STAT 201: Week 3

Estimating the Sampling Distribution with the Bootstrap

Vincenzo Coia
Monday, January 25, 2021

Summary of what we've learned so far

A central goal of this course:

Estimate a population parameter, ***along with how certain you are****.*

The sampling distribution shows us:

1. What point estimates are possible (even more: their probabilities of occurring, too)
2. Where the true parameter is (in most cases, lies at the mean of the sampling distribution)
    - More on this in a future week.

Summary of what we've learned so far

True or False:

Given a population and a sample size *n*, there is only one possible sampling distribution of a point estimator (e.g., doesn't depend on the actual sample drawn).

Summary of what we've learned so far

True or False:

Given a population and a sample size *n*, there is only one possible sampling distribution of a point estimator (e.g., doesn't depend on the actual sample drawn).

True!

Larger *n* implies:

1. narrower sampling distribution, and

2. for most point estimators (mean, median, variance, quantiles, etc.), the sampling distribution becomes more bell-shaped.

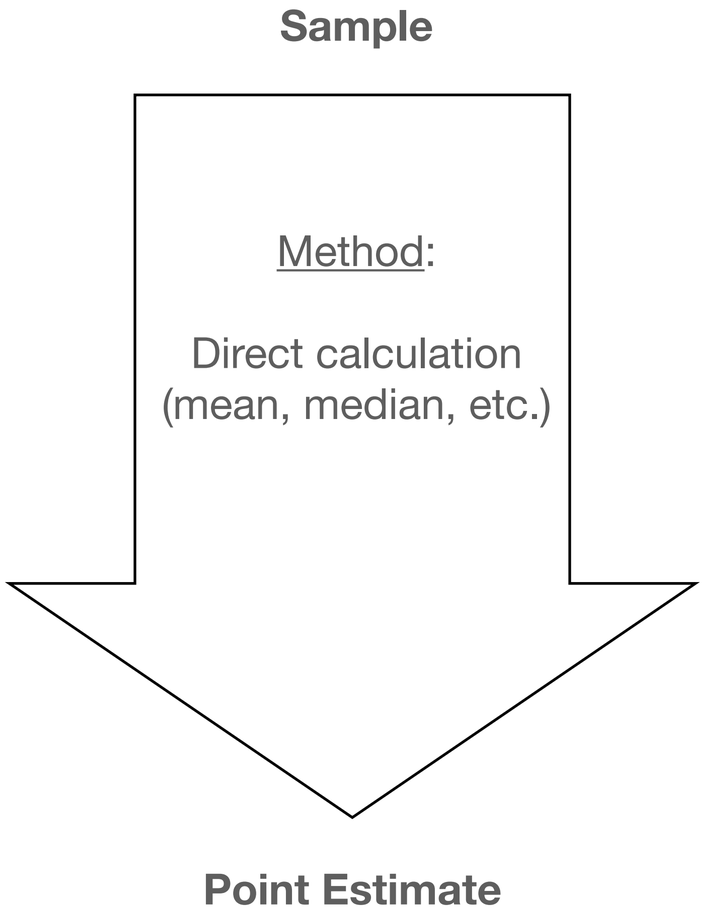
Motivation for Today's Topic

- The sampling distribution is never known in practice, just as the population parameter is unknown.
- If we knew the sampling distribution, there would be no need to estimate the population parameter. Why?

