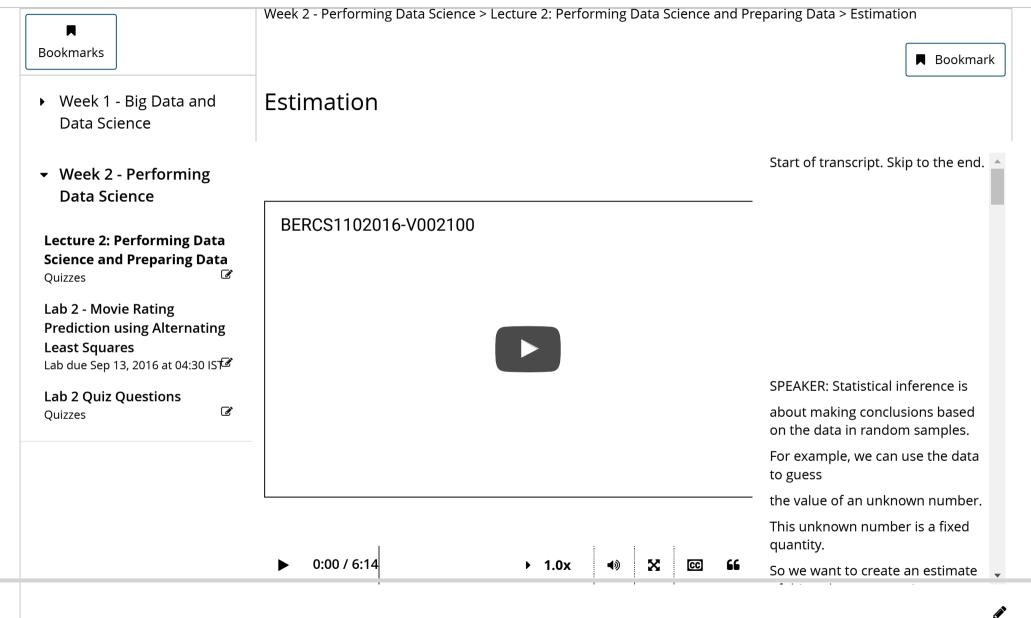


## BerkeleyX: CS110x Big Data Analysis with Apache Spark



Download video Download transcript .srt

The actual number of planes in this segment is 300. Our two population estimation methods are based on statistical approaches, such as the Method of moments and Maximum likelihood estimation (popularized by Fisher).

**A technical note**: the estimate based on doubling the average of the observed serial numbers is not unbiased. On average, it overestimates by exactly 1. For example, if N is 3, the average of draws from 1, 2, and 3 will be 2, and 2 times 2 is 4, which is one more than N. Twice the average minus 1 is an unbiased estimator of N.

# Estimation Using the Largest Serial Number

(1/1 point)

Which of the following choices is true for an estimate that is based on the largest serial number observed?

- The estimate will underestimate N
- The estimate has high variability
- The estimate has no bias

The estimate will always be very accurate

#### **EXPLANATION**

Unless the plane with serial number N is observed, the estimate will underestimate the true value of N. The estimate however, will have low variability and is biased to the left.

# Bias-Variance Tradeoff

(1/1 point)

Is it better to choose an estimation approach that has low variability but is biased, or an approach that has high variability but is unbiased?

- O It is always better to choose approaches with low variability over approaches that are unbiased
- It is always better to choose approaches that are unbiased over approaches with low variability
- The choice depends on the problem domain and whether (in the context of the problem domain)
  variability is more important than bias
- The choice is unimportant

### **EXPLANATION**

The choice of approaches and whether to prioritize bias or variability depends on the problem domain.



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.















