Confidence intervals

With bootstrapping

Prof. Dr. Jan Kirenz
HdM Stuttgart

We use **randomization** to **test claims**

Randomization tests are best suited for modeling experiments where

- the treatment (explanatory variable) has been **randomly assigned** to the observational units
- and there is an attempt to answer a simple **yes/no** research question

Examples:

1. Does this vaccine make it less likely that a person will get malaria?
2. Does drinking caffeine affect how quickly a person can tap their finger?
3. Can we predict whether candidate A will win the upcoming election?

Now we want to estimate **population parameters**

- Instead of testing a **claim** (yes/no),
- the goal now is to estimate the unknown value of a **population parameter**.

Examples:

1. How much less likely am I to get malaria if I get the vaccine?
2. How much faster (or slower) can a person tap their finger, on average, if they drink caffeine first?
3. What proportion of the vote will go to candidate A?

Bootstrapping is a simulation method

- The focus is on a **single proportion**
- Ideally, sample data was generated through **random sampling** from a population.
- our goal with bootstrapping is to **understand variability** of a **statistic**.

Randomization vs bootstrap

Randomization tests:

- modeled how the **statistic** would **change** if the treatment had been allocated differently

The **bootstrap**

- will model how a **statistic varies** from one **sample** to another **sample** taken from the population.
- This will provide information about **how different** the **statistic** is from the **parameter of interest**.

Medical consultant case study

Organ donation

👵 Medical consultant

Photo by Jafar Ahmed

👵 Claim of one consultant

- “**Average complication rate** for liver donor surgeries in the US is about **10%**”
- “my clients have had only **3 complications in the 62** liver donor surgeries I have facilitated”
- *“This is strong evidence that my work meaningfully contributes to reducing complications”.*

- **p**: represent the true complication:0,1
- **p’**: sample proportion for the complication rate: 3/62 = 0.048

Is it possible to assess the consultant’s claim (that the reduction in complications is due to her work) using the data?

We want to understand the consultant's true rate of complications (variability of the statistic).

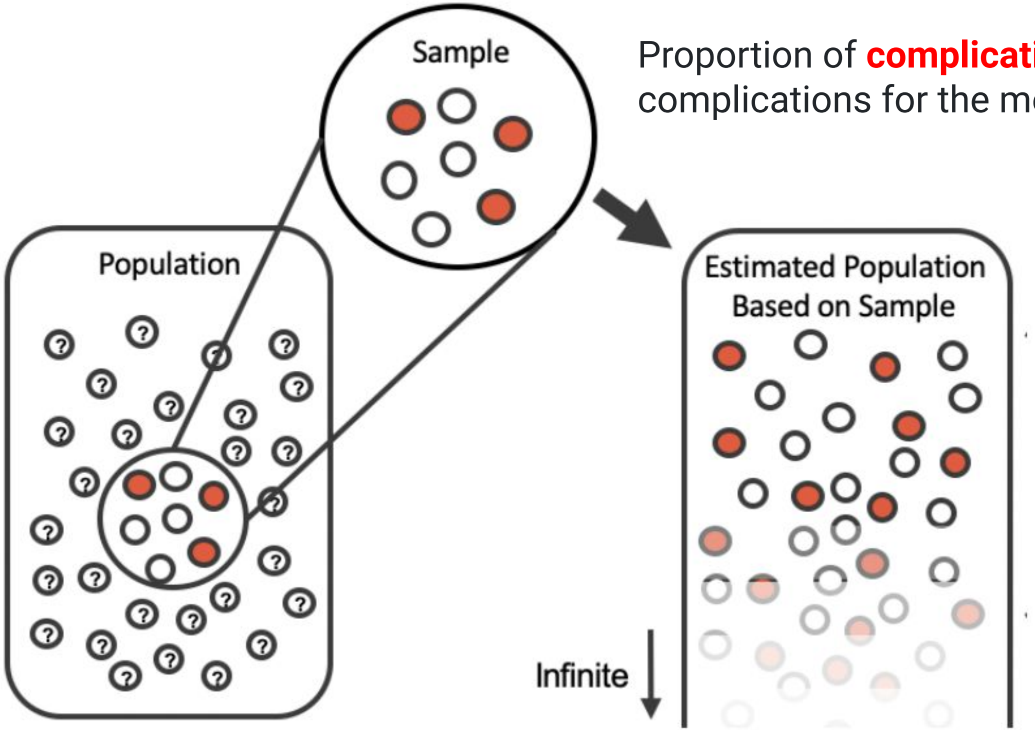
- A **parameter** is the "true" value of interest.
- We typically estimate the parameter using a **point estimate** from a sample of data.
- The point estimate is also known as the **statistic**.