Motivation for Today's Topic

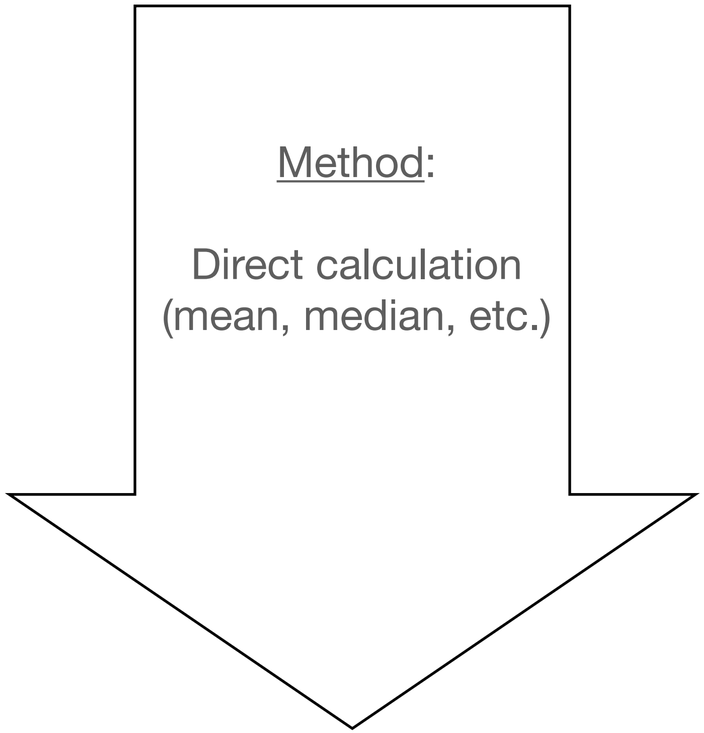
- The sampling distribution is never known in practice, just as the population parameter is unknown.
- If we knew the sampling distribution, there would be no need to estimate the population parameter. Why?
- ...Because we can extract the population parameter from the sampling distribution, and report 100% certainty.
- Today: estimate the sampling distribution, *and interpret it differently from the actual sampling distribution.*

Point Estimation



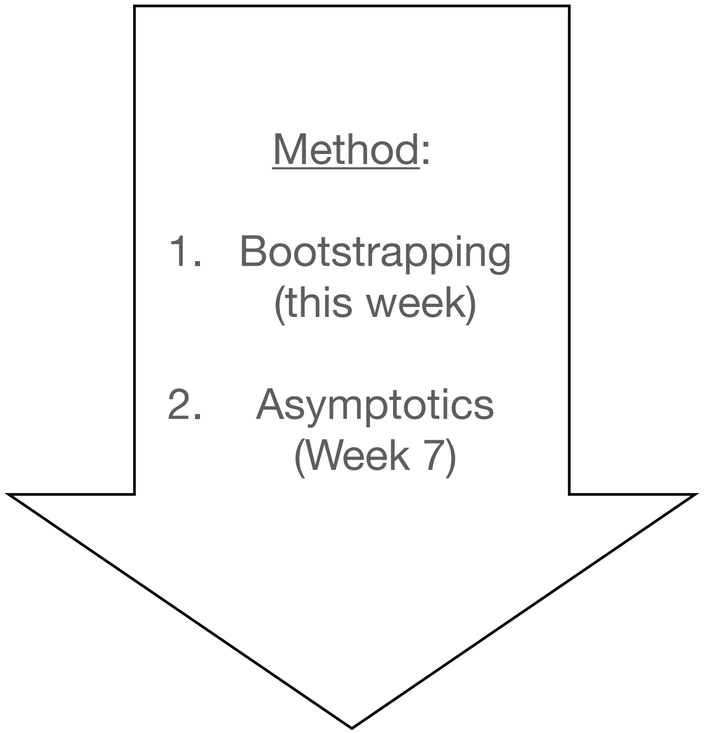
Point Estimation vs. Sampling Distribution Estimation

Sample



Point Estimate

Sample

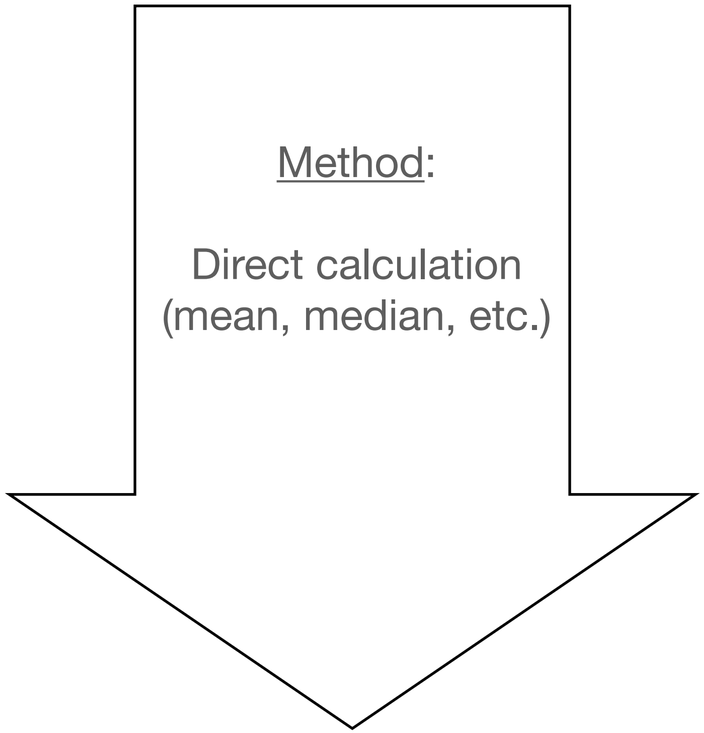


Sampling distribution estimate

Specifically:
Bootstrap distribution
(when Method 1 is used)

