Confidence Intervals for Normal Data

18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom

Agenda

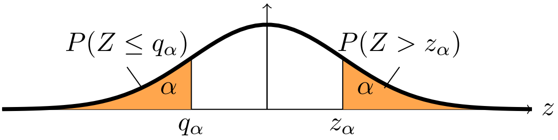
- Review of critical values and quantiles.
- Computing z, t, χ^2 confidence intervals for normal data.
- Conceptual view of confidence intervals.
- Confidence intervals for polling (Bernoulli distributions).
- CLT $\Rightarrow$ large sample confidence intervals for the mean.

Review of critical values and quantiles

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

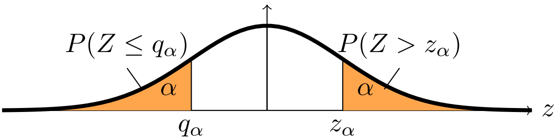
Letters for critical values:

- 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.

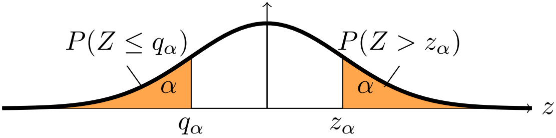
Concept question



1. $z_{.025} =$

(a) -1.96 (b) -.95 (c) .95 (d) 1.96 (e) 2.87

Concept question



1. $z_{.025} =$

(a) -1.96 (b) -.95 (c) .95 (d) 1.96 (e) 2.87

2. $-z_{.16} =$

(a) -1.33 (b) -.99 (c) .99 (d) 1.33 (e) 3.52

Solution on next slide.

Solution

1. $z_{.025} = 1.96$. By definition $P(Z > z_{.025}) = .025$. This is the same as $P(Z \le z_{.025}) = .975$. Either from memory, a table or using the R function `qnorm(.975)` we get the result.

2. $z_{.16} = .99$. We recall that $P(|Z| < 1) \approx .68$. Since half the leftover probability is in the right tail we have $P(Z > 1) \approx .16$. Thus $z_{.16} \approx 1$.

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)

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

- t confidence interval for the mean (σ unknown)

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

- χ^2 confidence interval for σ^2

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

- 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 1, 2, 3, 4 is drawn from $N(\mu, \sigma^2)$ 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

$\overline{x} = 2.5, \;\; s^2 = 1.667, \;\; s = 1.29$
$\sigma/\sqrt{n} = 1, \;\; s/\sqrt{n} = .645.$
1. $z_{.05} = 1.644$: z confidence interval is

$$2.5 \pm 1.644 \cdot 1 = [.856, 4.144]$$

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

$$2.5 \pm 2.353 \cdot .645 = [.982, \;\; 4.018]$$

3. $c_{.05} = 7.1814$, $c_{.95} = .352$ (3 degrees of freedom): χ^2 confidence interval is

$$\left[\frac{3 \cdot 1.667}{7.1814}, \;\; \frac{3 \cdot 1.667}{.352}\right] = [.696, \;\; 14.207].$$

4. Take the square root of the interval in 3. $[.593, \;\; 3.769]$.
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**
- The width and confidence level are measures of the precision and performance of the interval estimate; comparable to power and significance level in NHST.
- Confidence intervals are a frequentist method.
  - No need for a prior, only uses likelihood.
  - Frequentists never assign probabilities to unknown parameters: a 95% confidence interval of $[1.2, 3.4]$ for μ does **not** mean that $P(1.2 \leq \mu \leq 3.4) = .95$.
  - We will compare with Bayesian probability intervals next time.

In the applet, the confidence interval (**random interval**) covers the true mean $100(1-\alpha)\%$ of the times you hit '**generate data**':
http://ocw.mit.edu/ans7870/18/18.05/s14/applets/confidence-jmo.html

Table discussion

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

1. we increase n?
2. we increase c?
3. we increase μ?
4. we increase σ?

(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.

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:

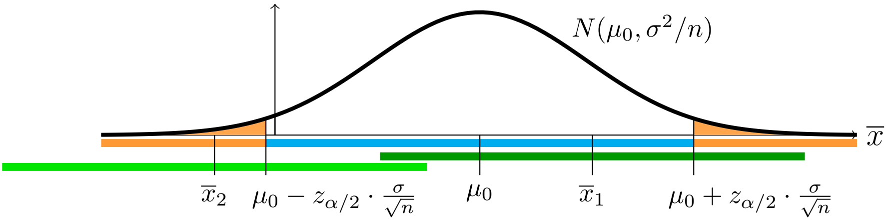
$$\mu_0 \text{ is in the first interval} \quad \Leftrightarrow \quad \overline{x} \text{ is in the second interval.}$$

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: binomial proportion confidence interval

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

A normal[†] $(1-\alpha)$ confidence interval for p 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 $\sigma = \sqrt{p(1-p)} \leq 1/2$.

Political polls often give a margin of error of $\pm 1/\sqrt{n}$, corresponding to a 95% confidence interval:

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

Conversely, a margin of error of $\pm.05$ means 400 people were polled.

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

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

Proof is in class 23 notes.

Board question

A $(1-\alpha)$ confidence interval for p 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 a table here.)

<u>**answer:**</u> 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 $n = 4106$.

Non-normal data

Suppose the data $x_1, x_2, \ldots, x_n$ is drawn from a distribution $f(x)$ that may not be normal or even parametric, but has finite mean, variance.

A version of the CLT says that for large n, the sampling distribution of the studentized mean is approximately standard normal:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \approx \quad \mathrm{N}(0, 1)$$

So for large n the $(1 - \alpha)$ confidence interval for μ is approximately

$$\left[\bar{x} - \frac{s}{\sqrt{n}} \cdot z_{\alpha/2},\ \ \bar{x} + \frac{s}{\sqrt{n}} \cdot z_{\alpha/2}\right] \qquad (1)$$

where $z_{\alpha/2}$ is the $\alpha/2$ critical value for $\mathrm{N}(0, 1)$.

This is called the *large sample confidence interval*.