Mathematics 231

Lecture 20

Liam O'Brien

Announcements

- Reading
  - Today M&M 6.1 353-363

Topics

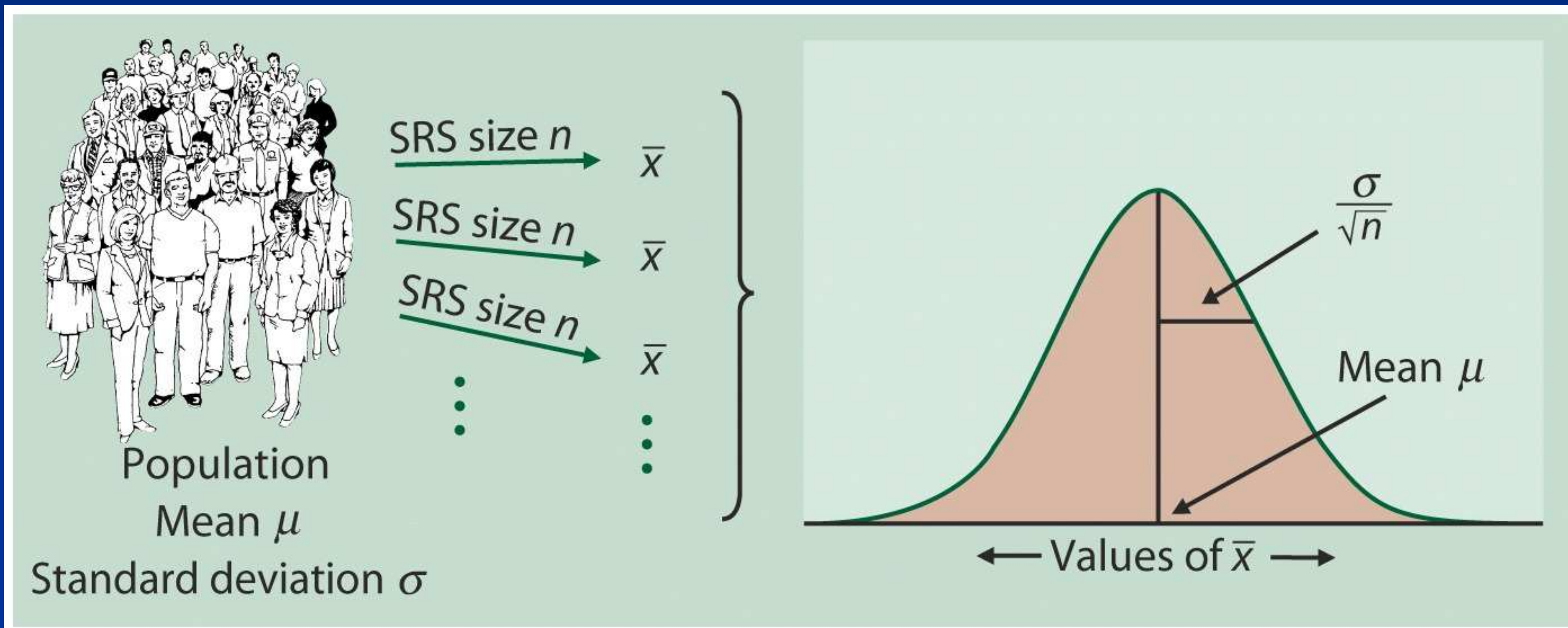
- Confidence intervals for μ with known σ

Inference

- **Inference**: From "part" (sample) infer about the "whole" (population).
- **Statistical inference**: Process of drawing conclusions about population characteristics based on information from a sample.
- Need to be able to quantify the uncertainty inherent in our inferences.

Confidence Intervals

- **Confidence intervals** for estimating a population parameter (e.g., the mean μ) are based on the sampling distribution of statistics.
- As a result, they report probabilities that state what would happen if we used the method many times.
- To introduce the notion of a confidence interval, we first make the unrealistic assumption that the standard deviation, σ, is known (we relax this later).


SRS size n
SRS size n
SRS size n
$\bar{x}$
$\bar{x}$
$\bar{x}$
Population
Mean μ
Standard deviation σ
$\frac{\sigma}{\sqrt{n}}$
Mean μ
Values of $\bar{x}$

Sampling Distribution of a Sample Mean

- Properties of sampling distribution of a sample mean:

1) $\mu_{\bar{X}} = \mu$

2) $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

3) The distribution is normal as $n \to \infty$

The third property is due to the Central Limit Theorem.

