Share

Making Population Inference Based on Only One Sample

UNIVERSITY OF MICHIGAN

Brady T. West

Research Associate Professor, Survey Research Center, Institute for Social Research

0:01 / 14:20

Lecture Overview

General approaches to making population inferences based on estimated features of sampling distributions

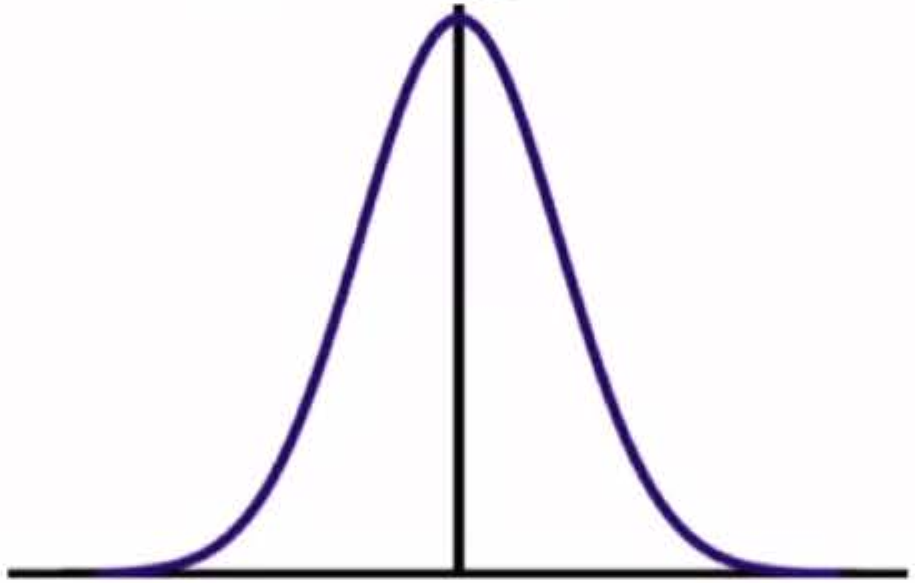
- Confidence Interval Estimate for Parameters of Interest
- Hypothesis Testing about Parameters of Interest

Examples of Parameters of Interest:
a mean, a proportion, a regression coefficient, an odds ratio, and many more!

Some examples of what we mean by a parameter of interest,

Key Assumption: Normality

These approaches assume that sampling distributions for the estimate are (approximately) normal, which is often met if sample sizes are "large"



Q: What if sampling distribution is not (approximately) normal?

A: Alternative inferential approaches discussed in later course

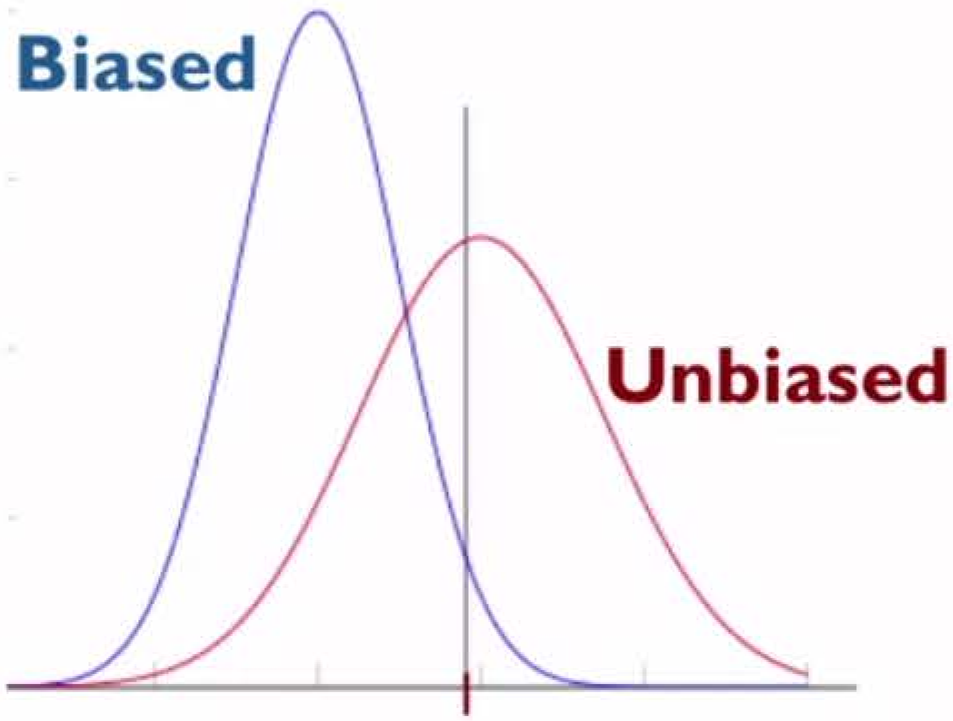
So, how do we go about doing this in step-by-step fashion.

