Lab 4 Assignment

Summer Heschong

2025-02-05

#Setup

```r
#load packages
library(here)
```

```
## here() starts at /Users/summerheschong/stats_spring25
```

```r
library(tidyverse)
```

```
## -- 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     1.0.2
```

```
## -- 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
```

```r
library(dplyr)
library(vcdExtra)
```

```
## Loading required package: vcd
## Loading required package: grid
## Loading required package: gnm
##
## Attaching package: 'vcdExtra'
##
## The following object is masked from 'package:dplyr':
##
##     summarise
```

```r
#load in data set
urchin_data <- read.csv(here('Data/Raw/urchins.csv'))
```

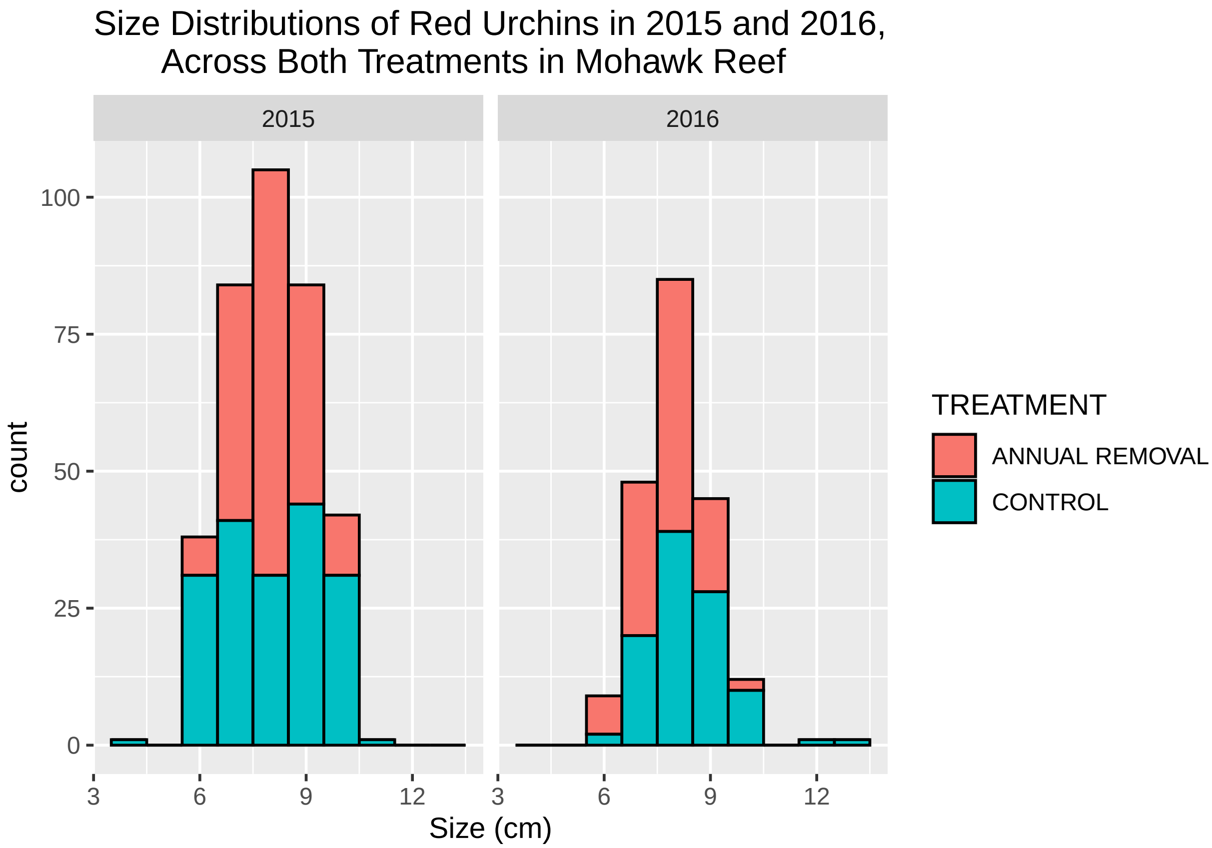
#1. Data Tidying and Visualization

##a. Trim and filter dataset and convert frequency table to record of individual observations

```r
#trim dataframe
urchin_data <- urchin_data %>%
  select(YEAR, MONTH, SITE, TREATMENT, SIZE, COUNT, COMMON_NAME)

#filter dataframe
red_urchin_data <- urchin_data %>%
  filter(SITE == 'MOHK' &
           COMMON_NAME == 'Red Urchin' &
           YEAR >=2015 & YEAR <= 2016)

#convert to record of individual observations
red_urchin_data <- expand.dft(red_urchin_data, freq = 'COUNT')

#filter and convert dataframe for purple urchins for later (question 2.b.)
purple_urchin_data <- urchin_data %>%
  filter(SITE == 'MOHK' &
           COMMON_NAME == 'Purple Urchin' &
           YEAR >=2015 & YEAR <= 2016)

purple_urchin_data <- expand.dft(purple_urchin_data, freq = 'COUNT')
```

