

Chapter 2 Data Sampling, Accuracy, and Precision

Angelo LaCommare-Soto

2025-02-12

Section 1 - Setup and Image Analysis

```
# 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)
```

```
## -- 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
```

Section 2A - Data Analysis, Importing Data

```
# Importing data  
canine <- read.csv("LaCommare-SotoAngelo_bobcat.csv",header=TRUE)
```

```
# viewing the data  
canine
```

```
##   image_number specimen_id tooth_side length_cm  
## 1             1      Bcat_L21      left    1.437  
## 2             2      Bcat_L21     right    1.473  
## 3             3      Bcat_L21      left    1.443  
## 4             4      Bcat_L21     right    1.482  
## 5             5      Bcat_L21      left    1.455  
## 6             6      Bcat_L21     right    1.481
```

Question Answers

- Each column in the dataset 'canine' represents a variable for canine images.
- Each row in the dataset 'canine' represents an observation of a canine image.

Section 2B - Summarizing Data

```
# summarizing the data
s <- summary(canine)
s
```

```
## image_number specimen_id tooth_side length_cm
## Min. :1.00 Length:6 Length:6 Min. :1.437
## 1st Qu.:2.25 Class :character Class :character 1st Qu.:1.446
## Median :3.50 Mode :character Mode :character Median :1.464
## Mean :3.50 Mean :1.462
## 3rd Qu.:4.75 3rd Qu.:1.479
## Max. :6.00 Max. :1.482
```

Question Answers

- Each column in the summary represents a variable in the dataset 'canine'.
- I did get the expected sample size of three measurements per canine by viewing the data frame.

Section 2C - Summary in a Different Way

```
# minimum length value
m1 <- min(canine$length_cm)
m1
```

```
## [1] 1.437
```

```
# maximum length value
m2 <- max(canine$length_cm)
m2
```

```
## [1] 1.482
```

```
# mean length value
m3 <- mean(canine$length_cm)
m3
```

```
## [1] 1.461833
```

```
# median length value
m4 <- median(canine$length_cm)
m4
```

```
## [1] 1.464
```

```
# filter data for splitting data table between right and left canines
lc <- canine %>% filter (tooth_side == 'left')
lc
```

```
##   image_number specimen_id tooth_side length_cm
## 1           1      Bcat_L21      left      1.437
## 2           3      Bcat_L21      left      1.443
## 3           5      Bcat_L21      left      1.455
```

```
rc <- canine %>% filter (tooth_side == 'right')
rc
```

```
##   image_number specimen_id tooth_side length_cm
## 1           2      Bcat_L21      right      1.473
## 2           4      Bcat_L21      right      1.482
## 3           6      Bcat_L21      right      1.481
```

```
# minimum length value per side
m1_lc <- min(lc$length_cm)
m1_lc
```

```
## [1] 1.437
```

```
m1_rc <- min(rc$length_cm)
m1_rc
```

```
## [1] 1.473
```

```
# maximum length value per side
m2_lc <- max(lc$length_cm)
m2_lc
```

```
## [1] 1.455
```

```
m2_rc <- max(rc$length_cm)
m2_rc
```

```
## [1] 1.482
```

```
# mean length value per side
m3_lc <- mean(lc$length_cm)
m3_lc
```

```
## [1] 1.445
```

```
m3_rc <- mean(rc$length_cm)
m3_rc
```

```
## [1] 1.478667
```

```
# median length value per side
m4_lc <- median(lc$length_cm)
m4_lc
```

```
## [1] 1.443
```

```
m4_rc <- median(rc$length_cm)
m4_rc
```

```
## [1] 1.481
```

Section 2D - Estimating Accuracy and Precision (All)

```
# true population mean length of bobcat canines
TV=(1.3775+1.2315)/2
TV
```

```
## [1] 1.3045
```

```
# accuracy
a <- abs(m3-TV)
a
```

```
## [1] 0.1573333
```

```
# precision
p <- m2-m1
p
```

```
## [1] 0.045
```

Question Answers

- In general, my canine length measurements were inaccurate (average of 1.5 mm off of true mean), which is greater than a small tick-mark on the reference ruler.
- However, my canine length measurements were generally precise (max-min = 0.45 mm), which is lower than a small tick-mark on the reference ruler.

Section 2E - Estimating Accuracy and Precision (Left and Right)

```
# True population mean length of left canine for bobcats: 1.3775 cm
# accuracy for left canine
a_lc <- abs(m3_lc-1.3775)
a_lc
```

```
## [1] 0.0675
```

```
# precision for left canine
p_lc <- m2_lc-m1_lc
p_lc
```

```
## [1] 0.018
```

```
# True population mean length of left canine for bobcats: 1.2315 cm
# accuracy for right canine
a_rc <- abs(m3_rc-1.2315)
a_rc
```

```
## [1] 0.2471667
```

```
# precision for right canine
p_rc <- m2_rc-m1_rc
p_rc
```

```
## [1] 0.009
```

Question Answers

- My measurements for the left canine are more accurate than my measurements for the right canine. Overall, I would consider my left canine measurements accurate, as they are only 0.675 mm off of the true mean, on average, which is a small percentage of the mean length. My right canine measurements are not accurate, as they are about 2.472 mm off of the true mean, which is a larger percentage of the mean length.
- However, my measurements for the left canine are two times less precise (max-min = 0.18 mm) than my measurements for the right canine (max-min = 0.09 mm). Nonetheless, I would consider my measurements for both left and right canines to be precise.
- Some factors that may have impacted the accuracy of my measurements was the placement of the ruler relative to the canine. The physical distance between canine and ruler in the image may have caused the gradations in the ruler to get slightly distorted relative to the tooth. Some factors that may have impacted the precision of my measurements was arbitrarily deciding to what point to extend my measurement from the point of the canine.

Stop, Think, Do

This portion of exploring the length measurements of the right upper canine for my carnivore was done throughout the code above.

Discussion Question Answers

- Highly accurate measurements can have low precision if they contain lots of outliers, but a majority of measurements hover over the true mean.
- If I estimate a difference between the maximum and minimum value of 5 cm and my classmate estimates a difference of 2 cm, my classmate's measurements are more precise.
- A source of sampling error in this activity is the decision on how to measure the canine from the point to the skull. Since the tooth is slightly curved, I measured a slightly diagonal line. A source of bias in this activity is the decision of which gradation marks to use for setting a pixel-to-length scale.