

A close-up photograph of a person's hand holding a brown trout in water. The fish is speckled with dark spots and has a blueish tint on its back. The hand is positioned on the left side of the frame, supporting the fish. The water is dark and rippled. A semi-transparent white banner is overlaid across the middle of the image, containing the title text.

The Bootstrap

Today:

- ▶ Introduce bootstrapping;
- ▶ Talk about confidence intervals (the most common application of the bootstrap);
- ▶ Examples!

Recall sampling: Definitions

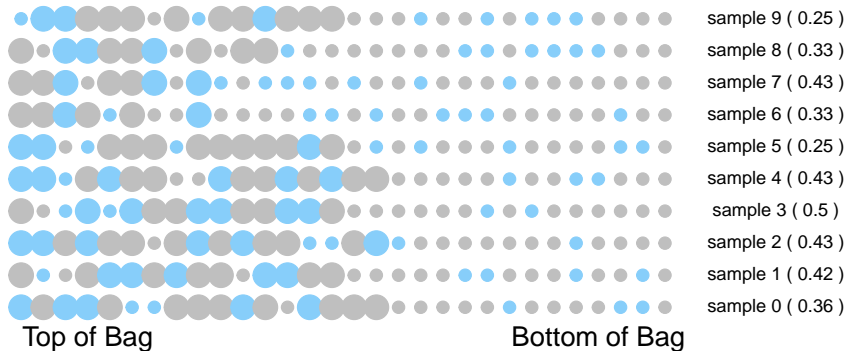
- ▶ **Population:** a 'complete' group of N objects, items, entities, or events of interest – e.g. all adults living in the US;
- ▶ **Sample:** a selected subset of n individuals from a population – e.g. 5,000 US adults appearing in a poll;
- ▶ **Summary Statistic:** a summary of the information in a set of observations – e.g. mean, median, mode, etc.;
- ▶ **Census:** a counting of all elements of the population;
- ▶ **Sampling:** the act of collecting a sample of size n from a population of size N ;
 - ▶ sample only when we can't perform a census;
 - ▶ typically the sample size $n \ll N$;
- ▶ **Sample Statistic:** a summary statistic computed from a sample that estimates the unknown population parameter.

Recall sampling: A simple example – marbles in a bag

- ▶ Consider a bag full of marbles;
 - ▶ the number of marbles in the bag is unknown;
 - ▶ there are multiple but unknown colors of marbles in the bag;
 - ▶ the fraction of any particular color of marbles in the bag is unknown;
- ▶ Questions we could ask:
 - ▶ How many marbles are in the bag?
 - ▶ How many colors of marbles are in the bag?
 - ▶ **What is the fraction of blue marbles in the bag?**
- ▶ Assume we can't just dump the bag out or remove marbles from it permanently – can we devise a process to answer any of these questions?
- ▶ What about the following:
 1. Stick a hand in the top of the bag and pull out handful of marbles;
 2. Observe them;
 3. Return them to the bag and then mix;
 4. Repeat.

Bootstrapping is REALLY similar to this!

What is the fraction of blue marbles?



What is bootstrapping?

- ▶ Let's return to the marbles in a bag experiment with a slight modification – suppose we know there are 30 marbles and we want to know the fraction of blue marbles;
 1. Sample marbles **with replacement**:
 - 1.1 Stick a hand in the bag and pull out a marble;
 - 1.2 Observe whether it is blue;
 - 1.3 Return it to the bag and mix;
 - 1.4 Repeat this 30 times – these 30 observations will be one bootstrap sample;
 2. Compute the fraction of blue marbles you observed in your bootstrap sample;
 3. Repeat steps 1 and 2 many (1000s of) times.
- ▶ This is all bootstrapping is – sample from your data with replacement, compute the statistic you are interested in, repeat!

Confidence Intervals – most common application of bootstrapping!

