Confidence Intervals for Normal Data

18.05 Spring 2018

Agenda

- Exam on Monday April 30.
- Practice questions posted.
- Friday's class is for review (no studio)

Today

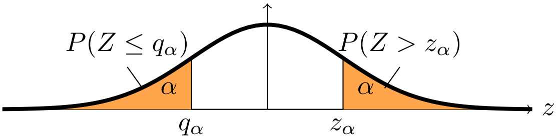
- Review of critical values and quantiles.
- Computing z, t, χ^2 confidence intervals for normal data.
- Conceptual view of confidence intervals.
- Confidence intervals for polling (Bernoulli distributions).

Review of critical values and quantiles

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

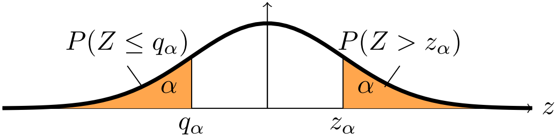
Letters for critical values:

- z_α for $\mathrm{N}(0, 1)$
- t_α for $t(n)$
- c_α, x_α all purpose



q_α and z_α for the standard normal distribution.

Concept question



1. $z_{.025} =$

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

2. $-z_{.16} =$

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

Solution on next slide.

Solution

1. **<u>answer:</u>** $z_{.025} = 1.96$. By definition $P(Z > z_{.025}) = 0.025$. This is the same as $P(Z \leq z_{.025}) = 0.975$. Either from memory, a table or using the R function `qnorm(.975)` we get the result.

2.**<u>answer:</u>** $-z_{.16} = -0.99$. We recall that $P(|Z| < 1) \approx .68$. Since half the leftover probability is in the right tail we have $P(Z > 1) \approx 0.16$. Thus $z_{.16} \approx 1$, so $-z_{.16} \approx -1$.

Computing confidence intervals from normal data

Suppose the data $x_1, \ldots, x_n$ is 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 $x_1, \ldots, x_n \sim \mathrm{N}(\mu, \sigma^2)$ 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 N(μ, σ^2) with μ unknown.

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

   For the remaining parts, suppose σ is unknown.

2. Find a 90% t confidence interval for μ.
3. Find a 90% χ^2 confidence interval for σ^2.
4. Find a 90% χ^2 confidence interval for σ.
5. Given a normal sample with $n = 100$, $\overline{x} = 12$, and $s = 5$, find the rule-of-thumb 95% confidence interval for μ.

Solution

$\overline{x} = 2.5, \;\; s^2 = 1.667, \;\; s = 1.29$
$\sigma/\sqrt{n} = 1, \;\; s/\sqrt{n} = 0.645.$
1. $z_{.05} = 1.644$: z confidence interval is

$$2.5 \pm 1.644 \cdot 1 = [0.856, 4.144]$$

2. $t_{.05} = 2.353$ (3 degrees of freedom): t confidence interval is

$$2.5 \pm 2.353 \cdot 0.645 = [0.982, \; 4.018]$$

3. $c_{0.05} = 7.1814$, $c_{0.95} = 0.352$ (3 degrees of freedom): χ^2 confidence interval is

$$\left[\frac{3 \cdot 1.667}{7.1814}, \; \frac{3 \cdot 1.667}{0.352}\right] = [0.696, \; 14.207].$$

4. Take the square root of the interval in 3. $[0.834, \; 3.769]$.
5. The rule of thumb is written for z, but with $n = 100$ the $t(99)$ and standard normal distributions are very close, so we can assume that $t_{.025} \approx 2$. Thus the 95% confidence interval is $12 \pm 2 \cdot 5/10 = [11, \; 13]$.

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
- Confidence intervals are a frequentist method.
  - No need for a prior, only uses likelihood.
  - Frequentists don't assign probabilities to hypotheses
  - A 95% confidence interval of $[1.2, 3.4]$ for μ doesn't mean that $P(1.2 \leq \mu \leq 3.4) = 0.95$.
- We will compare with Bayesian probability intervals later.

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

Table discussion

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.

