Inference of Population Means

STA 610 - Applied Statistics for Health Professions

Andrew DiLernia

## Learning Objectives

* State hypotheses for 1-sample and 2-sample -test
* State the linear model formulations of 1-sample and 2-sample -test
* State and check assumptions of 1-sample and 2-sample -test
* Obtain and interpret results of 1-sample and 2-sample -test
* Obtain and interpret confidence intervals for population means & mean differences

Let’s begin by loading some R packages for this activity using the code below. Note: if it is the first time you are using an R package, you may need to install it first using the install.packages() function.

*# Load necessary packages*  
library(tidyverse)  
library(ggthemes)  
library(flextable)

Next, we can set default theme settings for plots, and load some functions to simplify table customization and creation using the code below.

*# Set ggplot theme for visualizations*  
theme\_set(ggthemes::theme\_few())  
  
*# Set options for flextables*  
set\_flextable\_defaults(na\_str = "NA")  
  
*# Load function for printing tables nicely*  
source("https://raw.githubusercontent.com/dilernia/STA323/main/Functions/make\_flex.R")

## Inference of Population Means

**Statistical Inference** involves using data analysis to infer characteristics of a population of interest. Some of the most common tools of statistical inference include:

* Hypothesis testing
* Confidence intervals

## Univariate -test

The univariate -test, also called the one-sample -test, is an example of a hypothesis test for conducting statistical inference of a single quantitative variable. For example, we can use a one-sample -test in the following scenarios:

1. In June of 2022, the United States EPA set health advisory limits for the [PFAS](https://www.cdc.gov/biomonitoring/PFAS_FactSheet.html#:~:text=Print,in%20a%20variety%20of%20products.) hexafluoropropylene oxide dimer acid (HFPO-DA) to be 10 parts per trillion (ppt). We want to test if the average HFPO-DA level in water from residential homes in Grand Rapids, MI exceeds the level set by the EPA. ☣️🫗
2. Conventionally, it was believed that the average human body temperature was 98.6 degrees Fahrenheit, but more recently this notion has been met with skepticism. We want to test if the average temperature of people is truly 98.6 degrees Fahrenheit. 🌡
3. We want to test if the average height of starting NBA players is above 77 inches (6 ft 5 inches). 🏀
4. We want to test if the average number of pets each GVSU student has is different than 1. 🐕 🐈 🦜

➡️ What are the response variables in each scenario, and what is the type of each variable being as specific as possible?

1. *Response Variable:* PFAS value (in ppt)
2. *Response Variable:* Human body temperature
3. *Response Variable:* Height of each Player
4. *Response Variable:* Number of pets per student

➡️ Describe how we could obtain a random sample (or at least close to one) in each scenario.

*1. Sample : Collecting water samples from each individual house in the chosen area.*

*2. Sample : Take the temperature of people from different regions, sex and age groups.*

*3. Sample : Height of all NBA players from player database.*

*4. Sample : Querying each student of GVSU.*

### Assumptions

Some conditions must be met to conduct a valid one-sample -test:

1. Independent observations (commonly violated when observations consist of repeated measurements across time)
2. The [normality assumption](https://docs.google.com/document/d/1jVscyqIwKhuVfzlP60_8XNKoWQz7vvnN/edit#heading=h.1ksv4uv) must be met. That is, we need that the sampling distribution of the sample mean is normal. This can be met if we have normally distributed data, or a large enough sample size without extreme outliers.

### Hypotheses

For testing if the population mean, , is a specific value, , a formal statement of the hypotheses is:

or we can also test if , instead of .