- We estimate the probability **p** of a complication for a client of the medical consultant by examining the past complications rates of her clients:
- **p'** = 3/62 = 0.048

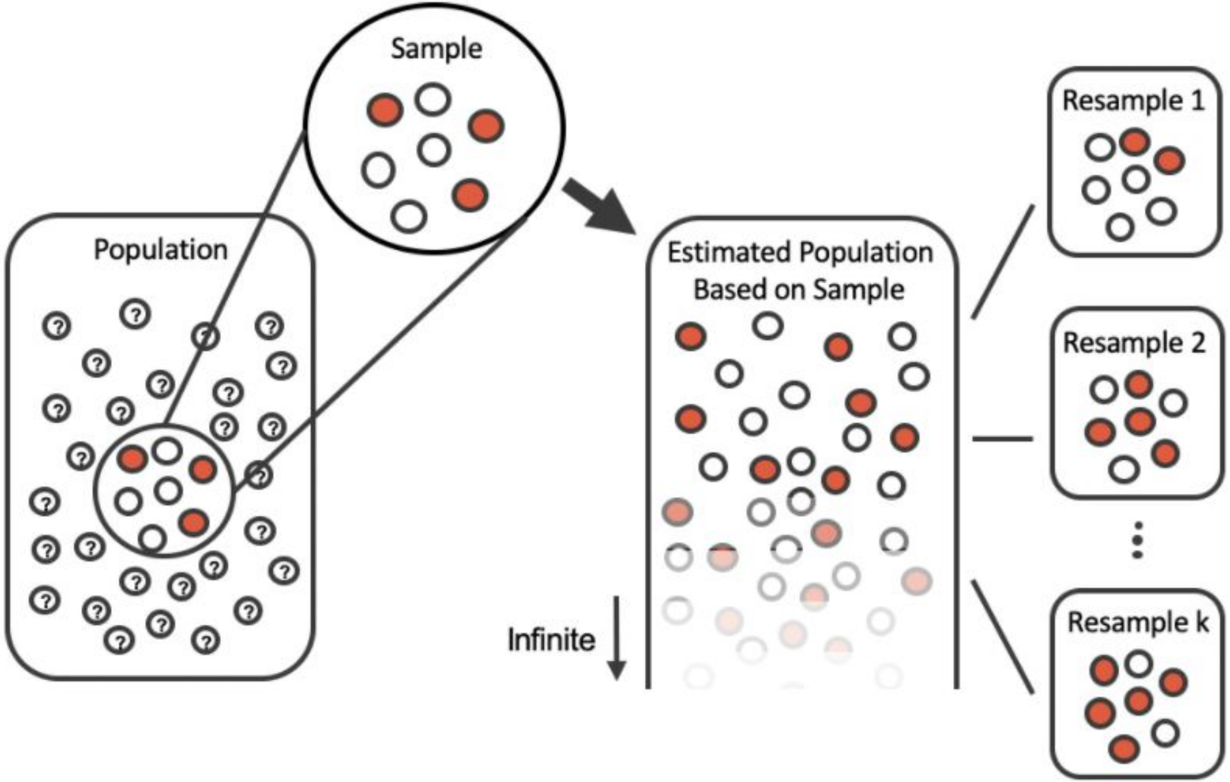
  is used to estimate p

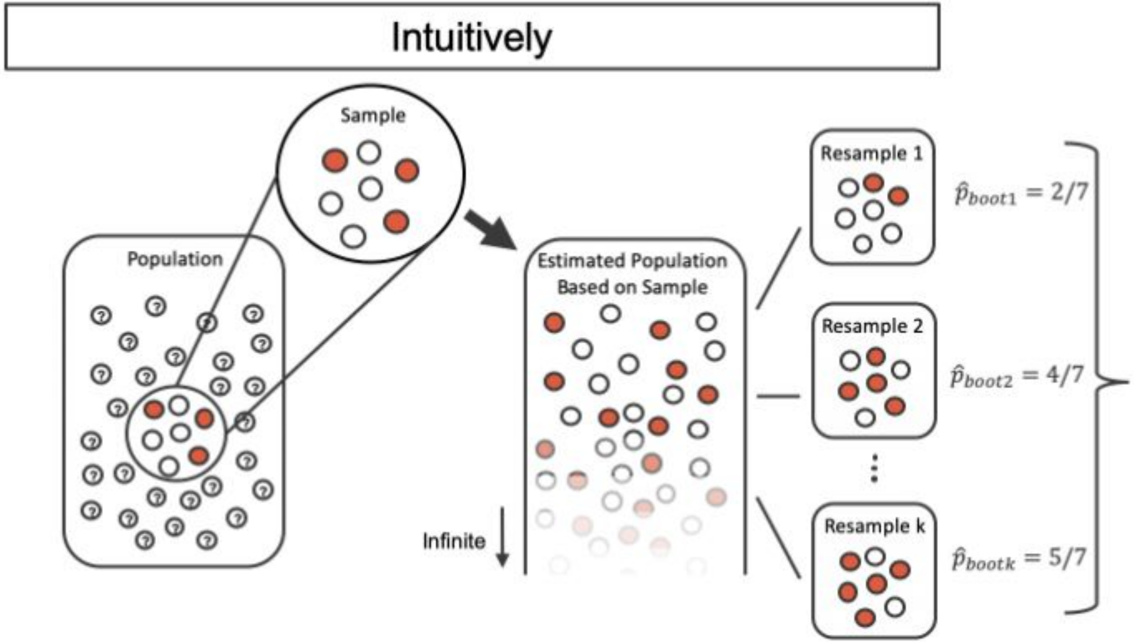
