

# 03\_Homework\_t\_test

Bill Perry

## Homework Week 3

The goal of this assignment is to begin doing statistical tests and we are starting with T Tests. From here on out there should be a rather standard format for these sorts of laboratory reports. I would like to see how you design your projects and begin to do the statistical tests. What I will be looking for is:

1. How you import the data and begin to assess what the data are
  - What types of variables are there? So run a `str(dataframe)` and report the variable types. Why - so many times the code fails because the data are coded incorrectly.
2. What are the basic summary stats for each variable by group if appropriate
  - response variable name with units
  - number
  - mean
  - standard deviation
  - standard error
3. One or more exploratory graphs
  - what is the structure of the data by groups
  - does it look like variance between groups is consistent
  - are there outliers
4. What test are you going to be doing?
  - what is the name of the test
  - what are the assumptions - really important
  - what is the hypothesis that will be tested
  - The test itself with the output
    - you should interpret the output briefly noting what it means
    - do you accept or reject the hypothesis
5. Tests of each assumption for the test
  - Not for each test if the assumptions are met or not
  - if assumptions are not met what do you do next
6. A final plot that would be closer to publication quality
  - this means that it should have proper axes labels, legend labels
  - a exported pdf of the plot using `ggsave`

## 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: 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

## 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.

- We need to explore the data graphically
  - as a whole (**1 points**)
    - box plots or
    - histograms
  - By mainland and island (**1 points**)
    - box plots or
    - histograms
- Generate summary statistics (**1 point**)
  - mean, standard deviation, standard error, N
- Do a one sample T test for each population comparing to a population mean of 19 g (**3 points**)
  - 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 (**4 points**)
  - 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    3.5.2      ✓ tibble     3.3.0
✓ lubridate  1.9.4      ✓ tidyr      1.3.1
✓ purrr      1.1.0

— 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
str(w5_df)
```

```
spc_tbl_ [62 × 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ sampling_site: chr [1:62] "Sidney Island" "Sidney Island" "Sidney Island" "Sidney Island" ...
 $ location      : chr [1:62] "island" "island" "island" "island" ...
 $ date_year     : num [1:62] 2021 2021 2021 2021 2021 ...
 $ mass_g        : num [1:62] 26 24 21.5 23 22 21 22 20 28 25 ...
 - attr(*, "spec")=
 .. cols(
```

```

..   sampling_site = col_character(),
..   location = col_character(),
..   date_year = col_double(),
..   mass_g = col_double()
.. )
- attr(*, "problems")=<externalptr>

```

Let's calculate some basic statistics for mice

```

# Calculate basic statistics
w5_stats <- w5_df %>%
  group_by(sampling_site) %>%
  summarize(
    n = sum(!is.na(mass_g)),
    mean_mass = mean(mass_g, na.rm = TRUE),
    sd_mass = sd(mass_g, na.rm = TRUE),
    se_mass = sd_mass / sqrt(n)
  )

# Display the statistics
w5_stats

```

```

# A tibble: 2 × 5
  sampling_site      n mean_mass sd_mass se_mass
  <chr>         <int>    <dbl>   <dbl>   <dbl>
1 Sidney Island    33     23.4     2.80    0.487
2 Vancouver       28     20.2     1.71    0.324

```

## Submission Guidelines

### What to turn in -

1. a quarto markdown file and dataframe. Note that your code should be able to run with what you turn in.
2. a self-contained word and html file showing the code and output
3. annotations in the quarto file that shows or tells what is being done in the r code chunks describing what you are trying to do - credit will be given even if it does not work as long as you detail what you are doing. As we start to move into more statistics you will be expected to interpret the results.

### Points

- summary stats - 1 points
- exploratory graphs - 1 point
- Hypotheses - 2 points
- Assumptions - 2 points
- Model and Test Results - 4
- interpretation - 3 points
- Assumption Tests - 4 points
- Final figure - 3 points