Introduction to Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

October 6, 2019

Topics

Objectives

Objectives:

- Distinguish between a point estimate and a confidence interval.
- Compute and interpret a confidence interval for a population mean when the population standard deviation is known.
- Describe what effect the level of confidence and sample size have on the width of a confidence interval.

Point Estimates and Confidence Intervals (8.1)

- Statistics like $\bar{x}$ and s, when used to **estimate** unknown population parameters like μ and σ, are sometimes called ***point estimates*** because they consist of *single values*.

Point Estimates and Confidence Intervals (8.1)

- Statistics like $\bar{x}$ and s, when used to **estimate** unknown population parameters like μ and σ, are sometimes called ***point estimates*** because they consist of *single values*.
- But a *point estimate* won't equal the *true value* exactly because of **sampling error.**

Point Estimates and Confidence Intervals (8.1)

- Statistics like $\bar{x}$ and s, when used to **estimate** unknown population parameters like μ and σ, are sometimes called ***point estimates*** because they consist of *single values*.
- But a *point estimate* won't equal the *true value* exactly because of **sampling error.**
- It's preferable to attach a ***margin of error*** to a point estimate indicating how large the sampling error might be.

- A ***confidence interval*** (or ***CI***) for a population parameter is an interval of the form

  $$\textbf{Point Estimate} \pm \textbf{Margin of Error}$$

  and is interpreted as a whole **range of plausible values** for the true (unknown) population parameter.

- A ***confidence interval*** (or ***CI***) for a population parameter is an interval of the form

$$\textbf{Point Estimate} \pm \textbf{Margin of Error}$$

  and is interpreted as a whole **range of plausible values** for the true (unknown) population parameter.

- Each CI has an associated ***level of confidence*** indicating **how sure we can be** that the true (unknown) value of the **population parameter** is **contained in the interval.**

- We ***choose*** the level of confidence ***prior*** to computing a CI.

- We ***choose*** the level of confidence ***prior*** to computing a CI.

  Our choice **affects how wide** the interval will be. (More on this later.)

Confidence Interval for μ when σ is Known (8.2, 8.3)

- Suppose we have a **random sample** from a **population** whose **mean** and **standard deviation** are μ and σ.

Confidence Interval for μ when σ is Known (8.2, 8.3)

- Suppose we have a **random sample** from a **population** whose **mean** and **standard deviation** are μ and σ.

  Suppose also that:

  σ is ***known***.
  μ is ***unknown***.

- We want to **estimate** μ using a **CI**, which will be of the form

$$\bar{x} \pm \text{Margin of Error}$$

- We want to **estimate** μ using a **CI**, which will be of the form

$$\bar{x} \pm \text{Margin of Error}$$

On the next slides, we'll determine how big the **margin of error** would need to be for us to be **95% confident** that the interval will contain μ.

- We know (Slides 13) that $\bar{x}$ follows a **normal** distribution with mean and standard error

$$\mu_{\bar{x}} = \mu \qquad \text{and} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- We know (Slides 13) that $\bar{x}$ follows a **normal** distribution with mean and standard error

$$\mu_{\bar{x}} = \mu \qquad \text{and} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(when either the sample is from a **normal** population or the sample size n is **large**).

- Thus the **standardized version of** $\bar{x}$,

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}},$$

follows a **standard normal** distribution (Slides 11) **...**

- Thus the **standardized version of $\bar{x}$**,

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}},$$

follows a **standard normal** distribution (Slides 11) **...**

... and therefore will lie between

$$-z_{0.025} = -1.96 \qquad \text{and} \qquad z_{0.025} = 1.96$$

95% of the time when we take a sample of size n.

- In other words,

$$-1.96 \;\le\; \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \;\le\; 1.96\,.$$

95% of the time.

- In other words,

$$-1.96 \;\leq\; \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \;\leq\; 1.96\,.$$

95% of the time.

- We'll "solve" for μ:

- In other words,

$$-1.96 \;\leq\; \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \;\leq\; 1.96\,.$$

95% of the time.

- We'll "solve" for μ:

Multiplying through by $\sigma/\sqrt{n}$, we can rewrite this as

$$-1.96\frac{\sigma}{\sqrt{n}} \;\leq\; \bar{x}-\mu \;\leq\; 1.96\frac{\sigma}{\sqrt{n}}\,,$$

- In other words,

$$-1.96 \;\leq\; \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \;\leq\; 1.96\,.$$

95% of the time.