Intuitively

Proportion of **complications** and no complications for the medical consultant

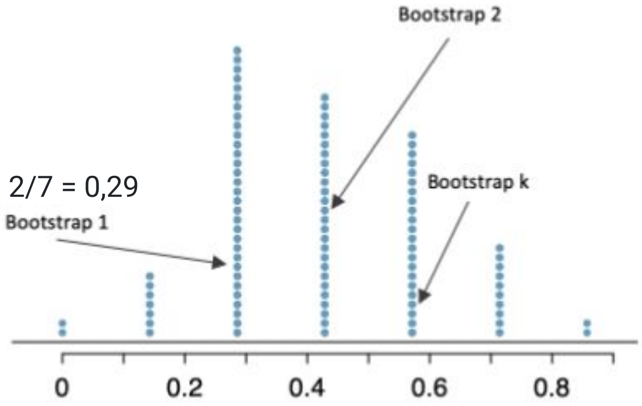


Intuitively



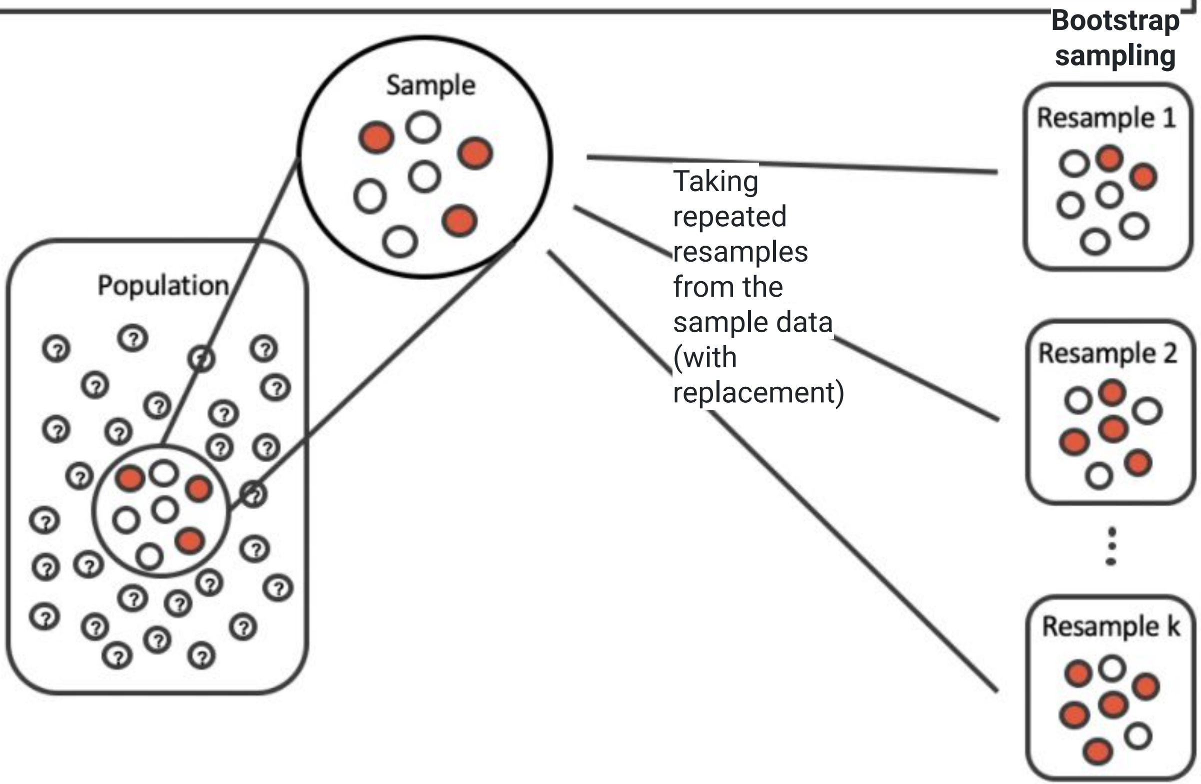


Bootstrap distribution

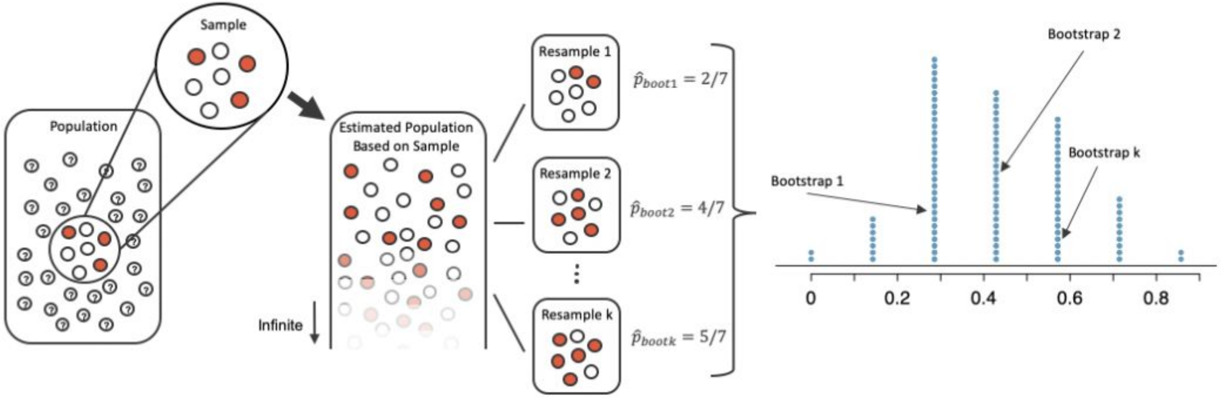