Answers

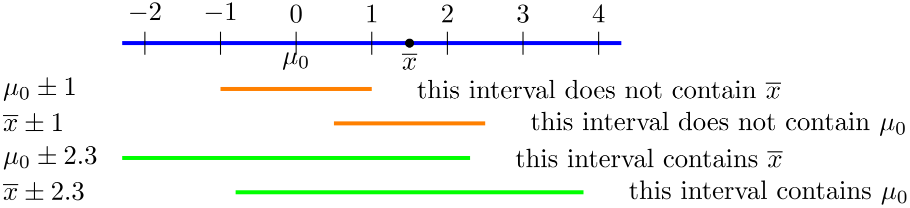
1. Narrower. More data decreases the variance of $\bar{x}$
2. Wider. Greater confidence requires a bigger interval.
3. No change. Changing μ will tend to shift the location of the intervals.
4. Wider. Increasing σ will increase the uncertainty about μ.

Intervals and pivoting

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

μ_0: hypothesized mean (not known)

Pivoting: $\overline{x}$ is in the interval $\mu_0 \pm 2.3 \Leftrightarrow \mu_0$ is in the interval $\overline{x} \pm 2.3$.



Algebra of pivoting:

$$\mu_0 - 2.3 < \overline{x} < \mu_0 + 2.3 \Leftrightarrow \overline{x} + 2.3 > \mu_0 > \overline{x} - 2.3.$$

Board question: confidence intervals, non-rejection regions

Suppose $x_1, \ldots, x_n \sim \mathrm{N}(\mu, \sigma^2)$ 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:

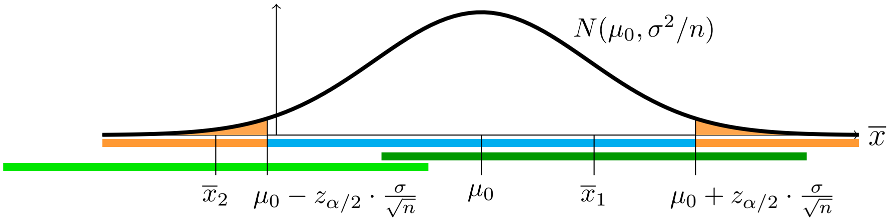
$$\mu_0 \text{ is in the first interval} \quad \Leftrightarrow \quad \overline{x} \text{ is in the second interval.}$$

Solution

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

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

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 21 notes.)

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

†There are many types of binomial proportion confidence intervals.

`http://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval`

Board question

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

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

1. How many people would you have to poll to have a margin of error of .01 with 95% confidence? (You can do this in your head.)

2. How many people would you have to poll to have a margin of error of .01 with 80% confidence. (You'll want R or other calculator here.)

3. If $n = 900$, compute the 95% and 80% confidence intervals for θ.
answer: *See next slide.*

answer: **1.** Need $1/\sqrt{n} = .01$ So $n = 10000$.

2. $\alpha = .2$, so $z_{\alpha/2} =$ `qnorm(.9)` $= 1.2816$. So we need $\dfrac{z_{\alpha/2}}{2\sqrt{n}} = .01$.

This gives $n = 4106$.

3. 95% interval: $\overline{x} \pm \dfrac{1}{\sqrt{n}} = \overline{x} \pm \dfrac{1}{30} = \overline{x} \pm .0333$

80% interval: $\overline{x} \pm z_{.1} \dfrac{1}{2\sqrt{n}} = \overline{x} \pm 1.2816 \cdot \dfrac{1}{60} = \overline{x} \pm .021$.

Concept question: overnight polling

During the presidential election season, pollsters often do 'overnight polls' and report a 'margin of error' of about $\pm 5\%$.

The number of people polled is in which of the following ranges?

(a) 0 – 50
(b) 50 – 100
(c) 100 – 300
(d) 300 – 600
(e) 600 – 1000

Answer: $5\% = 1/20$. So $20 = \sqrt{n} \;\Rightarrow\; n = 400$.

National Council on Public Polls: Press Release, Sept 1992

"The National Council on Public Polls expressed concern today about the current spate of overnight Presidential polls. [...] Overnight polls do a disservice to both the media and the research industry because of the considerable potential for the results to be misleading. The overnight interviewing period may well mean some methodological compromises, the most serious of which is.."

...what?

"...the inability to make callbacks, resulting in samples that do not adequately represent such groups as single member households, younger people, and others who are apt to be out on any given night. As overnight polls often result in findings that are less reliable than those from more carefully conducted polls, if the media reports them, it should be with great caution."

`http://www.ncpp.org/?q=node/42`