Chapter 7

Angelo LaCommare-Soto

2025-03-12

Size Structure of Round Stingrays

Research Question 1: What is the size structure of the population of round stingrays in Seal Beach?

Section 1 - 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
ray <- read.csv("ray.csv",header=TRUE)

# viewing the structure of the data
str(ray)
```

```
## 'data.frame':    2427 obs. of  11 variables:
##  $ year      : int  2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...
##  $ month     : chr  "April" "April" "April" "April" ...
##  $ day       : int  7 7 7 7 7 7 7 7 7 7 ...
##  $ date      : chr  "2000-04-07" "2000-04-07" "2000-04-07" "2000-04-07" ...
```

```
##  $ sex       : chr  "M" "F" "M" "M" ...
##  $ tag       : int  4 5 6 8 11 12 13 15 16 18 ...
##  $ size_class: chr  "Medium" "Medium" "Large" "Medium" ...
##  $ disc_width: num  16.3 14.1 17.8 13.3 13.1 14.3 20.4 17.1 13.4 16.5 ...
##  $ length    : num  23 19 25 17.5 17 24 26 28 23 26 ...
##  $ location  : chr  "Seal Beach" "Seal Beach" "Seal Beach" "Seal Beach" ...
##  $ tide      : chr  "Low" "Low" "Low" "Low" ...
```

```
head(ray)
```

```
##   year month day       date sex tag size_class disc_width length   location
## 1 2000 April   7 2000-04-07   M   4     Medium       16.3   23.0 Seal Beach
## 2 2000 April   7 2000-04-07   F   5     Medium       14.1   19.0 Seal Beach
## 3 2000 April   7 2000-04-07   M   6      Large       17.8   25.0 Seal Beach
## 4 2000 April   7 2000-04-07   M   8     Medium       13.3   17.5 Seal Beach
## 5 2000 April   7 2000-04-07   F  11     Medium       13.1   17.0 Seal Beach
## 6 2000 April   7 2000-04-07   F  12     Medium       14.3   24.0 Seal Beach
##   tide
## 1  Low
## 2  Low
## 3  Low
## 4  Low
## 5  Low
## 6  Low
```

Question Answers

a. The 'ray' dataset has 2427 observations of 11 variables.
b. Given the question 'What is the size structure of the population of round stingrays in Seal Beach?', we are interested in the 'disc_width' numerical variable.

Section 2 - Estimating the Distribution of Body Sizes

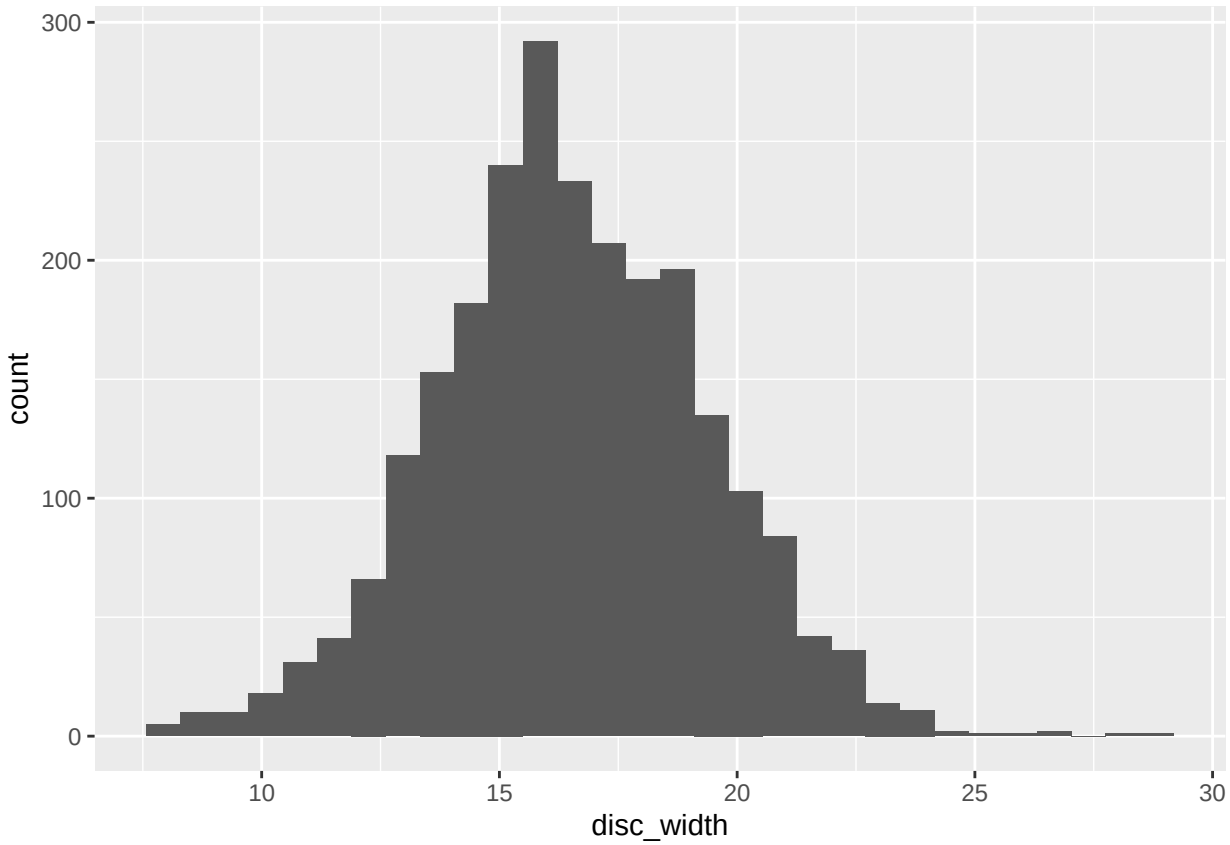
```
# summary of disc width
summary(ray$disc_width)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.60   14.60   16.40   16.49   18.50   28.50
```

Challenge: Create Two Useful Visualizations with this Data