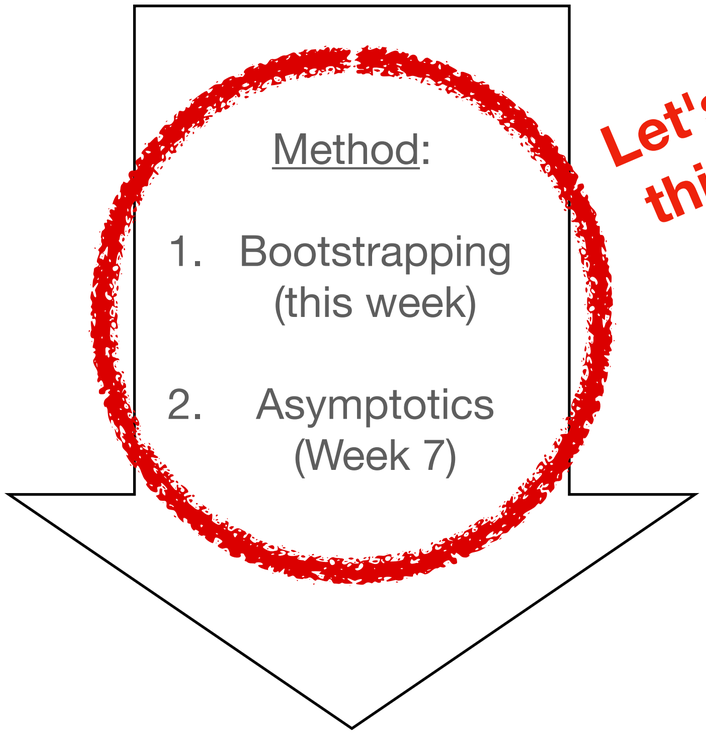
Point Estimation vs. Sampling Distribution Estimation

Sample



Point Estimate

Sample



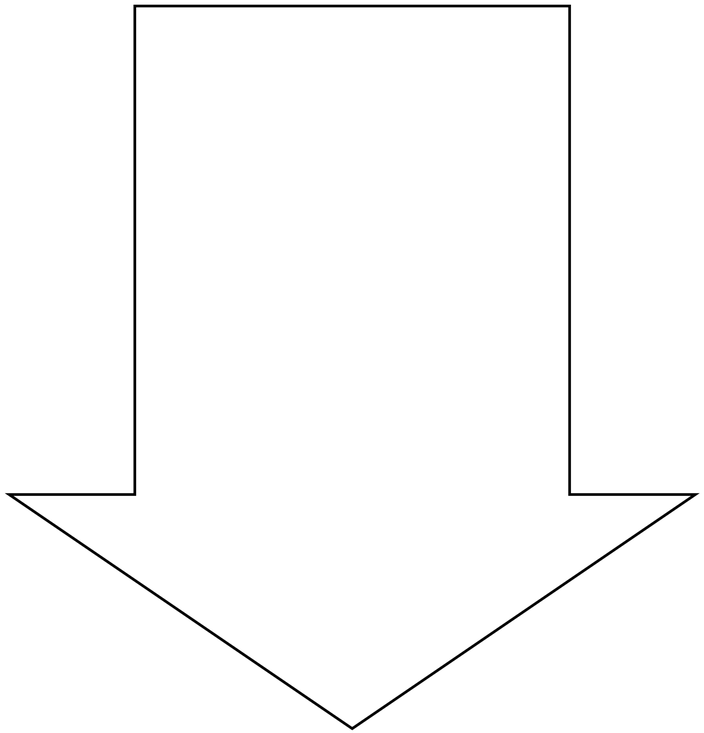
Let's fill this in!

