Bootstrap

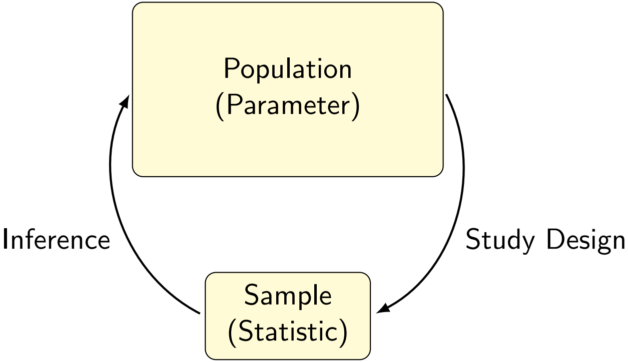
Grinnell College

March 1, 2024

Review

- **Standard deviation** (σ) and **standard eror** ($\sigma/\sqrt{n}$)
- A **sampling distribution**
- Point Estimate $\pm$ Margin of Error and **68-95-99 rule**
- A **confidence interval** is an interval with the properties that:
  - It is constructed according to a procedure or set of rules
  - It is intended to give plausible range of values for a *parameter* based on a *statistic*
  - It has no probability; the interval either contains the true value or it does not

The Statistical Framework



Repeated Samples

The confidence intervals we constructed of the form:

$$\text{Point Estimate} \pm \text{Margin of Error}$$

- Relied on assumptions (TBD) about our *sampling process*
- Examined what might happen if we could repeat sampling ad infinitum

There are, naturally, some limitations:

- We are limited to collecting a single sample
- Our assumptions may be tenuous

It would be helpful to have a more general method of constructing intervals with similar properties we had before

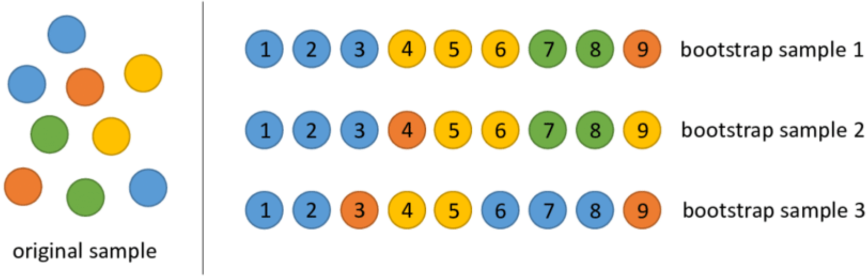
Bootstrapping

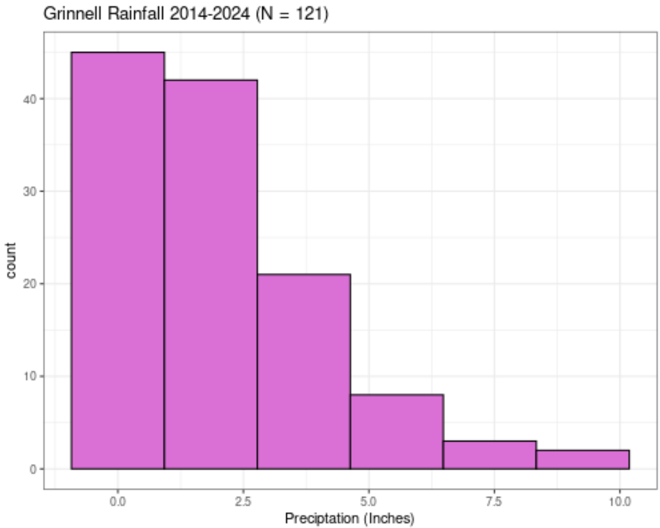
Somewhat amazingly, we can get around this problem with a technique known as **bootstrapping**

Instead of drawing more samples from our *original population*, we treat our sample as an *estimate* of the population and instead draw bootstrapped samples from our original sample

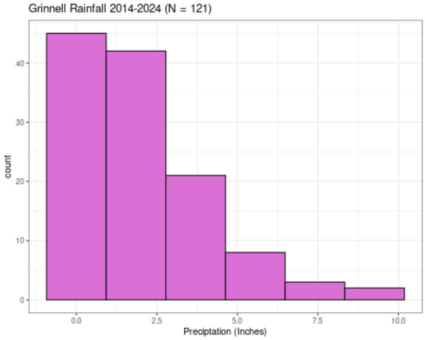
Bootstrapping

"Pick yourself up from bootstraps"