```
# Creating a frequency distribution of sizes
p1 <- ggplot(ray, aes(x=disc_width)) + geom_histogram()
p1
```

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



Question Answers

a. The distribution of disc width for round stingray does appear to be normally distributed based on the above quick and dirty histogram.
b. However, in order to properly determine the normality of the data, we need to calculate some statistics and run some statistical tests that require the mean and standard deviation of the dataset.

Section 3 - Fit a Normal Distribution to the Disc Width Data

```
# mean disc width
m <- mean(ray$disc_width)
m
```

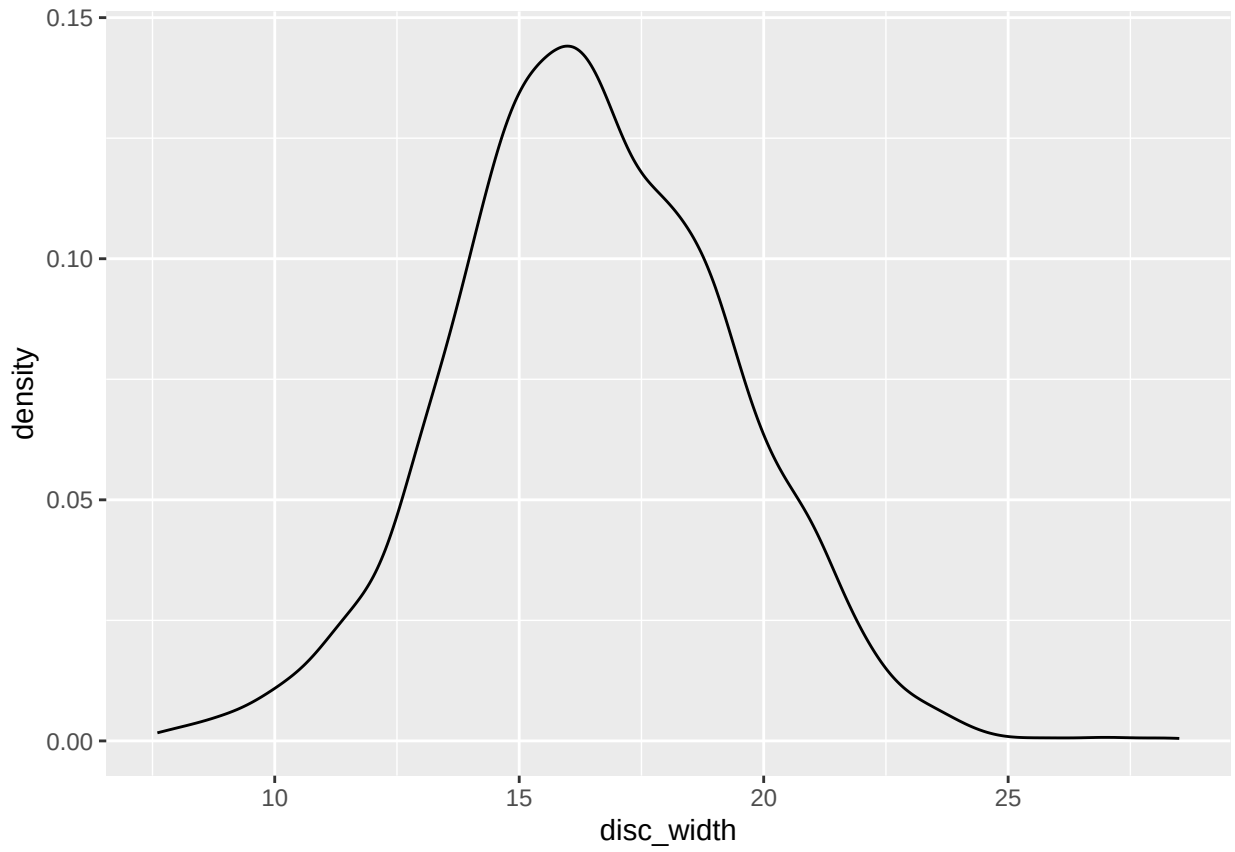
```
## [1] 16.49341
```