Sampling distribution estimate

Specifically:
Bootstrap distribution
(when Method 1 is used)

Bootstrap Estimation Procedure

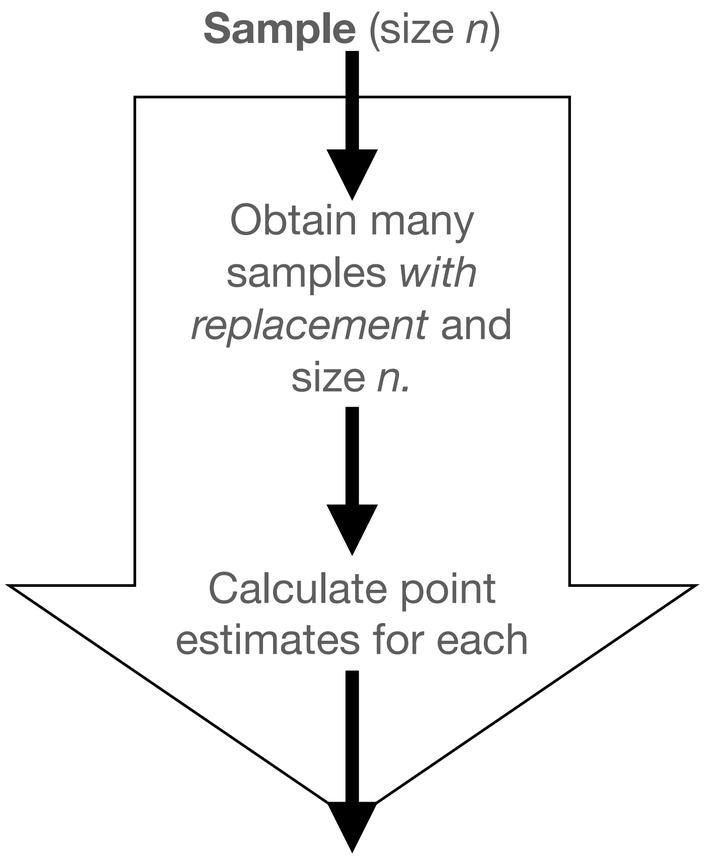
Sample (size *n*)



Sampling distribution estimate

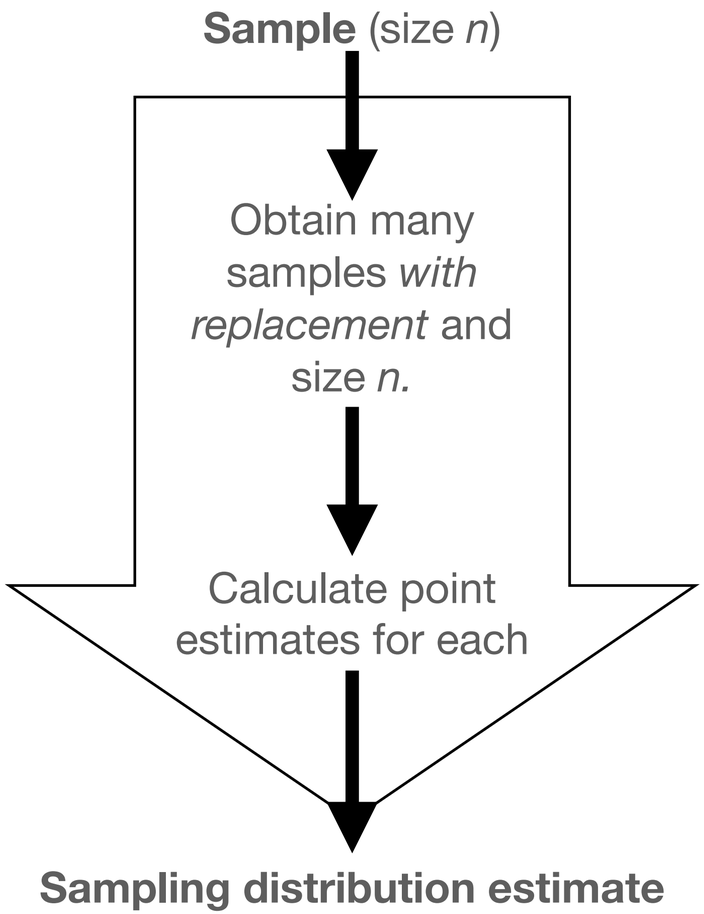
Specifically:
Bootstrap distribution
(when Method 1 is used)