### Fixed-effects model formulation

Linear model interpretation of the one-sample -test:

* : represents observation
* : overall grand mean
* : normally distributed error term for observation

### Test statistic

The test statistic for the univariate -test is:

where:

* : sample mean (sometimes this is denoted instead of )
* : sample size
* : sample standard deviation
* : null value

## Palmer penguins example: One-sample -test

In this activity we will analyze the Palmer Penguins data set: a data set consisting of measurements collected on 344 penguins from 3 islands in [Palmer Archipelago, Antarctica](https://www.google.com/maps/place/Palmer+Archipelago/@-64.1463814,-62.1800963,7z/data=!3m1!4b1!4m5!3m4!1s0xbc78dd6dc38c572b:0xe609367aeed33087!8m2!3d-64.1957848!4d-62.0153384).

|  |  |
| --- | --- |

*Artwork by @allison\_horst*

Let’s import data from the website GitHub to use for this activity.

*# Load Palmer penguins data*  
penguins <- readr::read\_csv("https://raw.githubusercontent.com/dilernia/STA323/main/Data/penguins.csv")

*Data dictionary for Palmer Penguins data set.*

| Variable | Description |
| --- | --- |
| species | Species of the penguin |
| island | Island the penguin was found on |
| bill\_length\_mm | Bill length (mm) |
| bill\_depth\_mm | Bill depth (mm) |
| flipper\_length\_mm | Flipper length (mm) |
| body\_mass\_g | Body mass (g) |
| sex | Sex of the penguin |
| year | Year data was collected |

As a researcher interested in penguins, we may be interested in the average flipper length (mm) of penguins from Palmer Archipelago, Antarctica. From our previous experience, we believe that penguins typically have an average flipper length of 200mm, but is this true for these penguins from Antarctica as well? We can explore this using a one-sample -test.

➡️ Formally state the hypotheses for our question of interest.

➡️ What is the null value, , for this test?

Let’s obtain some output to check whether or not the assumptions are met for the one-sample -test.

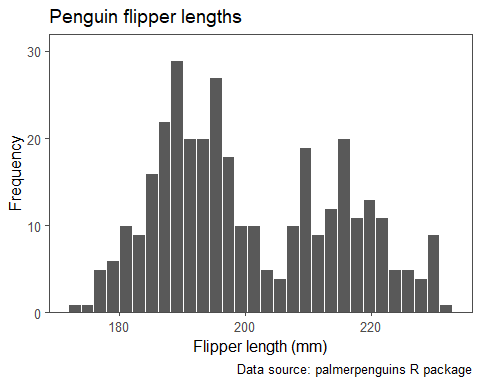
*# Calculating descriptive statistics*  
quant1Stats <- penguins %>%   
 dplyr::summarize(  
 Minimum = min(flipper\_length\_mm, na.rm = TRUE),  
 Q1 = quantile(flipper\_length\_mm, na.rm = TRUE, probs = 0.25),  
 M = median(flipper\_length\_mm, na.rm = TRUE),  
 Q3 = quantile(flipper\_length\_mm, na.rm = TRUE, probs = 0.75),  
 Maximum = max(flipper\_length\_mm, na.rm = TRUE),  
 Mean = mean(flipper\_length\_mm, na.rm = TRUE),  
 R = Maximum - Minimum,  
 s = sd(flipper\_length\_mm, na.rm = TRUE),  
 n = n()  
)  
  
*# Printing table of statistics*  
quant1Stats %>%   
 make\_flex(caption = "Quantitative summary statistics for penguin flipper lengths (mm).")

*Table 1: Quantitative summary statistics for penguin flipper lengths (mm).*

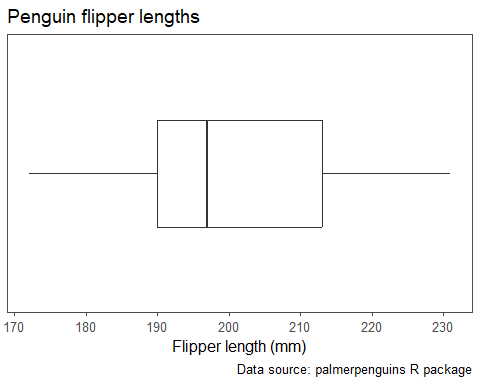
| Minimum | Q1 | M | Q3 | Maximum | Mean | R | s | n |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 172.00 | 190.00 | 197.00 | 213.00 | 231.00 | 200.92 | 59.00 | 14.06 | 344 |

