

Inference, Bootstrapping, and Randomization inference

Venki

PhD student

Public Policy and Political Economy, School of EPPS

Start recording!

Note to self

Outline

1

Defining Inference

2

Sampling distribution

3

Confidence intervals

4

Testing hypotheses about means

5

The dangers that lurk in the dark

6

Breaking normality of Sampling Distribution

7

Bootstrap demo

8

Inference problem in the causal setup

9

Bootstrap v Randomization inference

1. Defining Inference

Revisiting last workshop...

When you go to the field and perform an RCT and collect data what exactly are you observing?

Sample space

Events

Probability Density

Joint Density

Probability Distribution

Marginal Density

Random Variables

Conditional Density

Support

PMF

Realizations

CMF

1. Defining Inference

Estimation of Estimands is done with Estimators

Estimation

Generalizing a “quantity of interest” from the sample (using realizations) to the population is called estimation.

Quantity of interest could be any of the n^{th} order moments – Expectation Value, Variance, Skew...

Estimand

The “quantity of interest” is called estimand.

In this course - ATE, ATT and rarely ATC

When you propose your project on Week 8, tell us which estimand you are planning to get from your analysis.

$$E[X] \text{ or } \mu$$

Estimator

The mathematical function (or numerical algorithm) used to do estimation of an estimand is called an estimator.

There are different type of estimators to find the same estimand! Choose wisely!

Finally, output of an estimator is an estimate.

$$\hat{E}[X] \text{ or } \bar{X}$$

2. Sampling distribution

Will the grass in the other sample be greener?

We got our realizations.

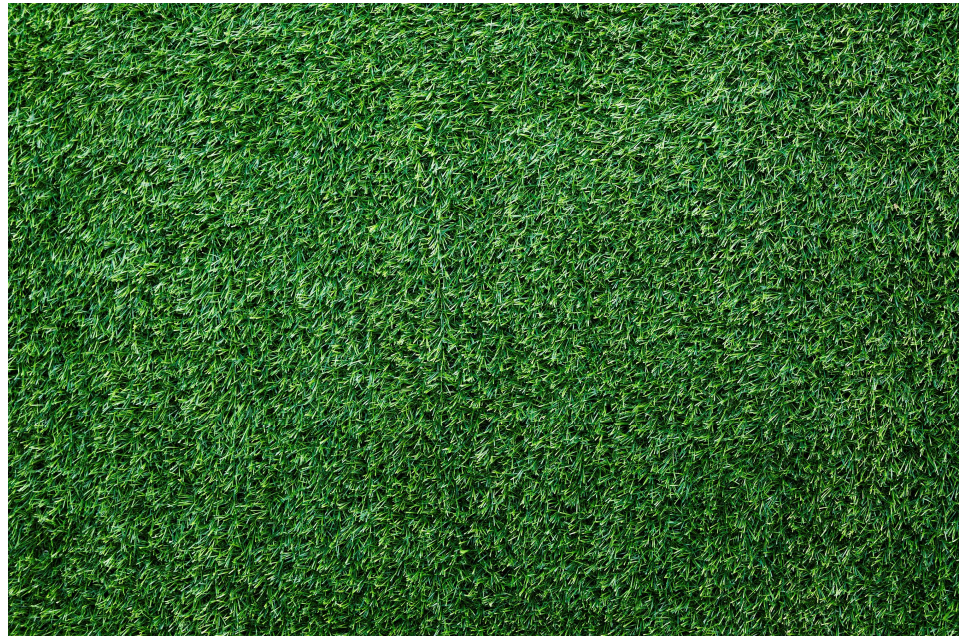
We got an estimate from it.

But are we content? No!

What about the counterfactual
variation across samples?

“But what if I get another sample? If I can only
get *MORE* funding...”

*Enter the sampling
distribution...*



2. Sampling distribution

Dice roll example

Let us consider what happens if we did collect a lot of samples from the same toy experiment.

Click to open an Excel file...

2. Sampling distribution

Estimating the SEM

But we can't actually get the sampling distribution! Even when we tried a 100 times we weren't certain of the normality.