- ▶ We'd like to measure the variation of a statistic (e.g. the β s in a regression);
 - ▶ Want to be able to say: 'effect of x on y is $\beta \pm z$;
 - ▶ Typical way to do this is to use a **confidence interval**;
- ▶ A 95% confidence interval is a region around the statistic, a plus/minus:
 - ▶ **Interpretation 1**: there is a 95% probability that the 95% confidence interval calculated using a future sample will include the true population parameter;
 - ▶ **Interpretation 2**: the region contains values that are not statistically significantly different from the point estimate at the 0.05 level;
- ▶ Rough rules of thumb:
 - ▶ The 95% confidence interval is twice the standard error;
 - ▶ If the 95% confidence interval does not include 0 then we reject the null of no effect.

	Estimate	Std. Error	t value	$\Pr(> t)$
(Intercept)	1.3127	1.0125	1.30	0.1963
percent	0.0428	0.0188	2.28	0.0237

Using Bootstrapping to Estimate Confidence Intervals

- ▶ Three steps:
 1. Sample from the original data (or bag of marbles) with replacement and compute the stat of interest 1000 times;
 2. Sort the results from smallest to largest;
 3. Take the 25th and 975th results from this sorted list – these are the boundaries of your confidence interval;
- ▶ But we have other ways of thinking about the error in our inferences (e.g. p-values) so why do this?

Using Bootstrapping to Estimate Confidence Intervals

- ▶ Three steps:
 1. Sample from the original data (or bag of marbles) with replacement and compute the stat of interest 1000 times;
 2. Sort the results from smallest to largest;
 3. Take the 25th and 975th results from this sorted list – these are the boundaries of your confidence interval;
- ▶ But we have other ways of thinking about the error in our inferences (e.g. p -values) so why do this?
 - ▶ Some statistics do not have things like p -values that are computable using equations (e.g. medians, model performance metrics);

Using Bootstrapping to Estimate Confidence Intervals

- ▶ Three steps:
 1. Sample from the original data (or bag of marbles) with replacement and compute the stat of interest 1000 times;
 2. Sort the results from smallest to largest;
 3. Take the 25th and 975th results from this sorted list – these are the boundaries of your confidence interval;
- ▶ But we have other ways of thinking about the error in our inferences (e.g. p -values) so why do this?
 - ▶ Some statistics do not have things like p -values that are computable using equations (e.g. medians, model performance metrics);
 - ▶ Some problems in regression will compromise the p -value but we can easily get a sense of whether to reject the null with bootstrapped CIs.