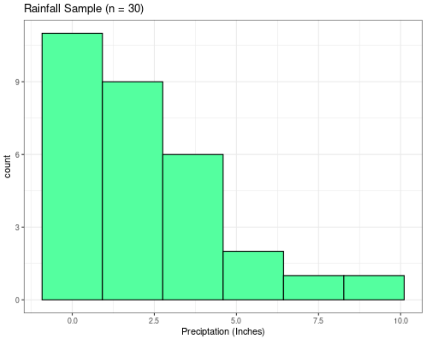




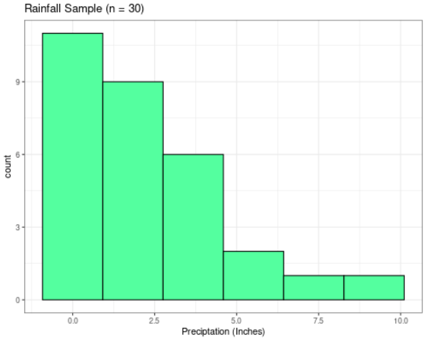
Grinnell Rainfall 2014-2024 (N = 121)
count
40
30
20
10
0
0.0
2.5
5.0
7.5
10.0
Preciptation (Inches)



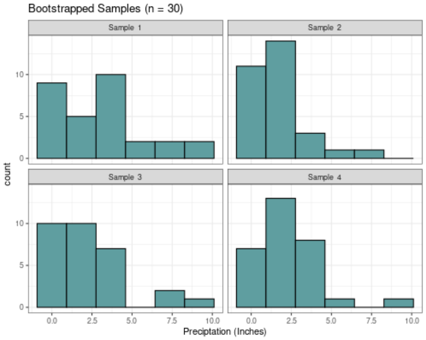
Grinnell Rainfall 2014-2024 (N = 121)
count
0
10
20
30
40
0.0
2.5
5.0
7.5
10.0
Preciptation (Inches)



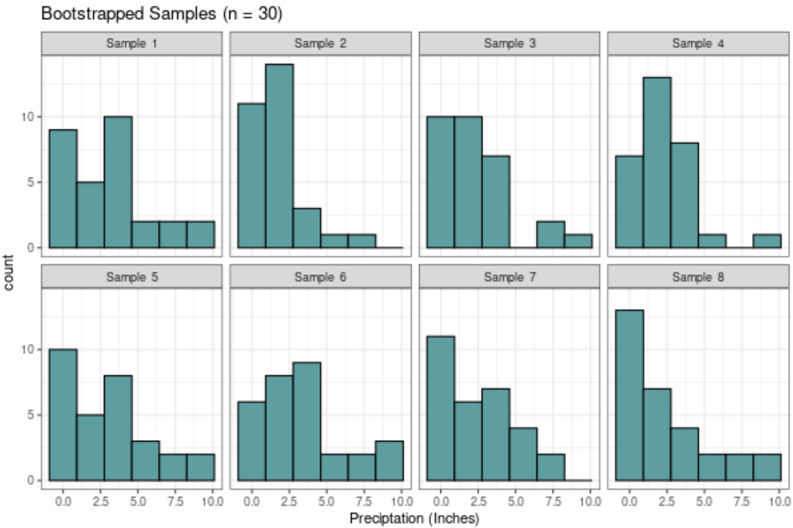
Rainfall Sample (n = 30)
count
0
3
6
9
0.0
2.5
5.0
7.5
10.0
Preciptation (Inches)