measure for how the proportions vary from sample to sample

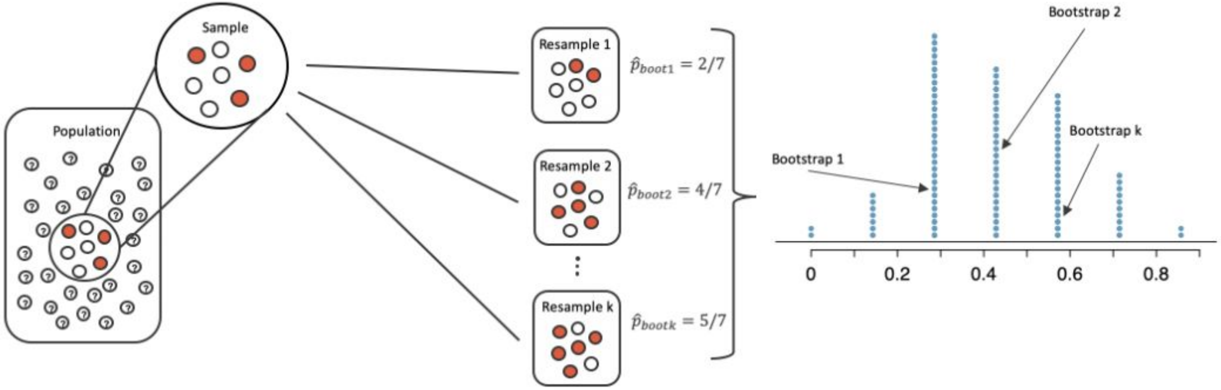
In Practice


Bootstrap sampling
Sample
Population
Taking repeated resamples from the sample data (with replacement)
Resample 1
Resample 2
Resample k

