Probability and Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

April 15, 2019

Notes

Topics

Notes

Objectives

Objectives:

- Calculate and interpret:
  - CI for a normal mean μ when σ is known.
  - CI for a general population mean μ when σ is unknown and n is large.
  - CI for a normal mean μ when σ is unknown.
- Determine the sample size n for attaining a CI width under each of these three scenarios.

Notes

Introduction to Confidence Intervals (7.1)

- A single-number estimate (such as $\bar{X}$) of a population parameter (such as μ) is called a ***point estimate***.
- The difference between the **point estimate** and the **true value** is called the ***sampling error*** of the estimate.

  In particular, the sampling error of the sample mean is

$$\textbf{Sampling Error} = \bar{X} - \mu .$$

- A point estimate, by itself, doesn't indicate how big the sampling error might be.

Notes

- It's preferable, therefore, to instead use an **interval estimate**, or ***confidence interval*** (or ***CI***), consisting of a range of estimates for the (unknown) population parameter.
- The first step in computing a CI is to choose a ***level of confidence***, which measures degree of reliability of the interval.

Notes

Example

In a study of sleep deprivation among students at a university, a sample of $n = 22$ were students asked students how many hours they sleep per night.

The mean was $\bar{x} = 5.77$ hours with a standard deviation of $s = 1.57$ hours.

Notes

A **95% confidence interval** for $\boldsymbol{\mu}$, the mean number of hours slept per night in the student population, is

$$\mathbf{(5.07,\ 6.47)}$$

Any value of μ in this range is considered plausible, at the 95% confidence level.

Notes

CI for a Normal Mean μ When σ is Known (7.1)

- Suppose
  1. $X_1, X_2, \ldots, X_n$ are a random sample from a **normal** population.
  2. The population mean $\boldsymbol{\mu}$ **is unknown** but the standard deviation $\boldsymbol{\sigma}$ **is known**.
- We'll first derive a CI for $\boldsymbol{\mu}$ using **level of confidence 95%**.

Notes

- We know that

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right),$$

so

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1).$$

Notes

- Recall that $z_{0.025} = 1.96$, so the area under the standard normal curve between -1.96 and 1.96 is **0.95**.
- Therefore,

$$\begin{aligned} 0.95 &= P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) \\ &= P\left(-1.96\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96\frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(-\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(\bar{X} + 1.96\frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right). \end{aligned}$$

Notes

- The last line above says that we can be **95% confident** that μ will be in the interval

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}.$$

95% Confidence Interval for a Normal Mean μ When σ is Known:

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}.$$

This CI is valid when the sample is from a normal population and σ is known.

Notes

- For other levels of confidence, we replace 1.96 by the appropriate **critical value** $z_{\alpha/2}$.

100$(1-\alpha)$% Confidence Interval for a Normal Mean μ When σ is Known:

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

This CI is valid when the sample is from a normal population and σ is known.

- Commonly used levels are **90%**, **95%**, and **99%**.

Notes

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

$$\bar{x} = 272$$

Notes

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% confidence interval** for μ is

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

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

Notes

If we use a **90% level of confidence** instead, the **critical value** is $z_{0.05} = 1.64$ and we end up with

$$(250.0,\ 294.0)$$

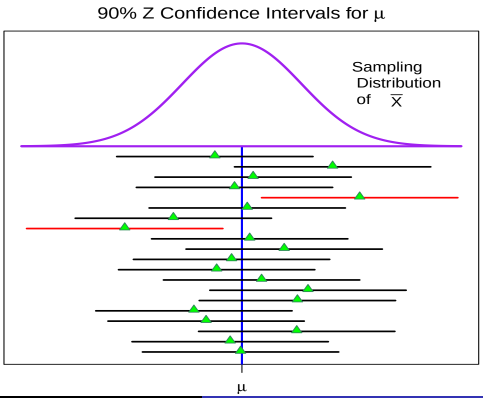
Note that this interval is **narrower** than the 95% interval (which was (245.7,   298.3)).