*# Creating a histogram*  
penguins %>%   
 ggplot(aes(x = flipper\_length\_mm)) +   
 geom\_histogram(color = "white") +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.10))) +  
 labs(title = "Penguin flipper lengths",  
 x = "Flipper length (mm)",  
 y = "Frequency",  
 caption = "Data source: palmerpenguins R package")

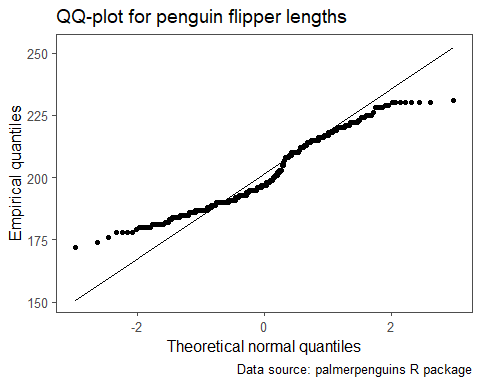
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



*# Creating a box plot*  
penguins %>%   
 ggplot(aes(x = flipper\_length\_mm)) +   
 geom\_boxplot() +  
 scale\_y\_discrete(breaks = NULL) +  
 labs(title = "Penguin flipper lengths",  
 x = "Flipper length (mm)",  
 caption = "Data source: palmerpenguins R package")



*# Creating a Quantile-Quantile (QQ) plot*  
penguins %>%   
 ggplot(aes(sample = flipper\_length\_mm)) +   
 stat\_qq\_line() +  
 stat\_qq() +  
 labs(title = "QQ-plot for penguin flipper lengths",  
 x = "Theoretical normal quantiles",  
 y = "Empirical quantiles",  
 caption = "Data source: palmerpenguins R package")



➡️ State each assumption for the one-sample -test, and indicate whether or not each assumption is met, citing specific evidence from the output obtained.

Next, we implement the one-sample -test.

*# Implementing a one-sample t-test using R*  
ttestRes <- t.test(penguins$flipper\_length\_mm,  
 mu = 200, conf.level = 0.95)  
  
*# Printing model output*  
ttestRes %>%   
 broom::tidy() %>%   
 make\_flex()

| estimate | statistic | p.value | parameter | conf.low | conf.high | method | alternative |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 200.92 | 1.20 | 0.23 | 341.00 | 199.42 | 202.41 | One Sample t-test | two.sided |

➡️ State our p-value, decision, and conclusion in the context of the problem testing at the significance level, citing specific evidence from the obtained output.

➡️ Provide and interpret a confidence interval for the average flipper length of penguins from Palmer Archipelago, Antarctica.

➡️ Do the results of the -test and confidence interval align? Why or why not?

## Two-sample -test

The two sample -test is used when we want to test if the average response value differs between two groups. There are multiple two-sample tests available: the paired samples -test and the 2 independent samples -test (pooled or unpooled also called Welch’s -test). In this activity, we will focus on the unpooled two-independent samples version of the test.

We could use an independent two-sample -test to explore the following:

1. Is the average lead level (parts per billion) in drinking water the same for residential households in Flint, MI as in Grand Rapids, MI? 🫗
2. Does the average pay of data scientists differ from statisticians at a large corporation? 💸
3. Is the average duration of survival of clinical trial participants longer for those given treatment A compared to treatment B? 💊

➡️ What are the response and explanatory variables in each scenario, and what is the type of each variable being as specific as possible?

➡️ Comment on how we could obtain a random sample (or at least close to one) in each scenario.