Intuitively



In Practice



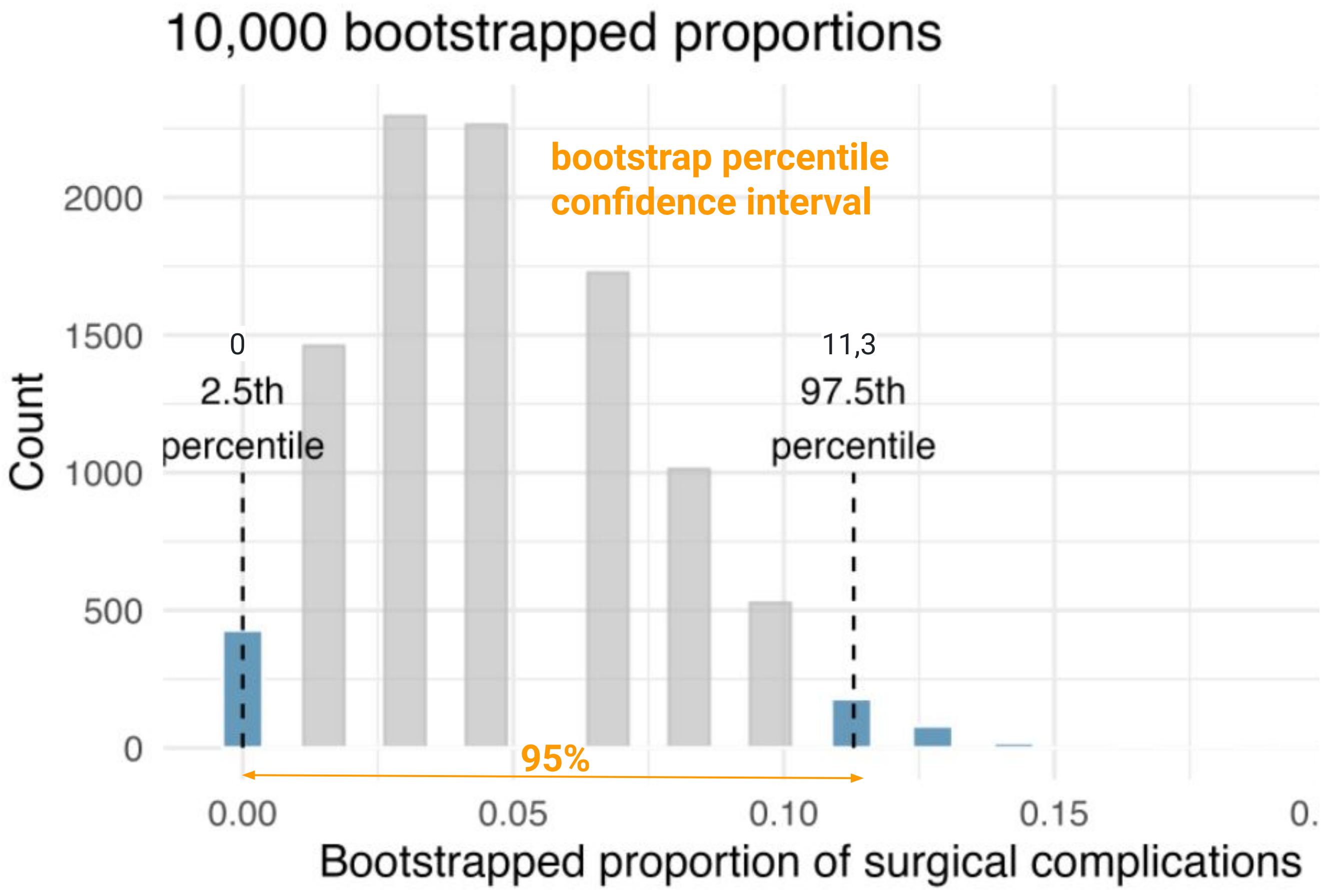
Medical consultant example

Consider each client to be one of the marbles in the bag.

There will be:

- 59 white marbles (no complication) and
- 3 red marbles (complication).

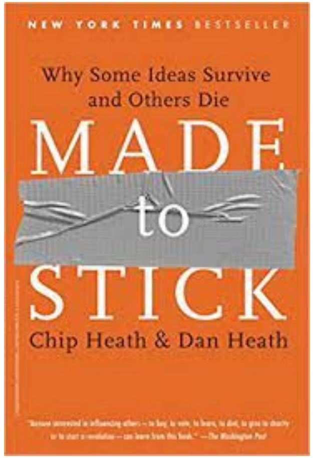
If we choose 62 marbles out of the bag (one at a time with replacement) and compute the proportion of simulated patients with complications, then this “bootstrap” proportion represents a single simulated proportion from the “resample from the sample” approach.