Step 1: sample w/ replacement from the marble bag 1000 times

```
[1] 0.2333 0.4667 0.3333 0.3667 0.3000 0.4000 0.3000 0.2000 0.5333 0.3667 0.3000 0.2333 0.3667 0.3000 0.2667 0.3333 0.1667 0.1667 0.3000 0.2000 0.4333 0.4667 0.3333 0.3333 0.3000 0.4000
[27] 0.3000 0.4000 0.4000 0.3333 0.2000 0.3667 0.3000 0.5000 0.3667 0.2000 0.3000 0.2333 0.3667 0.3333 0.4333 0.3667 0.2333 0.3000 0.2667 0.3000 0.1333 0.4667 0.3333 0.5000 0.4333 0.3333
[53] 0.3333 0.3000 0.3667 0.3667 0.3333 0.2333 0.3667 0.4667 0.2667 0.2333 0.3000 0.3667 0.2667 0.3333 0.3667 0.3000 0.2667 0.3000 0.4333 0.3333 0.2333 0.2333 0.2667 0.4333 0.4333 0.3667
[79] 0.1667 0.5000 0.3667 0.2667 0.4000 0.4000 0.3333 0.3667 0.3000 0.5000 0.4000 0.2667 0.2000 0.3333 0.2667 0.2667 0.3667 0.3667 0.2333 0.3667 0.2667 0.4000 0.2667 0.4000 0.2000 0.1667
[105] 0.3000 0.3000 0.3333 0.3333 0.3000 0.3667 0.4000 0.4000 0.2667 0.2333 0.2667 0.3667 0.3000 0.2667 0.4000 0.2667 0.3333 0.3333 0.3333 0.3000 0.2333 0.5000 0.4333 0.3667 0.3333 0.4333 0.3000
[131] 0.2667 0.1667 0.3667 0.2667 0.3000 0.3000 0.3667 0.3333 0.2667 0.3000 0.2333 0.2000 0.5333 0.3667 0.5333 0.4000 0.3667 0.2667 0.2667 0.3667 0.3000 0.4000 0.4000 0.3000 0.2000 0.4000
[157] 0.2333 0.3333 0.3000 0.4000 0.3000 0.3667 0.5000 0.3000 0.2333 0.2667 0.3667 0.3000 0.2333 0.3333 0.2000 0.1667 0.4333 0.4000 0.2667 0.1667 0.3333 0.4667 0.3333 0.3667 0.4333 0.3667
[183] 0.3333 0.2000 0.3667 0.3000 0.1333 0.3667 0.4000 0.4333 0.2667 0.2333 0.3333 0.5000 0.4667 0.4333 0.3667 0.2333 0.4667 0.3000 0.5000 0.3333 0.3667 0.2667 0.2667 0.3333 0.3667 0.4000
[209] 0.3333 0.3000 0.2333 0.3667 0.2000 0.4333 0.4000 0.3333 0.4000 0.2667 0.1667 0.3667 0.3000 0.2667 0.3667 0.3333 0.4667 0.4667 0.3667 0.3000 0.3000 0.3333 0.2667 0.4667 0.2667 0.3000
[235] 0.4333 0.3667 0.3333 0.3333 0.3333 0.3333 0.3667 0.3333 0.3333 0.1667 0.2333 0.4000 0.3667 0.4667 0.2000 0.2333 0.3667 0.4667 0.2667 0.3000 0.3000 0.4000 0.4000 0.2667 0.3000 0.3333
[261] 0.3000 0.3000 0.2000 0.3333 0.2333 0.2333 0.3000 0.3667 0.2333 0.2333 0.2667 0.3667 0.3000 0.3667 0.3333 0.2667 0.2667 0.3333 0.4000 0.3333 0.1333 0.1333 0.3667 0.3000 0.3000 0.3667
[287] 0.2667 0.2000 0.3333 0.4000 0.3333 0.3000 0.1333 0.4333 0.2333 0.3333 0.3000 0.2333 0.1333 0.3000 0.3667 0.4000 0.2667 0.3000 0.4000 0.3333 0.3000 0.2667 0.3667 0.2333 0.3667 0.4667
[313] 0.4000 0.3667 0.1667 0.3667 0.2667 0.2000 0.4000 0.3333 0.4000 0.3333 0.5333 0.4000 0.3667 0.3333 0.4333 0.3000 0.3333 0.3333 0.3667 0.3000 0.3000 0.3333 0.3000 0.3000 0.2667 0.2000