But we do know from CLT that it will be normal distribution ($n > 30$), eventually. Using that knowledge the following estimator of the SEM is known to be unbiased

$$SEM = SE_{\bar{x}} = \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

But we don't know population Sdev!

Sample Sdev can serve as a population Sdev *under certain assumptions...*

The spirit here is that the within-sample variability is a good proxy for the across samples variability.

But how close should other estimates be for us to know our sample is good?

What is the benchmark for “weirdness” of our sample?

2. Sampling distribution

Some benchmarks from the empirical normal distribution

Know that in a normal distribution...

About 68% of the data falls within 1 Sdev of the mean.

About 95% of the data falls within 2 Sdev of the mean.

About 99.7% of the data falls within 3 Sdev of the mean.

*Set your benchmark
for weirdness
accordingly*

3. Confidence intervals

So, we can get the SEM.

So what?

Take a step back. Remember...

We have: Sample mean; SEM (estimated using our sample sdev).

We want now: A plausible range of values for the population mean.

So, we want to estimate an interval of values for the expectation value (estimand).

3. Confidence intervals

Computing CI

We proceed as follows:

1. Using our SEM, draw a normal (or t for $n < 30$) distribution around the sample mean.
2. Based on the level of confidence we're interested in, select the range of values in that distribution which contain a certain amount of density and are centered at the sample mean.

For $n > 30$

$$\bar{x} \pm 1.96 \times SE_{\bar{x}}$$

So, what does the 95% confidence interval mean?

Do the sampling a 100 times. Construct a CI around each of those sample's mean. *On an average*, 95 of those CI would contain the sample mean.

3. Confidence intervals

What the CI doesn't mean!

“We are 95% sure our current sample contains the population parameter.”

NOPE! 

We have no way of knowing if the population mean is within our confidence interval.

“95% of the sample means will fall within this sample's 95% confidence interval”

NOPE! 

Since we don't have the sampling distribution*, we don't know if the current sample is within 95% of that distribution or it is in the weird 5%

The 95% is not about the interval – it is about the sampling process!

The sampling process used to create this sample, if repeated will create CIs that will, on an average, contain the sample mean, 95% of the time. 

* - Keep in mind that the example I showed was a special case where repeating the experiment is very easy with modern computational power! In general, you can't get the sampling distribution!

