Chapter 8: Interval Estimation of Parameters

LEARNING OBJECTIVES

1. Introduction
2. CI on μ of a $N(\mu, \sigma^2)$: σ^2 known
3. CI on μ of a $N(\mu, \sigma^2)$: σ^2 unknown
4. CI on μ of any distribution: large-sample
5. CI on σ^2 a normal distribution
6. CI for the proportion p: large-sample

INTRODUCTION

- A **confidence interval** estimate for is an interval of the form $l \leq \mu \leq u$, 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 values of random variables L and U, respectively.
- Suppose that we can determine values of L and U such that the following probability statement is true:

$$P(L \leq \mu \leq U) = 1 - \alpha$$

There is a probability of 1- α of selecting a sample for which the CI will contain the true value of μ

- After we have selected the sample: $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting **confidence interval** for μ is

$$l \leq \mu \leq u$$

Lower-confidence limit

Upper-confidence limit

How to find the random variables L and U ?

Since $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$, we can find a real number

$z_{\alpha/2}$ such that

=NORMSINV(1- $\alpha/2$)

$$P(-z_{\alpha/2} \le \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \le z_{\alpha/2}) = 1 - \alpha$$

or

$$P(\overline{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}})$$

L

U

Theorem

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance, $(1 - \alpha)$-CI on μ is given by

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Example

A college admissions director wishes to estimate the mean age of all students currently enrolled. In a random sample of 20 students, the mean age is found to be 22.9 years. From past studies, the standard deviation is known to be 1.5 years, and the population is normally distributed. Construct a 90% confidence interval of the population mean age.

Solution: We have $\bar{x} = 22.9$; $\sigma = 1.5$; $1 - \alpha = 0.9$;

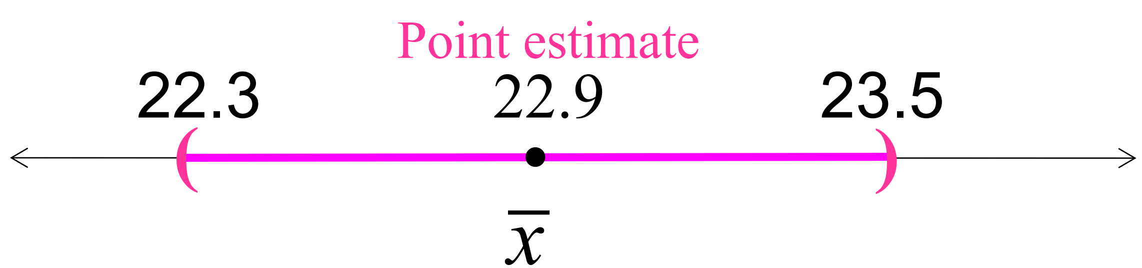
$$\alpha = 0.1 \Rightarrow z_{\alpha/2} = \text{NORMSINV}(1\text{- }0.1/2) = 1.645$$

Confidence interval:

$$22.9 - 1.645\frac{1.5}{\sqrt{20}} \le \mu \le 22.9 - 1.645\frac{1.5}{\sqrt{20}}$$

