2(\text{CulmenDepth}) + \beta_3(\text{FlipperLength}) + \beta_4(\text{SexMale}) \\
& + \beta_5(\text{SpeciesGentoo}) + \beta_6(\text{SpeciesChinstrap}) + \beta_7(\text{SexMale} \times \text{SpeciesGentoo}) \\
& + \beta_8(\text{SexMale} \times \text{SpeciesChinstrap}) + \beta_9(\text{SexMale} \times \text{FlipperLength}) \\
& + \beta_{10}(\text{SpeciesGentoo} \times \text{FlipperLength}) + \beta_{11}(\text{SpeciesChinstrap} \times \text{FlipperLength}) \\
& + \beta_{12}(\text{SexMale} \times \text{SpeciesGentoo} \times \text{FlipperLength}) \\
& + \beta_{13}(\text{SexMale} \times \text{SpeciesChinstrap} \times \text{FlipperLength}) + \varepsilon, \text{ where } \varepsilon \overset{\text{iid}}{\sim} \text{N}(0, \sigma).
\end{aligned}
$$

$$
\begin{aligned}
\text{Body}\hat{\text{M}}\text{ass} = {} & -650.13 + 22.33(\text{CulmenLength}) + 61.35(\text{CulmenDepth}) + 11.23(\text{FlipperLength}) \\
& + 517.47(\text{SexMale}) - 2390.3(\text{SpeciesGentoo}) - 276.05(\text{SpeciesChinstrap}) \\
& + 4276.54(\text{SexMale} \times \text{SpeciesGentoo}) - 6006.99(\text{SexMale} \times \text{SpeciesChinstrap}) \\
& - 0.34(\text{SexMale} \times \text{FlipperLength}) + 16.18(\text{SpeciesGentoo} \times \text{FlipperLength}) \\
& + 0.96(\text{SpeciesChinstrap} \times \text{FlipperLength}) - 19.7(\text{SexMale} \times \text{SpeciesGentoo} \times \text{FlipperLength}) \\
& + 28.3(\text{SexMale} \times \text{SpeciesChinstrap} \times \text{FlipperLength}).
\end{aligned}
$$

For each combination of species and sex, the model calculates a unique regression line. This is possible because of the three-way interaction term between sex, species, and flipper length, which modifies the slope and intercept of the relationship between flipper length and body mass based on the specific sex-species combination. Since there are three species and two sexes, this setup naturally leads to six distinct lines (3 species × 2 sexes = 6 lines). The inclusion of a three-way interaction term allows the model to account for how these differences change with other variables (in this case, flipper length) which ensures that the model accurately represents each morphometric characteristic that contributes to overall body mass.

3. Results

Question #1 (Cross Validation): How well does this model perform while predicting test data?

To assess the efficacy of our regression model in predicting data, we used a cross validation procedure. We used the trained model to generate predictions for the body masses on the separate test data divided during model selection. The effectiveness of the model was quantitatively assessed by calculating the Pearson correlation coefficient between the observed and predicted body mass values across the test data. Subsequently, the shrinkage value was used to further evaluate the model's performance.

We generally use shrinkage as a rough indicator of how well the model is doing at predicting the holdout sample. Shrinkage of 10% or less should not be a problem, but a shrinkage of more than 50% would be worrisome (Cannon et al., 2019). We computed a shrinkage value of 3.91%, which suggests that our model is efficient at predicting holdout data.

Question #2 (Confidence Intervals): Can we conclude differences in body masses across sexes of penguin species?

We subsetted the training data to only include the means of each predictor variable so we could construct 95% confidence intervals for the mean body mass of each species and sex of penguin. We are 95% confident that male adélie penguins with mean culmen lengths, culmen depths, and flipper lengths weigh on average between 3,958.91 and 4,111.08 grams, while female adélie penguins with mean culmen lengths, culmen depths, and flipper lengths weigh on average between 3,298.37 and 3,446.12 grams. The lack of overlap between the confidence intervals for the two sexes suggests strong sexual mass dimorphism across the adélie penguins in our sample, where males are heavier than females.

Likewise, we are 95% confident that male gentoo penguins with mean culmen lengths, culmen depths, and flipper lengths weigh on average between 5,375.96 and 5,535.23 grams, while female gentoo penguins with mean culmen lengths, culmen depths, and flipper lengths weigh on average between 4,600.01 and 4,763.95 grams. Like the adélie penguins, the lacking overlap between the confidence intervals for the two sexes suggests strong sexual mass dimorphism across the gentoo penguins in our sample, where again, males are heavier than females.

Finally, we are 95% that male chinstrap penguins with mean culmen lengths, culmen depths, and flipper lengths weigh on average between 3,824.82 and 4,042.56 grams, while female chinstrap penguins with mean culmen lengths, culmen depths, and flipper lengths weigh on average between 3,428.39 and 3,634.21 grams. Like their *Pygoscelis* relatives, chinstrap penguins lack an overlap between the confidence intervals for the two sexes, which suggests strong sexual mass dimorphism skewed towards the males in the sample.

Question #3 (Prediction Intervals): Given the same terms observed across each sex, what does our model predict for an individual penguin's body mass in that species?

