Making Population Inference Based on Only One Sample

Brady T. West

Lecture Overview

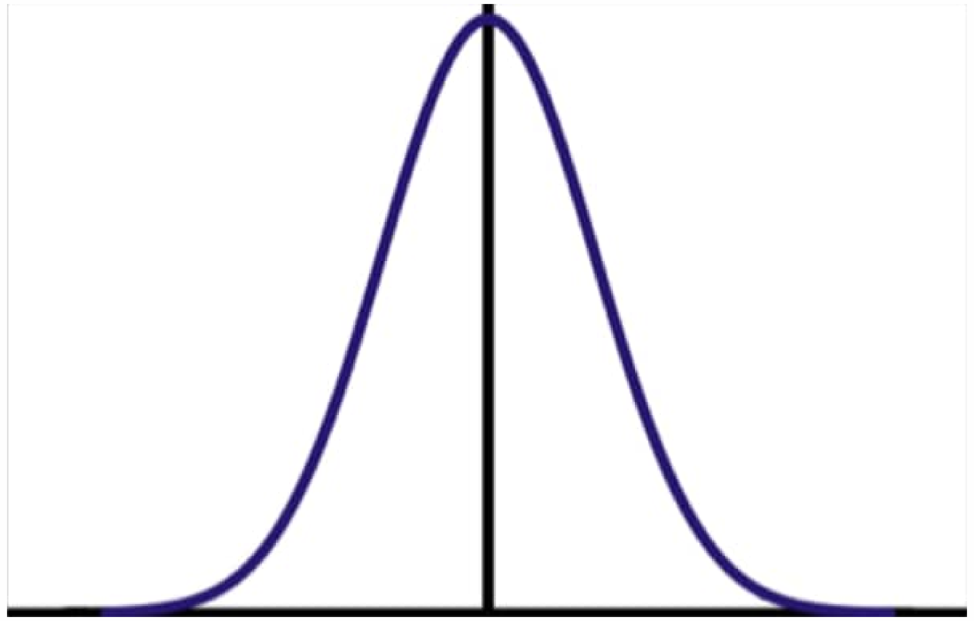
General approaches to making population inferences based on estimated features of sampling distributions

- 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!

Key Assumption: Normality

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



All possible values of estimate

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

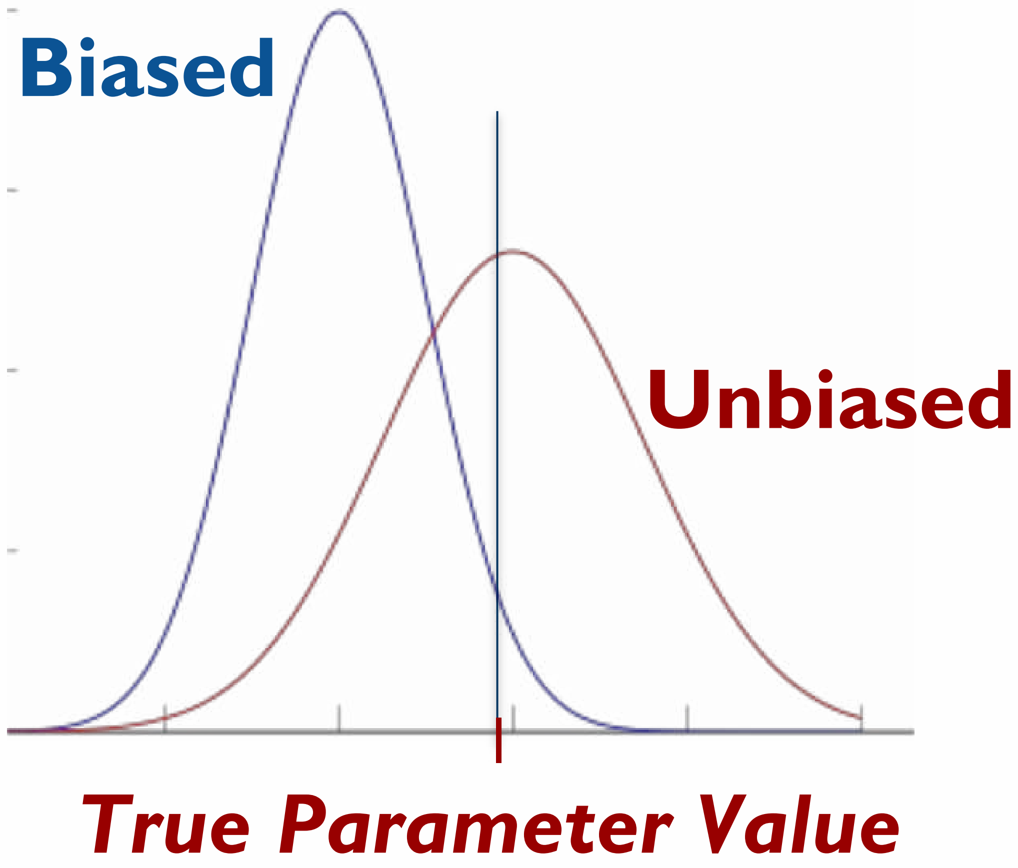
A: Alternative inferential approaches discussed in later course

