Confidence Intervals (1 of 2)

Contents of this chapter:

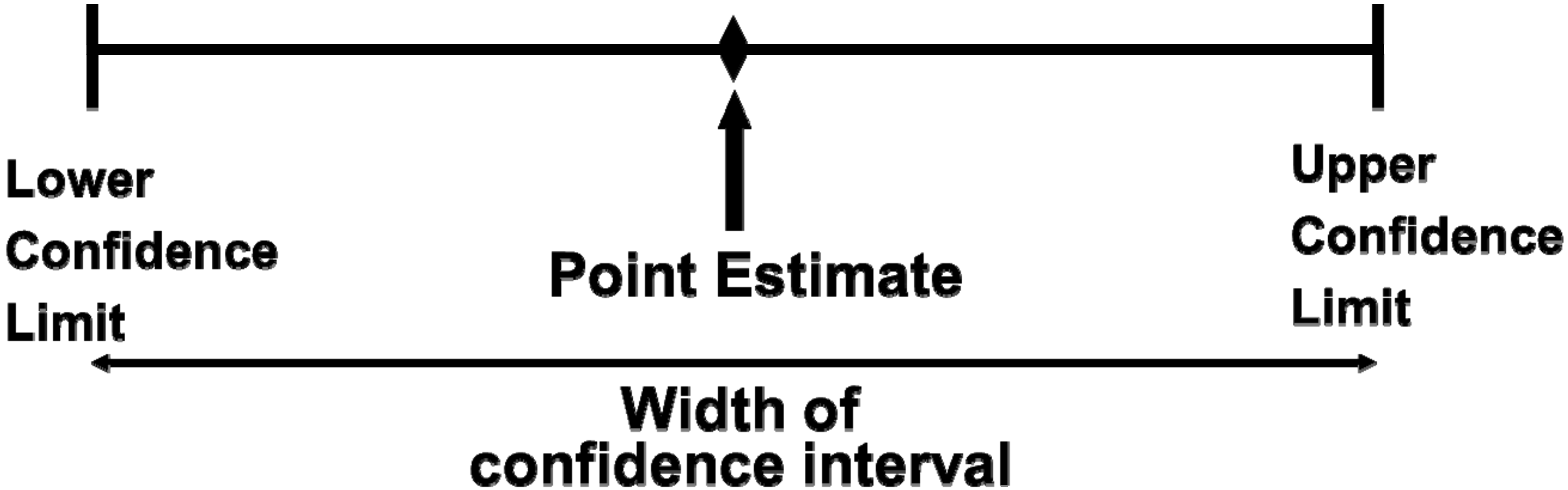
- Confidence Intervals for the Population Mean, μ
  - when Population Variance σ^2 is Known
  - when Population Variance σ^2 is Unknown
- Confidence Intervals for the Population Proportion, *P* (large samples)
- Confidence interval estimates for the variance of a normal population
- Finite population corrections
- Sample-size determination

Section 7.1 Properties of Point Estimators

- An estimator of a population parameter is
  - a random variable that depends on sample information . . .
  - whose value provides an approximation to this unknown parameter
- A specific value of that random variable is called an estimate

Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about variability



Point Estimates

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\overline{x}$
Proportion	P	$\hat{p}$

Unbiasedness (1 of 2)

