Sirindhorn International Institute of Technology
Thammasat University
School of Information, Computer and Communication Technology

IES302 2011/2 Part II.3 Dr.Prapun

13 Confidence Interval on the Mean of a Normal Distribution

13.1. Motivation: We are often involved in estimating parameters. You know that you can use the sample average $\overline{X}$ to estimate the (true) population mean μ. However, we also know that the population mean is unlikely to be exactly equal to your estimate. Reporting the estimate as a single number is unappealing, because there is nothing inherent in $\overline{X}$ that provides any information about how close it is to μ. Your estimate could be very close, or it could be considerably far from the true mean. A way to avoid this is to report the estimate in terms of a range of plausible values called a *confidence interval*. = interval estimate + confidence level

Definition 13.2. The following terms are of great importance in interval estimation:

- $[\ell, u]$ $[L, U]$ ***Interval Estimate***: A range of values within which the actual value of the population parameter may fall.
- ***Interval Limits***: The lower and upper values of the interval estimate.
- ***Confidence Interval***: An interval estimate for which there is a specified degree of certainty that the actual value of the population parameter will fall within the interval.

13.3. We cannot be certain that the interval contains the true, unknown population parameter. However, the confidence interval

steps: 1) Grab a sample (of size n)
2) Find $\overline{X}$ (and some other statistics)
3) Use results from (2) to construct the interval $[L, U]$

If we can repeat these steps many times.... (see Figure 20)

Quality/performance : "[L,U] contains μ" or "[L,U] does not contain μ"

is constructed so that we have high confidence that it does contain the unknown population parameter

A confidence interval always specifies a *confidence level*, usually 90%, 95%, or 99%, which is a measure of the *reliability* of the procedure.

Definition 13.4. A ***confidence interval*** (CI) estimate for the population mean μ is an interval of the form $\ell \leq \mu \leq u$, where the endpoints ℓ and u are computed from the sample data.

Because different samples will produce different values of ℓ and u, these end-points are random variables and hence we should write them as L and U, respectively. Suppose that we can determine values of L and U such that the following probability statement is true:

write "$100(1-\alpha)\%$ CI on μ"

$$P\left[L \leq \mu \leq U\right] = 1 - \alpha$$

For example, 95% CI on μ
$\Rightarrow \alpha = 1 - \frac{95}{100} = 0.05$

where $0 \leq \alpha \leq 1$. There is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ.

The end-points or bounds L and U are called the ***lower-*** and ***upper-confidence limits***, respectively, and $1 - \alpha$ is called the ***confidence coefficient***. When the confidence coefficient is stated as a percentage, we call it a ***confidence level***.

To illustrate these and several other terms discussed so far, we have provided their values in the following example, which is typical of published statistical findings.

Example 13.5. Consider the following study: "In our simple random sample of 2000 households, we found the average income to be $\overline{x} = \$65,000$, with a standard deviation, $s = \$12,000$. Based on these data, we have 95% confidence that the population mean is somewhere between \$64,474 and \$65,526."

In this study, we have

- Point estimate of μ: \$65,000 $\bar{x}$
- Point estimate of σ: \$12,000 s
- Interval estimate of μ: \$64,474 to \$65,526
  ℓ u

If we can repeat the same steps of finding the "interval estimate" many times, then we can imagine finding the proportion of times that the interval actually contain μ. ∞

