

Penguins

Your Name Here

May 2022

Introduction

This report analyzes the `penguins` data from the `palmerpenguins` package. It will settle the longstanding debate in the literature whether there is a positive correlation between the variables

- bill length
- bill depth

According to Horst, Hill, and Gorman (2020), the dataset contains size measurements for 344 penguins from three species observed on three islands in the Palmer Archipelago, Antarctica.

Distribution of species

We can determine the number of penguins for each species easily with functions from the `tidyverse` (Wickham et al. 2019):

```
penguins %>%  
  count(species)
```

However, using `datasummary_skim()` from the package `modelsummary` (Arel-Bundock 2022) for the same task results in a nicely formatted table that is automatically embedded in this report.

species	N	%
Adelie	152	44.2
Chinstrap	68	19.8
Gentoo	124	36.0

Visual exploration

At first glance, there seems to be a *negative* correlation between bill length and bill depth, if anything. However, broken down by species the plot does support the notion of a positive correlation. This phenomenon (an effect that vanishes or reverses when groups are combined) is an example of [Simpson's paradox](#).

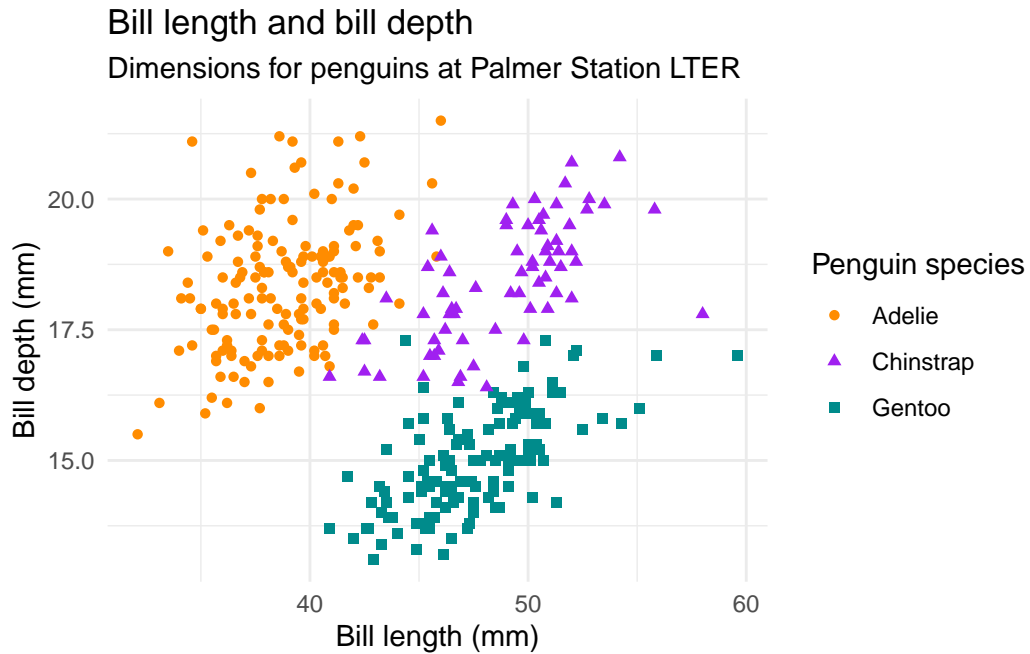


Figure 1: Simpson's paradox

In the next chapter, we will analyze the correlation depicted in Figure 1¹ using linear regression.

Statistical analysis

We formalize our model with the following equation:

$$depth_i = \beta_0 + \beta_1 length_i + \varepsilon_i.$$

Results are displayed in Table 1.

¹Labels should start with `fig-` for figures and `tbl-` for tables.

Table 1: Linear regression

	species: Adelie	species: Chinstrap	species: Gentoo
Bill length (mm)	0.179*** (0.036)	0.222*** (0.044)	0.205*** (0.021)
Num.Obs.	151	68	123
R2 Adj.	0.148	0.418	0.409
Std.Errors	Robust	Robust	Robust

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

References

- Arel-Bundock, Vincent. 2022. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. <https://vincentarelbundock.github.io/modelsummary/>.
- Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. <https://allisonhorst.github.io/palmerpenguins/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.