or

$$\mathbf{22.3 \le \mu \le 23.5}$$



CI ON μ OF A N(μ, σ^2): σ^2 KNOWN

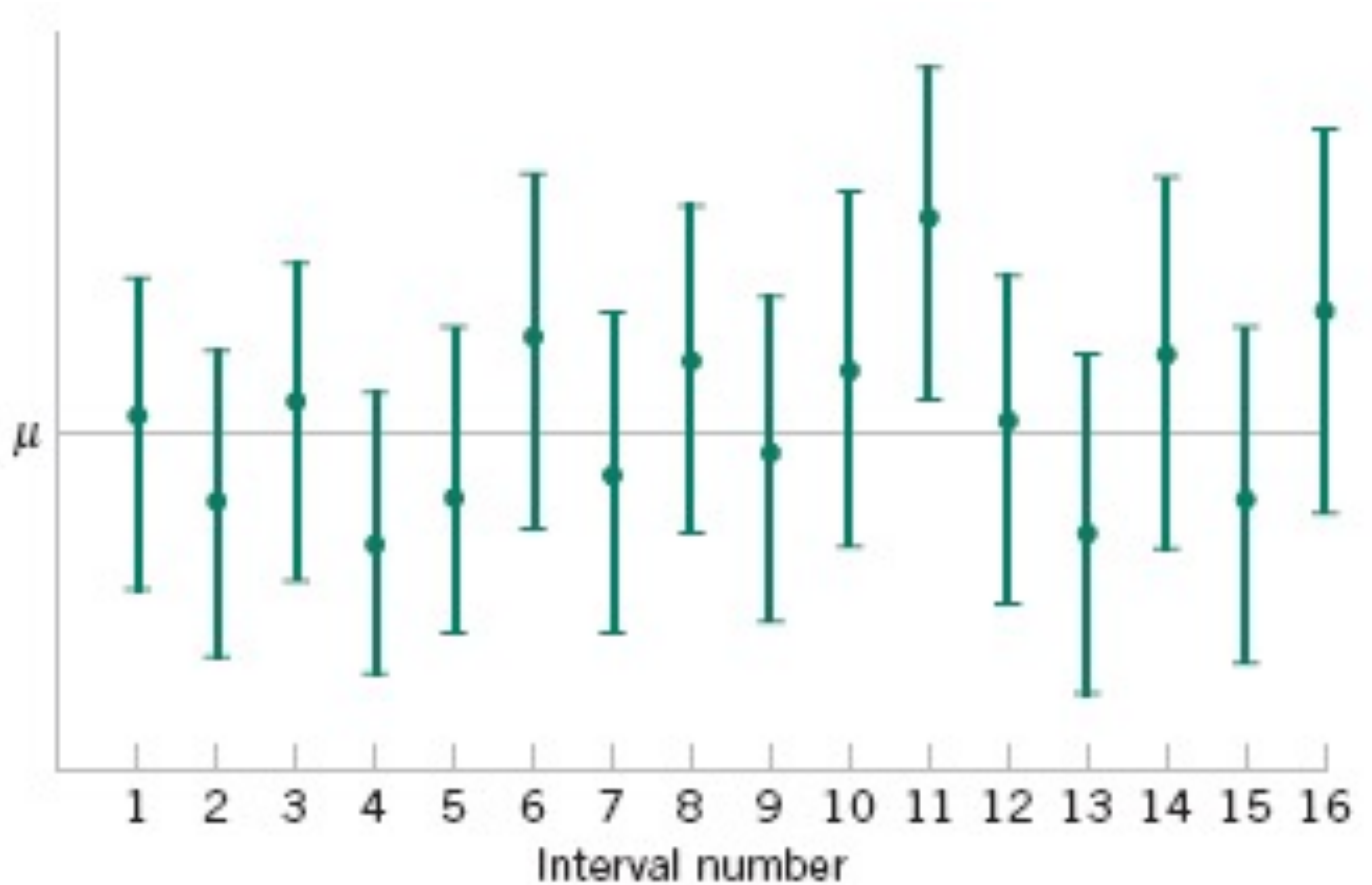


Figure 8-1 Repeated construction of a confidence interval for μ.

Confidence Level and Precision of Estimation

The length of a confidence interval is a measure of the precision of estimation.