Notes

Properties of Confidence Intervals (7.1)

- A confidence interval is a **random interval**.
- A confidence level of 90% implies that 90% of all samples would give an interval that contains μ.

Notes



Notes

- The **width** of a confidence interval indicates the **precision** of the estimate.

  The **width** is determined by:

$$2\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

- Confidence intervals have the following properties:
  1. Using a **larger sample size** n results in a **narrower** CI.
  2. A **smaller population standard deviation** σ leads to a **narrower** CI.
  3. Using a **lower level of confidence** results in a **narrower** CI.

Notes

Sample Size Determination (7.1)

- We can make the **width** of the confidence interval as **small** as we want by using a **large enough sample size** n.
- Suppose we want the **width** to be w. Then we'd need n to be large enough that

$$2\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = w.$$

  Solving for n gives the required sample size.

Notes

Sample Size for a Desired CI Width: The width of the $100(1-\alpha)\%$ confidence interval for μ will be w when

$$n = \left(2\,z_{\alpha/2}\cdot\frac{\sigma}{w}\right)^2.$$

This calculation is valid when the sample is to be from a normal population and σ is known.

Notes

Example (Cont'd)

Recall that scores on the National Assessment of Educational Progress Study quantitative skills test are a **normal** population with **known standard deviation** $\sigma = 60$, but with a mean μ whose value is **unknown**.

Suppose we want to carry out a study to estimate the population mean μ, and we want the **width** of a **95% confidence interval** for μ to be 10 units.

How big would n **need to be**?

Notes

We'd need

$$n = \left(2 \cdot 1.96 \cdot \frac{60}{10}\right)^2 = \mathbf{553.2},$$

which we **round up** to $n = \mathbf{554}$.

Notes

Large-Sample CI for a General Population Mean μ (7.2)

- When the **sample size** n **is large**, two things are true:
  1. Regardless of the shape of the population, by the Central Limit Theorem,
     $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{N}(0,\, 1) \qquad \text{(approximately)}$$
  2. The **sample standard deviation** S will remain fairly **constant** from one sample to the next, and **approximately equal to** σ.

Notes

- As a consequence, **when** n **is large**,
  $$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim \text{N}(0,\, 1) \qquad \text{(approximately)}$$
  even if the sample is from **non-normal** population.

  (Note that σ was replaced by S above.)
- Thus we can use the previously derived CI with σ **replaced by** S.

Notes

Large-Sample $100(1-\alpha)$% Confidence Interval for a Population Mean μ:

$$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}}.$$

This CI is valid when n is large, regardless of the shape of the population distribution.

- In practice, n is **large enough** if $n > 40$.

Notes

Sample Size Determination (7.2)

- We can determine the (approximate) **sample size n** needed for the CI to have **width w**.

Sample Size for a Desired CI Width: The width of the $100(1-\alpha)\%$ confidence interval for μ will be approximately w when

$$n = \left(2\, z_{\alpha/2} \cdot \frac{S}{w}\right)^2.$$

In practice we plug in an **educated guess for S**.

This calculation is valid when the sample is to be from a general population and σ is unknown.

Notes

CI for a Normal Mean μ When σ is Unknown (7.3)

- Suppose
  1. $X_1, X_2, \ldots, X_n$ are a random sample from a **normal** population.
  2. The population mean and standard deviation **μ and σ are both unknown**.
  3. **n isn't necessarily large**.
- To derive a CI for μ, we'll need a **new probability distribution**.

Notes

The t Distribution (7.3)

Proposition

Suppose $X_1, X_2, \ldots, X_n$ are a random sample from a **normal** population. Then the random variable

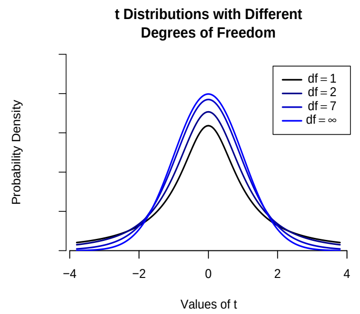
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows a t ***distribution*** with $n-1$ ***degrees of freedom***.