Bootstrap Estimation Procedure



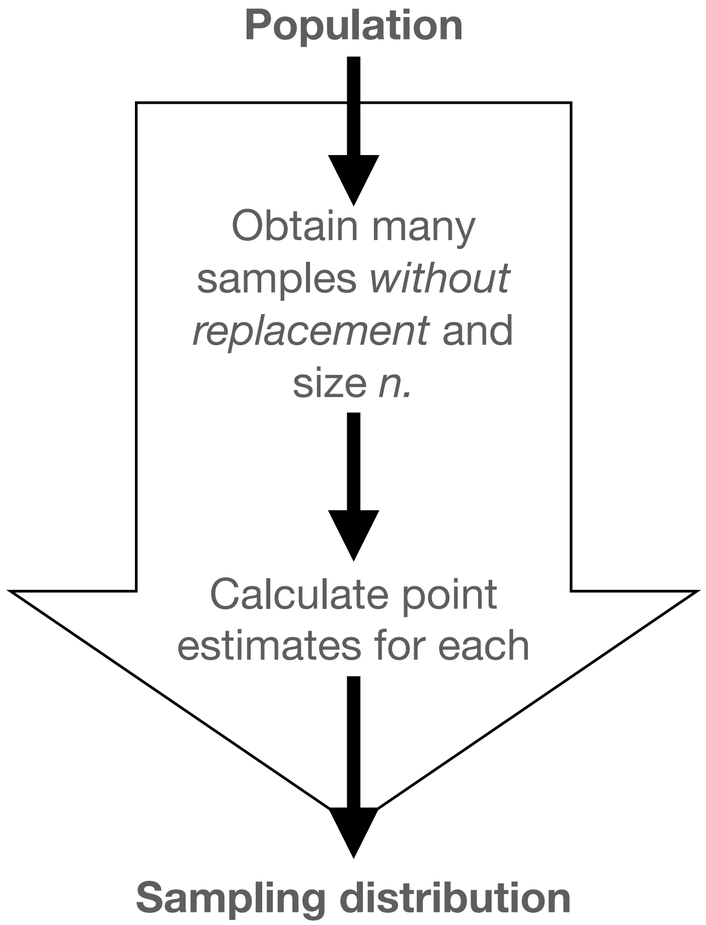
Specifically:
Bootstrap distribution
(when Method 1 is used)

Bootstrap Estimation Procedure

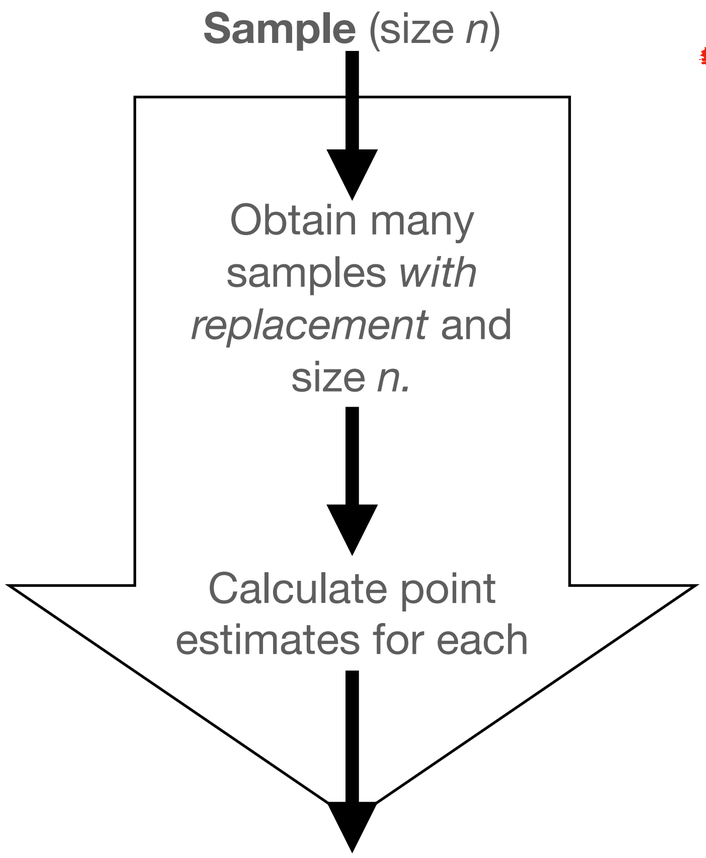


Specifically:
Bootstrap distribution
(when Method 1 is used)