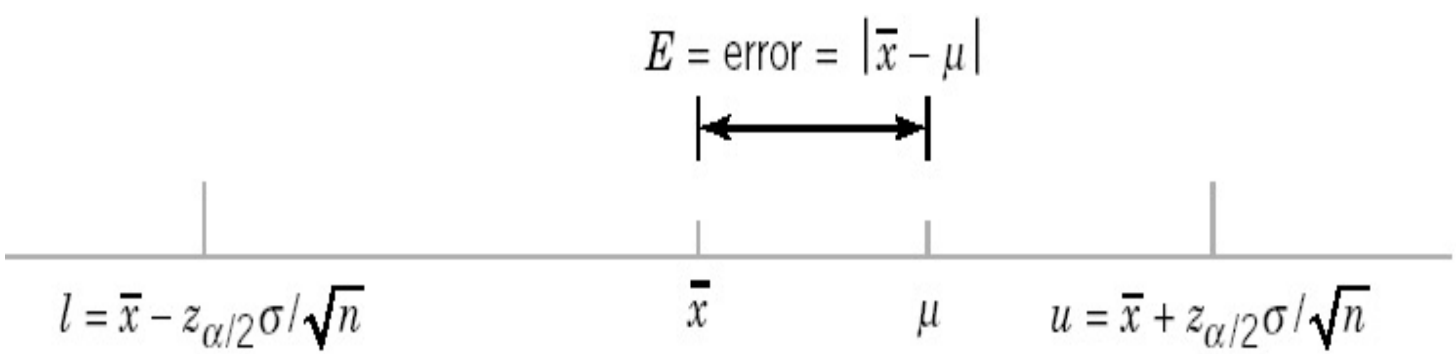


Figure 8-2 Error in estimating μ with $\bar{x}$.

Theorem — **Choice of Sample Size**

If $\bar{x}$ is used as an estimate of μ , we can be $(1 - \alpha)$-confident that the error $|\bar{x} - \mu|$ will not exceed a specified amount E when the sample size is

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

Theorem — **One-Sided Confidence Bounds**

A $(1 - \alpha)$ upper-confidence bound for μ is

$$\mu \le \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

A $(1 - \alpha)$ lower-confidence bound for μ is

$$\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}} \le \mu$$

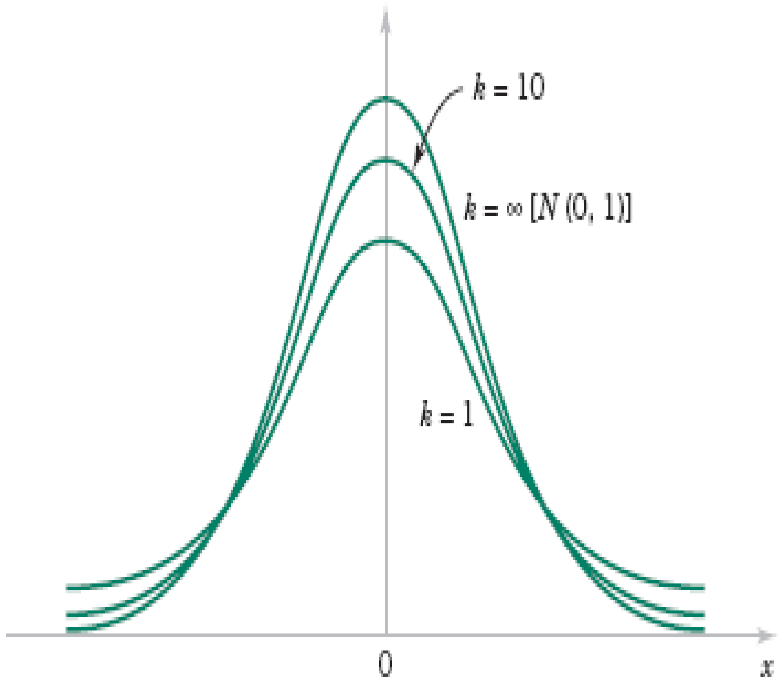
Definition Student Distribution

The variable random X with probability density function

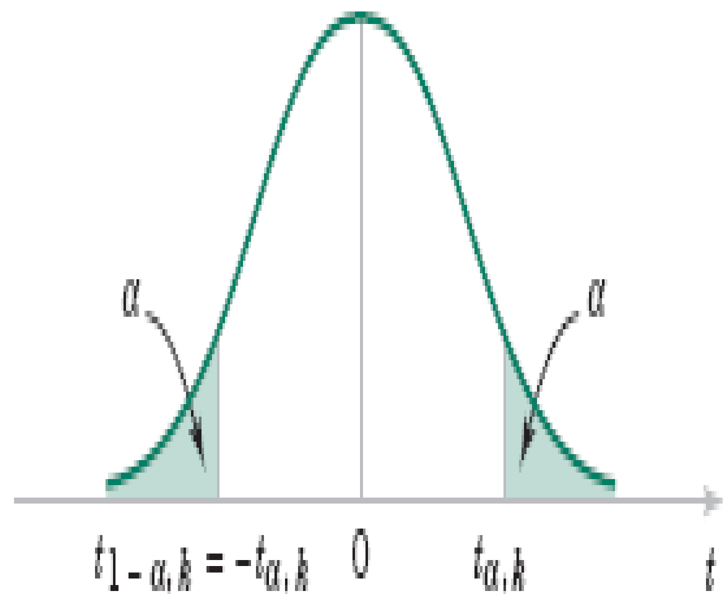
$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma(k/2)} \frac{1}{\left(\frac{x^2}{k}+1\right)^{\frac{k+1}{2}}}, \quad -\infty < x < +\infty$$

is called a Student variable random or *t* distribution with *k* degrees of freedom.