3. Confidence intervals

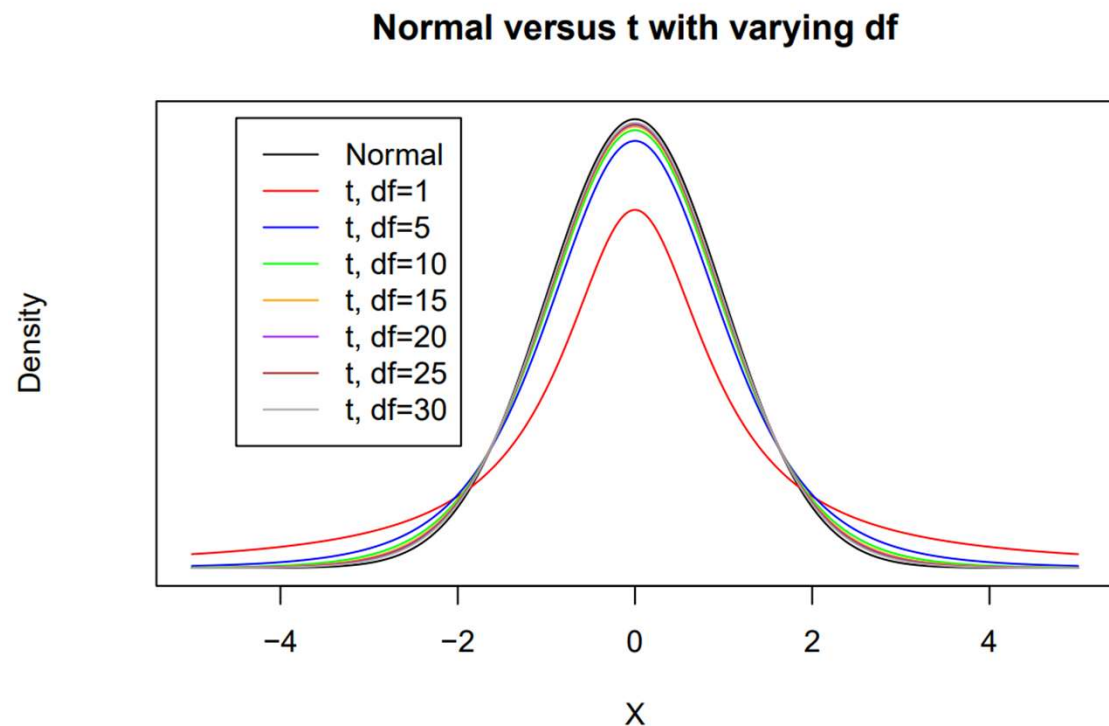
The t-distro

*Smaller sample. Then
fatter, flatter tails*

We need to go further out
into the tails to get the
same 95% data.

With the smaller samples, it becomes
more probable to get the weirder
mean values. So, things flatten out
more.

*It generalizes to
normal as degrees of
freedom (sample size)
goes beyond 30.*



4. Testing hypotheses about means

Checking claims about the estimand

Claim: Your intervention had no effect.

In other words, the mean of your treatment and control groups are actually indistinguishable.

So, estimand in the distribution of treatment and control group are the same.

Claim: Your distributional assumption is wrong.

For instance, we may be asked to test for normality.

In general, the claim can be about different estimands, not just means.

We will only talk about means now. But know that a bevy of tests exist for all sorts of estimands (like Chi-squared for proportions).

4. Testing hypotheses about means

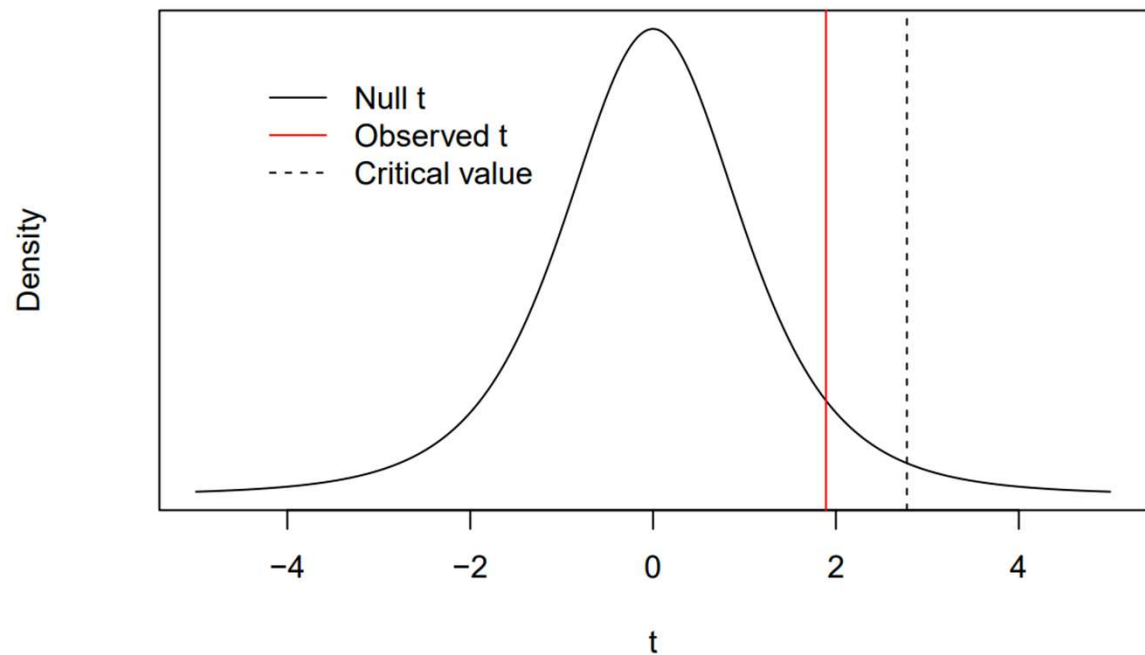
1 sample t-test

One sample t-test (df=4)

You have a hypothesis about the population mean. You are given one sample from that population to test the claim's validity.

$$H_0 : \mu_x = 0; H_a : \mu_x \neq 0$$

$$t_{\bar{x}} = \frac{\bar{x}}{SE_{\bar{x}}} = \frac{\bar{x}}{\frac{\sigma_x}{\sqrt{n}}} = \frac{\bar{x} \sqrt{n}}{\sigma_x}$$



Food for thought: What if in-sample variation doesn't proxy well for across samples-variation? When would that happen?