Compare: Sampling Distribution Procedure



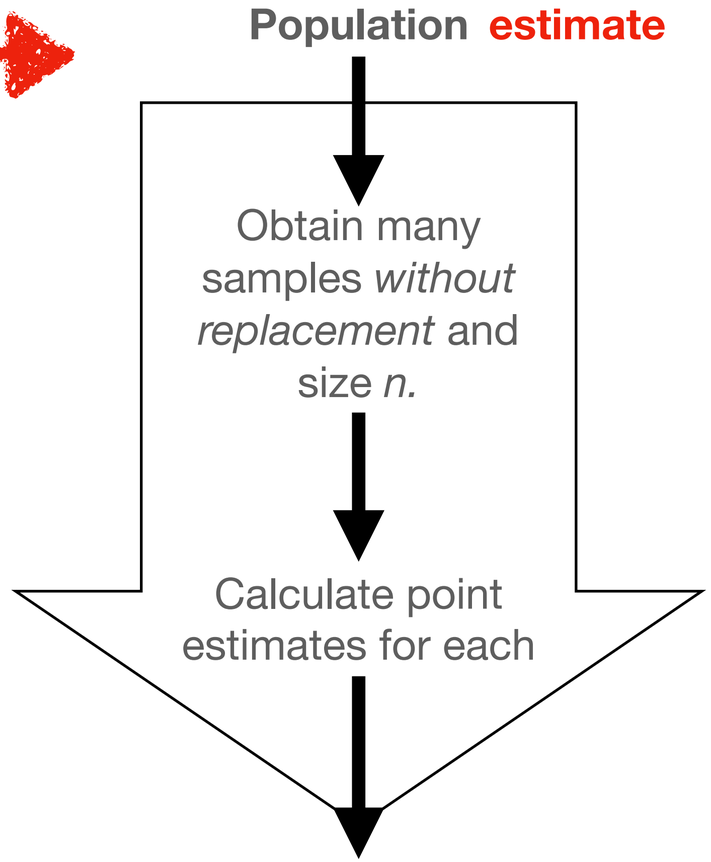
Justification for the Bootstrap Procedure



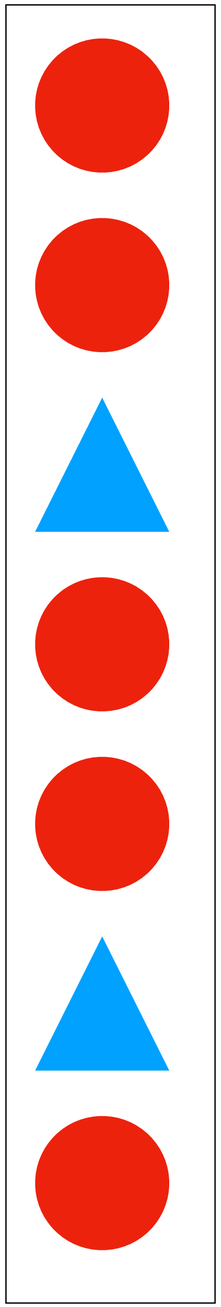
Sampling distribution estimate

Specifically:
Bootstrap distribution
(when Method 1 is used)

Approximate the population by repeating the sample many times.

