class10

February 3, 2021 10:18 AM

Outline

§8.1 Basic properties of confidence intervals

§8.2 Large sample confidence intervals for a population mean and proportion

Confidence interval for μ

1

Normal population with known σ

Consider a simple situation –

Suppose that the parameter of interest is a population mean μ and that

1. The population distribution is normal
2. the value of population standard deviation σ is known.

2

95% Confidence interval

If, after observing $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, we compute the observed sample mean $\bar{x}$ and then substitute $\bar{x}$ into the random interval in place of $\bar{X}$, the resulting **fixed** interval is called a **95% confidence interval for** $\boldsymbol{\mu}$. This CI can be expressed as

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) \text{ is a 95\% CI for } \mu$$

or as

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \text{ with 95\% confidence}$$

A concise expression for the interval is $\bar{x} \pm 1.96\sigma/\sqrt{n}$, where $-$ gives the left endpoint (lower limit) and $+$ gives the right endpoint (upper limit).

3

$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95$

$\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

Random interval.

$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$ confidence interval

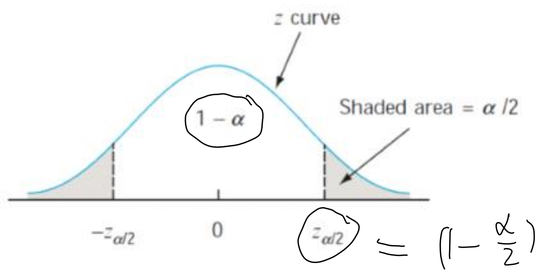
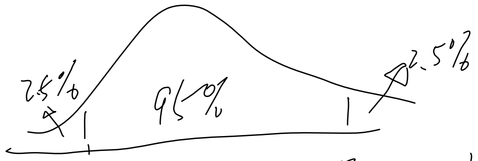


Figure 8.4 $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$

$= (1-\frac{\alpha}{2})100$th percentile

or $1-\frac{\alpha}{2}$ quantile



$Z_{.025} = 1.96$

4

CI for μ: Normal population with known σ

In general,

A **100(1-α)% confidence interval (CI)** for the mean μ of a normal population when the value of σ is known is given by

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

or, equivalently, by $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

In words,

point estimate $\pm$ (critical value)(standard error).

5

Traditionally, confidence levels are taken to be 90%, 95%, or 99%. The 90% level is the smallest respectable level, while the 95% level is the most often used.

The relationship between the levels and α are given in the table below:

Level	α	$\alpha/2$	z multiplier (critical value)
90%	0.1	0.05	1.645 or 1.64
95%	0.05	0.025	1.96
99%	0.01	0.005	2.576 or 2.58

$\frac{19}{20} = .95.$

90% $1-\alpha = .9,\ \alpha = .1.$

$\frac{\alpha}{2} = .05$

$z_{\alpha/2} = z_{.05} = 1.645$

$z_{\frac{\alpha}{2}} =$ qnorm(1 − $\frac{\alpha}{2}$) in R

= qnorm($\frac{\alpha}{2}$, lower.tail = F)

6

Example. The average zinc concentration recovered from a sample of zinc measurements in 36 different locations is found to be 2.6 grams per milliliter. Find a **99%** confidence interval for the mean zinc concentration in the river. Assume that the population distribution is normal and that the population standard deviation is 0.3.

$n = 36,\ \bar{x} = 2.6,\ \sigma = .3$

99% $\bar{x} \pm 2.576 \frac{.3}{\sqrt{36}}$

$= 2.6 \pm 2.576 \frac{.3}{\sqrt{36}}$

95% $\bar{x} \pm 1.96 \frac{.3}{\sqrt{36}}$

We are 99% confident that $2.47 < \mu < 2.73$. (Wider than the 95% CI 2.6 ± 0.10!)

How do you interpret the resulting confidence interval, or the level of confidence used?

7

- If we think of the **width** of a CI as specifying its **precision**, then **the confidence level (or reliability) of the interval is inversely related to its precision**.

  (For the same sample/data), the larger the confidence level, the wider the confidence interval; the gain in reliability entails a loss in precision.
- An appealing strategy is to specify both the desired confidence level and interval width and then determine the necessary sample size. For a given level, the larger the sample size, the narrower the interval.

100% CI $= (-\infty, \infty)$!