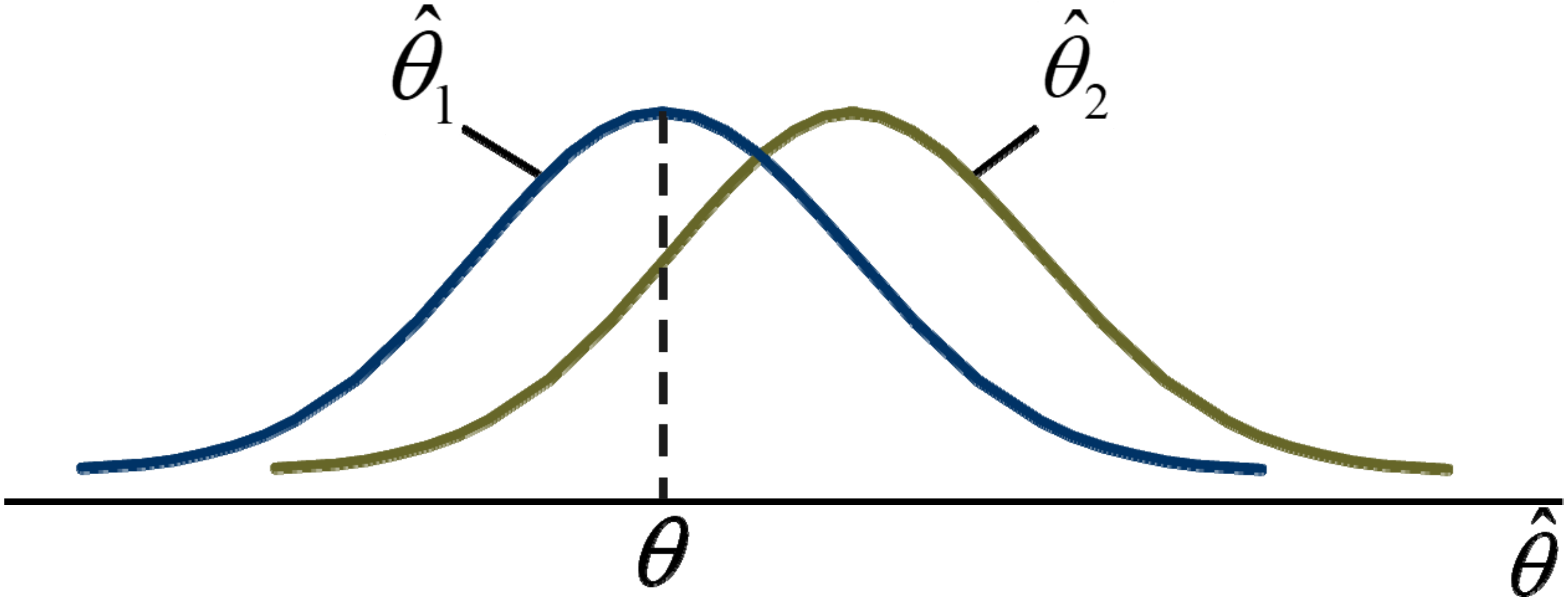
- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if its expected value is equal to that parameter:

$$E(\hat{\theta}) = \theta$$

- Examples:
  - The sample mean $\overline{x}$ is an unbiased estimator of μ
  - The sample variance s^2 is an unbiased estimator σ^2
  - The sample proportion $\hat{p}$ is an unbiased estimator of *P*

Unbiasedness (2 of 2)

- $\hat{\theta}_1$ is an unbiased estimator, $\hat{\theta}_2$ is biased:



Bias

- Let $\hat{\theta}$ be an estimator of $\boldsymbol{\theta}$
- The bias in $\hat{\theta}$ is defined as the difference between its mean and $\boldsymbol{\theta}$

$$\text{Bias}(\hat{\theta}) = E(\theta) - \theta$$

- The bias of an unbiased estimator is 0

Most Efficient Estimator

- Suppose there are several unbiased estimators of $\boldsymbol{\theta}$
- The most efficient estimator or the minimum variance unbiased estimator of $\boldsymbol{\theta}$ is the unbiased estimator with the smallest variance
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators $\boldsymbol{\theta}$, based on the same number of sample observations. Then,
  - $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if $Var(\hat{\theta}_1) < Var(\theta_2)$
  - The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is the ratio of their variances:

$$\text{Relative Efficiency} = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$$

Confidence Interval Estimation

- How much uncertainty is associated with a point estimate of a population parameter?
- An interval estimate provides more information about a population characteristic than does a point estimate
- Such interval estimates are called confidence interval estimates

