

Lab 4: Confidence Intervals and T-tests

Elizabeth Miles-Flynn

2026-02-11

Set Up

Data Tidying and Visualization

```
# load necessary data
urchins.csv <- read.csv(file = here("Week4/urchins.csv"),
                        stringsAsFactors = TRUE)
str(urchins.csv)

## 'data.frame':    12066 obs. of  23 variables:
##  $ YEAR      : int   2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 ...
##  $ MONTH     : int    10 10 10 10 10 10 10 10 10 10 ...
##  $ DATE      : Factor w/ 323 levels "2008-10-10","2008-10-22",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ SITE      : Factor w/ 5 levels "AQUE","CARP",...: 5 5 5 5 5 5 5 5 5 5 ...
##  $ TRANSECT  : int     3 3 3 3 3 3 3 3 3 3 ...
##  $ TREATMENT : Factor w/ 2 levels "ANNUAL","CONTROL": 2 2 2 2 2 2 2 2 2 2 ...
##  $ SP_CODE   : Factor w/ 2 levels "SFL","SPL": 1 1 1 1 1 1 1 1 1 1 ...
##  $ SIZE      : num    2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 ...
##  $ COUNT     : int    34 14 13 8 6 5 3 2 2 3 ...
##  $ AREA      : int    80 80 80 80 80 80 80 80 80 80 ...
##  $ NOTES     : Factor w/ 23 levels "", "2 swap 940: 18",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ SCIENTIFIC_NAME: Factor w/ 2 levels "Mesocentrotus franciscanus",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COMMON_NAME : Factor w/ 2 levels "Purple Urchin",...: 2 2 2 2 2 2 2 2 2 2 ...
##  $ TAXON_KINGDOM : Factor w/ 1 level "Animalia": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TAXON_PHYLUM  : Factor w/ 1 level "Echinodermata": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TAXON_CLASS   : Factor w/ 1 level "Echinoidea": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TAXON_ORDER   : Factor w/ 1 level "Camarodonta": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TAXON_FAMILY  : Factor w/ 1 level "Strongylocentrotidae": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TAXON_GENUS   : Factor w/ 2 levels "Mesocentrotus",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ GROUP         : Factor w/ 1 level "INVERT": 1 1 1 1 1 1 1 1 1 1 ...
##  $ SURVEY        : Factor w/ 1 level "URCHINS": 1 1 1 1 1 1 1 1 1 1 ...
##  $ MOBILITY      : Factor w/ 1 level "MOBILE": 1 1 1 1 1 1 1 1 1 1 ...
##  $ GROWTH_MORPH  : Factor w/ 1 level "SOLITARY": 1 1 1 1 1 1 1 1 1 1 ...

# Trim the original dataset so that only the Year, Month, Site, Treatment, Size, Count, and
# Common Name are included. Then, filter the dataset further to only include data from the
# Mohawk (MOHK) reef for red urchins in 2015 and 2016.
urchins_clean <- urchins.csv %>%
  select(YEAR, MONTH, SITE, TREATMENT, SIZE, COUNT, COMMON_NAME) %>%
  filter(SITE == "MOHK",
```

```

COMMON_NAME == "Red Urchin",
YEAR %in% c(2015, 2016))

# convert this data from a frequency table to a record of individual observations
# (hint: try the uncount() function in tidyr)
urchins_uncount <- urchins_clean %>%
  uncount(weights = COUNT, .id = "id")

## Bin sizes with Rice Rule  $2n^{(1/3)} = 18.2115$ , rounded to 18 per bin,  $n = 755$ 
 $2*755^{(1/3)}$ 

## [1] 18.2115

urchin_size <- ggplot(data = urchins_uncount, aes(x = SIZE, fill = TREATMENT)) +
  geom_histogram(bins = 18, color = "black") +
  facet_wrap(vars(YEAR)) +
  labs(x = "Size (cm)",
       y = "Red Urchin Count (n = 755)",
       title = "Size Distributions of Red Urchins in 2015 and 2016 Across
Different Treatments Within Mohawk Reef",
       caption = "Figure 1. Data derived from Santa Barbara Channel LTER
measuring red (Mesocentrotus franciscanus) sea urchins size frequency
from 2015 and 2016. Annual removal and Control (no removal) were two
treatments conducted at Mohawk Reef.")

urchin_size

