

AE 11: Modelling penguins

In this application exercise we will be studying penguins. The data can be found in the **palmerpenguins** package and we will use **tidyverse** and **tidymodels** for data exploration and modeling, respectively.

```
library(tidyverse)
library(tidymodels)
library(palmerpenguins)
```

Please read the following context and take a **glimpse** at the data set before we get started.

This data set comprising various measurements of three different penguin species, namely Adelie, Gentoo, and Chinstrap. The rigorous study was conducted in the islands of the Palmer Archipelago, Antarctica. These data were collected from 2007 to 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network. The data set is called **penguins**.

```
# add code here
```

Our goal is to understand better how various body measurements and attributes of penguins relate to their body mass. First, we are going to investigate the relationship between a penguins' flipper lengths and their body masses.

- **Question:** Based on our research focus, which variable is the response variable?

Add response here

- **Demo:** Visualize the relationship between flipper length and body mass of penguins.

```
# add code here
```

Correlation

- **Your turn (5 minutes):**
 - What is correlation? What values can correlation take?
Strength and direction of a linear relationship. It's bounded by -1 and 1.
 - Are you good at guessing correlation? Give it a try! <https://www.rossmanchance.com/applets/2021/guesscorrelation/GuessCorrelation.html>
- **Demo:** What is the correlation between flipper length and body mass of penguins?

```
# add code here
```

Defining, fitting, and summarizing a model

- **Demo:** Write the population model below that explains the relationship between body mass and flipper length.

add math text here

- **Demo:** Fit the linear regression model and display the results. Write the estimated model output below.

```
# add code here
```

add math text here

- **Your turn:** Interpret the slope and the intercept in the context of the data.
 - Intercept: *Add your response here*
 - Slopes: *Add your response here*
- **Your turn:** Recreate the visualization from above, this time adding a regression line to the visualization `geom_smooth(method = "lm")`.

```
# add code here
```

- What is the estimated body mass for a penguin with a flipper length of 210?

```
# add code here
```

- What is the estimated body mass for a penguin with a flipper length of 100?

```
# add code here
```

Another model

- **Demo:** A different researcher wants to look at body weight of penguins based on the island they were recorded on. How are the variables involved in this analysis different?

Add response here

- **Demo:** Make an appropriate visualization to investigate this relationship below. Additionally, calculate the mean body mass by island.

```
# add code here
```

```
# add code here
```

- **Demo:** Change the geom of your previous plot to `geom_point()`. Use this plot to think about how R models these data.

```
# add code here
```

- **Your turn:** Fit the linear regression model and display the results. Write the estimated model output below.

```
# add code here
```

Add math text here

- **Demo:** Interpret each coefficient in context of the problem.
 - Intercept: *Add your response here*
 - Slopes: *Add your response here*
- **Demo:** What is the estimated body weight of a penguin on Biscoe island? What are the estimated body weights of penguins on Dream and Torgersen islands?

```
# add code here
```