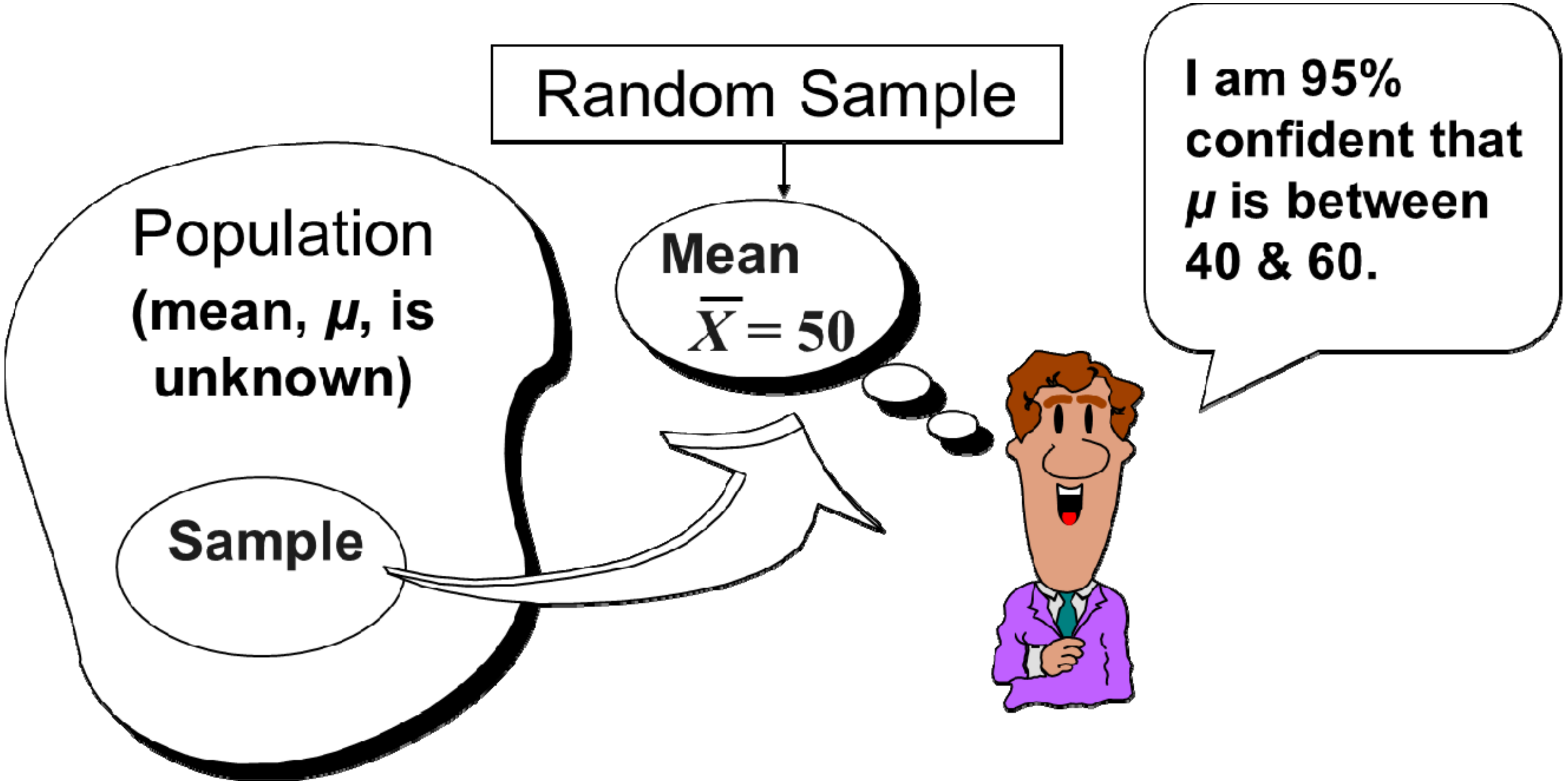
Confidence Interval Estimate

- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observation from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - Can never be 100% confident

Confidence Interval and Confidence Level

- If $P(a < \theta < b) = 1 - \alpha$ then the interval from *a* to *b* is called a $100(1-\alpha)\%$ confidence interval of θ.
- The quantity $100(1-\alpha)\%$ is called the confidence level of the interval
  - α is between 0 and 1
  - In repeated samples of the population, the true value of the parameter θ would be contained in $100(1-\alpha)\%$ of intervals calculated this way.
  - The confidence interval calculated in this manner is written as $a < \theta < b$ with $100(1-\alpha)\%$ confidence

Estimation Process



Confidence Level, Left Parenthesis 1 Minus Alpha Right Parenthesis

- Suppose confidence level = 95%
- Also written $(1-\alpha)=0.95$
- A relative frequency interpretation:
  - From repeated samples, 95% of all the confidence intervals that can be constructed of size *n* will contain the unknown true parameter
- A specific interval either will contain or will not contain the true parameter
  - No probability involved in a specific interval