CI ON μ OF A N(μ, σ²): σ² UNKNOWN



Probability density functions of several *t* distributions.



Percentage points of the *t* distribution.

Theorem — *t* distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t distribution with n - 1 degrees of freedom.

Theorem

If $\bar{x}$ and *s* are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2,

- A (1 – *α*)-**percent confidence interval** on μ is given by

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

=TINV(*α*,*n*-1)

- A (1 – *α*) **upper-confidence bound** for μ is

$$\mu \le \bar{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}$$

- A (1 – *α*) **lower-confidence bound** for μ is

$$\bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}} \le \mu$$

Example

An article in the journal *Materials Engineering* describes the results of tensile adhesion tests on 22 U-700 alloy specimens. The load at specimen failure is as follows:

19.8 10.1 14.9 7.5 15.4 15.4 15.4 18.5 7.9 12.7 11.9 11.4 11.4 14.1 17.6 16.7 15.8 19.5 8.8 13.6 11.9 11.4

Find a 95% CI on μ.

Solution: We have $\bar{x} = 13.71$ and $s = 3.55$

$$\alpha = 1 - 0.95 = 0.05;\ \ n = 22 \Rightarrow t_{\alpha/2,n-1} = \text{TINV}(0.05,21) = 2.08$$

A 95% CI on μ is $\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n}$

$$12.14 \le \mu \le 15.28$$

ARBITRARY DISTRIBUTION: A LARGE-SAMPLE

Theorem

When n is large, the quantity

$\boxed{n \geq 40}$

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

has an approximate $N(0, 1)$. Consequently,

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

is a **large sample confidence interval** for μ, with confidence level of approximately $(1 - \alpha)$.

Example

A sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values are

1.230	0.490	0.490	1.080	0.590	0.280	0.180	0.100	0.940
1.330	0.190	1.160	0.980	0.340	0.340	0.190	0.210	0.400
0.040	0.830	0.050	0.630	0.340	0.750	0.040	0.860	0.430
0.044	0.810	0.150	0.560	0.840	0.870	0.490	0.520	0.250
1.200	0.710	0.190	0.410	0.500	0.560	1.100	0.650	0.270
0.270	0.500	0.770	0.730	0.340	0.170	0.160	0.270	

Find an approximate 95% CI on μ.

