

Chapter 8 Comparing Two Means: The t-test

Angelo LaCommare-Soto

2025-03-19

Detecting Sexual Dimorphism Using a t-test

Research Question 1: Is there a difference in mean body length between male and female *Tamias minimus*?

Section 1 - Analyzing *Tamias minimus* Digital Images

Completed - see 'LaCommare-SotoAngelo_chipmunk.csv' below.

Section 2 - Importing Data

```
# set working directory for all chunks in this file (default working directory is wherever Rmd file is)
getwd()
```

```
## [1] "C:/Users/Angelo L/Documents/GitHub/BIOL710/RCode710/RCode/working_directory"
```

```
library(tidyverse)
```

```
## Warning: package 'purrr' was built under R version 4.4.3
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.4      v readr      2.1.5
```

```
## v forcats   1.0.0      v stringr   1.5.1
```

```
## v ggplot2    3.5.1      v tibble    3.2.1
```

```
## v lubridate  1.9.4      v tidyr     1.3.1
```

```
## v purrr      1.0.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Importing data
```

```
chipmunk <- read.csv("LaCommare-SotoAngelo_chipmunk.csv",header=TRUE)
```

```
# viewing the structure of the data
```

```
str(chipmunk)
```

```
## 'data.frame': 30 obs. of 4 variables:
## $ image_number: int 1 1 2 2 3 3 4 4 5 6 ...
## $ specimen_id: int 408 425 416 421 420 423 426 8764 427 429 ...
## $ sex : chr "F" "F" "F" "F" ...
## $ length_cm : num 13.5 12.1 12.7 12.4 13.2 ...
```

```
head(chipmunk)
```

```
## image_number specimen_id sex length_cm
## 1 1 408 F 13.543
## 2 1 425 F 12.121
## 3 2 416 F 12.736
## 4 2 421 F 12.366
## 5 3 420 F 13.202
## 6 3 423 F 12.754
```

Question Answers

- H₀: There is no sexual dimorphism between body lengths of male and female *Tamias minimus*. H_A: There is sexual dimorphism between body lengths of male and female *Tamias minimus*.
- A two-sample t-test is most appropriate to answer the research question of 'Is there a difference in mean body length between male and female *Tamias minimus*?'.

Section 3 - Estimating the t-statistic

```
# Calculating important statistics for male and female group of chipmunks
male <- filter(chipmunk, sex == 'M')
m_mean <- mean(male$length_cm)
sd_male <- sd(male$length_cm)
m_df <- nrow(male)-1
```

```
female <- filter(chipmunk, sex == 'F')
f_mean <- mean(female$length_cm)
sd_female <- sd(female$length_cm)
f_df <- nrow(female)-1
```

```
# Calculating the pooled sample variance
psv <- ((m_df*(sd_male^2))+(f_df*(sd_female^2)))/(m_df+f_df)
```

```
# Calculating the Standard Error
SE_xm_xf <- sqrt(psv*((1/nrow(male))+(1/nrow(female))))
```

```
# Calculating the t-statistic
t <- (m_mean-f_mean)/SE_xm_xf
t
```

```
## [1] -4.033601
```

Question Answer

- a. When looking at the critical values of the single tail t-distribution in a statistical table (statsexamples.com), we find that for $df=28$ and a t-statistic of -4.034 , the p-value is less than 0.001 , which, adjusting for a two tails makes the p-value less than 0.002 . Therefore, we reject the null hypothesis ($p < 0.05$) and can state that there is a significant difference in mean body length between males and females. More specifically, females were longer than males.