```
# standard deviation of disc width
sd <- sd(ray$disc_width)
sd
```

```
## [1] 2.830471
```

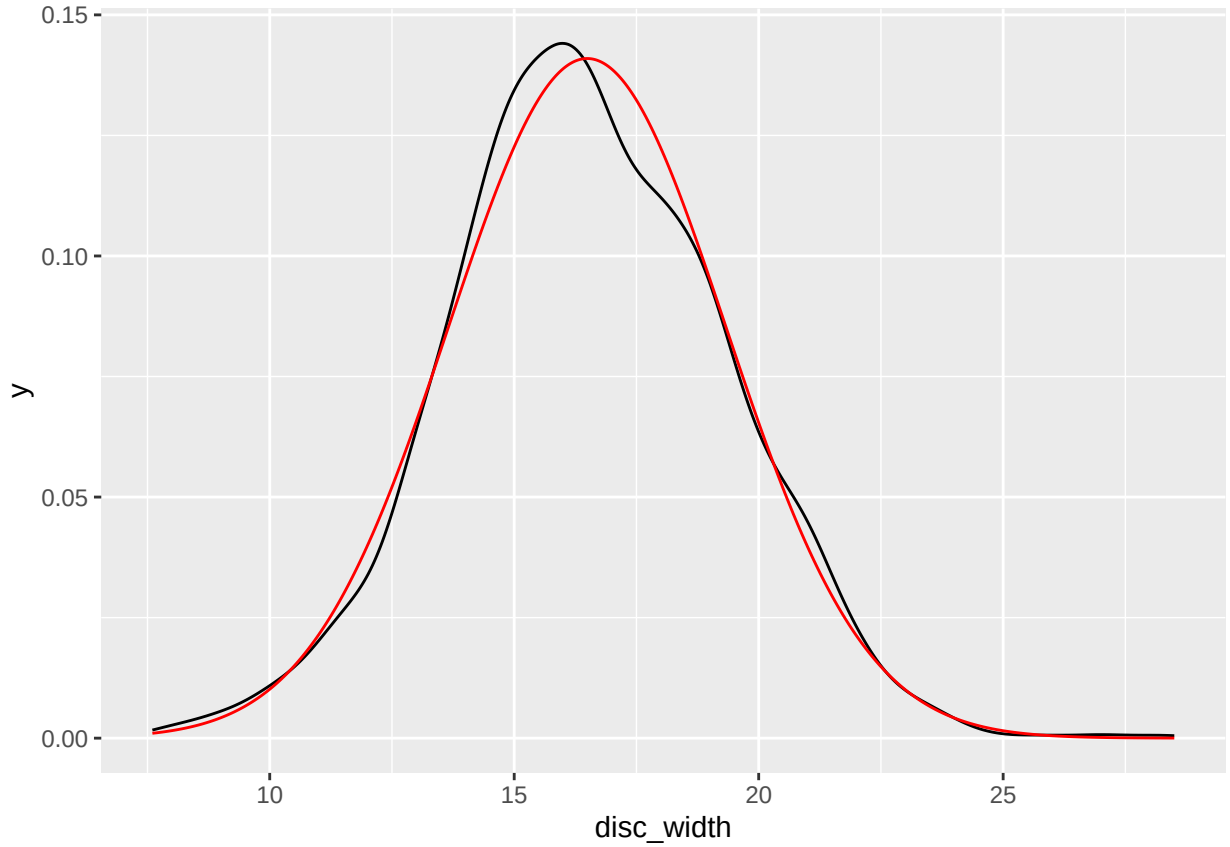
```r
# plotting the density probability of disc width
p2 <- ggplot(ray,aes(x=disc_width)) +
geom_density()