##b. Visualize size distributions

```r
#create histogram
fig1 <- ggplot(red_urchin_data, aes(x = SIZE, fill = TREATMENT)) +
  geom_histogram(bins = 39, binwidth = 1, color = 'black') +
  facet_wrap(~YEAR) +
  labs(title = 'Size Distributions of Red Urchins in 2015 and 2016,
       Across Both Treatments in Mohawk Reef',
       x = 'Size (cm)')
fig1
```



#2. Confidence Intervals

##a. Calculate the 95%, two-sided CI for red urchin sizes in 2015 across both treatment types

```
red_urchin_test <-t.test(red_urchin_data$SIZE,
                         subset = red_urchin_data$YEAR == 2015)
red_urchin_test
```

```
##
##  One Sample t-test
##
## data:  red_urchin_data$SIZE
## t = 174.15, df = 555, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  8.136500 8.322133
## sample estimates:
## mean of x
##  8.229317
```

Answer: The 95% two-sided CI for red urchin sizes in 2015 across both treatment types is [8.1, 8.3]

##b. Calculate the 95%, two-sided CI for purple urchin sizes in 2016 across both treatment types

```
purple_urchin_test <- t.test(purple_urchin_data$SIZE,
                             subset = purple_urchin_data$YEAR ==2016)
purple_urchin_test
```

```
##
##  One Sample t-test
##
## data:  purple_urchin_data$SIZE
## t = 160.2, df = 616, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  5.364542 5.497695
## sample estimates:
## mean of x
##  5.431118
```

Answer: The 95% CI for purple urchin sizes in 2016 across both treatment types is [5.4, 5.5]

#3. One-sample Hypothesis Tests

##a. Is the mean red urchin size 9cm? Perform a one-sample test using 2015 data and one using 2016 data to investigate.

```r
#Null Hypothesis: the mean red urchin size is 9cm
#Alternative hypothesis: the mean red urchin size is not 9cm

#one-sample t-test for 2015
urchin_ttest_15 <- t.test(red_urchin_data$SIZE,
                         subset = red_urchin_data$YEAR==2015,
                         mu = 9,
                         alternative = 'two.sided')
urchin_ttest_15
```

```
##
##  One Sample t-test
##
## data:  red_urchin_data$SIZE
## t = -16.31, df = 555, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 9
## 95 percent confidence interval:
##  8.136500 8.322133
## sample estimates:
## mean of x
##  8.229317
```

```r
#one-sample t-test for 2016
urchin_ttest_16 <- t.test(red_urchin_data$SIZE,
                         subset = red_urchin_data$YEAR==2016,
                         mu = 9,
                         alternative = 'two.sided')
urchin_ttest_16
```

```
##
##  One Sample t-test
##
## data:  red_urchin_data$SIZE
## t = -16.31, df = 555, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 9
```

```
## 95 percent confidence interval:
##  8.136500 8.322133
## sample estimates:
## mean of x
##  8.229317
```

Answer: For both 2015 and 2016 the p value is less than 0.05 therefore we reject the null hypothesis and retain the alternative hypothesis. The mean red urchin size for 2015 and 2016 is not 9cm.

