Chapter 8 - Lecture 2(1) Large-Sample Confidence Intervals for a Population Mean

Yuan Huang

February 11, 2013

1. Review: $N(\mu, \sigma^2)$ with known σ^2
2. General population with known σ^2
3. General population with unknown σ^2
4. One-Sided Confidence Intervals (Confidence Bounds)

Review

- A Random sample $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, σ^2 is known.

  Two key assumptions:

  1. Normality
  2. σ^2 is known.

- We have seen this last lecture: the $(1-\alpha)\%$ Confidence Interval for μ is

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

Think what we can do if dropping normality assumption, but still assume σ^2 is known.

In this case, suppose we want to give an approximate 95% CI of the population mean μ based on a large sample ($n > 30$).

Solution:

CLT $\sqrt{n}(\bar{X} - \mu)/\sigma \approx N(0,1)$;

$P(-1.96\sigma/\sqrt{n} < \bar{X} - \mu < 1.96\sigma/\sqrt{n}) \approx 0.95$

Approximate 95% CI for μ is $(\bar{x} - 1.96\sigma/\sqrt{n}, \bar{x} + 1.96\sigma/\sqrt{n})$.

Example 1. The sweetness of the apples in a farm has mean μ and variance 0.2^2. To estimate μ, 100 apples are selected randomly and the sweetness of each of them is measured. Suppose the average sweetness is 0.5.

1. Please give a point estimate of μ;
2. Please give an approximate 95% confidence interval of μ.

Solution:

1. A sensible point estimate of μ is $\bar{x} = 0.5$.
2. An approximate 95% confidence interval of μ is

$$
\begin{aligned}
& (\bar{x} - 1.96\sigma/\sqrt{n}, \bar{x} + 1.96\sigma/\sqrt{n}) \\
= \; & (0.5 - 1.96 \times 0.2/\sqrt{100}, 0.5 + 1.96 \times 0.2/\sqrt{100}) \\
= \; & (0.4608, 0.5392)
\end{aligned}
$$

Be aware of the difference:

1. only works when $n > 30$;
2. The $100(1-\alpha)\%$ CI is approximate, so that the actual confidence is not exactly $100(1-\alpha)$.

Example 2. With the new drug, the time to recovery from a disease has the population mean μ and variance 3.5^2. Suppose 49 patients received the new drug and the sample average of recovery time is 5. Based on these data please give an approximate 90% CI for μ.

Solution: For the two sided CI, $z_{\frac{\alpha}{2}} = 1.645$.

An approximate 95% confidence interval of μ is

$$\begin{aligned} & (\bar{x} - 1.645\sigma/\sqrt{n}, \bar{x} + 1.645\sigma/\sqrt{n}) \\ = & \ (5 - 1.645 \times 3.5/\sqrt{49}, 5 + 1.645 \times 3.5/\sqrt{49}) \\ = & \ (4.1775, 5.8225) \end{aligned}$$

So far:

- When the random data is normally distributed with known variance σ^2, the exact $100(1-\alpha)\%$ CI for the mean μ is:

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

- Without normality assumption, we use CLT to get an approximate CI for μ when the sample size is large and σ^2 is known. The approximate $100(1-\alpha)\%$ CI for the mean μ is:

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

How about if we also drop the assumptions of known σ^2 ?

We know when n is large, we have CLT and consistency of S^2.

$$\sqrt{n}(\bar{X} - \mu)/\sigma \approx N(0, 1)$$

$$S^2 \approx \sigma^2$$

$$\rightarrow \sqrt{n}(\bar{X} - \mu)/S \approx N(0, 1)$$

.

The result is: without normality assumption, when n is large and σ^2 is unknown, we still get approximate Z CI for the population mean μ.

An approximate $100(1-\alpha)\%$ confidence interval for μ is:

$$(\bar{x} - z_{\alpha/2} s/\sqrt{n}, \bar{x} + z_{\alpha/2} s/\sqrt{n})$$

.

Exercises 8.12

A random sample of 110 lighting flashes in a region resulted in a sample average radar echo duration of 0.81 second and a sample standard deviation 0.34 second. Calculate a 99% (two-sided) CI for the true average echo duration μ.

Solution:

For the two sided CI, $z_{\frac{\alpha}{2}} = z_{\frac{1-0.99}{2}} = 2.575829$.
An approximate 99% confidence interval of μ is

$$
\begin{aligned}
& (\bar{x} - 2.575829s/\sqrt{n}, \bar{x} + 2.575829s/\sqrt{n}) \\
= & (0.81 - 2.575829 \times 0.34/\sqrt{110}, 0.81 + 2.575829 \times 0.34/\sqrt{110}) \\
= & (0.726, 0.894)
\end{aligned}
$$

The confidence intervals discussed thus far give both a lower confidence bound and an upper confidence bound.

- **Bound on the error of estimation associated with** $(1-\alpha)$ **CI** is the half-width $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ or $z_{\alpha/2}\frac{s}{\sqrt{n}}$
- $\alpha/2$ in the formula is due to two sided.

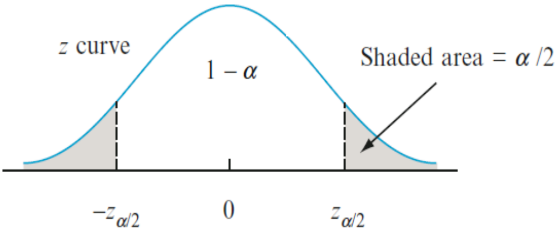

Figure 8.4 $P(-z_{a/2} \leq Z \leq z_{a/2}) = 1-\alpha$

Figure: Modern Mathematical Statistics with Applications, 2rd, P387

In two-sided CI, depends on the conditions, we have

- upper bound $\bar{x} + z_{\alpha/2} s/\sqrt{n}$; $\bar{x} + z_{\alpha/2}\sigma/\sqrt{n}$
- lower bound $\bar{x} - z_{\alpha/2} s/\sqrt{n}$; $\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}$

However, for one-sided CI, we only need one of these two types of bound. Thus we assign all the α to one side. For example, an approximate 95% one-sided CI μ is

$$P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} < 1.645\right) \approx 0.95$$

Proposition

A large-sample upper confidence bound for μ is

$$\mu < \bar{x} + z_\alpha \frac{s}{\sqrt{n}} \quad \text{or written as} \quad (-\infty, \bar{x} + z_\alpha \frac{s}{\sqrt{n}})$$

A large-sample lower confidence bound for μ is

$$\mu > \bar{x} - z_\alpha \frac{s}{\sqrt{n}} \quad \text{or written as} \quad (\bar{x} - z_\alpha \frac{s}{\sqrt{n}}, \infty)$$

Example 8.10

A random sample of 50 patients who had been seen at an outpatient clinic was selected, and the waiting time to see a physician was determined for each one, resulting in a sample mean time of 40.3 min and a sample standard deviation of 28 min. What is the upper confidence bound for true average waiting time with a confidence level roughly 95% ?

Solution:

- $n = 50$
- $\bar{x} = 40.3$
- $s = 28$
- $z_{0.05} = 1.645$
- upper confidence bound is $\bar{x} + z_\alpha \frac{s}{\sqrt{n}} = 40.3 + 1.645 \times 28/\sqrt{50} = 46.8$