### Hypotheses

A general form of the hypotheses for an independent two-sample -test is:

Testing equality of the two groups corresponds to (most common)

or can test if , instead of .

### Assumptions

* **Random sample** / independent observations for each group
* [**Normality assumption**](https://docs.google.com/document/d/1jVscyqIwKhuVfzlP60_8XNKoWQz7vvnN/edit#heading=h.1ksv4uv): We need that the sampling distribution of the difference in the sample means, , is normal. This can be met if we have normally distributed data for each group, or large enough sample sizes without extreme outliers for each group.
* **Homoskedasticity**: Equal variances is assumed for the pooled two-sample -test (yields greater statistical power when variances are truly equal), but not for the unpooled version (more flexible / less assumptions). In this class, ✨**we’ll always use the unpooled version** which does **not** require homoskedasticity✨.

### Fixed-effects model formulation

Linear model interpretation of two-sample -test:

* : represents observation of group
* : group-specific mean for group
* : normally distributed error term for observation of group

### Test statistic

Test statistic for Welch’s -test (unpooled):

* : sample mean of group
* : sample size of group
* : sample variance of group
* : null value (commonly 0)

### Checking assumptions

* **Independent observations**: random sample, not temporally correlated (related across time or repeated measures)
* [**Normality**](https://docs.google.com/document/d/1jVscyqIwKhuVfzlP60_8XNKoWQz7vvnN/edit#heading=h.30j0zll): Consider sample size and check histograms and QQ-plots. Check for outliers using box plots, summary statistics.

## Palmer penguins example: Two-sample -test

As an ecological researcher, we may be interested in whether or not penguins from the Biscoe island weigh the same as penguins from the Dream island, on average. To explore this question, let’s conduct a two-sample -test.

### Model statement

Our model statement is given by

where

* is the body mass (g) for the penguin from the Dream island
* is the body mass (g) for the penguin from the Biscoe island
* is the average body mass (g) of penguins from the Dream () and Biscoe island ().
* are the independent and normally distributed error terms

### Hypotheses

➡️ Provide a formal statement of our hypotheses for our question of interest.

➡️ What is the null value, , in this scenario?

➡️ What would the statement of hypotheses be for testing if Biscoe penguins have a greater average body mass in grams than penguins from the Dream island?

### Checking Assumptions

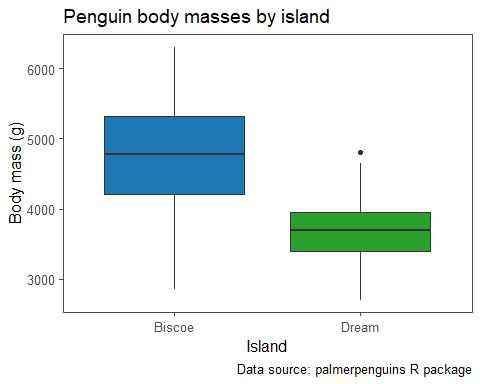
*# Calculating descriptive statistics*  
quant2Stats <- penguins %>%   
 dplyr::filter(island %in% c("Biscoe", "Dream")) %>%   
 group\_by(island) %>%   
 summarize(  
 Minimum = min(body\_mass\_g, na.rm = TRUE),  
 Q1 = quantile(body\_mass\_g, na.rm = TRUE, probs = 0.25),  
 M = median(body\_mass\_g, na.rm = TRUE),  
 Q3 = quantile(body\_mass\_g, na.rm = TRUE, probs = 0.75),  
 Maximum = max(body\_mass\_g, na.rm = TRUE),  
 Mean = mean(body\_mass\_g, na.rm = TRUE),  
 R = Maximum - Minimum,  
 s = sd(body\_mass\_g, na.rm = TRUE),  
 n = n()  
)  
  
*# Printing table of statistics*  
quant2Stats %>%   
 make\_flex(caption = "Summary statistics for penguin body masses (g) by species.")