We are confident (with 95% confidence) that, in the population, the true probability of a complication is between 0% and 11.3%.

Tappers and listeners case study

MADE to STICK
SUCCESs Model

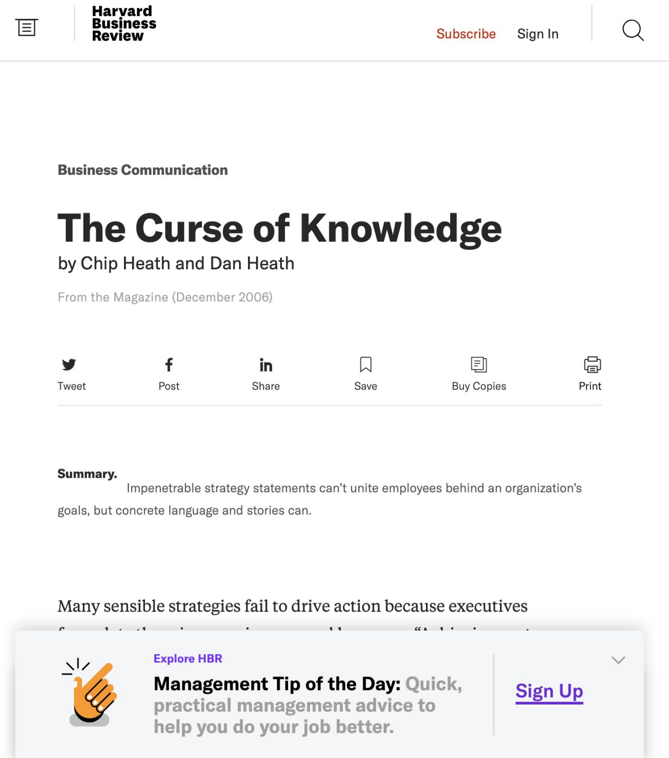


A sticky idea is understood, it's remembered, and it changes something. Sticky ideas of all kinds—ranging from the "kidney thieves" urban legend to JFK's "Man on the Moon" speech—have six traits in common. If you make use of these traits in your communication, you'll make your ideas stickier. (You don't need all 6 to have a sticky idea, but it's fair to say the more, the better!)

