The penguins of Antarctica

Selina Baldauf

25/3/2022

# Introduction

There are three main penguin species in Antarctica (*Chinstrap*, *Gentoo*, *Adelie*). You can see them in the following figure:



Artwork by [Allison Horst](https://twitter.com/allison_horst)

In this paper we want to answer the following questions

1. How bill depth depends on bill length?
2. Which penguin species has the highest body mass?

# Methods

All analysis was done using R version 4.1.3 (R Core Team 2022) and the R markdown package (Allaire et al. 2021).

## The data

The data was collected on islands in Antarctica and published by Gorman, Williams, and Fraser (2014). You can find the original paper with the title “Ecological sexual dimorphism and environmental variability within a community of Antarctic penguins (genus *Pygoscelis*)” (Gorman, Williams, and Fraser 2014) in PLoS ONE[[1]](#footnote-23)

The data is published via the palmerpenguins R package (Horst, Hill, and Gorman 2020) which you can find [on this website](https://allisonhorst.github.io/palmerpenguins/).

**The data contains (among others) the following measurements:**

* bill length
* bill depth
* body mass
* sex
  + male
  + female

## The analysis

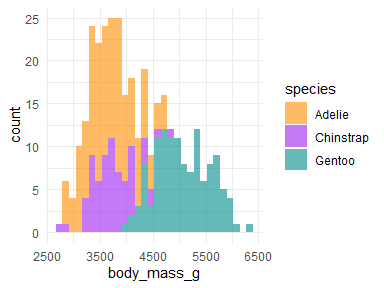
We did some plots, calculated some summary statistics and a linear model of the form

# Results

The mean weight of all penguin species is 4201.754386. *Gentoo* penguins have an average weight of 5076 g, *Adelie* penguins of 3701 g and *Chinstrap* penguins of 3733 g.

The figure below shows that *Gentoo* penguins have the highest body mass.

ggplot(penguins, aes(x = body\_mass\_g, fill = species)) +  
 geom\_histogram(alpha = 0.6) +  
 scale\_fill\_manual(values = c("darkorange", "purple", "cyan4")) +  
 theme\_minimal()



Histogram of weight of the three penguin species.

There is a positive relationship between bill length and bill depth for all 3 species, as the figure below shows.

ggplot(  
 data = penguins,  
 aes(  
 x = bill\_length\_mm,  
 y = bill\_depth\_mm,  
 color = species,  
 shape = species  
 )  
) +  
 geom\_point(size = 3, alpha = 0.8) +  
 geom\_smooth(method = "lm", se = FALSE) +  
 scale\_color\_manual(values = c("darkorange", "purple", "cyan4")) +  
 theme\_minimal()

In general, it looks like the body characteristics differ between the sexes but also between the penguin species, as the table below illustrates:

penguins %>%   
 filter(!(is.na(sex))) %>%   
 group\_by(species, sex) %>%   
 summarize(  
 bill\_length = mean(bill\_length\_mm, na.rm = TRUE),  
 bill\_depth = mean(bill\_depth\_mm, na.rm = TRUE),  
 flipper\_length = mean(flipper\_length\_mm, na.rm = TRUE),  
 body\_mass = mean(body\_mass\_g, na.rm = TRUE)  
 )

## # A tibble: 6 x 6  
## # Groups: species [3]  
## species sex bill\_length bill\_depth flipper\_length body\_mass  
## <fct> <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 Adelie female 37.3 17.6 188. 3369.  
## 2 Adelie male 40.4 19.1 192. 4043.  
## 3 Chinstrap female 46.6 17.6 192. 3527.  
## 4 Chinstrap male 51.1 19.3 200. 3939.  
## 5 Gentoo female 45.6 14.2 213. 4680.  
## 6 Gentoo male 49.5 15.7 222. 5485.

The linear model analysis shows that both bill length and species have a significant effect on the bill depth of the penguins.

# Calculate the linear model  
lm\_penguins <- lm(bill\_depth\_mm ~ bill\_length\_mm + species, data = penguins)  
# Create the summary table  
summary(lm\_penguins)

##   
## Call:  
## lm(formula = bill\_depth\_mm ~ bill\_length\_mm + species, data = penguins)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.4529 -0.6864 -0.0508 0.5519 3.5915   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.59218 0.68302 15.508 < 2e-16 \*\*\*  
## bill\_length\_mm 0.19989 0.01749 11.427 < 2e-16 \*\*\*  
## speciesChinstrap -1.93319 0.22416 -8.624 2.55e-16 \*\*\*  
## speciesGentoo -5.10602 0.19142 -26.674 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9533 on 338 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.769, Adjusted R-squared: 0.7669   
## F-statistic: 375.1 on 3 and 338 DF, p-value: < 2.2e-16

# References

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2021. *Rmarkdown: Dynamic Documents for r*. <https://github.com/rstudio/rmarkdown>.

Gorman, Kristen B., Tony D. Williams, and William R. Fraser. 2014. “Ecological Sexual Dimorphism and Environmental Variability Within a Community of Antarctic Penguins (Genus Pygoscelis).” Edited by André Chiaradia. *PLoS ONE* 9 (3): e90081. <https://doi.org/10.1371/journal.pone.0090081>.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. <https://doi.org/10.5281/zenodo.3960218>.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

1. paper available [here](https://doi.org/10.1371/journal.pone.0090081). [↑](#footnote-ref-23)