Step 1: Compute the Point Estimate

Compute an **unbiased point estimate** of the parameter of interest

Unbiased Point Estimate:
average of all possible values for point estimate
(a.k.a. *expected value of the point estimate*)
is **equal to true parameter value**

The sampling distribution is centered at the truth!



True Parameter Value
All possible values of
point estimate

Step 1: Compute the Point Estimate

Compute an **unbiased point estimate** of the parameter of interest

Key Idea: want estimate to be **unbiased *with respect to sample design***!

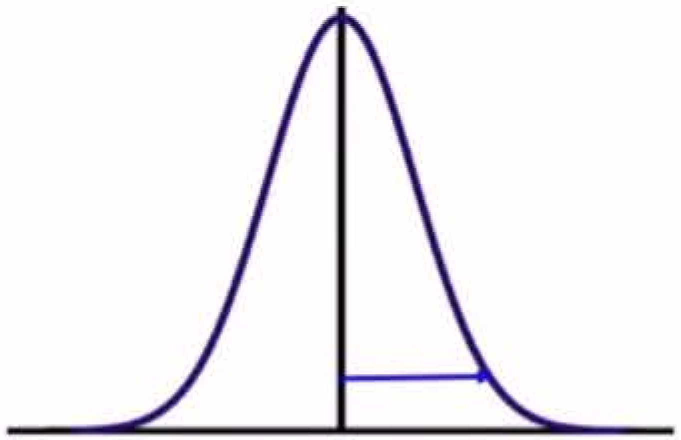
If cases had unequal probabilities of selection, those weights need to be used when computing the point estimate!

Step 2: Estimate the Sampling Variance of the Point Estimate

Compute an **unbiased estimate of the variance** of the sampling distribution for the particular point estimate

Unbiased Variance Estimate:
Correctly describes variance of the sampling distribution *under the sample design used*

Square root of variance = **Standard Error of the Point Estimate**



All possible values of estimate

To Form a Confidence Interval

Best Estimate ± Margin of Error

Best Estimate = Unbiased Point Estimate

Margin of Error = "a few" Estimated Standard Errors

"**a few**" = multiplier from appropriate distribution based on desired confidence level and sample design

95% Confidence Level ↔ 0.05 Significance

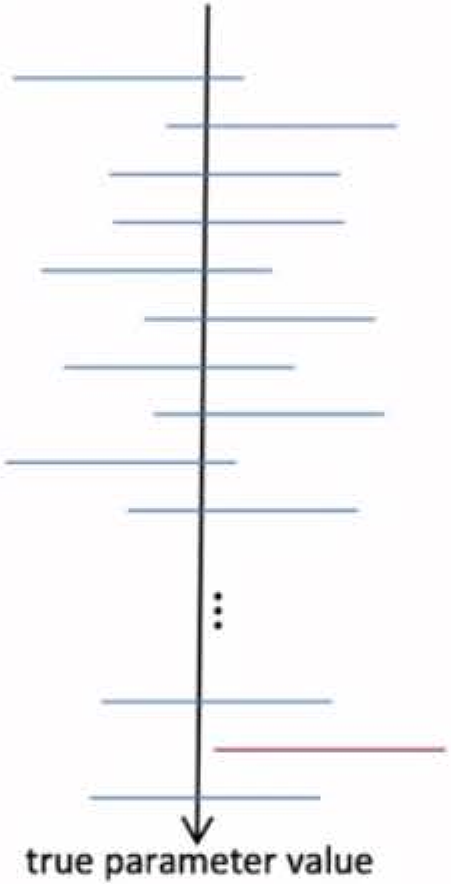
What that equates to is a 0.05 level of significance.

9:28 / 14:20

To Form a Confidence Interval

Best Estimate ± Margin of Error

Key Idea: 95% confidence level
→ expect 95% of intervals
will cover true population value
(if computed in this way in repeated samples)



cover that black line.

To Form a Confidence Interval

Best Estimate ± Margin of Error

Caution: important to get all 3 pieces right for correct inference!

If best estimate is *not unbiased point estimate*
OR if margin of error does *not use correct multiplier*
or does *not use unbiased estimate of the standard error*
→ confidence interval will not have the advertised **coverage!**

the confidence interval will not have

Suppose that an analyst correctly computes an estimate of sampling variance and a correct multiplier for computing a 95% confidence interval but forgets to account for the sample design features (captured in the survey weights) when computing the estimate. What is the problem with the resulting 95% confidence interval?

- The confidence interval will be too narrow.
- The confidence interval will not be centered at an unbiased estimate of the parameter of interest, meaning that inference about the parameter based on the confidence interval may be biased.

Correct

If the sample design (via the survey weights) is not accounted for when computing the population estimate, the estimate may be biased, and the confidence interval may be shifted away from the true population parameter. Inferences about the true population parameter may be biased as a result.

- Nothing: 95% of estimates will fall within the resulting interval, as the sample design only needs to be accounted for when estimating the sampling variance.
- Nothing: 95% of intervals computed this way across repeated samples will cover the true parameter, as the sample design only needs to be accounted for when estimating the sampling variance.