General Formula

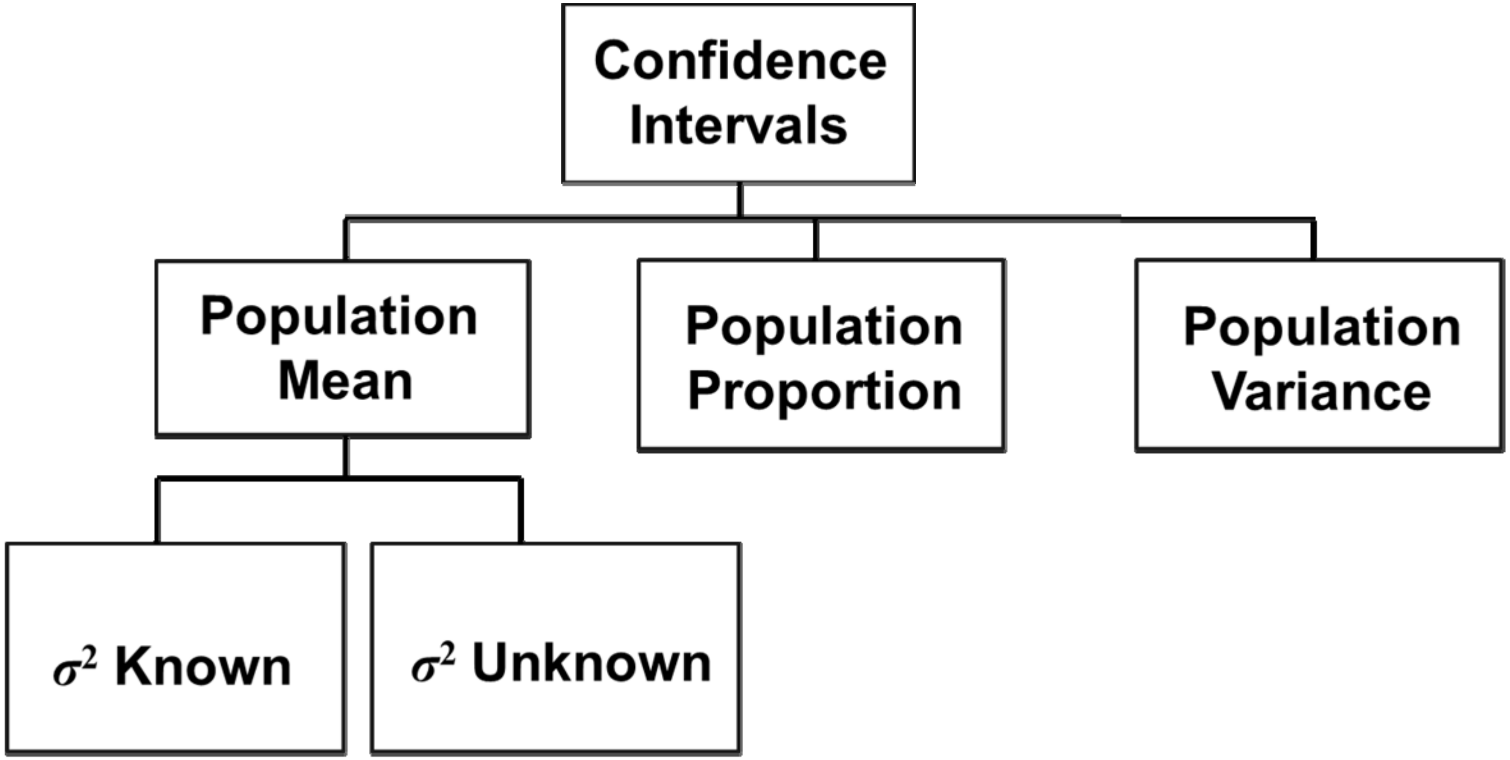
- The general form for all confidence intervals is:

$$\hat{\theta} \pm ME$$

$$\textbf{Point Estimate} \pm \textbf{Margin of Error}$$

- The value of the margin of error depends on the desired level of confidence

Confidence Intervals (2 of 2)



(From normally distributed populations)

Section 7.2 Confidence Interval Estimation for the Mean (Sigma Squared Known)

- Assumptions
  - Population variance σ^2 is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

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

(where $z_{\frac{\alpha}{2}}$ is the normal distribution value for a probability of $\frac{\alpha}{2}$ in each tail)

Confidence Limits

- The confidence interval is

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

- The endpoints of the interval are

$$\text{UCL} = \overline{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$ Upper confidence limit

$$\text{LCL} = \overline{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$ Lower confidence limit

Margin of Error (1 of 2)