- We write this as

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Notes



Notes

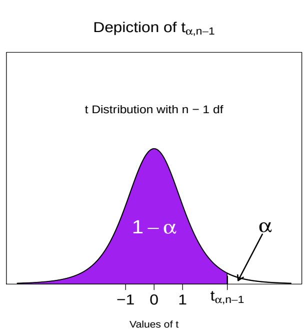
CI for a Normal Mean μ When σ is Known
Large-Sample CI for a General Population Mean μ
CI for a Normal Mean μ When σ is Unknown
CI for a Normal Mean μ When σ is Unknown
The t Distribution
The One-Sample t CI
Sample Size Determination

- **Properties of the t Distribution**:
  1. The t distribution is centered on 0 and resembles the N$(0, 1)$ distribution, but has "heavier" tails.
  2. It has one **parameter**, its **degrees of freedom** (or **df**).
  3. As the df increases, the t-distribution gets closer and closer to an exact N$(0, 1)$ distribution.
- We use $t_{\alpha,n-1}$ to denote the t ***critical value*** that has area α to its **right** under the $t(n-1)$ curve:

Nels Grevstad

Notes

CI for a Normal Mean μ When σ is Known
Large-Sample CI for a General Population Mean μ
CI for a Normal Mean μ When σ is Unknown
CI for a Normal Mean μ When σ is Unknown
The t Distribution
The One-Sample t CI
Sample Size Determination



Nels Grevstad

Notes

CI for a Normal Mean μ When σ is Known
Large-Sample CI for a General Population Mean μ
CI for a Normal Mean μ When σ is Unknown
CI for a Normal Mean μ When σ is Unknown
The t Distribution
The One-Sample t CI
Sample Size Determination

- Values of $t_{\alpha,n-1}$ are obtained from a t ***distribution table***.

Nels Grevstad

Notes

The One-Sample t CI (7.3)

- When the sample is from a **normal** population and σ **is unknown**, the CI formula is derived as before, but using the $t(n-1)$ distribution instead of the **N**$(0,1)$ distribution:

$$
\begin{aligned}
0.95 &= P\left(-t_{0.025,n-1} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{0.025,n-1}\right) \\
&\vdots \\
&= P\left(\bar{X} - t_{0.025,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{0.025,n-1}\frac{S}{\sqrt{n}}\right)
\end{aligned}
$$

Notes

$100(1-\alpha)$% Confidence Interval for a Normal Mean μ:

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}.$$

This is called the ***one-sample t confidence interval***.

It's valid when the sample is from a normal population and σ is unknown.

Notes

Example (Cont'd)

Recall that in the sleep deprivation study, a sample of $n = 22$ were students asked students how many hours they sleep per night.

The sample mean and standard deviation were

$$
\begin{aligned}
\bar{x} &= 5.77 \\
s &= 1.57
\end{aligned}
$$

We'll estimate μ, the population mean number of hours slept per night, using **95% one-sample t CI**.

Notes

The CI is

$$
\begin{aligned}
\bar{X} \pm t_{0.025,21}\cdot\frac{S}{\sqrt{n}} &= 5.77 \pm 2.080\cdot\frac{1.57}{\sqrt{22}} \\
&= 5.77 \pm 0.70 \\
&= (5.07,\ 6.47)
\end{aligned}
$$

where the value $t_{0.025,21} = 2.080$ was obtained from a t **distribution table**.

Values of μ in this interval are considered **plausible**, and we can be **95% confident** that μ is in the interval somewhere.

Notes

Sample Size Determination (7.3)

- To determine an (approximate) **sample size** for the CI to have **width** w, we use

$$n = \left(2\, z_{\alpha/2} \cdot \frac{S}{w}\right)^2$$

as before, plugging in an **educated guess for** S.

(Note that we use $z_{\alpha/2}$, not $t_{\alpha/2,n-1}$, because $t_{\alpha/2,n-1}$ depends on n.)

Notes

Notes

Notes

Notes