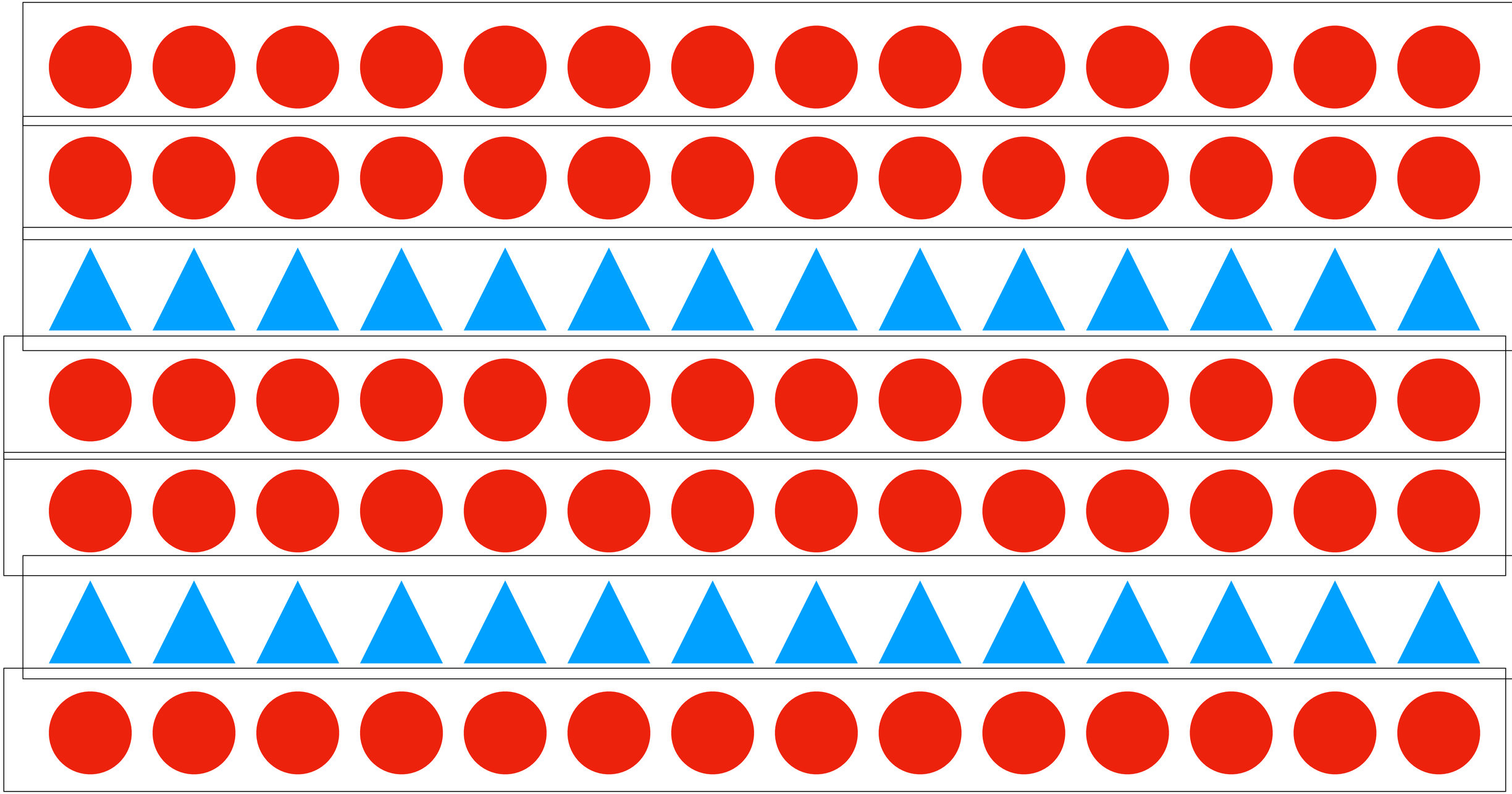


Sampling distribution estimate

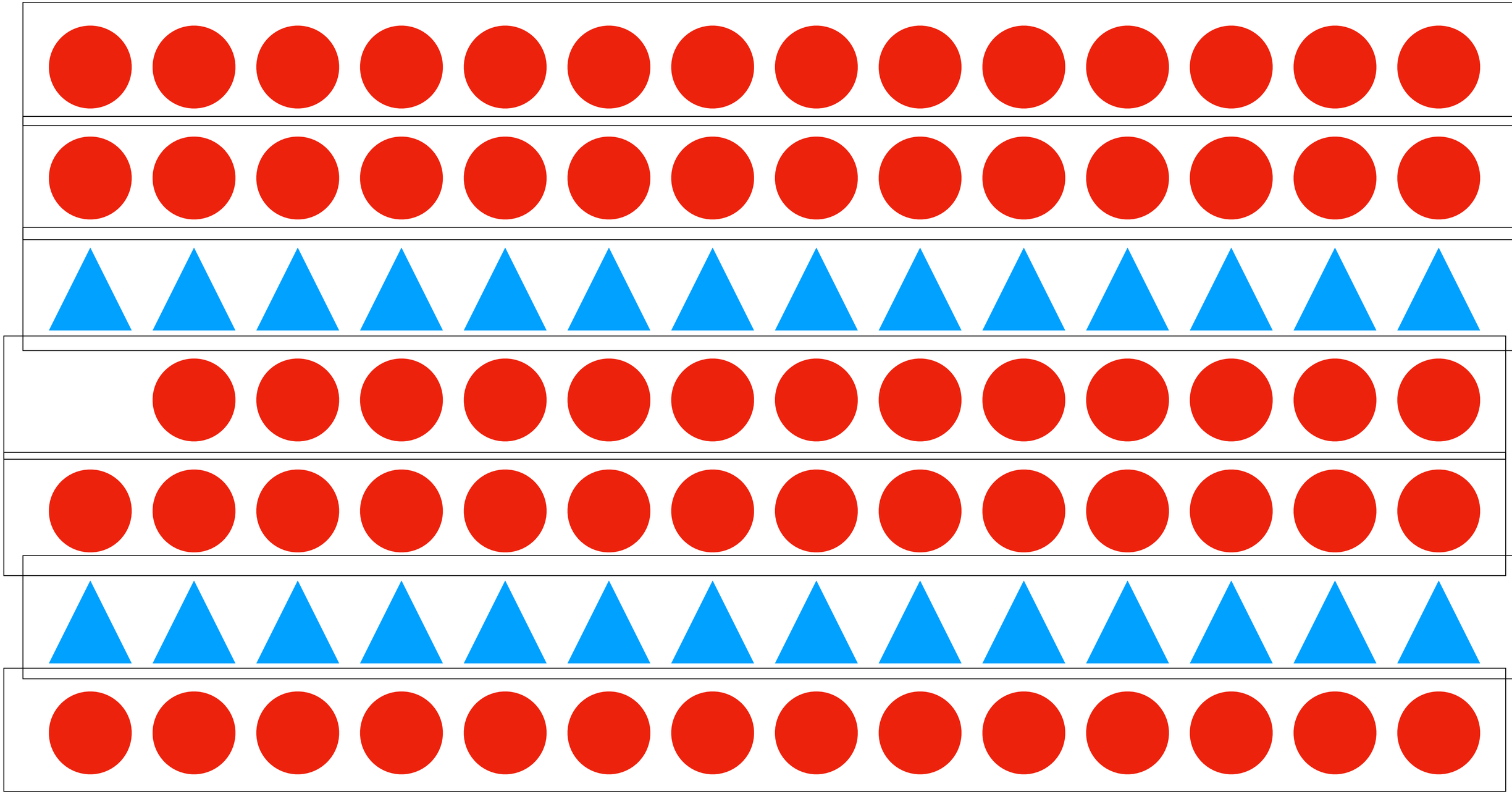


Sample

Repeat many times...



Draw 1 at random... that item's probability of being selected again still remains approximately the same!



Do not repeat the sample in practice.

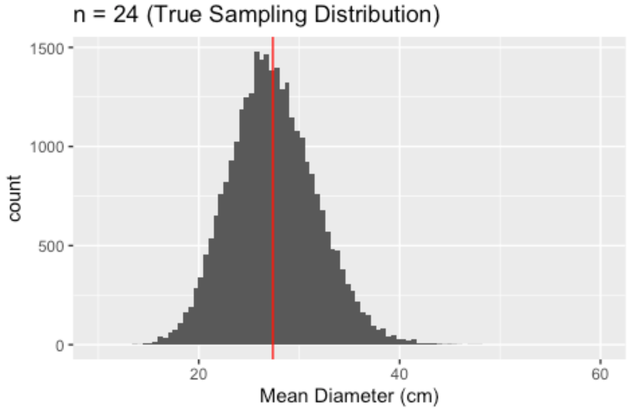
Just use sampling with replacement!

Interpreting a Sampling Distribution Estimate

Not bootstrap-specific -- more on this next week!

	What the **Sampling distribution** tells us (unachievable in practice)	What the **Sampling distribution *estimate*** tells us
On point estimates:	Tells us what point estimates are most likely to occur.	Nothing / not useful!
On the population parameter:	Tells us where the population parameter is (usually near the middle, and often the distribution's mean)	**The sampling distribution will realistically "cover" the population parameter.**

Interpreting a Sampling Distribution Estimate
(Preview of Worksheet 4)



Bootstrap Sampling Distributions (of Sample Means)