PRINCIPLE 1	PRINCIPLE 2	PRINCIPLE 3	PRINCIPLE 4	PRINCIPLE 5	PRINCIPLE 6	
S	U	C	C	E	S	s
SIMPLE	**UNEXPECTED**	**CONCRETE**	**CREDIBLE**	**EMOTIONAL**	**STORIES**	
Simplicity isn't about dumbing down, it's about prioritizing. (Southwest will be THE low-fare airline.) What's the core of your message? Can you communicate it with an analogy or high-concept pitch?	To get attention, violate a schema. (The Nordie who ironed a shirt…) To hold attention, use curiosity gaps. (What are Saturn's rings made of?) Before your message can stick, your audience has to want it.	To be concrete, use sensory language. (Think Aesop's fables.) Paint a mental picture. ("A man on the moon…") Remember the Velcro theory of memory—try to hook into multiple types of memory.	Ideas can get credibility from outside (authorities or anti-authorities) or from within, using human-scale statistics or vivid details. Let people "try before they buy." (Where's the Beef?)	People care about people, not numbers. (Remember Rokia.) Don't forget the WIIFY (What's In It For You). But identity appeals can often trump self-interest. ("Don't Mess With Texas" spoke to Bubba's identity.)	Stories drive action through simulation (what to do) and inspiration (the motivation to do it). Think Jared. Spring-board stories (See Denning's World Bank tale) help people see how an existing problem might change.	

www.MADE to STICK.com

The curse of knowledge


Harvard Business Review

Subscribe Sign In

Business Communication

The Curse of Knowledge

by Chip Heath and Dan Heath

From the Magazine (December 2006)

Tweet Post Share Save Buy Copies Print

Summary. Impenetrable strategy statements can't unite employees behind an organization's goals, but concrete language and stories can.

Many sensible strategies fail to drive action because executives

Experiment:

- 120 Listeners, 120 Tappers
- Expectation: 50% guess correct
- Observation: 3 out of 120 (**p'=0.025**)
- *Question: what is the* ***true proportion (p)*** *of people who can guess the tune?*

Sampling 5 tapper-listener pairs
(from the bag of 3 red and 117 white marbles)

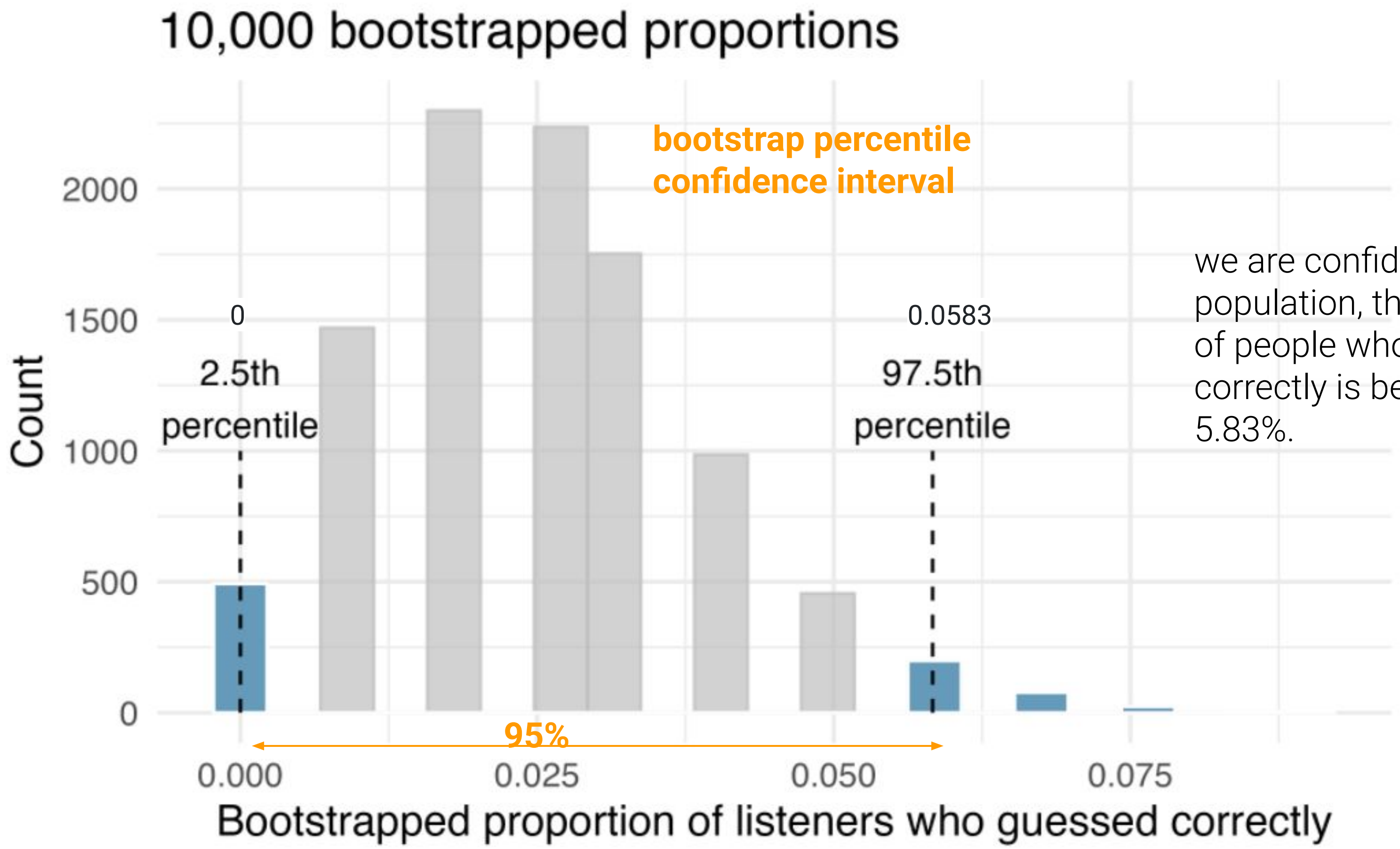
W	W	W	R	W
Wrong	Wrong	Wrong	Correct	Wrong