Challenge - Check work with `t.test()`

```
t.test(male$length_cm, female$length_cm, alternative = "two.sided", var.equal = TRUE)

##
## Two Sample t-test
##
## data: male$length_cm and female$length_cm
## t = -4.0336, df = 28, p-value = 0.000384
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.0828239 -0.6798427
## sample estimates:
## mean of x mean of y
## 11.22860 12.60993
```

Section 4 - Visualizing the Data

```
# Creating a table of descriptive statistics using R package rstatix.
# installing the package
install.packages("rstatix", repos="http://cran.us.r-project.org")

## Installing package into 'C:/Users/Angelo L/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)

## package 'rstatix' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\Angelo L\AppData\Local\Temp\RtmpSaSwgs\downloaded_packages

# loading package
library(rstatix)

## Warning: package 'rstatix' was built under R version 4.4.3

##
## Attaching package: 'rstatix'

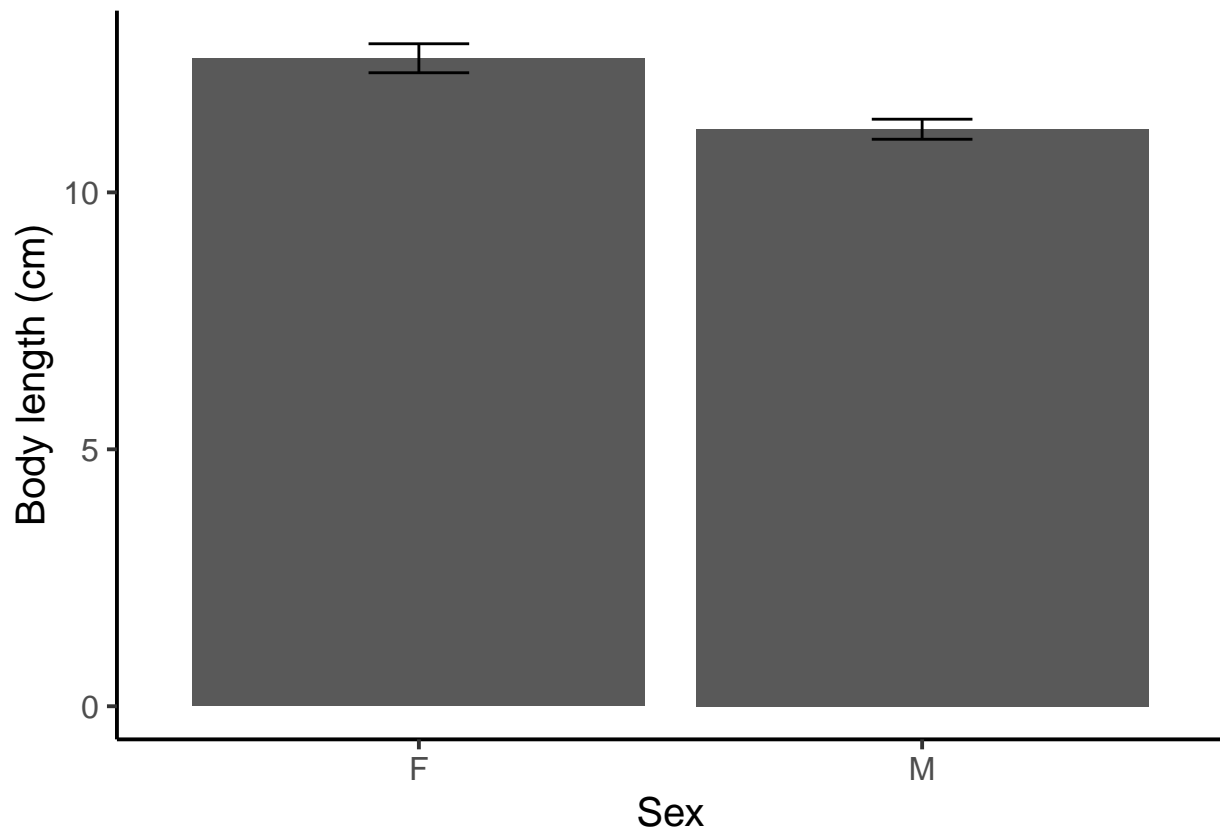
## The following object is masked from 'package:stats':
##
## filter
```

```

# For this, we use the function get_summary_stats() where the first argument is the response variable,
# summary statistics
sum <- chipmunk %>%
  group_by(sex) %>%
  get_summary_stats(length_cm)

# bar plot with error bars for standard error
p1 <- ggplot(sum, aes(x=sex, y=mean)) +
  geom_bar(stat="identity") +
  geom_errorbar(aes(min=mean-se, max=mean+se), width=0.2) +
  theme_classic(15) +
  ylab("Body length (cm)") +
  xlab("Sex")
p1

