6 One-Sample Confidence Intervals

MTH 3240 Environmental Statistics

Spring 2020

Notes

Objectives

Objectives:

- Distinguish between (standardized) t-scores and z-scores.
- Compute and interpret one-sample z and one-sample t confidence intervals for a population mean.

Notes

Introduction

- In most practical problems, the values of **population parameters** such as μ and σ **won't be known** and need to be **estimated** from sample data.
- Estimates come in two forms:
  - ***Point estimates*** (single values).
  - ***Confidence intervals*** (ranges of values).

Notes

- Examples of **point estimates**:

	Sample Estimate	Population Parameter
Mean	$\bar{X}$	μ
Std Dev	S	σ
Median	$\tilde{X}$	$\tilde{\mu}$
Proportion	$\hat{P}$	p

Notes

- A ***confidence interval*** (***CI***) is an entire **range** of **values**, all of which are considered to be reasonable **estimates** of the population parameter.
- We'll focus (for now) on the **CI** for an *unknown* **population mean** $\boldsymbol{\mu}$.
- The **CI** for μ is computed differently depending on whether
  - The population standard deviation $\boldsymbol{\sigma}$ is **known**.
  - The population standard deviation $\boldsymbol{\sigma}$ is **unknown**.

Notes

- The first step in computing a **CI** will be to *choose* a ***level of confidence*** (e.g. 95%).

  It represents the degree to which we want to be sure that the interval will contain $\boldsymbol{\mu}$.

Notes

One Sample Z CI for μ

- The **one-sample z CI for $\boldsymbol{\mu}$** is used when $\boldsymbol{\sigma}$ **is known**.

Notes

- Suppose we have a random sample from a population.

  If either
  1. The population is normal, or
  2. The sample size n is large,

  then

$$\bar{X} \sim \text{N}(\mu, \sigma_{\bar{X}}) \qquad \text{where} \qquad \boldsymbol{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

  In this case, converting $\bar{\boldsymbol{X}}$ to a $\boldsymbol{z}$**-score** gives

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \sim \text{N}(0, 1).$$

Notes

- The 2.5th and 97.5th percentiles of the N$(0, 1)$ distribution are **-1.96** and **1.96**, so there's a **95% chance** that

$$-1.96 < \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < 1.96$$

which, after a little algebra, is the same as

$$\bar{X} - 1.96\,\sigma_{\bar{X}} < \mu < \bar{X} + 1.96\,\sigma_{\bar{X}}.$$

- Thus we can be **95% confident** that μ will be contained in the interval

$$\bar{X} \pm 1.96\,\sigma_{\bar{X}} \qquad \text{where} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

This is the ***95% one-sample z* CI *for* μ**.

Notes

Example

In a study of truck emissions and air quality in California, the engine idle time (minutes) per day was recorded for $n = 13$ trucks.

The sample mean was

$$\bar{X} = 29.6$$

Suppose that in the **population** of trucks, the **standard deviation** of idle times is $\sigma = 10.0$ minutes.

Notes

The *standard error* of $\bar{X}$ is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10.0}{\sqrt{13}} = \mathbf{2.8}$$

The **95% z CI** for the **true** (unknown) **population mean μ** is

$$\begin{aligned} \bar{X} \pm 1.96\,\sigma_{\bar{X}} &= 29.6 \pm 1.96 \times 2.8 \\ &= 29.6 \pm 5.5 \\ &= (\mathbf{24.1},\ \mathbf{35.1}). \end{aligned}$$

We can be **95% confident** that μ is in this range (somewhere).

Notes

- For other levels of confidence (e.g. 90% or 99%), the so-called *z* ***critical value*** will be something other than 1.96.

- We'll denote a **generic level of confidence** by $100(1 - \alpha)\%$, where α is the value such that

$$\textbf{Level of Confidence} = 100(1 - \alpha)\%.$$

Example: $\alpha = 0.05$ for a 95% level of confidence.

Notes

One-Sample Z CI: A $100(1-\alpha)\%$ ***one-sample z CI for μ*** is

$$\bar{X} \pm z_{\alpha/2}\sigma_{\bar{X}} \quad \text{where} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

The $z_{\alpha/2}$ value is discussed on the next few slides.

- The CI is valid if either
  1. The population is normal, or
  2. The sample size n is large.

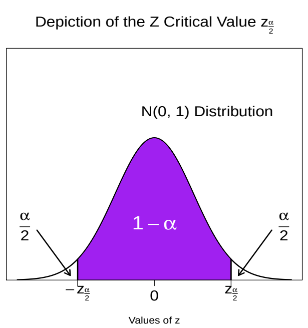
Notes

- $z_{\alpha/2}$ denotes the so-called z **critical value** associated with a $100(1-\alpha)$**% level of confidence**.

  It's the $100(1-\alpha/2)$***th percentile*** of the **standard normal** distribution.

  Example: For a 95% level of confidence, $\alpha = 0.05$ and $z_{0.025} = 1.96$ is the 97.5th percentile.

Notes



Notes

- For three commonly used confidence levels, the z **critical values** are

$$z_{0.05} = 1.64 \quad \text{(for a } 90\% \text{ confidence level)}$$
$$z_{0.025} = 1.96 \quad \text{(for a } 95\% \text{ confidence level)}$$
$$z_{0.005} = 2.58 \quad \text{(for a } 99\% \text{ confidence level)}$$

Notes

- The "plus or minus" part is called the ***margin of error***.

