Introduction to Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

October 6, 2019

Notes

Topics

Notes

Objectives

Objectives:

- Determine the sample size required for the margin of error in a CI for a population mean to be no bigger than some specified value.
- Distinguish between the t distribution and the standard normal distribution.
- Compute and interpret a CI for a population mean when the population standard deviation isn't known.

Notes

Determining the Required Sample Size for Estimating μ (8.2)

- Recall that the **margin of error** in a **one-mean z CI** for μ is

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

We can make the **margin of error as small as we want** by using a **large enough** n.

Notes

- Suppose we want the **margin of error** to **no bigger than** some value E.

  Solving
  $$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq E$$
  for n gives the ***required sample size*** (see the next slide).

Notes

Determining the Required Sample Size for Estimating μ
Confidence Interval for μ when σ is Unknown

Sample Size for Estimating μ: The sample size required for the margin of error in a CI for μ to be no bigger than some value E is

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2,$$

which we **round up** to the nearest integer.

In practice, we usually don't know the value of σ, so we plug in a guess for it's value (e.g. based on prior studies).

Nels Grevstad

Notes

Determining the Required Sample Size for Estimating μ
Confidence Interval for μ when σ is Unknown

Example

Suppose we want to conduct a survey to find out how much time Denver residents spend exercising, on average, per week.

Our goal is to **estimate** the true (unknown) **mean** amount of time μ (in minutes) using a **99% CI**.

A similar study in Seattle suggests that a reasonable guess for the Denver population standard deviation σ is **50** minutes.

How large should our sample size be if we want the **margin of error** to be **no bigger than 10** minutes?

Nels Grevstad

Notes

Determining the Required Sample Size for Estimating μ
Confidence Interval for μ when σ is Unknown

We have

$$\begin{aligned} E &= 10 \\ z_{\alpha/2} &= z_{0.005} = 2.58 \\ \sigma &= 50 \end{aligned}$$

so we'd need

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{2.58(50)}{10}\right)^2 = 166.41,$$

which we **round up** to **167**.

Nels Grevstad

Notes

Exercise

Suppose, as before, that we want to **estimate** the true (unknown) **mean** weekly exercise time μ of Denver residents using a **99% CI**, and that a reasonable guess for σ is **50** minutes.

Suppose now, though, that we only need the **margin of error** to be **no bigger than 15** minutes.

a) Will the **sample size** required for a **15-minute margin of error** be **larger** or **smaller** than the one required for a **10-minute margin of error**?

Notes

b) Calculate the **sample size** required for a **15-minute margin of error**. Compare it to the one required for a **10-minute margin of error**.

Notes

Confidence Interval for μ when σ is Unknown (8.3)

Introduction

- Usually the population standard deviation σ ***isn't* known**, so we **can't** use the *one-mean z CI* for μ.

  Instead, we **estimate** σ by the sample standard deviation s, and then use the ***one-mean t CI*** described ahead.

Notes

Introduction to the t Distribution

- For the ***one-mean t CI***, we'll need a new probability distribution, the t **distribution**.

Notes

- Recall that the *standardized version* of the sample mean,

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

follows a **standard normal** distribution (when the sample is from a *normal* population or n is *large*).

- If we use the **estimate s in place of σ**, the *standardized version* of $\bar{x}$,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}},$$

instead follows a so-called ***t distribution*** with $n-1$ ***degrees of freedom*** (or ***df***).

Notes

- There's a different t **distribution** for each value of the **degrees of freedom**, $n-1$.

  Several t distribution curves are shown on the next slide.

Notes

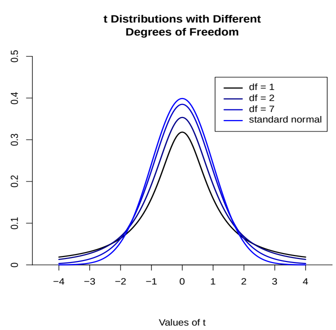


Figure: t distributions with different df along with the standard normal (z) curve.

Notes

- **Properties of the t Distributions**
  - They're **centered on zero** and resemble the standard normal distribution, but are more spread out in the tails.
  - As the **df increases**, the t **curves** get closer and closer to the **standard normal z curve**.

    When the **df** is larger than about **40**, the t and z **curves** are practically **indistinguishable**.

Notes

The One-Mean t Confidence Interval for μ

- The next slide gives the procedure for computing a CI for μ when σ ***isn't* known**.

Notes

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

$$\bar{x} \; \pm \; t_{\alpha/2} \frac{s}{\sqrt{n}},$$

with ***margin of error***

$$\text{Margin of Error} \;=\; t_{\alpha/2} \frac{s}{\sqrt{n}},$$

where $t_{\alpha/2}$ is a t **critical value**, obtained from the t distribution with $n-1$ df, and α is either **0.10**, **0.05**, or **0.01**, depending on the level of confidence (see the next slides).