[339] 0.2667 0.2333 0.3333 0.2333 0.4000 0.4667 0.2333 0.3667 0.3333 0.4000 0.4000 0.4000 0.2333 0.3667 0.2333 0.3333 0.4000 0.5000 0.1667 0.1667 0.2333 0.4000 0.3333 0.2667 0.3333
[365] 0.4000 0.3667 0.3667 0.3333 0.3667 0.3000 0.3000 0.4000 0.3000 0.2333 0.3000 0.3333 0.2000 0.3667 0.2333 0.3667 0.3333 0.3333 0.3333 0.2333 0.4667 0.4333 0.4000 0.5667 0.5000 0.3000
[391] 0.4000 0.3000 0.3333 0.2333 0.2000 0.4000 0.4333 0.3333 0.3000 0.2333 0.1667 0.4000 0.4333 0.3000 0.3000 0.4667 0.4333 0.4667 0.2000 0.3333 0.3000 0.3667 0.3667 0.3333 0.2667
[417] 0.3000 0.3333 0.2667 0.4000 0.4667 0.3667 0.3000 0.2667 0.4000 0.3667 0.3333 0.4667 0.2667 0.4000 0.3333 0.4667 0.3000 0.2667 0.3667 0.5000 0.4333 0.2667 0.4000 0.5667 0.3667
[443] 0.2667 0.3000 0.3333 0.2000 0.3667 0.3333 0.4333 0.2667 0.3000 0.3667 0.4000 0.3667 0.3000 0.4333 0.2667 0.3333 0.3000 0.3333 0.3667 0.3000 0.4000 0.3667 0.3667 0.4667 0.4000
[469] 0.3000 0.3000 0.3333 0.3000 0.4000 0.4000 0.2333 0.4333 0.2333 0.2000 0.3667 0.2333 0.3000 0.2667 0.4333 0.3333 0.4000 0.4333 0.2333 0.2000 0.2333 0.4667 0.3333 0.2667 0.2333 0.3333
[495] 0.4667 0.3333 0.2667 0.3667 0.3667 0.3333 0.3667 0.2667 0.3667 0.4000 0.2667 0.2000 0.4000 0.2667 0.3667 0.4333 0.5000 0.2333 0.2000 0.4333 0.4000 0.4333 0.2667 0.2333 0.3333 0.3667
[521] 0.4000 0.4000 0.4667 0.2667 0.2333 0.4667 0.4000 0.3333 0.3000 0.2000 0.2333 0.3000 0.4000 0.2667 0.4667 0.4000 0.3000 0.4000 0.3000 0.4000 0.4000 0.2667 0.3333 0.3667 0.3000 0.4000
[547] 0.4000 0.3333 0.4000 0.3000 0.4000 0.2333 0.4333 0.4000 0.2667 0.3000 0.1667 0.2333 0.2333 0.3333 0.3000 0.5000 0.5667 0.4000 0.3333 0.4333 0.2333 0.2333 0.4333 0.3000 0.3000 0.3000
[573] 0.3667 0.5333 0.3667 0.3667 0.2667 0.3333 0.4000 0.5000 0.4333 0.1667 0.3667 0.2333 0.4000 0.2333 0.3333 0.4000 0.3667 0.5667 0.4333 0.3000 0.1667 0.3667 0.2333 0.3667 0.4000 0.3000
[599] 0.3333 0.3333 0.2333 0.2667 0.4000 0.3667 0.3667 0.4000 0.3333 0.2667 0.3667 0.4000 0.3333 0.4667 0.3333 0.3667 0.3000 0.3667 0.3000 0.4333 0.1667 0.2333 0.3667 0.2333 0.3667 0.2000
[625] 0.2667 0.4667 0.2000 0.3000 0.3667 0.2333 0.2667 0.3667 0.3333 0.3667 0.3667 0.2333 0.4333 0.2667 0.3333 0.1333 0.4000 0.2000 0.4000 0.2667 0.3333 0.3333 0.3333 0.3000 0.3667
[651] 0.3667 0.3667 0.4000 0.3000 0.2333 0.5667 0.2667 0.3000 0.3333 0.3333 0.3667 0.2000 0.4667 0.3000 0.2333 0.3333 0.3000 0.2667 0.2333 0.3667 0.3000 0.3667 0.4000 0.3000 0.3000 0.3333