Solution: We have $\bar{x} = 0.5250$ and $s = 0.3486$

$\alpha = 1- 0.95 = 0.05$; $z_{\alpha/2} = 1.96$

ARBITRARY DISTRIBUTION: A LARGE-SAMPLE

Distribution of mercury concentration is not normal

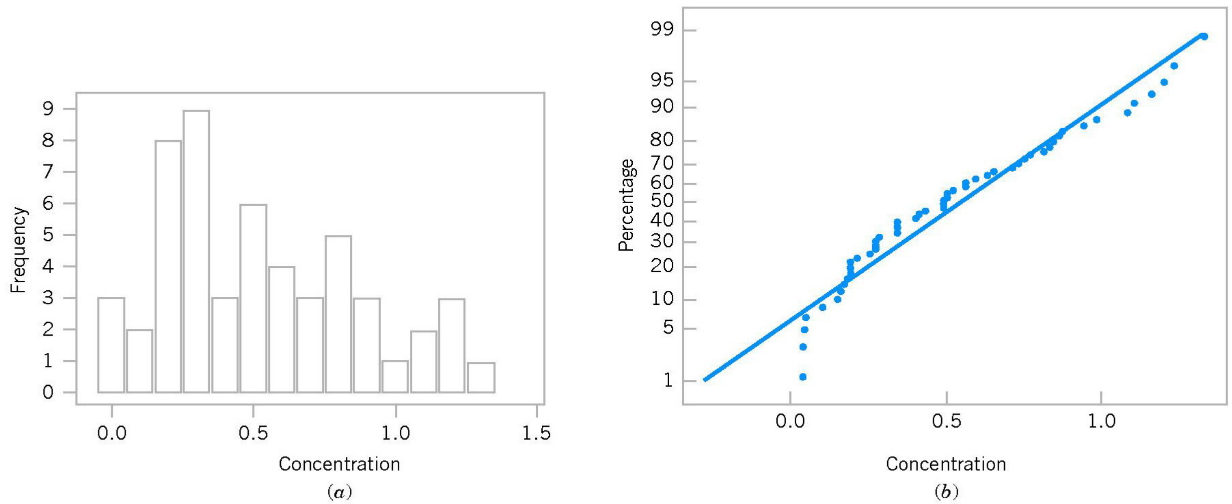


Figure 8-3 Mercury concentration in largemouth bass (a) Histogram. (b) Normal probability plot.

Because $n > 40$, the approximate 95% CI on is

$$\bar{x} - z_{0.025}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{0.025}\frac{s}{\sqrt{n}}$$

$$0.5250 - 1.96\frac{0.3486}{\sqrt{53}} \le \mu \le 0.5250 + 1.96\frac{0.3486}{\sqrt{53}}$$

$$0.4311 \le \mu \le 0.6189$$

Definition — **Chi-square Distribution**

The variable random X with probability density function

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{\frac{-x}{2}} \quad , \qquad x > 0$$

is called a χ^2 variable random with k degrees of freedom

Theorem

The random variable

$$\frac{(n-1)S^2}{\sigma^2}$$

has a χ^2 distribution with n -1 degrees of freedom.

CI ON σ^2 OF A NORMAL

Theorem

- A $(1-\alpha)$-**percent confidence interval** on σ^2 is given by

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$

=CHIINV(1- *α*/2,*n*-1)

- A $(1-\alpha)$ **upper-confidence bound** for σ^2 is

$$\sigma^2 \le \frac{(n-1)s^2}{\chi_{1-\alpha,n-1}}$$

- A $(1-\alpha)$ **lower-confidence bound** for σ^2 is

$$\frac{(n-1)s^2}{\chi_{\alpha,n-1}} \le \sigma^2$$

Theorem — Normal approximation

If n is large, the distribution of

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

is approximately $N(0, 1)$.

We can find $z_{\alpha/2}$ such that $P\left(-z_{\alpha/2} \le \dfrac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \le z_{\alpha/2}\right) \simeq 1-\alpha$

or

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le p \le \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \simeq 1-\alpha$$

Theorem CI for p

Let $\hat{p}$ is a point estimation for the proportion p of the population based on a random sample of size n, an approximate $(1 - \alpha)$-confidence interval on p is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Remark: $n\hat{p} > 5$ and $n(1-\hat{p}) > 5$

Example

In a survey of 1219 U.S. adults, 354 said that their favorite sport to watch is football. Construct a 95% confidence interval for the proportion of adults in the United States who say that their favorite sport to watch is football.

Solution: The point estimation for p:

$$\hat{p} = \frac{354}{1219} \simeq 0.29$$

Because $n\hat{p} \approx 354 > 5$ and $n(1-\hat{p}) \approx 865 > 5$, a 95% CI for p is:

$$0.29 - 1.96\sqrt{\frac{0.29(1-0.29)}{1219}} \le p \le 0.29 - 1.96\sqrt{\frac{0.29(1-0.29)}{1219}}$$

$$0.265 \le p \le 0.315$$

Theorem — Choice of Sample Size

If $\hat{p}$ is used as an estimate of p, we can be $(1-\alpha)$-confident that the error $|\hat{p} - p|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$$

$\leq 1/4$

An upper bound on n is given by

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

Theorem One-Sided Confidence Bounds

A $(1-\alpha)$ **upper-confidence bound** for p is

$$p \leq \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A $(1-\alpha)$ **lower-confidence bound** for p is

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p$$

SUMMARY

We have studied:

1.CI on μ of a $N(\mu, \sigma^2)$: σ^2 known

2. CI on μ of a $N(\mu, \sigma^2)$: σ^2 unknown

3. CI on μ of any distribution: large-sample

4. CI on $\sigma2$ a normal distribution

5. CI for the proportion *p*: large-sample

Homework: Read slides of the next lecture.