- We'll "solve" for μ:

Multiplying through by $\sigma/\sqrt{n}$, we can rewrite this as

$$-1.96\frac{\sigma}{\sqrt{n}} \;\leq\; \bar{x}-\mu \;\leq\; 1.96\frac{\sigma}{\sqrt{n}}\,,$$

which says the ***sampling error*** **won't be bigger than** $\mathbf{1.96\sigma/\sqrt{n}}$ **95% of the time.**

- (cont'd)

  Subtracting $\bar{x}$ from all three terms gives

$$-\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \;\leq\; -\mu \;\leq\; -\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}.$$

- (cont'd)

  Subtracting $\bar{x}$ from all three terms gives

$$-\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \;\leq\; -\mu \;\leq\; -\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}.$$

  Multiplying each of the three terms above by -1 (which changes the direction of the inequalities) gives

$$\bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \;\geq\; \mu \;\geq\; \bar{x} - 1.96\frac{\sigma}{\sqrt{n}}.$$

- (cont'd)

  Subtracting $\bar{x}$ from all three terms gives

  $$-\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \;\leq\; -\mu \;\leq\; -\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}.$$

  Multiplying each of the three terms above by -1 (which changes the direction of the inequalities) gives

  $$\bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \;\geq\; \mu \;\geq\; \bar{x} - 1.96\frac{\sigma}{\sqrt{n}}.$$

  Finally, reordering the terms, we get that **95% of the time**,

  $$\boldsymbol{\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \;\leq\; \mu \;\leq\; \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}}.$$

- (cont'd)

  Thus we can be **95% confident that μ will be between**

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \qquad \text{and} \qquad \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}.$$

95% One-Mean z CI for μ: When the population standard deviation σ **is known**, a 95% confidence interval for the **unknown** population **mean** μ is:

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

and the **margin of error** is

$$\text{Margin of Error} \ = \ 1.96 \frac{\sigma}{\sqrt{n}}.$$

(The CI and margin of error are valid when either the sample is from a **normal** population or the sample size n is **large**.)

- For **other levels of confidence**, we replace 1.96 by the appropriate so-called ***z critical value***:

Commonly Used Z Critical Values:

$$
\begin{aligned}
z_{0.05} &= 1.645 && \text{for a 90\% level of confidence} \\
z_{0.025} &= 1.96 && \text{for a 95\% level of confidence} \\
z_{0.005} &= 2.58 && \text{for a 99\% level of confidence}
\end{aligned}
$$

These $z_{\alpha/2}$ values are obtained from Table II.

One-Mean z CI for μ: When the population standard deviation σ **is known**, the ***one-mean z confidence interval*** for the **unknown** population **mean** μ is

$$\bar{x} \ \pm \ z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

with ***margin of error***

$$\text{Margin of Error} \ = \ z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

where $z_{\alpha/2}$ is a z **critical value** and α is either **0.10**, **0.05**, or **0.01**, depending on the level of confidence (see the next slide).

$\alpha = \mathbf{0.10}$ for a **90%** level of confidence (so $1-\alpha = \mathbf{0.90}$, $\alpha/2 = \mathbf{0.05}$, and $z_{\mathbf{0.05}} = \mathbf{1.645}$)

$\alpha = \mathbf{0.05}$ for a **95%** level of confidence (so $1-\alpha = \mathbf{0.95}$, $\alpha/2 = \mathbf{0.025}$, and $z_{\mathbf{0.025}} = \mathbf{1.96}$)

$\alpha = \mathbf{0.01}$ for a **99%** level of confidence (so $1-\alpha = \mathbf{0.99}$, $\alpha/2 = \mathbf{0.005}$, and $z_{\mathbf{0.005}} = \mathbf{2.58}$)

(The CI and margin of error are valid when either the sample is from a **normal** population or the sample size n is **large**.)

Example

The National Assessment of Educational Progress Study examined quantitative skills of young adult Americans. Men aged 21 to 25 years were given a short test of their quantitative skills. Scores on the test range from 0 to 500.

In a sample of $n = 20$ young men who took the test, the sample mean score was

$$\bar{x} = 272$$

Suppose it's reasonable to assume that the distribution of scores in the population is **normal** with **known standard deviation** $\sigma = 60$, but with a mean μ whose value is **unknown.**

Suppose it's reasonable to assume that the distribution of scores in the population is **normal** with **known standard deviation** $\sigma = 60$, but with a mean μ whose value is **unknown.**

A **95% CI** for μ is

$$\bar{X} \pm z_{0.025} \frac{\sigma}{\sqrt{n}} \quad = \quad 272 \pm 1.96 \times \frac{60}{\sqrt{20}}$$

Suppose it's reasonable to assume that the distribution of scores in the population is **normal** with **known standard deviation** $\sigma = 60$, but with a mean μ whose value is **unknown.**