[677] 0.3333 0.3000 0.4000 0.1667 0.2333 0.2667 0.3000 0.4000 0.4000 0.3333 0.4000 0.4333 0.4000 0.2667 0.4333 0.3000 0.2667 0.2333 0.5333 0.2667 0.3000 0.3333 0.3667 0.3333 0.3333 0.3333
[703] 0.4333 0.4333 0.5000 0.5667 0.2333 0.3667 0.3000 0.5333 0.3667 0.4000 0.4667 0.3667 0.3333 0.3333 0.1667 0.2667 0.3333 0.4667 0.3333 0.5000 0.3000 0.2667 0.3667 0.2333 0.2333 0.3333
[729] 0.5333 0.3000 0.4333 0.5667 0.4667 0.3000 0.1333 0.2667 0.4000 0.5000 0.5000 0.3000 0.4667 0.3667 0.2000 0.3000 0.5333 0.4333 0.2333 0.3333 0.4667 0.3333 0.4000 0.4000 0.2000 0.4667
[755] 0.3000 0.3667 0.2333 0.3000 0.2000 0.2333 0.2667 0.2667 0.2667 0.4000 0.3333 0.3333 0.3000 0.1667 0.2667 0.4000 0.4000 0.3333 0.4000 0.3667 0.2667 0.5000 0.3333 0.3667 0.1667 0.3000
[781] 0.4000 0.1667 0.2333 0.2000 0.2333 0.2667 0.2000 0.3333 0.4000 0.3667 0.3667 0.3000 0.4000 0.3333 0.3000 0.3000 0.4000 0.3000 0.2667 0.3000 0.3333 0.4333 0.4000 0.3333 0.3333 0.4667
[807] 0.3667 0.3000 0.3000 0.3000 0.3667 0.3000 0.3667 0.3667 0.3333 0.3000 0.4667 0.2333 0.4000 0.3667 0.4000 0.5000 0.2667 0.3333 0.3333 0.2333 0.3333 0.3667 0.2333 0.3333 0.4333 0.2333
[833] 0.4667 0.4000 0.4333 0.4333 0.2000 0.3333 0.4000 0.3667 0.4000 0.3667 0.1333 0.4333 0.2000 0.3667 0.3000 0.3667 0.2000 0.2333 0.2667 0.3000 0.2000 0.3667 0.3333 0.2000 0.3333 0.3667
[859] 0.3667 0.3000 0.2667 0.4333 0.3333 0.3333 0.1333 0.3667 0.3667 0.2667 0.3000 0.4000 0.3667 0.3000 0.2667 0.3667 0.4333 0.3000 0.2667 0.3000 0.3000 0.3000 0.2667 0.2667 0.3667 0.3667
[885] 0.3000 0.3667 0.3333 0.2333 0.3333 0.3333 0.4000 0.3333 0.4333 0.4333 0.5000 0.4667 0.4000 0.1667 0.2667 0.3000 0.2333 0.3000 0.4667 0.2000 0.3333 0.4667 0.3667 0.3667 0.3667 0.3667
[911] 0.3333 0.3333 0.4333 0.2333 0.4667 0.2667 0.1667 0.2000 0.4000 0.3667 0.2333 0.4000 0.2667 0.3667 0.2667 0.2667 0.3333 0.3667 0.3667 0.3333 0.3667 0.5000 0.3667 0.3000 0.2333 0.2000
[937] 0.3667 0.4333 0.4000 0.3667 0.4000 0.4000 0.4333 0.4000 0.3000 0.2667 0.5333 0.3667 0.4667 0.3000 0.0667 0.2333 0.3667 0.3667 0.3000 0.4667 0.3667 0.2333 0.4333 0.2667 0.4000 0.2333
[963] 0.3000 0.2667 0.3333 0.4000 0.4333 0.1667 0.3000 0.5000 0.3333 0.4000 0.3333 0.3667 0.2333 0.2000 0.4000 0.3000 0.2333 0.2333 0.2333 0.4000 0.4000 0.4333 0.0333 0.4667 0.2333 0.4667
[989] 0.3000 0.3333 0.3000 0.4333 0.2667 0.3667 0.1667 0.3000 0.2667 0.3000 0.4333 0.2333
```

Step 2: sort

[illegible]

Step 3: take the 25th and 975th – $[0.1667, 0.5]$

[illegible]

Why should we care?

Bootstrapping is a powerful—and pretty easy—way to make robust inferences.