```

Size Distributions of Red Urchins in 2015 and 2016 Across Different Treatments Within Mohawk Reef

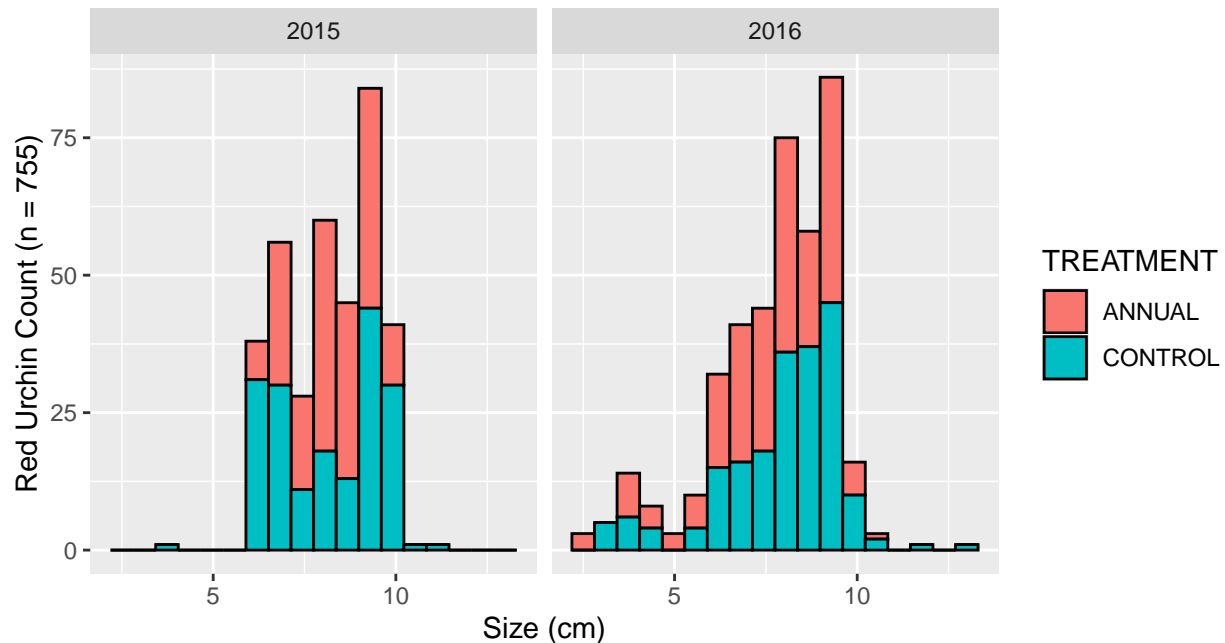


Figure 1. Data derived from Santa Barbara Channel LTER measuring red (*Mesocentrotus franciscanus*) sea urchins size frequency from 2015 and 2016. Annual removal and Control (no removal) were two treatments conducted at Mohawk Reef.

Confidence Intervals

```
urchin2015 <- urchins_uncount %>%
  filter(YEAR == 2015)

ttest2015 <- t.test(urchin2015$SIZE,
  conf.level = 0.95)

ttest2015
```

```
##
## One Sample t-test
##
## data: urchin2015$SIZE
## t = 131.47, df = 354, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 8.075943 8.321240
## sample estimates:
## mean of x
## 8.198592
```

```
urchin2016 <- urchins_uncount %>%
  filter(YEAR == 2016)
```

```
ttest2016 <- t.test(urchin2016$SIZE,
                    conf.level = 0.95)
ttest2016

##
## One Sample t-test
##
## data: urchin2016$SIZE
## t = 94.791, df = 399, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  7.569683 7.890317
## sample estimates:
## mean of x
##      7.73
```

Question 2 Answers

- 2a. Approximately 131 of the 95% CIs will include the true value of μ ($n = 355$).
- 2b. Approximately 94 of the 95% CIs will include the true value of μ ($n = 400$).