4. Testing hypotheses about means

2 sample t-test

*You have 2 samples.
Are these samples
from the same
population or are
they from different
populations?*

This is the whole point of this class – is the mean of the treated group different from control group?!

Our confidence interval is:

$$(\mu_{treated} - \mu_{control}) \pm t_{df} \times SE_{dim}$$

We need to define two quantities here:

$$SE_{dim} = \sqrt{\frac{\sigma_{treated}^2}{n_{treated}} + \frac{\sigma_{control}^2}{n_{control}}}$$

$$df = n_{treated} + n_{control} - 2$$

(This simple degrees of freedom estimate is a simplification. Wait till class tonight...)

5. The dangers that lurk in the dark

Ghostbusters

Hypothesis testing

You cannot prove a null hypothesis is true/false

You can, with evidence, say “The null hypothesis is sufficiently unlikely that I reject it”. This does not make it false.

You cannot say “The null hypothesis is sufficiently likely that I accept it”.

You can say “I am unable to reject the null hypothesis”.

You cannot prove an alternative hypothesis is true/false.

The p-value is..

NOT a probability that our data are true.

NOT a probability that the null hypothesis is true

NOT a probability that the alternative hypothesis is true

NOT a probability that the effects were produced by random chance

NOT describing the size of the effect

NOT telling us if an effect is “significant”

Outline

1

Defining Inference ✓

2

Sampling distribution ✓

3

Confidence intervals ✓

4

Testing hypotheses about means ✓

5

The dangers that lurk in the dark ✓

6

Breaking normality of Sampling Distribution

7

Bootstrap demo

8

Inference problem in the causal setup

9

Bootstrap v Randomization inference

6. Breaking normality of Sampling Distribution

But what if the estimand wasn't mean?

Click to open a new Excel file with the same Dice rolling example.

6. Breaking normality of Sampling Distribution

Dealing with estimands other than mean

In the case I showed, the sampling distribution should be analytically solvable. But in many cases, we don't even know if that is possible!