Margin of Error: For the one-sample z CI,

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

- A **smaller margin of error** indicates that $\bar{X}$ is a **more precise estimate** of μ.
- The margin of error will be **small** if either:
  - σ is **small**, or
  - n is **large**.

Notes

Properties and Interpretation of CIs

- **Properties and interpretation of CIs**:
  1. A **CI** for μ gives a set of **plausible values** for μ.
  2. A **higher level of confidence** will result in a **wider CI**.
  3. A **larger sample size** will result in a **narrower CI**.

Notes

One Sample t CI for μ

- The **one-sample t CI for μ** is used when σ **is unknown**.
- We replace σ in the CI formula by its **estimate** S.

  But then (as we'll see) we also need to replace the z critical value by a t **critical value**.

Notes

- Suppose we have a random sample from a population.

  If either
  1. The population is normal, or
  2. The sample size n is large,

  then $\bar{X}$ can be converted to a so-called ***t-score*** via

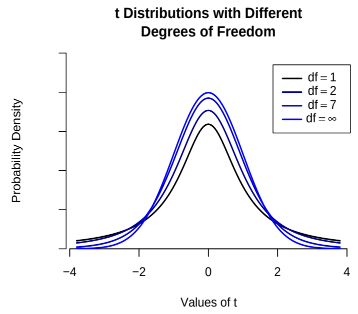
$$T = \frac{\bar{X}-\mu}{S_{\bar{X}}} \quad \text{where} \quad S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

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

  We write this as

$$T = \frac{\bar{X}-\mu}{S_{\bar{X}}} \sim \text{t}(n-1).$$

Notes



Notes

- **Properties of the t distribution**:
  1. The t curve's tails extend farther away from zero than the z (standard normal) curve's.
  2. The t curve gets closer and closer to the z curve as the **df** increases.

     If the **df** are more than about 40, the t and z curves are indistinguishable.

Notes

One-Sample t CI: A $100(1-\alpha)\%$ ***one-sample t CI for μ*** is

$$\bar{X} \pm t_{\alpha/2,n-1} S_{\bar{X}} \qquad \text{where} \qquad S_{\bar{X}} = \frac{S}{\sqrt{n}}.$$

The $t_{\alpha/2,n-1}$ value is discussed on the next few slides.

- The CI is valid if either
  1. The population is normal, or
  2. The sample size n is large.

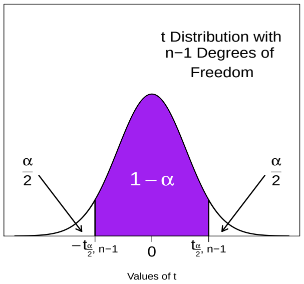
Notes

- $t_{\alpha/2,n-1}$ denotes the t **critical value** associated with a **$100(1-\alpha)\%$ level of confidence**.

  It's the $100(1-\alpha/2)$***th percentile*** of the $t(n-1)$ distribution.

  (It can be obtained from a **t distribution table**.)

Notes



Notes

- The ***margin of error*** is:

Margin of Error: For the one-sample t confidence interval,

$$\text{Margin of Error} = t_{\alpha/2,n-1} S_{\bar{X}} = t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}.$$

Notes

Exercise

Rocky Flats was a nuclear weapons production plant located 16 miles northwest of Denver that was in operation from 1952 until 1989.

Its hazardous waste spills and leaking barrels of radioactive waste contaminated soil in the area.

Cleanup of the site by government agencies took 10 years and cost $7 billion.

Notes

Public concern about the cleanup prompted an independent assessment by a private contractor in 2000.

For comparison, the contractor obtained **background soil radiation** levels from $n = 10$ sites along the Front Range of the Colorado Rocky Mountains.

The table below shows the concentrations of the plutonium isotope ^{239,240}Pu (in Bq/kg).

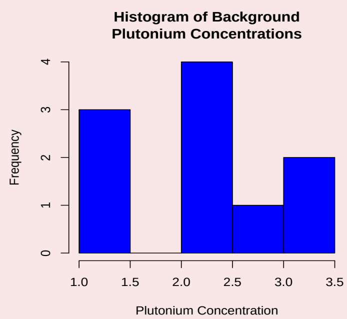
Notes

Background Soil Radiation Concentrations

Site	^{239,240}Pu
Z01	1.20
Z02	2.10
Z03	1.46
Z04	2.10
Z05	2.10
Z06	1.14
Z07	3.29
Z08	3.22
Z09	2.07
Z10	2.70

A histogram of the data is below.

Notes



Notes

The sample size is $n = 10$ and the **sample mean** and **standard deviation** are

$$\bar{X} = 2.14 \quad \text{and} \quad S = 0.76.$$

Thus the (point) **estimate** of the **true** (unknown) **population mean background radiation level** μ is $\bar{X} = 2.14$.

We'll compute and interpret a **95% CI** for μ.

Notes

a) The **one-sample t CI** is valid if either
   - The population is normal, or
   - The sample size n is large.

   Based on the shape of the histogram, is the **normality assumption** met?

b) Compute and interpret a **95% one-sample t CI** for μ.

   Hint: You should get $2.14 \pm 0.54 = (1.60,\ 2.68)$.

c) Is it **plausible** that μ is as low as 2.0 Bq/kg? As low as 1.0 Bq/kg?

d) How big is the **margin of error** in the point estimate, $\bar{X}$, of μ?

Notes