13: INTRO TO THE BOOTSTRAP

Stat250 S25

Prof Amanda Luby

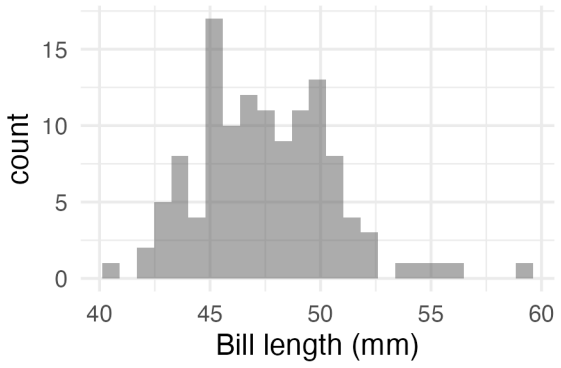
---

1 Roadmap

We want to develop an interval estimate of a population parameter:

1. *Exact method:* Find the sampling distribution in closed form (Ch 4). REquires knowledge of the distribution of the data
2. **Bootstrap Method**: Use the sample to approximate the population and simulate a sampling distribution (Ch5)
3. *Asymptotic method:* Use large-sample theory to approximate the sampling distribution (e.g., appeal to the CLT; Ch7)

2 Example: gentoo penguin bill length



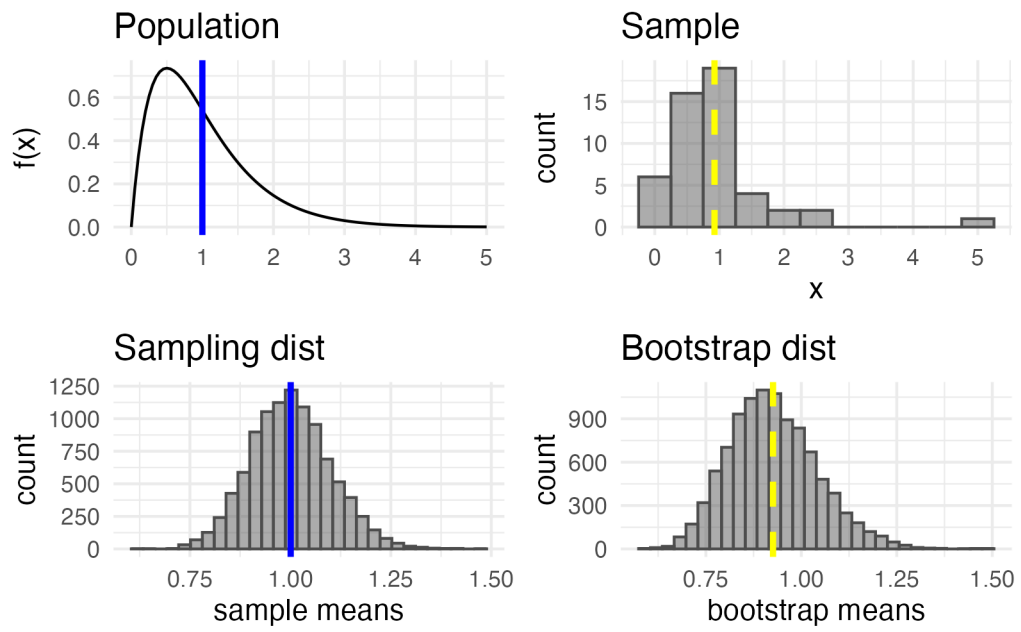
min	40.90
Q1	45.30
median	47.30
Q3	49.55
max	59.60
mean	47.50
sd	3.08
n	123.00
missing	1.00

2.1 The one-sample bootstrap algorithm

Given a sample of size n from a population,

1. Draw a resample of size n, **with replacement**, from the sample.
2. Compute the statistic of interest.
3. Repeat this resampling process (steps 1-2) many times, say 10,000.
4. Construct the bootstrap distribution of the statistic.

3 How does the bootstrap work?



	Mean	SD	Bias
Population	1	0.5	
Sample			
Sampling distribution			
Bootstrap distribution			

4 R Implementation

```
y <- gentoo$bill_length_mm # original sample
n <- nrow(gentoo)          # sample size
N <- 10^4                  # desired no. resamples
boot_means <- numeric(N)   # a place to store the bootstrap stats

# Resampling from the sample
for (i in 1:N) {
  x <- sample(y, size = n, replace = TRUE)
  boot_means[i] <- mean(x, na.rm = TRUE)  # you can choose other statistics
}
# Calculate a 95% percentile interval
quantile(boot_means, probs = c(0.025, 0.975))
```

```
    2.5%    97.5%
46.96693 48.04798
```