Notes

$\alpha = \mathbf{0.10}$ for a **90%** level of confidence (so $1-\alpha = \mathbf{0.90}$, $\alpha/2 = \mathbf{0.05}$, and the critical value is $t_{0.05}$)

$\alpha = \mathbf{0.05}$ for a **95%** level of confidence (so $1-\alpha = \mathbf{0.95}$, $\alpha/2 = \mathbf{0.025}$, and the critical value is $t_{0.025}$)

$\alpha = \mathbf{0.01}$ for a **99%** level of confidence (so $1-\alpha = \mathbf{0.99}$, $\alpha/2 = \mathbf{0.005}$, and the critical value is $t_{0.005}$)

(The actual values of $t_{0.05}$, $t_{0.025}$, and $t_{0.005}$ will depend on the df.)

Notes

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

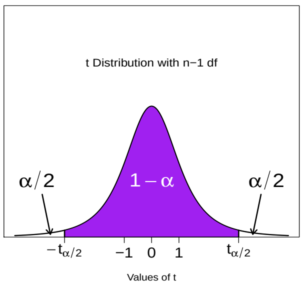
Notes

- The ***t critical value***, $t_{\alpha/2}$, is the t value for which the area to its **right** under the t distribution with $n-1$ degrees of freedom is $\alpha/2$.

Notes

Determining the Required Sample Size for Estimating μ
Confidence Interval for μ when σ is Unknown

Depiction of $t_{\alpha/2}$



Nels Grevstad

Notes

Determining the Required Sample Size for Estimating μ
Confidence Interval for μ when σ is Unknown

Example

In a study of the behavior of bats, **n = 11** observations of the distances (in cm) at which bats first detect insects were recorded.

The **sample mean** and **sample standard deviation** of the **11** observations are

$$\bar{x} = \mathbf{48.4} \qquad \text{and} \qquad s = \mathbf{18.1}$$

The ***population*** **mean** μ and **standard deviation** σ are both ***unknown***.

Nels Grevstad

Notes

Determining the Required Sample Size for Estimating μ
Confidence Interval for μ when σ is Unknown

One goal of the study was to **estimate** the **true (unknown) population mean** insect-detection-distance μ using a **CI**.

A **95% CI** for μ is

$$\begin{aligned} \bar{x} \pm t_{0.025}\frac{s}{\sqrt{n}} &= 48.4 \pm 2.228 \times \frac{18.1}{\sqrt{11}} \\ &= 48.4 \pm 12.2 \\ &= \mathbf{(36.2,\ 60.6)} \end{aligned}$$

where the t **critical value**, $t_{0.025} = \mathbf{2.228}$, was obtained from **Table IV** using $n - 1 = \mathbf{10}$ **df**.

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

Nels Grevstad

Notes

Is it plausible, based on the CI, that the true (unknown) mean distance μ is **as large as 55** cm?

Answer: Yes because 55 is contained in the interval.

Is it plausible that μ is **as large as 65** cm?

Answer: No because 65 isn't contained in the interval.

Notes

Note (two slides back) that the **margin of error** in the estimate, $\bar{x} = 48.4$ cm, is **12.2** cm.

We **interpret** this value as a measure of **how reliable** the estimate is.

More precisely, it tells us that the **sampling error** is **not likely** to be **larger** than **12.2** cm.

Notes

If we use a **99% confidence level**, the t **critical value** (from **Table IV** using $\mathbf{n-1=10}$ **df**) is $\mathbf{t_{0.005} = 3.169}$, and the **CI** for μ is

$$\begin{aligned}\bar{x} \pm t_{0.005}\frac{s}{\sqrt{n}} &= 48.4 \pm 3.169 \times \frac{18.1}{\sqrt{11}} \\ &= 48.4 \pm 17.3 \\ &= \mathbf{(31.1,\ 65.7)}\end{aligned}$$

Note that the **99%** CI is **wider** than the **95%** CI.

Notes

Exercise

A highway safety researcher studying the design of a highway sign is interested in the **population mean distance** μ at which drivers are first able to read the sign.

A sample of $n = 16$ drivers reported that they first read the sign at the following distances (in ft):

440	490	600	540	540	600	380	440
360	510	490	400	490	540	440	490

Notes

The **sample mean** and **sample standard deviation** of the **16** observations are

$$\bar{x} = 484.4 \qquad s = 71.3$$

a) **Compute** and **interpret** a **95% CI** for the **true (unknown) population mean** sign-readability-distance μ.

b) **Is it plausible**, based on the CI, that μ is **as large as 510** ft? Explain.

c) How big is the **margin of error** in the estimate, $\bar{x} = 484.4$, of μ? How do you **interpret** this value?

Notes

d) Now compute a **99% CI** for μ.

e) Which CI is **wider**, the **95%** interval or the **99%** interval?

Notes

Notes

Notes