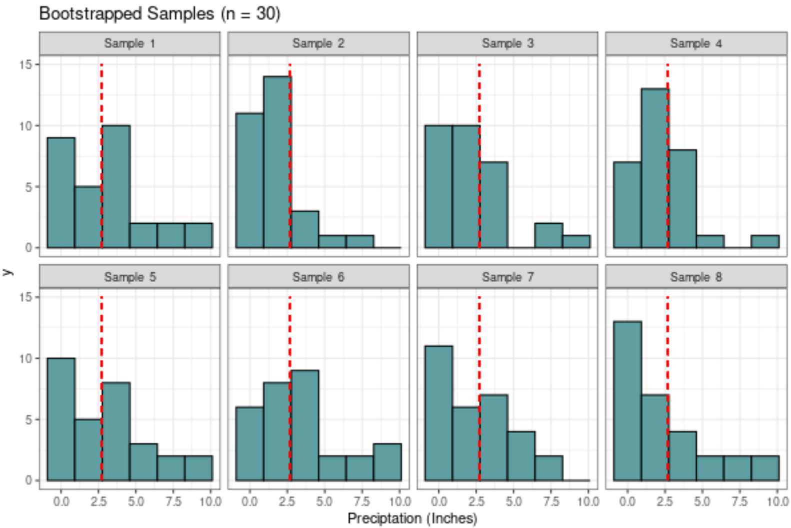


Rainfall Sample (n = 30)
count
9
6
3
0
0.0
2.5
5.0
7.5
10.0
Preciptation (Inches)



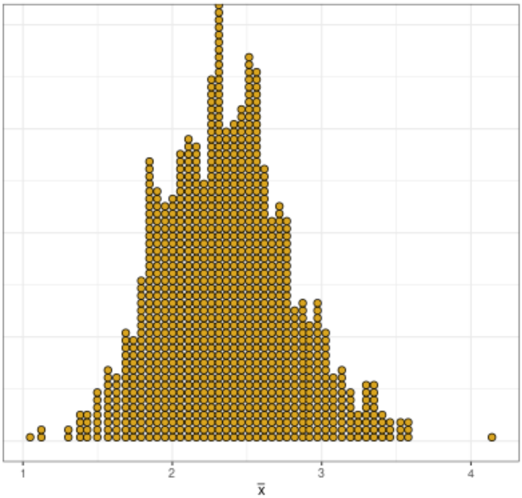
Bootstrapped Samples (n = 30)
Sample 1
Sample 2
Sample 3
Sample 4
count
10
5
0
0.0
2.5
5.0
7.5
10.0
Preciptation (Inches)


Bootstrapped Samples (n = 30)
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
Sample 7
Sample 8
count
10
5
0
0.0
2.5
5.0
7.5
10.0
Preciptation (Inches)



Bootstrapped Samples (n = 30)
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
Sample 7
Sample 8
y
0
5
10
15
0.0
2.5
5.0
7.5
10.0
Preciptation (Inches)

Bootstrapped Sample Means



Bootstrap Sampling Distribution

This sampling distribution is *identical* to what we should expect to find if we were to collect 1,000 random samples and compute the mean

That it looks the same is a consequence of finding the *mean*, not of the bootstrap process itself

While we could use the Point Estimate $\pm$ Margin of Error method here, we have something a little more robust

Percentiles

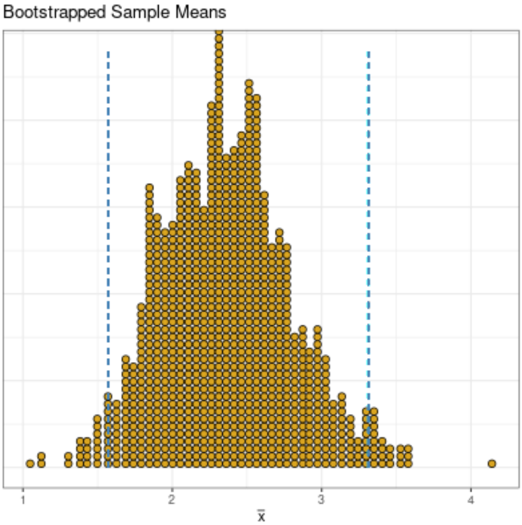
Recall what *percentiles* tell us about the distribution of data

- Median
- IQR

In that same fashion, we can find a 95% confidence interval by noting the percentiles 2.5% and 97.5%

The range between these will constitute the middle 95% of our sampling distribution

95% Confidence Intervals



Review

Bootstrapping involves the process of *resampling with replacement* from our original sample

When we compute a statistic on our bootstrapped sample (i.e., sample mean), we have a *bootstrapped sample statistic*

Repeating this process many many times gives us an estimate of the *sampling distribution*

Percentiles can be used on this bootstrapped sampling distribution without needing any further assumptions