Probability

confidence level

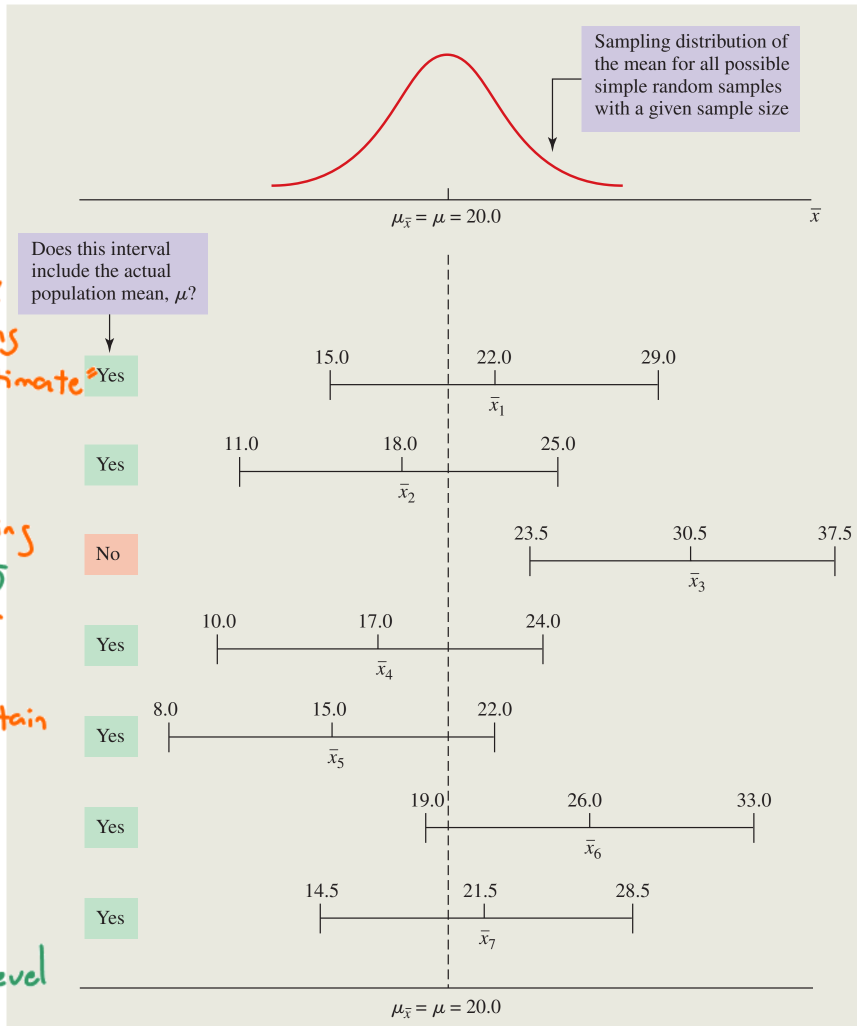

Figure 20: Examples of seven different interval estimates for a population mean, with each interval based on a separate simple random sample from the population. Six of the seven interval estimates include the actual value of μ

- Lower and upper interval limits for μ: \$64,474 and \$65,526 (ℓ, u)
- Confidence coefficient: 0.95 $= 1-\alpha$
- Confidence level: 95% $= 100(1-\alpha)\%$

When constructing a confidence interval for the mean, a key consideration is whether we know the actual value of the population standard deviation (σ). This will determine whether the normal distribution or the t distribution will be used in determining the appropriate interval.

Two techniques: know σ / do not know σ

13.1 Confidence Interval on the Mean of a Normal Distribution, Variance Known

In this subsection, we assume that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2.

13.6. Recall that $\overline{X}$ is normally distributed with mean μ and variance σ^2/n. Theorem 1: $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

From $\overline{X}$, find L, U such that $P[L \le \mu \le U] = 1-\alpha$

Start with $P\left[-c \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le c\right] = \Phi(c) - \Phi(-c) = \Phi(c) - (1-\Phi(c))$

$= 2\Phi(c) - 1$

$= 1-\alpha$

$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$\Downarrow$

Want c such that $\Phi(c) = \frac{2-\alpha}{2} = 1 - \frac{\alpha}{2}$

$$P\left[-c\frac{\sigma}{\sqrt{n}} \le \overline{X}-\mu \le c\frac{\sigma}{\sqrt{n}}\right]$$

$$P\left[\underbrace{\overline{X} - c\frac{\sigma}{\sqrt{n}}}_{L} \le \mu \le \underbrace{\overline{X} + c\frac{\sigma}{\sqrt{n}}}_{U}\right]$$

Example

confidence level	α	$1-\frac{\alpha}{2}$	$\xrightarrow{\Phi \text{ table}}$ c
90%	0.1	0.95	1.645
95%	0.05	0.975	1.960
99%	0.01	0.995	2.576

These c are called "$z_{\alpha/2}$"