p2
```




```r
# fitting a normal distribution
p3 <- p2 + stat_function(fun = dnorm, n = 2427, args = list(mean = m, sd = sd),colour="red")

p3
```




```
# Quick calculations
m-sd
```

```
## [1] 13.66294
```

```
m+sd
```

```
## [1] 19.32388
```

```
m-2*sd
```

```
## [1] 10.83247
```

```
m+2*sd
```

```
## [1] 22.15435
```

```
SE <- sd/sqrt(nrow(ray))
SE
```

```
## [1] 0.05745446
```

Question Answers

a. Assuming a normal distribution, the disc width values from 13.66cm to 19.32cm make up 68.3% of the data.
b. Assuming a normal distribution, the disc width values from 10.83cm to 22.15cm make up 95% of the data.
c. The standard error (precision) for the mean disc width of round stingrays is 0.057cm.

Section 4 - Estimating the Standard Normal Distribution for Size

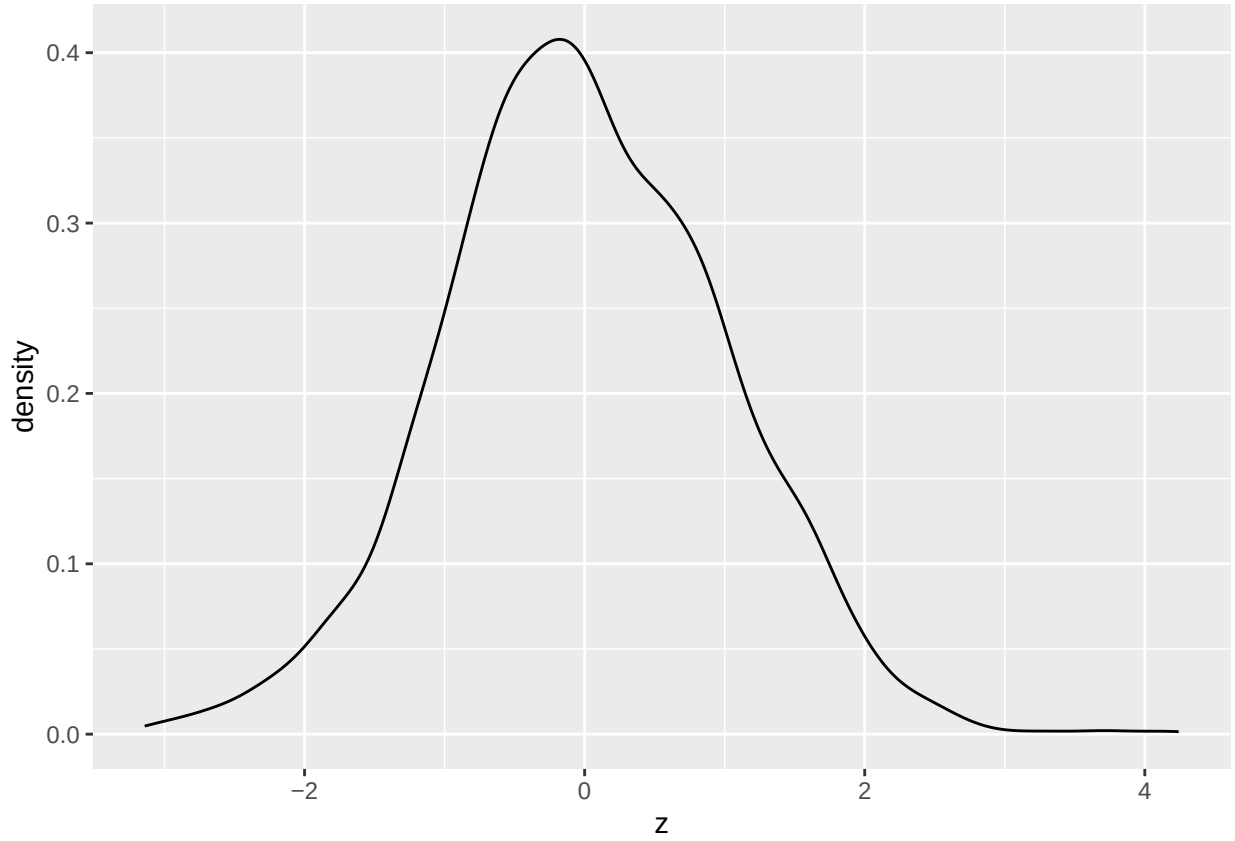
```
# creating a new row in the dataframe not using mutate
# estimating Z for disc width
ray$z <- (ray$disc_width-m)/sd