A **95% CI** for μ is

$$\begin{aligned} \bar{X} \pm z_{0.025} \frac{\sigma}{\sqrt{n}} &= 272 \pm 1.96 \times \frac{60}{\sqrt{20}} \\ &= 272 \pm 26.3 \end{aligned}$$

Suppose it's reasonable to assume that the distribution of scores in the population is **normal** with **known standard deviation** $\sigma = 60$, but with a mean μ whose value is **unknown.**

A **95% CI** for μ is

$$\begin{aligned}\bar{X} \pm z_{0.025} \frac{\sigma}{\sqrt{n}} &= 272 \pm 1.96 \times \frac{60}{\sqrt{20}} \\ &= 272 \pm 26.3 \\ &= \mathbf{(245.7,\ 298.3)}\end{aligned}$$

Suppose it's reasonable to assume that the distribution of scores in the population is **normal** with **known standard deviation $\sigma = 60$**, but with a mean μ whose value is **unknown.**

A **95% CI** for μ is

$$\begin{aligned} \bar{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} &= 272 \pm 1.96 \times \frac{60}{\sqrt{20}} \\ &= 272 \pm 26.3 \\ &= \mathbf{(245.7,\ 298.3)} \end{aligned}$$

We can be **95% confident** that the true (unknown) mean μ is in this interval somewhere.

If we had used a **90% level of confidence** instead, the **critical value** would've been $z_{0.05} = 1.64$ and we'd have ended up with

$$(\mathbf{250.0},\ \ \mathbf{294.0})$$

Note that this interval is **narrower** than the 95% interval (which was (245.7, 298.3)).

Properties and Interpretation of Confidence Intervals

(8.1, 8.2, 8.3)

- **Interpretation of Confidence Intervals**
  - A CI provides a range of **plausible values** for the **unknown population parameter** (e.g. μ).

Properties and Interpretation of Confidence Intervals

(8.1, 8.2, 8.3)

- **Interpretation of Confidence Intervals**
  - A CI provides a range of **plausible values** for the **unknown population parameter** (e.g. μ).
  - The **level of confidence** says how **confident** we can be that the (unknown) value of the population parameter (e.g. μ) is **within the CI**.

Properties and Interpretation of Confidence Intervals

(8.1, 8.2, 8.3)

- **Interpretation of Confidence Intervals**
  - A CI provides a range of **plausible values** for the **unknown population parameter** (e.g. μ).
  - The **level of confidence** says how **confident** we can be that the (unknown) value of the **population parameter** (e.g. μ) is **within the CI**.

    More precisely, it's the percentage of samples (of a given size) from the population that would produce a CI that contains the **population parameter** (e.g. μ).

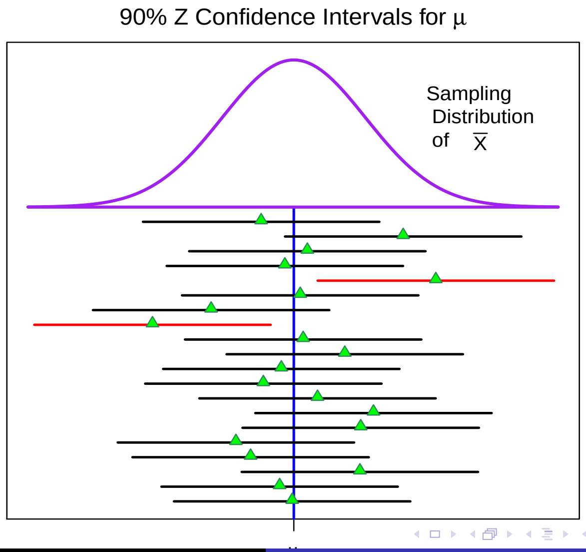
Properties and Interpretation of Confidence Intervals

(8.1, 8.2, 8.3)

- **Interpretation of Confidence Intervals**
  - A CI provides a range of **plausible values** for the **unknown population parameter** (e.g. μ).
  - The **level of confidence** says how **confident** we can be that the (unknown) value of the **population parameter** (e.g. μ) is **within the CI**.

    More precisely, it's the percentage of samples (of a given size) from the population that would produce a CI that contains the **population parameter** (e.g. μ).

    For example, a confidence level of 90% implies that 90% of all samples would give a CI that contains μ.


90% Z Confidence Intervals for μ
Sampling Distribution of $\overline{X}$

- **Properties of Confidence Intervals**
  - A **higher level of confidence** leads to a **wider CI** (so that we can be *more confident* that it contains μ).

- **Properties of Confidence Intervals**
  - A **higher level of confidence** leads to a **wider CI** (so that we can be *more confident* that it contains μ).
  - A **larger sample size** n leads to a **narrower CI** (because the *margin of error* will be *smaller*).