Confidence Intervals

- Consider a population with mean μ and standard deviation σ.
- Assume unrealistically that σ is known.
- Take an SRS of size n from this population: from this sample we calculate the sample mean.
- Given the sample mean, what can we say about the population mean, μ?

Example: Housing Prices

- Want to estimate the mean housing price for an area of coastal Maine.
- The true mean, μ, in the population is unknown.
- Assume σ is known and is 46 (unrealistically)
- Take an SRS of size n = 100 from this population and calculate the sample mean.
- Given the sample mean, what can we say about μ?

Unbiased but Variable

- Sample mean is an unbiased estimator of μ (from property 1 of sampling distributions).
- But how precise of an estimate does it provide?
- Would a second sample of size 100 produce the same estimate of μ?
- To answer this question we must consider variability.

Construction of Confidence Intervals

- Take an SRS of size n from a population with mean μ and (assume) known standard deviation σ.

Sample: SRS of size $n \rightarrow \bar{x}$

$$\bar{x} \sim N\left(\mu, \sigma/\sqrt{n} \right)$$

Recall: Empirical Rule says that with probability

close to 0.95 the sample mean will be $2\frac{\sigma}{\sqrt{n}}$ points

(or 2 SD of $\bar{x}$) of the population mean μ.

Construction of Confidence Intervals

Note: to say $\bar{x}$ is within $2\frac{\sigma}{\sqrt{n}}$ points (or 2 SD of $\bar{x}$) of the population mean μ is equivalent to saying that μ is within $2\frac{\sigma}{\sqrt{n}}$ points of $\bar{x}$.

So, in approximately 95% of all samples of size n, the interval $\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$ will cover μ.

The Gory Details

$$\text{We know } P\left(\mu - 2\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

$$= P\left(\bar{x} < \mu + 2\frac{\sigma}{\sqrt{n}} \text{ AND } \bar{x} > \mu - 2\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\bar{x} - 2\frac{\sigma}{\sqrt{n}} < \mu \text{ AND } \bar{x} + 2\frac{\sigma}{\sqrt{n}} > \mu\right)$$

$$= P\left(\bar{x} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2\frac{\sigma}{\sqrt{n}}\right)$$

Main Point

$$P\left(\mu - 2\frac{\sigma}{\sqrt{n}} \quad < \quad \bar{x} \quad < \quad \mu + 2\frac{\sigma}{\sqrt{n}}\right)$$

fixed random fixed

$$P\left(\bar{x} - 2\frac{\sigma}{\sqrt{n}} \quad < \quad \mu \quad < \quad \bar{x} + 2\frac{\sigma}{\sqrt{n}}\right)$$

random fixed random

Statement of Confidence

Based on the sampling distribution of $\bar{x}$ and the empirical rule, we can now state that we are about 95% confident that the interval $\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$ will cover μ.

This corresponds to a statement of our uncertainty in using $\bar{x}$ to estimate μ.

Increased Accuracy

The empirical rule is only approximate and is never used in reality by a statistician.

Using normal tables, a 95% confidence interval for μ is $\bar{x} \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$.

