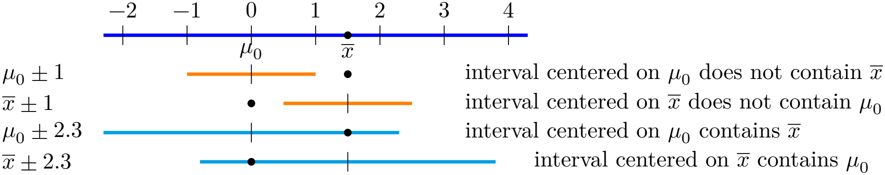
Confidence Intervals for Normal Data
18.05 Spring 2022



Announcements/Agenda

Exam 2

- Next Thursday in class
- Covers classes 10-20
- You can use a cheat sheet - 1 side of a 8×11 sheet
- We'll give you probability tables – these are in the review materials
- Class on Tuesday will be review
- Review materials are on MITx

Agenda

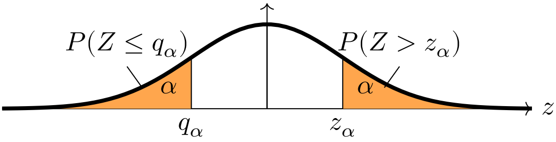
- Review of critical values and quantiles.
- Computing z, t, χ^2 confidence intervals for normal data.
- Conceptual view of confidence intervals.
- Confidence intervals for polling (Bernoulli distributions).

Review of critical values and quantiles

- **Quantile:** left tail $P(X < q_\alpha) = \alpha$
- **Critical value:** right tail $P(X > c_\alpha) = \alpha$

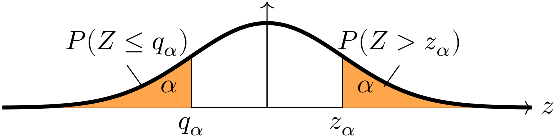
Letters for critical values:

- z_α for N$(0, 1)$
- t_α for t distributions
- c_α, x_α all purpose



q_α and z_α for the standard normal distribution.

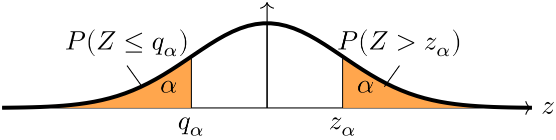
Concept question: critical values



1. $z_{0.025} =$

(a) -1.96 (b) -0.95 (c) 0.95 (d) 1.96 (e) 2.87

Concept question: critical values



1. $z_{0.025} =$

(a) -1.96 (b) -0.95 (c) 0.95 (d) 1.96 (e) 2.87

2. $-z_{0.16} =$

(a) -1.33 (b) -0.99 (c) 0.99 (d) 1.33 (e) 3.52

Computing confidence intervals from normal data

Suppose the data $x_1, \ldots, x_n$ is independently drawn from $\mathrm{N}(\mu, \sigma^2)$

Confidence level $= 1 - \alpha$

- z confidence interval for the mean (σ known)

$$\left[\overline{x} \; - \; \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \;\; \overline{x} \; + \; \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right]$$

- t confidence interval for the mean (σ unknown)

$$\left[\overline{x} \; - \; \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}, \;\; \overline{x} \; + \; \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}\right]$$

- χ^2 confidence interval for σ^2

$$\left[\frac{n-1}{c_{\alpha/2}} s^2, \;\; \frac{n-1}{c_{1-\alpha/2}} s^2\right]$$

- t and χ^2 have $n - 1$ degrees of freedom.

z rule of thumb

Suppose $x_1, \dots, x_n \sim \mathsf{N}(\mu, \sigma^2)$ with σ known.

The rule-of-thumb 95% confidence interval for μ is:

$$\left[\bar{x} - 2\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 2\frac{\sigma}{\sqrt{n}}\right]$$

A more precise 95% confidence interval for μ is:

$$\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

Board question: computing confidence intervals

The data 4, 1, 2, 3 is drawn from $\mathrm{N}(\mu, \sigma^2)$ with μ unknown.

(a) Find a 90% z confidence interval for μ, given that $\sigma = 2$.

For the remaining parts, suppose σ is unknown.

(b) Find a 90% t confidence interval for μ.

(c) Find a 90% χ^2 confidence interval for σ^2.

(d) Find a 90% χ^2 confidence interval for σ.

(e) Given a normal sample with $n = 100$, $\overline{x} = 12$, and $s = 5$, find the rule-of-thumb 95% confidence interval for μ.

Conceptual view of confidence intervals

- Computed from data $\Rightarrow$ **interval statistic**
- 'Estimates' a parameter of interest $\Rightarrow$ **interval estimate**
- Width $=$ measure of precision
- Confidence level $=$ measure of performance
- Connected to NHST: Given a test statistic x, the confidence interval for $\theta =$ all values θ_0 for which we wouldn't reject the null hypothesis $\theta = \theta_0$.
- Confidence intervals are a frequentist method.
  - No need for a prior, only uses likelihood.
  - Frequentists do not assign probabilities to hypothetical values of unknown parameters.
  - A 95% confidence interval of $[1.1, 3.3]$ for μ **does not mean** that $P(1.1 \leq \mu \leq 3.3) = 0.95$.
- We will compare with Bayesian probability intervals later.