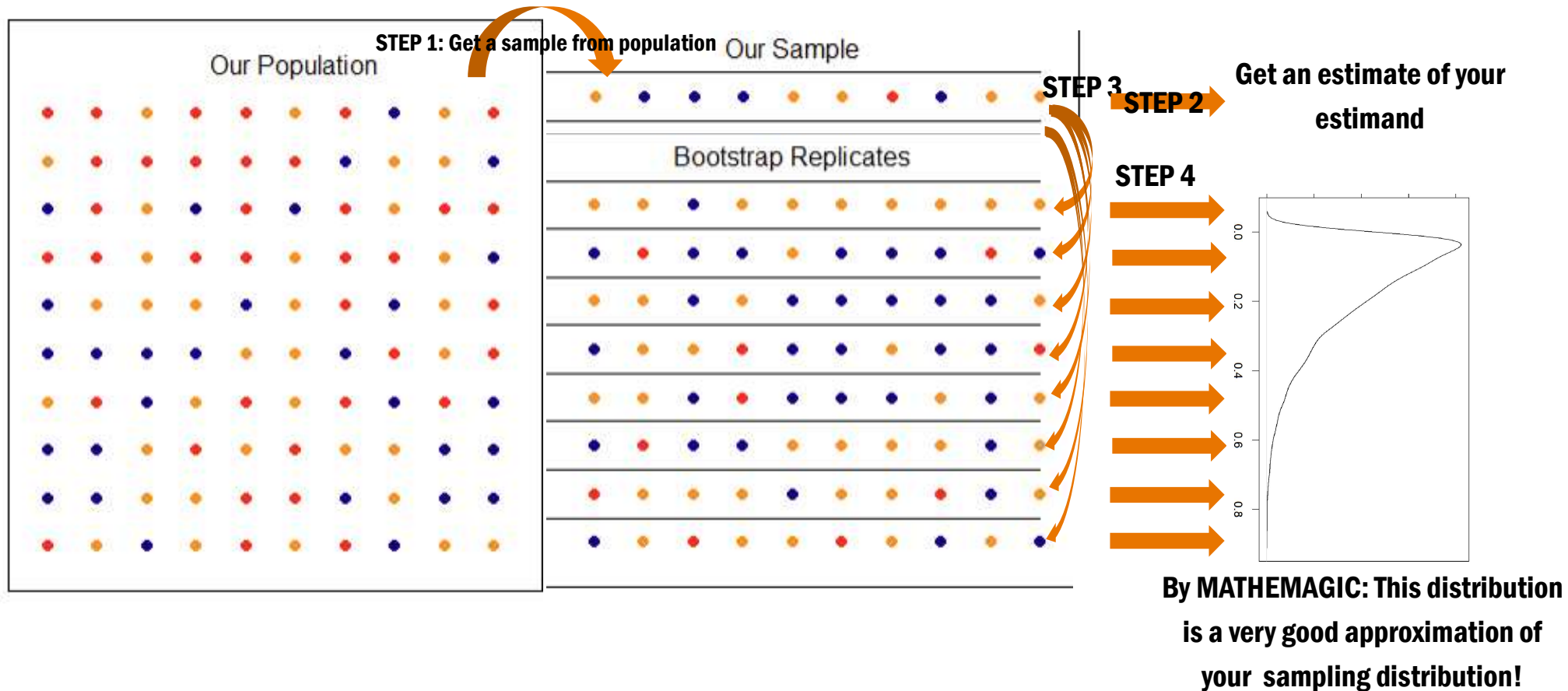
So even if you collect infinite samples, I can't tell you what sampling distribution it has to be!

We can't say stuff like 'Within 1.96 SD of the sampling distribution, 95% of the sample estimands should be available' like I did with means!

What now?! ENTER BOOTSTRAP...

7. Bootstrap demo

The mathemagic of Bootstrap



Demonstration

*Let us look at some R code to see how it works
with some real data, at the end if time permits...*

8. Bootstrap demo

Things to keep in mind

Remember...

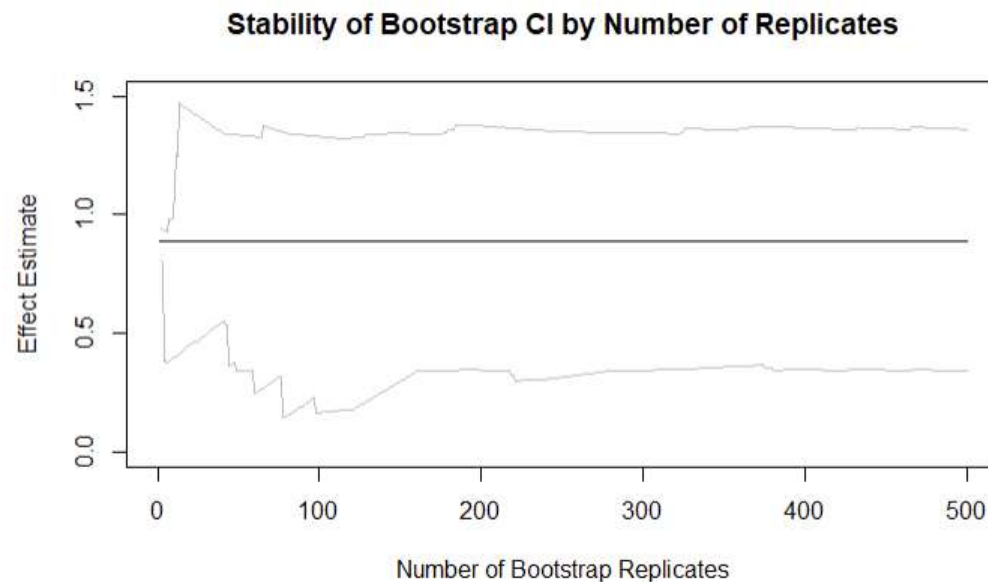
Your initial sample must be representative of the population

It should have all the events from the population

Sampling is with replacement and same size as your sample

In the code, we did 500 replicates. Is that enough?