One-sample Hypothesis Tests

```
urchin2015_control <- urchins_uncount %>%
  filter(YEAR == 2015, TREATMENT == "CONTROL")

ttest2015_control <- t.test(urchin2015_control$SIZE,
                           mu = 9,
                           conf.level = 0.95,
                           alternative = "two.sided")
ttest2015_control
```

```
##
## One Sample t-test
##
## data: urchin2015_control$SIZE
## t = -7.9726, df = 179, p-value = 1.77e-13
## alternative hypothesis: true mean is not equal to 9
## 95 percent confidence interval:
##  7.977734 8.383377
## sample estimates:
## mean of x
##  8.180556
```

```
urchin2016_control <- urchins_uncount %>%
  filter(YEAR == 2016, TREATMENT == "CONTROL")

ttest2016_control <- t.test(urchin2016_control$SIZE,
                           mu = 9,
```

```

                                conf.level = 0.95,
                                alternative = "two.sided")
ttest2016_control

##
## One Sample t-test
##
## data: urchin2016_control$SIZE
## t = -9.5138, df = 199, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 9
## 95 percent confidence interval:
##  7.656909 8.118091
## sample estimates:
## mean of x
##      7.8875

```

Question 3 Answers

3. Null hypothesis is that urchin size mean is 9 cm. Alternative hypothesis is that urchin size mean is not 9 cm. Both 2015 AND 2016 urchin true size mean is not equal to 9 cm because we rejected the null hypothesis, since p-values were close to zero and smaller than the significance level of 0.05.

Two-Sample Hypothesis Tests

```
t.test(SIZE ~ TREATMENT, data = urchin2015)
```

```

##
## Welch Two Sample t-test
##
## data: SIZE by TREATMENT
## t = 0.29451, df = 313.51, p-value = 0.7686
## alternative hypothesis: true difference in means between group ANNUAL and group CONTROL is not equal
## 95 percent confidence interval:
## -0.2078411 0.2810157
## sample estimates:
## mean in group ANNUAL mean in group CONTROL
##      8.217143      8.180556

```

```
t.test(SIZE ~ TREATMENT, data = urchin2016)
```

```

##
## Welch Two Sample t-test
##
## data: SIZE by TREATMENT
## t = -1.938, df = 397.51, p-value = 0.05333
## alternative hypothesis: true difference in means between group ANNUAL and group CONTROL is not equal
## 95 percent confidence interval:
## -0.634535032 0.004535032
## sample estimates:
## mean in group ANNUAL mean in group CONTROL
##      7.5725      7.8875

```

```
urchin_control <- urchins_uncount %>%
  filter(TREATMENT == "CONTROL")

t.test(SIZE ~ YEAR, data = urchin_control)
```

```
##
## Welch Two Sample t-test
##
## data: SIZE by YEAR
## t = 1.8823, df = 375.86, p-value = 0.06056
## alternative hypothesis: true difference in means between group 2015 and group 2016 is not equal to 0
## 95 percent confidence interval:
## -0.01306923 0.59918034
## sample estimates:
## mean in group 2015 mean in group 2016
## 8.180556 7.887500
```

Question 4 Answers

- 4a. Null hypothesis is that for both treatments, ANNAUL and CONTROL, urchin size mean is the same. Alternative hypothesis is that for both treatments, urchin size mean is not equal. The p-value is 0.7686 (2015) and 0.05333 (2016) is larger than the significance level of 0.05. This means we fail to reject the null hypothesis for both years (2015 and 2016) and claim that the true size mean difference between treatments is the same.
- 4b. Null hypothesis is that urchin mean size is the same between two years. Alternative hypothesis is that urchin mean size is not equal between two years. The p-value is 0.06056 and greater than the significance level of 0.05. This means we fail to reject the null hypothesis and claim that the true urchin mean sizes in 2015 and 2016 are the same.