*Table 3: Summary statistics for penguin body masses (g) by species.*

| island | Minimum | Q1 | M | Q3 | Maximum | Mean | R | s | n |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Biscoe | 2,850.00 | 4,200.00 | 4,775.00 | 5,325.00 | 6,300.00 | 4,716.02 | 3,450.00 | 782.86 | 168 |
| Dream | 2,700.00 | 3,400.00 | 3,687.50 | 3,956.25 | 4,800.00 | 3,712.90 | 2,100.00 | 416.64 | 124 |

*# Creating side-by-side box plots*  
penguins %>%   
 dplyr::filter(island %in% c("Biscoe", "Dream")) %>%   
 ggplot(aes(x = island, y = body\_mass\_g, fill = island)) +   
 geom\_boxplot() +   
 scale\_fill\_manual(values = c("#1F77B4", "#2CA02C")) +  
 labs(title = "Penguin body masses by island",  
 x = "Island",  
 y = "Body mass (g)",  
 caption = "Data source: palmerpenguins R package") +  
 theme(legend.position = "none")

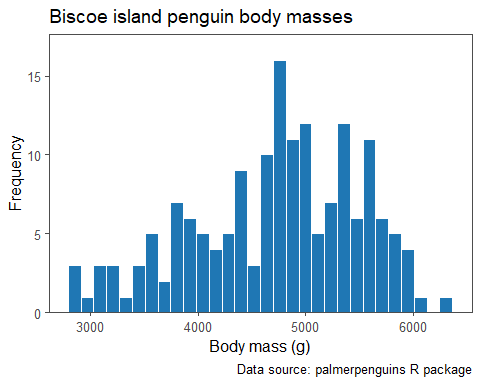
## Warning: Removed 1 rows containing non-finite values (`stat\_boxplot()`).



➡️ Describe what the box plots show us in terms of the center and symmetry of the body masses for both the Biscoe and Dream island penguins.

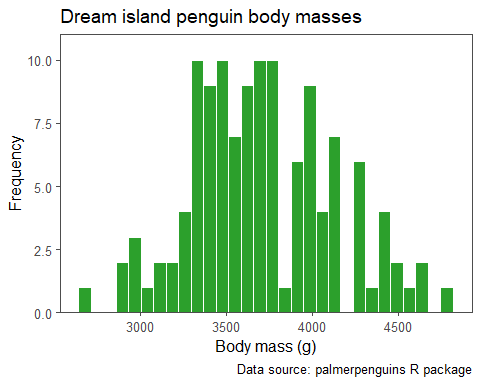
*# Creating a histogram*  
penguins %>%   
 dplyr::filter(island %in% c("Biscoe")) %>%   
 ggplot(aes(x = body\_mass\_g)) +   
 geom\_histogram(color = "white", fill = "#1F77B4") +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.10))) +  
 labs(title = "Biscoe island penguin body masses",  
 x = "Body mass (g)",  
 y = "Frequency",  
 caption = "Data source: palmerpenguins R package")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



*# Creating a histogram*  
penguins %>%   
 dplyr::filter(island %in% c("Dream")) %>%   
 ggplot(aes(x = body\_mass\_g)) +   
 geom\_histogram(color = "white", fill = "#2CA02C") +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.10))) +  
 labs(title = "Dream island penguin body masses",  
 x = "Body mass (g)",  
 y = "Frequency",  
 caption = "Data source: palmerpenguins R package")

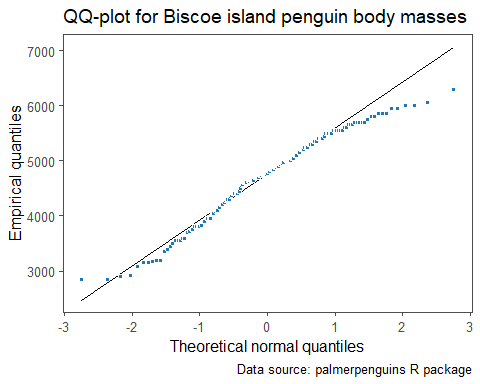
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