```



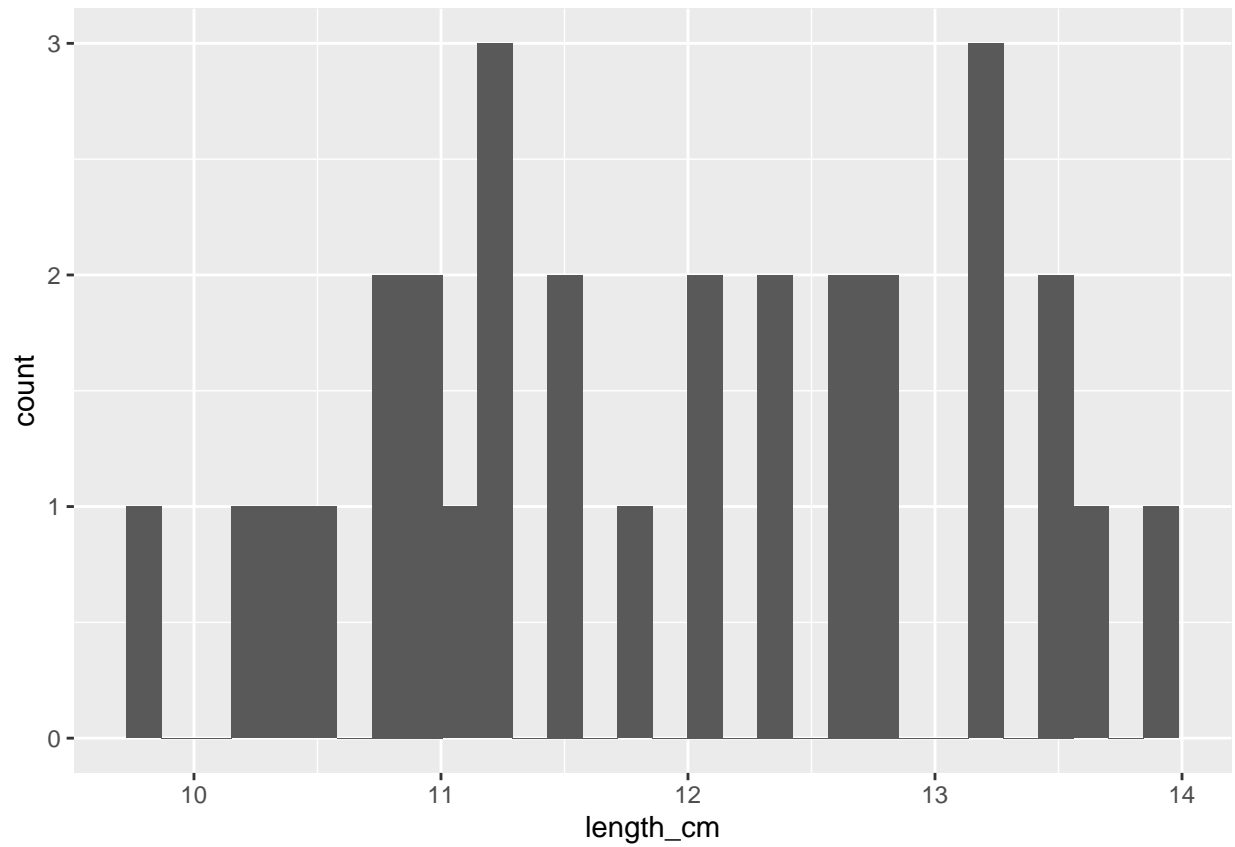
Section 5 - Checking the Model Assumptions of Normality

```

# histogram of all body lengths
p2 <- ggplot(chipmunk, aes(x=length_cm)) +
  geom_histogram()
p2

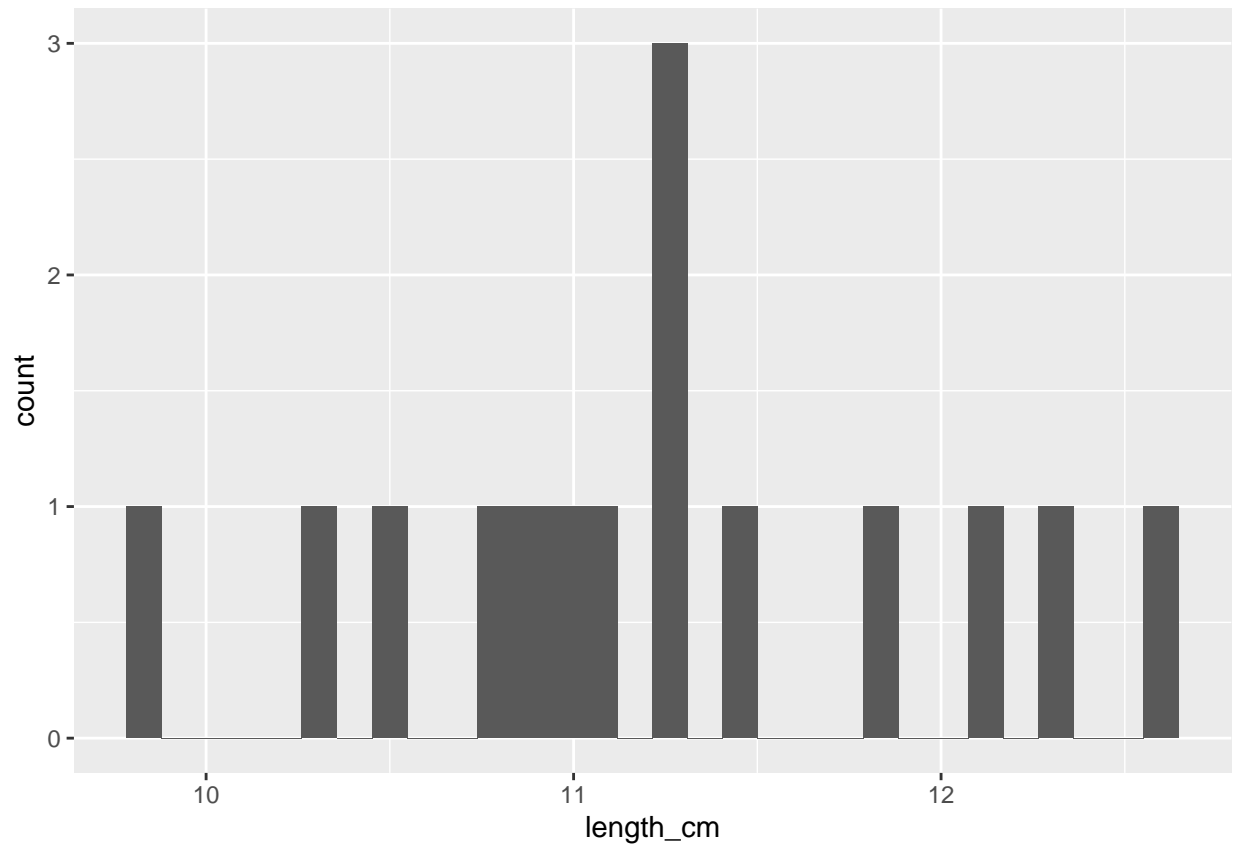
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



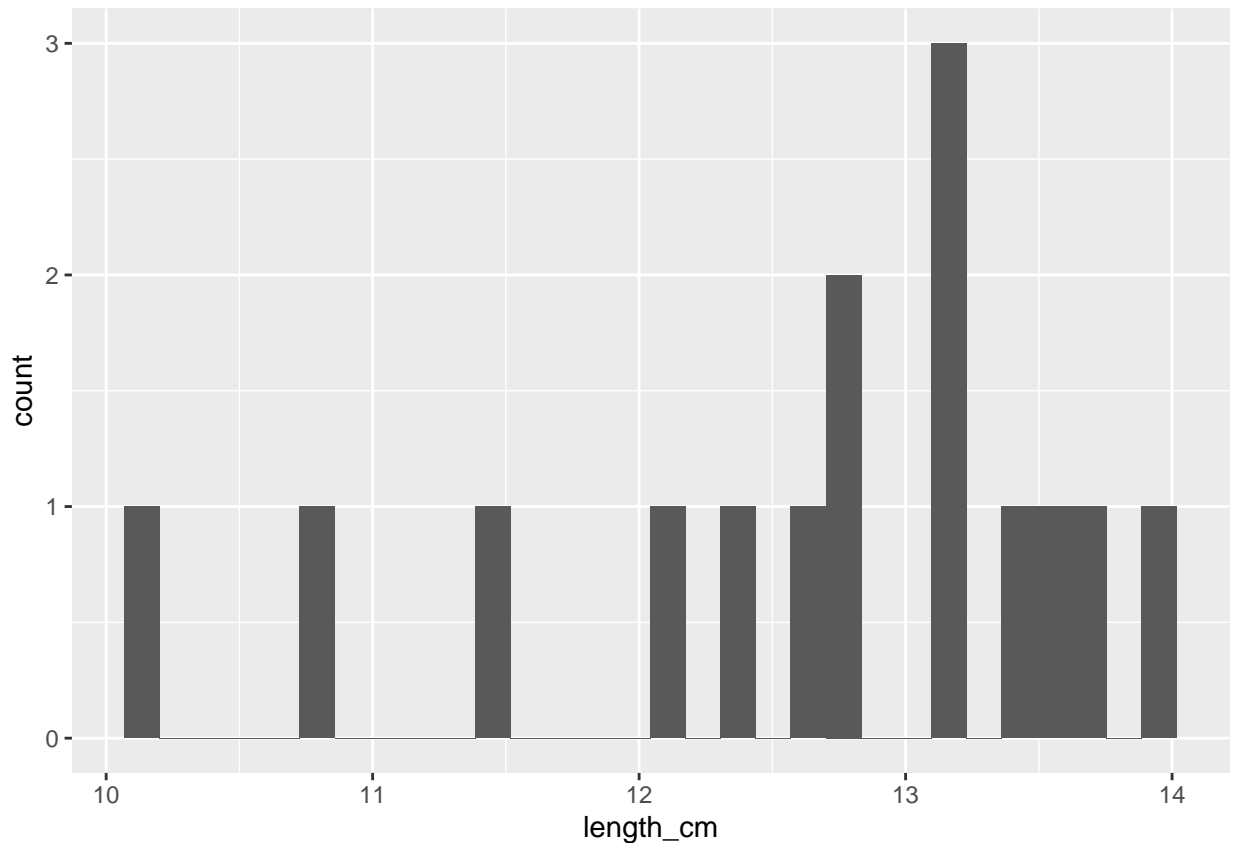
```
# histogram of male body lengths
p3 <- ggplot(male, aes(x=length_cm)) +
  geom_histogram()
p3
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



```
# histogram of female body lengths  
p4 <- ggplot(female, aes(x=length_cm)) +  
  geom_histogram()  
p4
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Question Answer

- a. The assumption of normality per sex group does hold for males, but not females. The male body length histogram has a clear and centered peak at around 11.3cm, and uniform tails. The female body length histogram has a slight left skew centered at about 13cm. However, given the small sample sizes, it is difficult to confidently determine normality.

Discussion Question Answers

- a. The t-statistic of -4.0336 with a df=28 indicates an extreme result that suggest a statistically significant difference between the mean body length of male chipmunk and the mean body length of female chipmunks.
- b. The 95% CI of -2.083 and -0.680 indicates a statistically significant difference between male and female body lengths since it does not include zero, as well as an estimate of the degree and direction of difference in mean body length between the sexes for the chipmunk population. In this case, we are 95% confident that female chipmunks are 0.680cm to 2.083cm larger than male chipmunks.
- c. If I was interested in testing whether the mean body length of my sample of 15 females is different from the known mean body length of another sampled population, I would use a one-sample t-test. This is because the known mean body length of a larger sample population that is more representative of the true body lengths would be a single known value to compare with my sample.