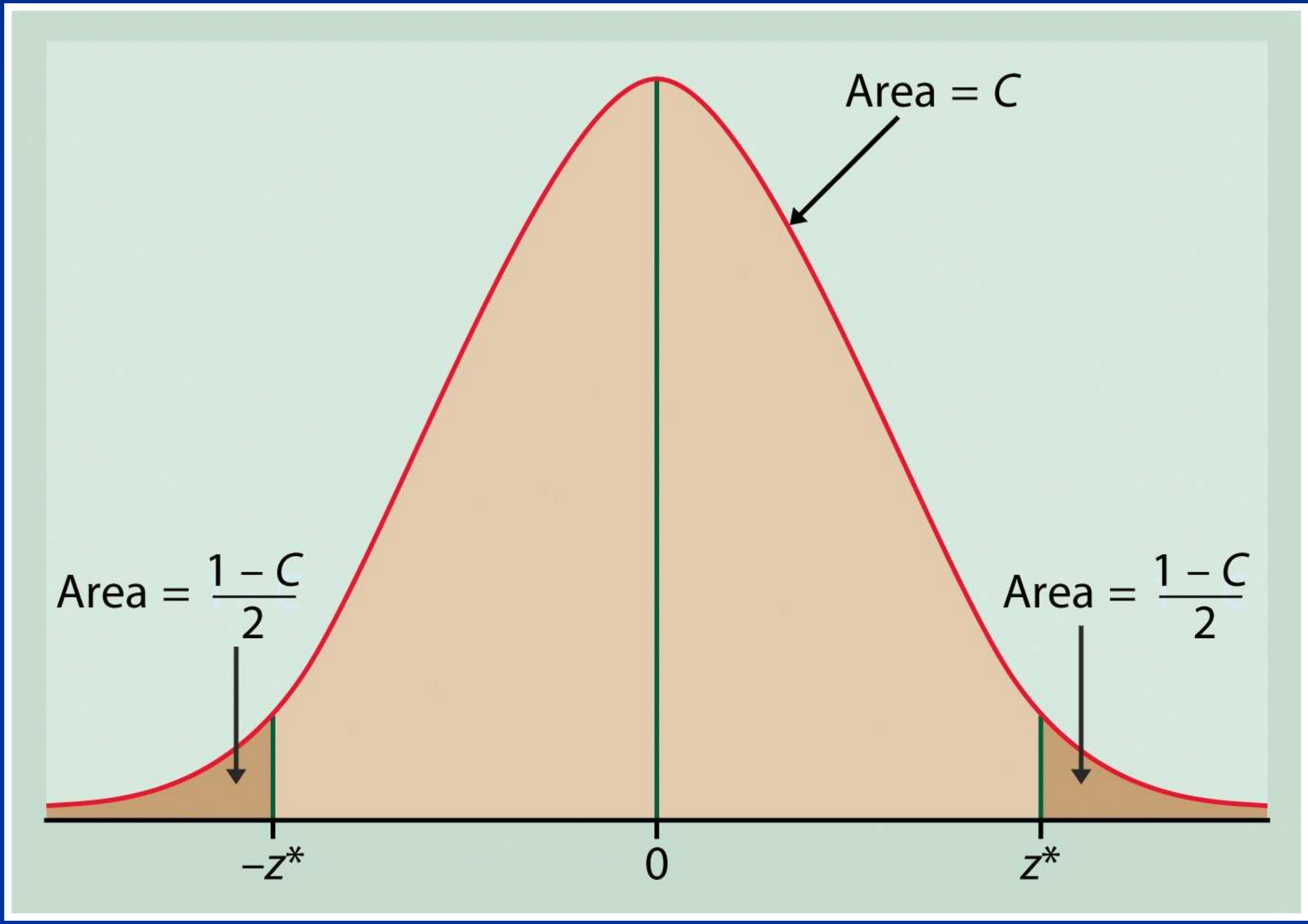
Similarly, a 99% confidence interval for μ is

$$\bar{x} \pm 2.57 \frac{\sigma}{\sqrt{n}}.$$

Confidence interval for μ has the form $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

where $z^* = 1.96$ for 95%.

In general, CI has the form: estimate $\pm\ z^* \sigma_{\text{estimate}}$



Example: Housing Prices

- Want to estimate the mean housing prices in an area of coastal Maine.
- The true mean, μ, in the population is unknown.
- Assume σ is known to be 46 (unrealistic).
- Take an SRS of size n = 100: mean = 220.
- What is the 95% confidence interval for μ?

Example: Housing Prices

95% confidence interval for $\mu : \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