Applet:
https://mathlets.org/mathlets/confidence-intervals/

Discussion: Width of confidence intervals

The quantities n, $c =$ confidence, $\overline{x}$, σ all appear in the z confidence interval for the mean.

How does the width of a confidence interval for the mean change if:

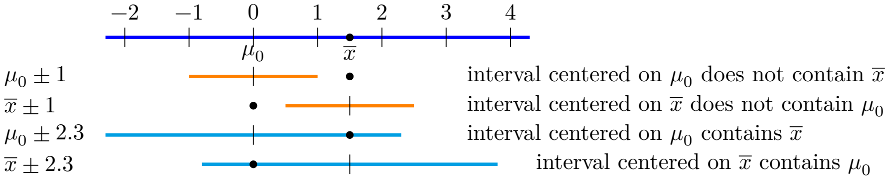
1. We increase n and leave the others unchanged?
2. We increase c and leave the others unchanged?
3. We increase μ and leave the others unchanged?
4. We increase σ and leave the others unchanged?

(A) it gets wider (B) it gets narrower (C) it stays the same.

Intervals and pivoting

$\overline{x}$: sample mean (statistic)

μ_0: hypothesized mean (not known)



Algebra of pivoting:

$$\mu_0 - 2.3 \leq \overline{x} \leq \mu_0 + 2.3 \qquad \overline{x} \text{ in non-rejection region}$$
$$\Leftrightarrow \overline{x} + 2.3 \geq \mu_0 \geq \overline{x} - 2.3 \qquad \mu_0 \text{ in confidence interval}$$
$$\Leftrightarrow |\overline{x} - \mu_0| \leq 2.3 \qquad \text{distance apart} \leq 2.3.$$

(Can also do for non-symmetric intervals.)

Board question: confidence intervals and non-rejection regions

Suppose $x_1, \dots, x_n \sim \mathrm{N}(\mu, \sigma^2)$ with σ known.

Consider two intervals:

1. The z confidence interval around $\overline{x}$ at confidence level $1 - \alpha$.
2. The z non-rejection region for $H_0 : \mu = \mu_0$ at significance level α.

Compute and sketch these intervals to show that:

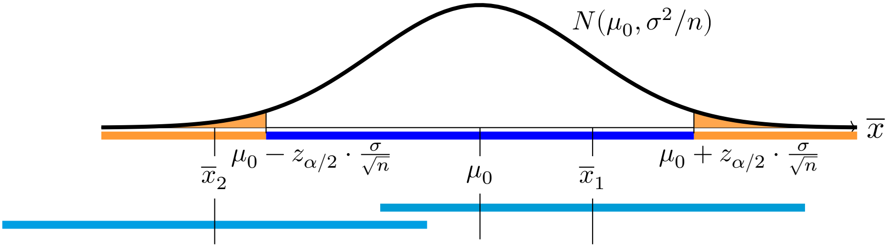
$$\mu_0 \text{ is in the first interval} \quad \Leftrightarrow \quad \overline{x} \text{ is in the second interval.}$$

Solution

Confidence interval: $\overline{x} \pm z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$

Non-rejection region: $\mu_0 \pm z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$

Since the intervals are the same width they either both contain the other's center or neither one does.



Polling: a binomial proportion confidence interval

Data $x_1, \ldots, x_n$ from a Bernoulli(θ) distribution with θ unknown.

A conservative normal[†] $(1-\alpha)$ confidence interval for θ is given by

$$\left[\,\bar{x} - \frac{z_{\alpha/2}}{2\sqrt{n}},\ \bar{x} + \frac{z_{\alpha/2}}{2\sqrt{n}}\,\right].$$

Proof uses the CLT and the observation $\sigma = \sqrt{\theta(1-\theta)} \leq 1/2$.

Political polls often give a margin-of-error of $\pm 1/\sqrt{n}$. This **rule-of-thumb** corresponds to a 95% confidence interval:

$$\left[\,\bar{x} - \frac{1}{\sqrt{n}},\ \bar{x} + \frac{1}{\sqrt{n}}\,\right].$$

(The proof is in the class 23 notes.)

Conversely, a margin of error of ± 0.05 means 400 people were polled.

[†] There are many types of binomial proportion confidence intervals.

https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval

Board question: Polling

For a poll to find the proportion θ of people supporting X we know that a $(1-\alpha)$ confidence interval for θ is given by

$$\left[\bar{x} - \frac{z_{\alpha/2}}{2\sqrt{n}},\ \bar{x} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right].$$

(a) How many people would you have to poll to have a margin of error of 0.01 with 95% confidence? (You can do this in your head.)

(b) How many people would you have to poll to have a margin of error of 0.01 with 80% confidence. (You'll want R or other calculator here.)

(c) If $n = 900$, compute the 95% and 80% confidence intervals for θ.