See plot



Uncertainty – what even is it?

In the traditional approach, uncertainty is “sampling-based”.

The counterfactual here is what would the estimate (of some estimand like mean) be *if we got a different sample of the same size from the population.*

But in the causal setting, uncertainty is “design-based”!

The counterfactual here is what would the estimate (of some estimand like ATT) be *if the treatment assignment was done on different units within the same sample.*

Please read...

Abadie, A., Athey, S., Imbens, G.W. and Wooldridge, J.M. (2020), Sampling-Based versus Design-Based Uncertainty in Regression Analysis. *Econometrica*, 88: 265-296.

<https://doi.org/10.3982/ECTA12675>

No escaping Design-based uncertainty!

Say, you can run an experiment with the entire population data, relevant to your question.

Do you still need to provide a CI for ATT or ATE?

YES! Why?

You got rid of sampling uncertainty by taking your whole population. But you still need to contend with design-based uncertainty!

Said differently...

Now, your hypothesis test is NOT whether your sample estimate can reject the null hypothesis about the population estimand – we have the population already!

But it is about whether your treatment can reject the null hypothesis about there being no treatment effect under random treatment assignment.

9. Inference problem in the causal setup

Randomization inference

TABLE I
SAMPLING-BASED UNCERTAINTY (✓ IS OBSERVED, ? IS MISSING)

Unit	Actual Sample			Alternative Sample I			Alternative Sample II			...
	Y_i	Z_i	R_i	Y_i	Z_i	R_i	Y_i	Z_i	R_i	
1	✓	✓	1	?	?	0	?	?	0	...
2	?	?	0	?	?	0	?	?	0	...
3	?	?	0	✓	✓	1	✓	✓	1	...
4	?	?	0	✓	✓	1	?	?	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...
n	✓	✓	1	?	?	0	?	?	0	...

TABLE II
DESIGN-BASED UNCERTAINTY (✓ IS OBSERVED, ? IS MISSING)

Unit	Actual Sample			Alternative Sample I			Alternative Sample II			...
	$Y_i^s(1)$	$Y_i^s(0)$	X_i	$Y_i^s(1)$	$Y_i^s(0)$	X_i	$Y_i^s(1)$	$Y_i^s(0)$	X_i	
1	✓	?	1	✓	?	1	?	✓	0	...
2	?	✓	0	?	✓	0	?	✓	0	...
3	?	✓	0	✓	?	1	✓	?	1	...
4	?	✓	0	?	✓	0	✓	?	1	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...
n	✓	?	1	?	✓	0	?	✓	0	...

*But how to get
design-based
uncertainty?
Randomization
inference!*

In a regression setting, it is as though one of your regressors is manipulable and you want the regression coefficient to reflect the uncertainty of that manipulation.

10. Bootstrap v Randomization inference

Bootstrap is different from randomization inference!

Bootstrap, while it is a relatively new technique, is still in the old sampling-based uncertainty realm!

So, it is apples v oranges to compare bootstrap and randomization inference.

It looks the same since we are using only the sample in both case.

But it is a completely different philosophy of what uncertainty even means altogether!

BOTTOMLINE

You will see a lot of different techniques in this course. All of them have R packages that report SE of the estimand.

What SE are these packages reporting? Is it sampling-based or design-based? PAY ATTENTION!

I didn't! But I'm going to from now on!

Outline

1

Defining Inference ✓

2

Sampling distribution ✓

3

Confidence intervals ✓

4

Testing hypotheses about means ✓

5

The dangers that lurk in the dark ✓

6

Breaking normality of Sampling Distribution ✓

7

Bootstrap demo ✓

8

Inference problem in the causal setup ✓

9

Bootstrap v Randomization inference ✓

Everything all at once

Summary

- Estimation
- Estimator
- Estimand

Defining Inference

- Counterfactual sample variance
- SEM

Sampling distribution

- Do the sampling a 100 times. Construct a CI around each of those sample's mean. *On an average*, 95 of those CI would contain the population mean.