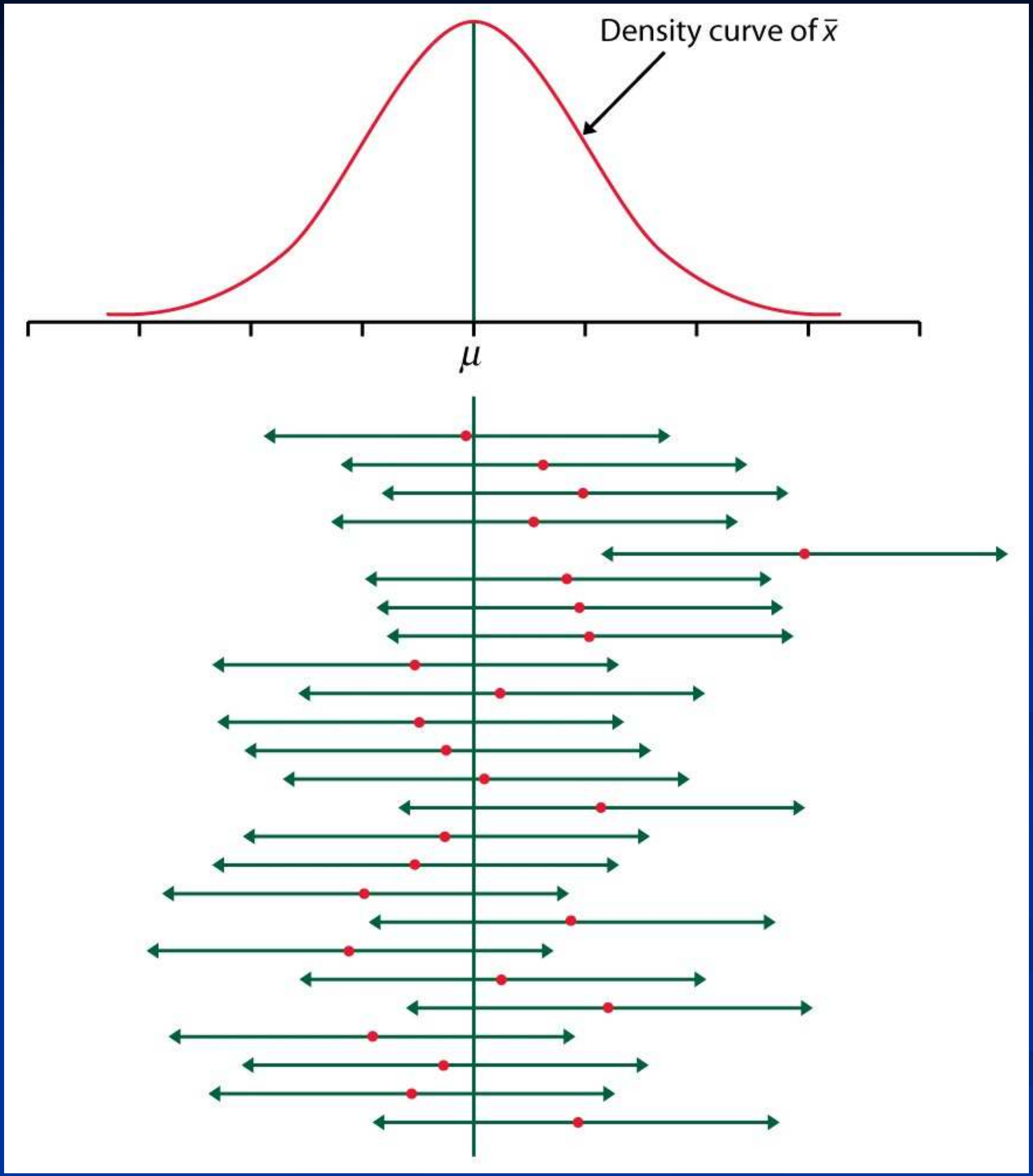
95% confidence interval for μ:

$$220 \pm 1.96 \frac{46}{\sqrt{100}} = 220 \pm 1.96(4.6)$$

$$(210.98, 229.02)$$

Interpreting Confidence Intervals

- Does this mean that the probability that the population mean, μ, is between \$210.98 thousand and \$229.98 thousand is 0.95?
- NO!!!!!!!
- If we drew 100 random samples of size 100 and calculated a 95% confidence intervals for each (such that we have 100 intervals), then about 95 of those intervals would cover the true population mean, μ.


Density curve of $\bar{x}$
μ

Interpretation

- Why are the following incorrect?
  - "The interval (197,233) is a 95% interval for the sample mean."
  - "In 95% of all possible samples, the sample mean will lie in the interval (197,233)"
  - "There is a 95% probability that the population mean lies in the interval (197,233)."

Interpretation

- Correct Interpretations
  - "There is a 95% probability that the interval generated from a random sample will contain the population mean."
  - "A plausible range of values for the true population mean is (197,233)."
  - "If we repeatedly calculate confidence intervals using this procedure, with different random samples each time, 95% of these intervals will cover the population mean."

Some Features of Confidence Intervals

- **Confidence level** gives the probability the interval covers the population parameter (e.g., μ). Conventionally 95% is chosen, but any value can be used.

95% CI: $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

99% CI: $\bar{x} \pm 2.57 \frac{\sigma}{\sqrt{n}}$

Some Features of Confidence Intervals

- **Margin of error:**

Probability is 0.95 that the interval $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$

contains the population mean, μ.

Put another way, $\bar{x}$ is an estimate of μ and the

margin of error is $1.96\frac{\sigma}{\sqrt{n}}$.

How to Reduce the MOE

1. Increase the sample size.

2. Use a lower level of confidence.

3. Reduce σ.

Sample Size and CI's

The margin of error of the 95% CI

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} \text{ is } 1.96 \frac{\sigma}{\sqrt{n}}$$

To obtain a certain margin of error, m,

$$\text{set } m = 1.96 \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{1.96\sigma}{m} \right)^2$$

Note it is the size of the sample, not of the popluation that determines the MOE.