# checking the new column in "ray"
head(ray)
```

```
##   year month day       date sex tag size_class disc_width length   location
## 1 2000 April   7 2000-04-07   M   4     Medium       16.3   23.0 Seal Beach
## 2 2000 April   7 2000-04-07   F   5     Medium       14.1   19.0 Seal Beach
## 3 2000 April   7 2000-04-07   M   6      Large       17.8   25.0 Seal Beach
## 4 2000 April   7 2000-04-07   M   8     Medium       13.3   17.5 Seal Beach
## 5 2000 April   7 2000-04-07   F  11     Medium       13.1   17.0 Seal Beach
## 6 2000 April   7 2000-04-07   F  12     Medium       14.3   24.0 Seal Beach
##   tide           z
## 1  Low -0.06833051
## 2  Low -0.84558640
## 3  Low  0.46161669
## 4  Low -1.12822491
## 5  Low -1.19888454
## 6  Low -0.77492678
```

```
# plotting z
p4 <- ggplot(ray,aes(x=z)) +
geom_density()

p4
```




```
# probability of getting a disc width equal to or less than 13 cm under a normal distribution
p_small <- pnorm(13, mean = m, sd = sd)
p_small
```

```
## [1] 0.1085615
```

Question Answer

a. The probability of a randomly sampled round stingray having a disc width of 13cm or less is 0.109.

Stop, Think, Do: Estimate the probability that a randomly sampled round stingray in Seal Beach is small, medium, or large.

```
# Given that the maximum size considered small is 13cm, we already calculated the probability
#for randomly capturing a small round skate
# Checking the maximum size that the researchers categorize as medium
med_ray <- filter(ray, size_class == 'Medium')
head(med_ray)
```

```
##   year month day       date sex tag size_class disc_width length   location
## 1 2000 April   7 2000-04-07   M   4     Medium       16.3   23.0 Seal Beach
## 2 2000 April   7 2000-04-07   F   5     Medium       14.1   19.0 Seal Beach
```

```
## 3 2000 April   7 2000-04-07   M   8     Medium       13.3   17.5 Seal Beach
## 4 2000 April   7 2000-04-07   F  11     Medium       13.1   17.0 Seal Beach
## 5 2000 April   7 2000-04-07   F  12     Medium       14.3   24.0 Seal Beach
## 6 2000 April   7 2000-04-07   F  16     Medium       13.4   23.0 Seal Beach
##   tide           z
## 1  Low -0.06833051
## 2  Low -0.84558640
## 3  Low -1.12822491
## 4  Low -1.19888454
## 5  Low -0.77492678
## 6  Low -1.09289510
```

```
max(med_ray$disc_width)
```

```
## [1] 16.5
```

```
# Calculating the probability of a medium round stingray under a normal distribution
p_sm_md <- pnorm (16.5, mean = m, sd = sd)
p_medium = p_sm_md - p_small
p_medium
```

```
## [1] 0.3923677
```

```
# Calculating the probability of a large round stingray under a normal distribution
p_large <- pnorm (16.5, mean = m, sd = sd, lower.tail = FALSE)
p_large
```

```
## [1] 0.4990708
```

```
p_small + p_medium + p_large
```

```
## [1] 1
```

The following are the probabilities of capturing a round stingray of small, medium, and large sizes based on the normal distribution model: Small: 0.109 Medium: 0.392 Large: 0.499

Discussion Question Answers

a. The normal distribution is a probability distribution of a continuous variable with a symmetrical bell shape centered at the mean with a spread that is dictated by the standard deviation.
b. Conversely, the standard normal distribution is a transformed normal distribution for a random variable centered at a mean z-score of zero with a standard deviation of one. The standardization of a normal distribution via the conversion into a z-score allows for the comparison of different normal distributions on different scales.
c. Sample size does not determine individual z-scores. However, larger sample sizes increase the precision of the data by reducing the standard error.