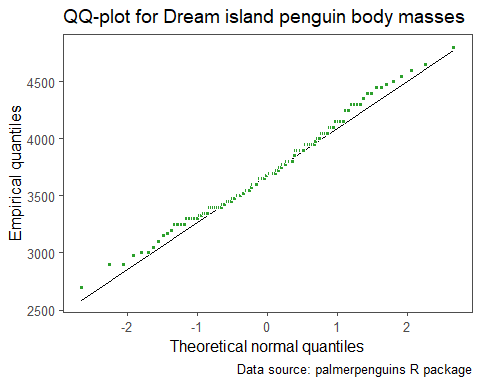
➡️ Describe what the histograms show us in terms of the body masses for the penguins.

*Histogram is unimodal and fairly symmetric.*

*# Creating a Quantile-Quantile (QQ) plot*  
penguins %>%   
 dplyr::filter(island %in% c("Biscoe")) %>%   
 ggplot(aes(sample = body\_mass\_g)) +   
 stat\_qq\_line() +  
 stat\_qq(color = "white", fill = "#1F77B4", pch = 21) +  
 labs(title = "QQ-plot for Biscoe island penguin body masses",  
 x = "Theoretical normal quantiles",  
 y = "Empirical quantiles",  
 caption = "Data source: palmerpenguins R package")



*# Creating a Quantile-Quantile (QQ) plot*  
penguins %>%   
 dplyr::filter(island %in% c("Dream")) %>%   
 ggplot(aes(sample = body\_mass\_g)) +   
 stat\_qq\_line() +  
 stat\_qq(color = "white", fill = "#2CA02C", pch = 21) +  
 labs(title = "QQ-plot for Dream island penguin body masses",  
 x = "Theoretical normal quantiles",  
 y = "Empirical quantiles",  
 caption = "Data source: palmerpenguins R package")



➡️ Provide our decision for the normality assumption for this data set, citing specific evidence from output obtained.

### Two-sample -test

Let’s conduct the unpooled independent samples -test:

*# Creating vectors of values*  
dreamMasses <- dplyr::filter(penguins, island == "Dream") %>%   
 dplyr::pull(body\_mass\_g)  
  
biscoeMasses <- dplyr::filter(penguins, island == "Biscoe") %>%   
 dplyr::pull(body\_mass\_g)  
  
*# Implementing the two-sample t-test*  
ttestRes2 <- t.test(x = dreamMasses,  
 y = biscoeMasses,  
 mu = 0,  
 conf.level = 0.95)  
  
*# Printing model output*  
ttestRes2 %>%   
 broom::tidy() %>%   
 make\_flex(ndigits = 1)

| estimate | estimate1 | estimate2 | statistic | p.value | parameter | conf.low | conf.high | method | alternative |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| -1,003.1 | 3,712.9 | 4,716.0 | -14.1 | <2e-16 | 264.8 | -1,143.3 | -862.9 | Welch Two Sample t-test | two.sided |

#### Understanding the output

➡️ What is our estimate of , denoted by , for this data set?

➡️ What would be our estimate of the average body mass in grams of all penguins across the Dream and Biscoe islands, assuming an equal number of penguins living on each island?

➡️ How many penguins are there in this data set from each island?

*n values from summary statistics*

➡️ What is ?

*s square ~ 174000 from summary statistics*

#### Implementing the test

➡️ Provide our test statistic, p-value, and decision at the significance level citing evidence from the output.

➡️ Provide our interpretation of the result of the two-sample -test based on our decision.

➡️ Provide and interpret the confidence interval for the difference in the means between the two groups.

➡️ Which is more useful for describing the difference, if any, between Dream island and Biscoe island penguins in terms of their average body masses - the two-sample -test or the confidence interval? Why?