… after selecting 120 marbles, we counted 2 red for p’=0.0167

In order to understand how far the observed proportion of $p' = 0.025$ might be from the true parameter (p), we should generate more simulations.

Here we've repeated the entire simulation ten times:

$$0.0417 \quad 0.025 \quad 0.025 \quad 0.0083 \quad 0.05 \quad 0.0333 \quad 0.025 \quad 0 \quad 0.0083 \quad 0$$



we are confident that, in the population, the true percent of people who can guess correctly is between 0% and 5.83%.

Do the data provide convincing evidence against the claim that 50% of listeners can guess the tapper’s tune?

Confidence intervals

Plausible range of values for the population parameter

Using only a single **point estimate** is like fishing in a murky lake with a spear

Using a **confidence interval** is like fishing with a net

If we want to be very certain we capture the population parameter, should we use a wider interval (e.g., 99%) or a smaller interval (e.g., 80%)?

Bootstrap confidence interval

The **95% bootstrap confidence interval** for the parameter **p** can be obtained directly using the ordered **p'** boot values

Consider the sorted p' boot values: Call the 2.5% bootstrapped proportion value "lower," and call the 97.5% bootstrapped proportion value "upper."

The 95% confidence interval is given by: (lower, upper)

Summary

Bootstrap process

1. Frame the **research question** in terms of a parameter to estimate.
2. **Collect data** with an observational study or experiment.
3. Model the **randomness** by using the data values as a proxy for the population (with **bootstrapping**).
4. Create the **confidence interval**.
5. Form a **conclusion**.

Summary of bootstrapping as an inferential statistical method

Question	**Answer**
What does it do?	**Resamples** (with replacement) from the observed data to mimic the sampling variability found by collecting data from a population
What is the random process described?	Random sampling from a population
What other random processes can be approximated?	Can also be used to describe random allocation in an experiment
What is it best for?	**Confidence intervals** (can also be used for bootstrap hypothesis testing for one proportion as well).
What physical object represents the simulation process?	Pulling marbles from a bag with replacement

Terms you should know

bootstrap percentile confidence interval	parameter	statistic
bootstrap sample	point estimate	
bootstrapping	sampling with replacement	