Confidence intervals

- T distro
- One sample ttest
- Two sample ttest

Testing Hypothesis about means

- Caveats on Hypothesis testing and p-values

The dangers that lurk in the dark

- But weird estimands can break normality or even be not analytically solvable!

Breaking normality of Sampling distribution

- How bootstrap works (without math derivations)

Bootstrap demo

- Sampling v Design-based uncertainty

Inference problem in the causal setup

- Pay attention to different SE that R packages report.

Bootstrap v Randomization inference

The Dunning Kruger effect



If you feel less confident after these workshops, its natural!

Venkatesh Subramanian

Venkatesh.Subramanian@utdallas.edu

See you in next workshop!

Appendix



1. Defining Inference

Note on notation

Estimand	Estimand (Greek)	Estimator (Latin)	Estimate (with Hat)
$\mathbb{V}[X]$	σ^2	s^2	$\widehat{\mathbb{V}[X]}$ or $\widehat{\sigma^2}$
$\mathbb{E}[X]$	μ	$\bar{x}$	$\widehat{\mathbb{E}[X]}$ or $\hat{\mu}$

1. Defining Inference

Is this an Estimator/Estimand/Estimate? (1)

$$\sigma^2 = \sum_{x \in X} (x - \mu)^2 p(x)$$

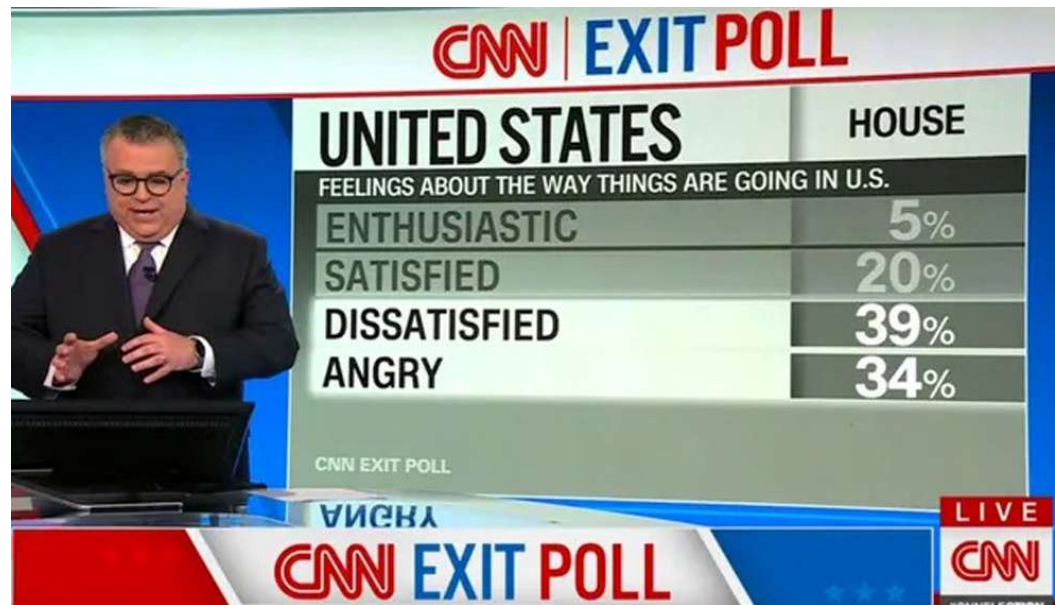
1. Defining Inference

Is this an Estimator/Estimand/Estimate? (2)

$$s^2 = \frac{1}{N} \sum_{i=1}^n (X_i - \bar{X})^2$$

1. Defining Inference

Is this an Estimator/Estimand/Estimate? (3)



1. Defining Inference

Is this an Estimator/Estimand/Estimate? (4)

$$s^2 = \frac{1}{N-1} \sum_{i=1}^n (X_i - \bar{X})^2$$