We also used our model to create prediction intervals on the untrained testing data for each species of penguin by sex. We randomly selected one adélie, chinstrap, and gentoo penguin from our testing data. For each species, we constructed a prediction interval such that the male and female intervals used the same culmen length, culmen depth, and flipper length from our randomly selected penguins so that we could compare intraspecies intervals.

We are 95% confident that with a culmen length of 37.7 millimeters, culmen depth of 18.7 millimeters, and flipper length of 180 millimeters, a female adélie penguin will have a body mass between 2,782.88 and 3,938.54 grams and a male adélie penguin will have a body mass between 3,230.49 and 4,403.29 grams. Our model predicts that a male adélie penguin with these body measurements will weigh between 447.6 and 464.76 grams more than a comparable female adélie penguin.

Similarly, we are 95% confident that with a culmen length of 50.8 millimeters, culmen depth of 19 millimeters, and flipper length of 210 millimeters, a female chinstrap penguin will have a body mass between 3,274.92 and 4,595.17 grams and a male chinstrap penguin will have a body mass between 3,707.86 and 4,925.53 grams. Our model predicts that a male chinstrap penguin with these body measurements will weigh between 330.36 and 432.94 grams more than a comparable female chinstrap penguin.

Finally, we are 95% confident that with a culmen length of 45.2 millimeters, culmen depth of 13.8 millimeters, and flipper length of 215 millimeters, a female gentoo penguin will have a body mass between 4,139.79 and 5,279.93 grams and a male gentoo penguin will have a body mass between 4,615.73 and 5,772.85 grams. Our model predicts that a male gentoo penguin with these body measurements will weigh between 475.94 and 492.92 grams more than a comparable female gentoo penguin.

Although these are prediction intervals for a single penguin, we observe that the males weigh more than the females across the species at comparable morphometrics, which is what we would expect. However, we cannot generalize this observation for every penguin in our testing sample.

4. Further Discussion

In our study, we compared models from both forwards selection and backwards elimination processes. While the forwards selection suggested a reduced model, the backwards elimination recommended the full model, which we ultimately adopted. This decision was driven by our goal to thoroughly explore the interactions between sex, species, and flipper lengths, particularly in light of (Gorman et al.'s, 2014) findings on strong sexual mass dimorphism within the gentoo and adélie species. Our full model aims to capture the flipper length dynamics between different species and sexes of penguins. These interaction terms captured the trend in the differences in species compared to our exploratory data analysis, where it was not immediately apparent.

We encountered challenges in our statistical analysis with missing data, but we ultimately omitted non-applicable (N/A) values. However, we acknowledge that this approach might introduce bias into our analysis if non-applicable values were not completely random. Additionally, we identified an anomaly in the raw dataset, where a “full-stop” character in the sex column of observation 338 disrupted the integrity of the model. To maintain the reliability of our analysis, we made the decision to exclude this observation as well.

Finally, our findings are somewhat consistent with those of Gorman et al.'s (2014). We found that all three *Pygoscelis* species exhibited sexual mass dimorphism as well, while Gorman et al. (2014) only found such patterns across adélie and gentoo species. Additionally, they gathered data on Nitrogen isotope concentrations for each individual, which provides information about the protein-richness of an individual's forage, and further separated their analyses from ours. Gorman et al. (2014) concluded that chinstrap penguins were the only species to exhibit intersexual differences in Nitrogen isotope concentrations. However, the difference was not large enough to suggest that the sexes occupy different trophic niches. While our study differed from Gorman et al.'s (2014) study in scope, an important similarity between the studies is the fact that the data we shared were all collected during the breeding season. As a result, we can only draw conclusions about these penguins during this portion of the annual cycle.

5. Conclusion

Our full model allowed for the widest range of inference from our data, and it satisfied all the conditions for linear regression: independence, linearity, constant variance, zero-mean, and normality. Our cross-validation analysis produced a shrinkage value of 3.19%, which suggests that our model performed well on our holdout data. When applied to our last two questions, our model suggests that there are strong intersexual differences in mean body masses for adélie, chinstrap, and gentoo penguins with mean culmen lengths, culmen depths, and flipper lengths. Further, the model suggests that if an individual of each sex in a single species possess the same measurements for culmen length, culmen depth, and flipper length, then the males will still be heavier than the females.

References

Cannon, Ann R., et al. *STAT2: Modeling with Regression and ANOVA.* 2nd ed., W. H. Freeman, 2019.

Gorman, Kristen B., et al. "Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus *Pygoscelis*)." *PLoS ONE*, vol. 9, no. 3, Mar. 2014, pp. e90081. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pone.0090081.

Hutchinson, G. E. "Concluding Remarks." *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 22, no. 0, Jan. 1957, pp. 414–27. *DOI.org (Crossref)*, https://doi.org/10.1101/SQB.1957.022.01.039.

Pérez-Barbería, Javier. "Sexual Segregation in Vertebrates: Ecology of the Two Sexes. *Based on a Workshop Held in Cambridge, September 2002. Edited by K. E. Ruckstuhl and P. Neuhaus*." *The Quarterly Review of Biology*, vol. 81, no. 4, Dec. 2006, pp. 424–25. *DOI.org (Crossref)*, https://doi.org/10.1086/511617.