#4. Two-Sample Hypothesis Test

##a. In 2015 was the mean red urchin size significantly different between the two treatments? In 2016?

```
#Prepare data for test:

#1. Create a dataset for Control 2015
red_urchin_C15_data <- red_urchin_data %>%
  filter(TREATMENT == 'CONTROL'&
         YEAR == 2015)

#2. Create dataset for Annual Removal 2015
red_urchin_AR15_data <- red_urchin_data %>%
  filter(TREATMENT == 'ANNUAL REMOVAL'&
         YEAR == 2015)

#3. Test for equal variances in 2015
var.test(x = red_urchin_C15_data$SIZE,
         y = red_urchin_AR15_data$SIZE)
```

```
##
##  F test to compare two variances
##
## data:  red_urchin_C15_data$SIZE and red_urchin_AR15_data$SIZE
## F = 2.2319, num df = 179, denom df = 174, p-value = 1.556e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.659458 2.999889
## sample estimates:
## ratio of variances
##          2.231887
```

```
#Result: variances are not equal

#1.Create a dataset for Control 2016
red_urchin_C16_data <- red_urchin_data %>%
  filter(TREATMENT == 'CONTROL'&
         YEAR == 2016)

#2,Create dataset for Annual Removal 2016
red_urchin_AR16_data <- red_urchin_data %>%
  filter(TREATMENT == 'ANNUAL REMOVAL'&
         YEAR == 2016)

#3. Test for equal variances in 2016
```

```r
var.test(x = red_urchin_C16_data$SIZE,
         y = red_urchin_AR16_data$SIZE)
```

```
##
##  F test to compare two variances
##
## data:  red_urchin_C16_data$SIZE and red_urchin_AR16_data$SIZE
## F = 1.5464, num df = 100, denom df = 99, p-value = 0.03084
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.041326 2.295565
## sample estimates:
## ratio of variances
##           1.546409
```

```r
#Result: variances are not equal

#Perform 2-sample t-test:

#Null Hypothesis: the mean red urchin size is not significantly different
#between the two treatments
#Alternative Hypothesis: the mean red urchin size is significantly different
#between the two treatments

#2-sample t-test for 2015
urchin_2ttest_15 <- t.test(red_urchin_C15_data$SIZE,
                           red_urchin_AR15_data$SIZE)

urchin_2ttest_15
```

```
##
##  Welch Two Sample t-test
##
## data:  red_urchin_C15_data$SIZE and red_urchin_AR15_data$SIZE
## t = -0.29451, df = 313.51, p-value = 0.7686
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2810157  0.2078411
## sample estimates:
## mean of x mean of y
##  8.180556  8.217143
```

```r
#2-sample t-test for 2016
urchin_2ttest_16 <- t.test(red_urchin_C16_data$SIZE,
                           red_urchin_AR16_data$SIZE)

urchin_2ttest_16
```

```
##
##  Welch Two Sample t-test
##
## data:  red_urchin_C16_data$SIZE and red_urchin_AR16_data$SIZE
```

```
## t = 4.0132, df = 190.99, p-value = 8.597e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2768581 0.8120528
## sample estimates:
## mean of x mean of y
##  8.554455  8.010000
```

Answer: The mean red urchin size at Mohawk Reef for 2015 is not significantly different for the Control group (mean +- standard deviation = 8.18 +-1.38, n = 180) and the Annual Removal group (mean +- standard deviation = 8.21 +- 0.92, n = 175), as determined by a two-sample, two-sided t-test(t = -0.28, df = 313.51, p = 0.77, alpha = 0.05)

The mean red urchin size at Mohawk Reef for 2016 is significantly different for the Control group (mean +- standard deviation = 8.55 +-1.06, n = 101) and the Annual Removal group (mean +- standard deviation = 8.01 +- 0.85, n = 100), as determined by a two-sample, two-sided t-test(t = 4.01, df = 190.99, p = 8.60 e-5, alpha = 0.05)

##b. Was the mean red urchin size in the control data significantly different between the two years?

```r
#First test for equal variances
var.test(x = red_urchin_C15_data$SIZE,
         y = red_urchin_C16_data$SIZE)
```

```
##
##  F test to compare two variances
##
## data:  red_urchin_C15_data$SIZE and red_urchin_C16_data$SIZE
## F = 1.691, num df = 179, denom df = 100, p-value = 0.004139
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.184118 2.373493
## sample estimates:
## ratio of variances
##           1.691033
```

```r
#Result: variances are not equal

#2-sample t-test for control data:

#Null Hypothesis: The mean red urchin size in the control group is not significantly
#different between 2015 and 2016
#Alternative Hypothesis: The mean red urchin size in the control group
#is significantly different between 2015 and 2016

urchin_2ttest_C <- t.test(red_urchin_C15_data$SIZE,
                          red_urchin_C16_data$SIZE)
urchin_2ttest_C
```

```
##
##  Welch Two Sample t-test
##
## data:  red_urchin_C15_data$SIZE and red_urchin_C16_data$SIZE
## t = -2.5383, df = 252.7, p-value = 0.01174
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.66399679 -0.08380299
## sample estimates:
## mean of x mean of y
##  8.180556  8.554455
```

Answer: The mean red urchin size at Mohawk Reef for the Control Group is significantly different for 2015 (mean +- standard deviation = 8.18 +-1.38, n = 180) and 2016 (mean +- standard deviation = 8.55 +-1.06, n = 101), as determined by a two-sample, two-sided t-test(t = -2.54, df = 252.7, p = 0.01, alpha = 0.05)