8

Sample size calculation

Example 9.4. Extensive monitoring of a computer time-sharing system has suggested that response time to a particular editing command is normally distributed with standard deviation 25 millisec. A new operating system has been installed, and we wish to estimate the true average response time μ for the new environment. Assuming that response times are still normally distributed with the same standard deviation, what sample size is necessary to ensure that the resulting 95% CI has a width of (at most) 10?

$\sigma = 25$

$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Width: $\boxed{2 z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le 10}$

95%: $z_{\alpha} = 1.96$ $\sigma = 25$

$95\%: \; z_{\frac{\alpha}{2}} = 1.96. \; \sigma = 25$

The width of a 95% CI is

$$w = 2(1.96)\frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(2z_{\alpha/2}\frac{\sigma}{w}\right)^2$$

The sample size n must satisfy

$$2(1.96)\frac{25}{\sqrt{n}} \leq 10$$

$$n \geq \left(2 \cdot 1.96 \cdot \frac{25}{10}\right)^2 = 96.04 \approx \mathbf{97}(?).$$

round up .

$(2)(1.96)\frac{25}{\sqrt{n}} = 10$

$\sqrt{n} = (2)(1.96)\, 25/10$

Confidence level, precision and sample size

The sample size n necessary to ensure an interval width w is obtained from $w = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ as

$$n = \left(2z_{\alpha/2}\frac{\sigma}{w}\right)^2$$

Half of the width of a confidence interval is called the **margin of error (ME)** (in text, **bound on the error of estimation**).

The sample size n necessary to ensure a desired margin of error is obtained from $\text{ME} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ as

$$n = \left(z_{\alpha/2}\frac{\sigma}{\text{ME}}\right)^2$$

$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ ← standard error

$z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ ← margin of error

$z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = ME$

In Example 9.4, what sample size is necessary to ensure that the resulting 99% CI has a margin of error =10?

$z_{\frac{\alpha}{2}} = 2.576$ or 2.58

$$n = \left(z_{\alpha/2}\frac{\sigma}{\text{ME}}\right)^2 = \left(2.58\frac{25}{10}\right)^2 = 41.60 \approx 42$$

Outline

§8.1 Basic properties of confidence intervals

§8.2 Large sample confidence intervals for a population mean and proportion

Confidence interval for μ

Outline

§8.1 Basic properties of confidence intervals

§8.2 Large sample confidence intervals for a population mean and proportion

Confidence interval for μ

general

"n is large"

Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and standard deviation σ.

Whatever the population distribution is, the Central Limit Theorem says that $\bar{X}$ is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$, as long as n is "large".

It follows that $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ is approximately standard normal. Accordingly

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

as a random variable here

As σ is unknown, we replace σ with the sample standard deviation S in the standardized random variable. It turns out that $Z = \frac{\bar{X}-\mu}{S\sqrt{n}}$ is also approximately standard normal. Then

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

$$P\left(-z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \bar{X} - \mu < z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) \doteq 1-\alpha$$

Based on this probability statement, we obtain an approximate CI for μ:

$$P\left(\bar{X} - z\frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

CI for μ:

$P(\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}) = 1-\alpha$

16

Large sample CI for μ

If n is sufficiently large, the standardized variable

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has approximately a standard normal distribution. This Implies that

$$\begin{array}{cccc} \bar{x} & \pm & z_{\alpha/2} & \frac{s}{\sqrt{n}} \\ \text{estimate} & \pm & \text{(critical value)} & \text{standard error} \end{array}$$

σ in the place of σ in previous formula.

is a **large sample confidence interval** for μ with confidence level approximately 100(1-α)%. This formula is valid regardless of the shape of the population distribution.

As a rule of thumb, $n \geq 30$ (40 in text)

17

Example. In a study of 100 $= n$ insurance sales reps from a $= \bar{X}$ certain large city, the average age of the group was 48.6, and the standard deviation was 4.1 years $= S$.

1. Find a 90% confidence interval of the average age of all insurance sales reps in the city. (48.6 ± 0.67) $= \mu$
2. Explain the 90% confidence level through a hypothetical situation of repeated samples.

If we repeat the study a large number of times, we expect 95% of CIs would include the true value of μ.

90% CI : $z_{.05} = 1.645$

$\bar{X} \pm z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

$= 48.6 \pm (1.645) \frac{4.1}{\sqrt{100}}$

$= 48.6 \pm .67$

$= (47.93, 49.27)$

18

Example. The average weight of 40 $= n$ randomly selected school buses was 4150 pounds $= \bar{X}$. The standard deviation was 480 $= S$ pounds. Construct a 95% confidence interval of the true mean weight of the buses.

19