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!



All possible values of the 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***!

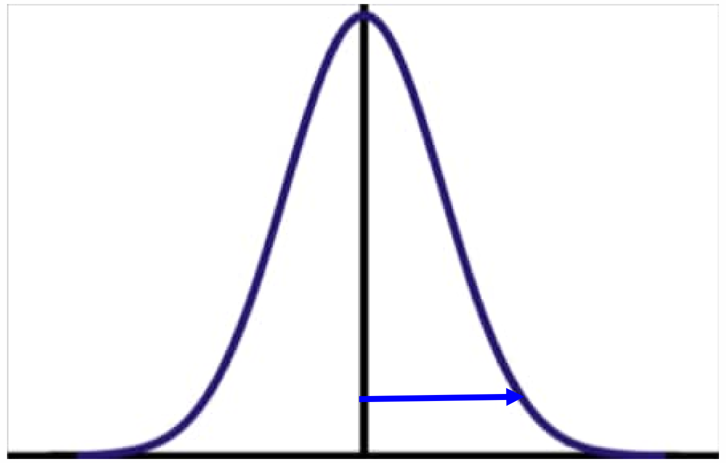
If cases had unequal probabilities of selection, those weights need to be used when computing the point estimate!

Step 2: Estimate the Sampling Variance of the Point Estimate

Compute an **<u>unbiased estimate of the variance</u>** 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 <u>estimate</u>

To Form a Confidence Interval

Best Estimate ± Margin of Error

Best Estimate = Unbiased Point Estimate

Margin of Error = "a few" Estimated Standard Errors

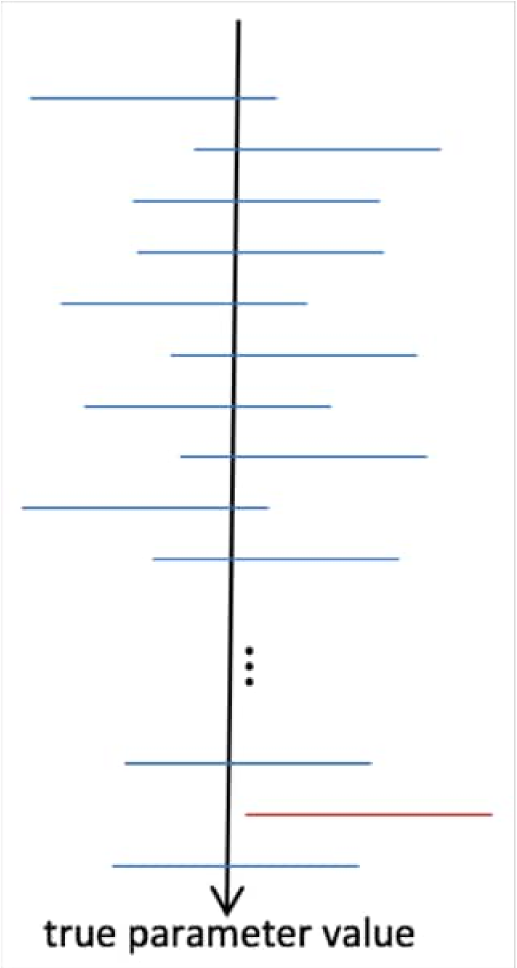
"**a few**" = multiplier from appropriate distribution based on desired confidence level and sample design

95% Confidence Level **0.05 Significance**

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)



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!**

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

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!

Important Reminder!

These inferential procedures are valid
if probability sampling was used!

What if data from a non-probability sample?

Inference approaches generally rely on modeling
and combinations of data with other probability samples!