To Form a Confidence Interval

Best Estimate ± Margin of Error

Key Idea:

Interval = *range of reasonable values* for parameter

If hypothesized value for parameter lies *outside* confidence interval,
we don't have evidence to support that value
at corresponding significance level

we generally don't have enough statistical evidence to support

To Test Hypotheses

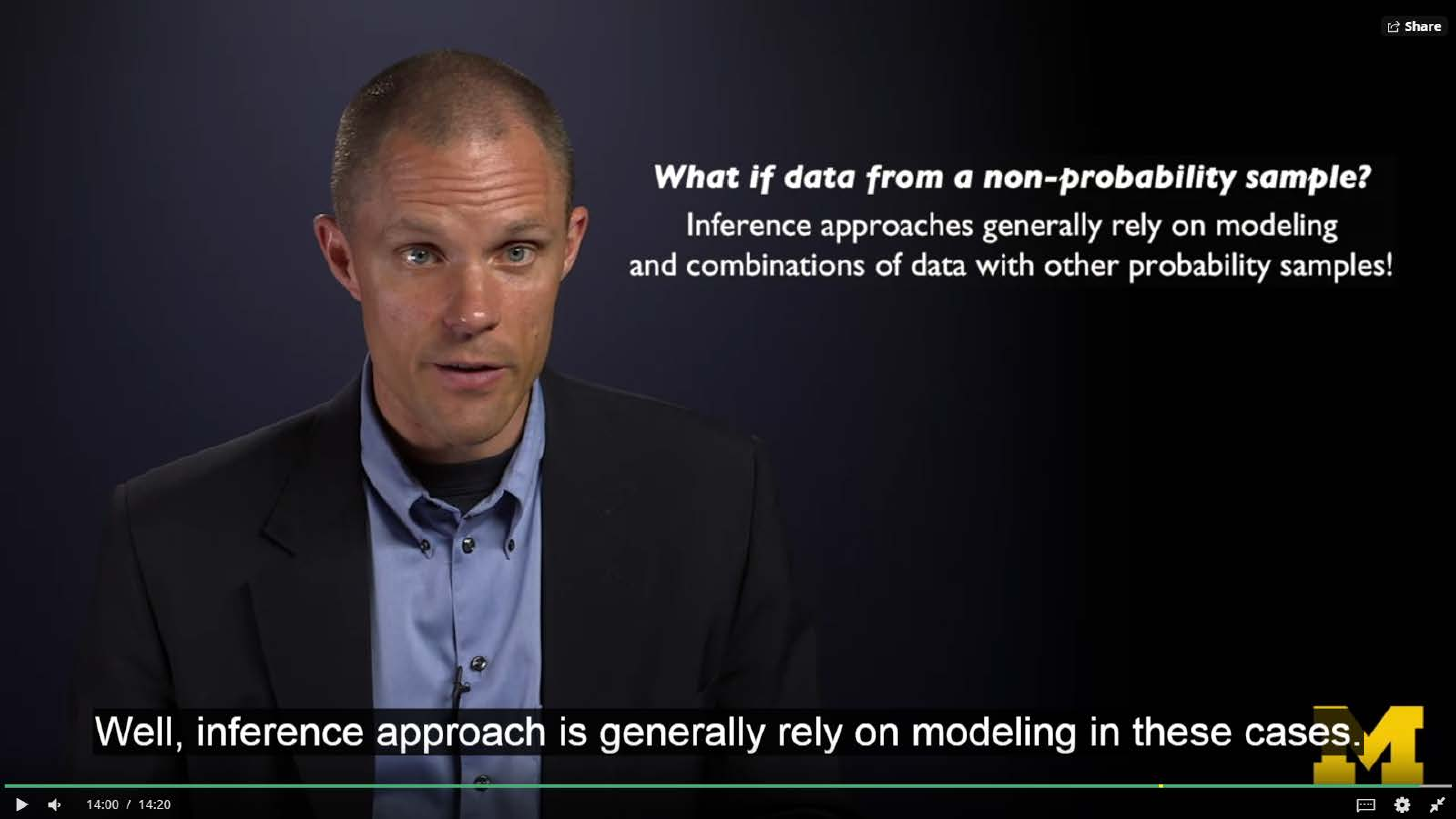
hypothesized or 'null' value

- Hypothesis: Could the value of the parameter be ________?
- Is point estimate for parameter close to this null value or far away?
- Use standard error of point estimate as yardstick

$$Test\ Statistic = \frac{(estimate - null\ value)}{standard\ error}$$

- If the null is true, what is the probability of seeing a test statistic this extreme (or more extreme)? If probability small, reject the null!


Share
Important Reminder!
These inferential procedures are valid
if probability sampling was used!
All the techniques that we just described rely on the idea of a probability
sample.
13:49 / 14:20


Share
What if data from a non-probability sample?
Inference approaches generally rely on modeling
and combinations of data with other probability samples!
Well, inference approach is generally rely on modeling in these cases.
14:00 / 14:20