05\_Homework

Bill Perry

# Homework Week 3

This is an assignment for you to practice coding and redo the work we do in class with a few twists on a new dataframe practicing to create new projects and writing new code. I realize you could copy the code from lecture and although you will get the code right, I urge you to retype it form scratch as it will be learned so much faster. This is a new language for you and if you dont “type” == “speak” the language you would remember it…. really and try breaking things. Dont be afraid you can download a new version or fix it… that is how we learn.

# Background

There is a general hypothesis out there that animals are larger on islands. This is summarized as The Island Rule: The “island rule” posits that when mainland animals colonize islands, small species tend to evolve larger bodies, and large species tend to evolve smaller bodies (insular dwarfism). Why Rodents Grow Larger:

* Reduced Predation: Islands often have fewer or no native predators of rodents, allowing them to evolve larger sizes without the selective pressure of being hunted.
* Reduced Competition: Islands may have fewer species competing for resources, which can allow rodents to exploit a wider range of resources and potentially grow larger.
* Evolutionary Advantage: Larger size can provide advantages in some island environments, such as better thermoregulation, increased competitive ability, and the ability to exploit larger prey or food sources.

[Foster’s Rule](https://en.wikipedia.org/wiki/Foster%27s_rule): J. Bristol Foster, a biologist, first described these trends in 1964, and his observations led to the formulation of “Foster’s rule” or the “island effect”.



[from this site](https://payseur.genetics.wisc.edu/research/research-island-syndrome-and-the-evolution-of-extremes/)

# Objectives and goals

In this active learning module, we’ll explore statistical inference using mice data from the San Juan Islands off of Vancouver, Canada.

Today we will test this on the mice from

* We need to explore the data graphically
  + as a whole
    - box plots
    - histograms
  + By mainland and island
    - box plots
    - histograms
* Generate summary statistics
  + mean, standard deviation, standard error, N
* Do a one sample T test for each population comparing to a population mean of 19 g
  + Question do the island or mainland differ from the population mean of 19 grams
  + State the Null and Alternate Hypothesis
  + Be sure to state and test assumptions
  + Report the statistical results as if you were writing a scientific results section
* Do a two sample T Test
  + Question does the island population differ in mass from the mainland?
  + State the Null and Alternate Hypothesis
  + Be sure to state and test assumptions
  + Report the statistical results as if you were writing a scientific results section

We’ll use the tidyverse package for data manipulation and visualization, along with patchwork for combining plots.

## Setup

First, let’s load the packages we need and the dataset:

# Load required packages  
library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
✔ purrr 1.0.4   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(patchwork)  
  
# Read in the data file  
w5\_df <- read\_csv("data/mice\_weights.csv")

Rows: 62 Columns: 4  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (2): sampling\_site, location  
dbl (2): date\_year, mass\_g  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Look at the first few rows  
head(w5\_df)

# A tibble: 6 × 4  
 sampling\_site location date\_year mass\_g  
 <chr> <chr> <dbl> <dbl>  
1 Sidney Island island 2021 26   
2 Sidney Island island 2021 24   
3 Sidney Island island 2021 21.5  
4 Sidney Island island 2021 23   
5 Sidney Island island 2021 22   
6 Sidney Island island 2021 21

Let’s calculate some basic statistics for mice

# Calculate basic statistics   
w5\_stats <- w5\_df %>%   
 group\_by(sampling\_site) %>%   
 summarize(  
 mean\_mass = mean(mass\_g, na.rm = TRUE),  
 sd\_mass = sd(mass\_g, na.rm = TRUE),  
 n = sum(!is.na(mass\_g)),  
 se\_mass = sd\_mass / sqrt(n)  
 )  
  
# Display the statistics  
w5\_stats

# A tibble: 2 × 5  
 sampling\_site mean\_mass sd\_mass n se\_mass  
 <chr> <dbl> <dbl> <int> <dbl>  
1 Sidney Island 23.4 2.80 33 0.487  
2 Vancouver 20.2 1.71 28 0.324