- The confidence interval,

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

- Can also be written as $\bar{x} \pm ME$

  where *ME* is called the margin of error

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

- The interval width, *w*, is equal to twice the margin of error

Reducing the Margin of Error

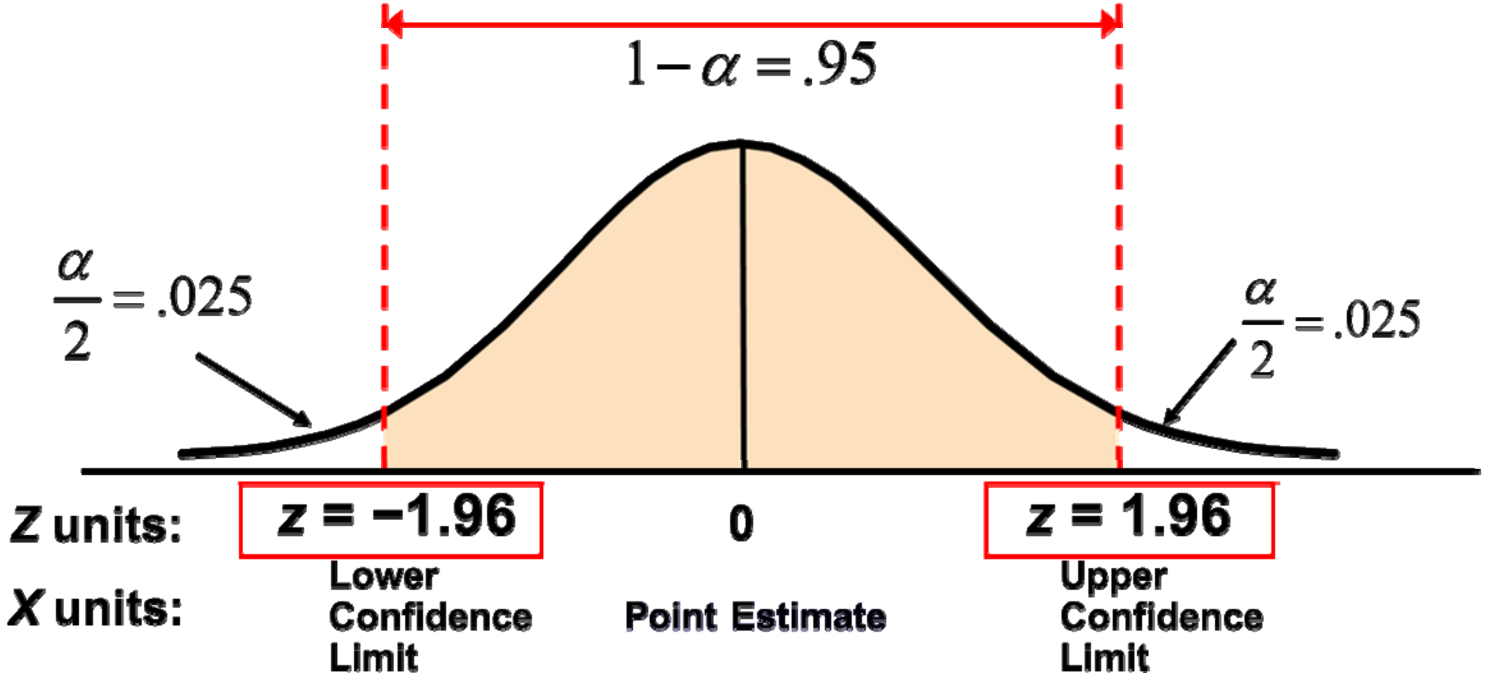
$$ME = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

The margin of error can be reduced if

- the population standard deviation can be reduced $(\sigma \downarrow)$
- The sample size is increased $(n \uparrow)$
- The confidence level is decreased, $(1 ? \alpha) \downarrow$

Finding Z of Start Expression Start Fraction Alpha over 2 End Fraction End Expression

- Consider a 95% confidence interval:



- Find $z_{.025} = \pm 1.96$ from the standard normal distribution table

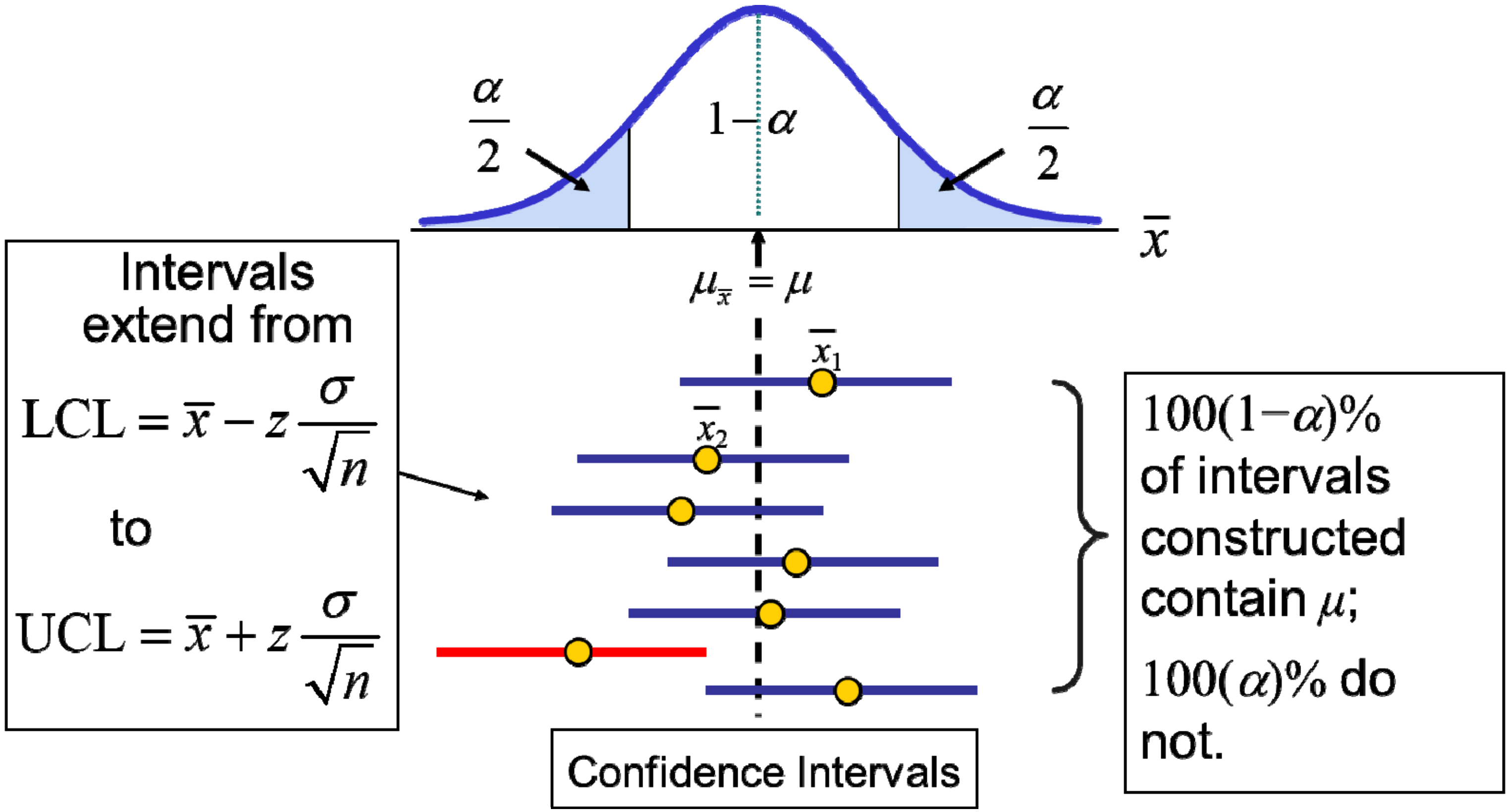
Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, 98%, and 99%

Confidence Level	Confidence Coefficient, $1-\alpha$	$Z_{\frac{\alpha}{2}}$ value
80%	.80	1.28
90%	.90	1.645
95%	.95	1.96
98%	.98	2.33
99%	.99	2.58
99.8%	.998	3.08
99.9%	.999	3.27

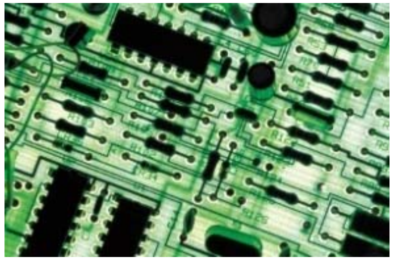
Intervals and Level of Confidence

Sampling Distribution of the Mean



Example 1 (1 of 2)

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.
- Determine a 95% confidence interval for the true mean resistance of the population.



Example 1 (2 of 2)

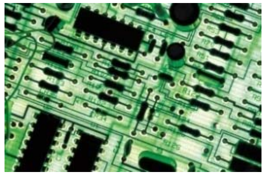
- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is .35 ohms.
- Solution:

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

$$= 2.20 \pm 1.96\left(\frac{.35}{\sqrt{11}}\right)$$

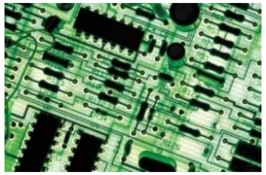
$$= 2.20 \pm .2068$$

$$1.9932 < \mu < 2.4068$$

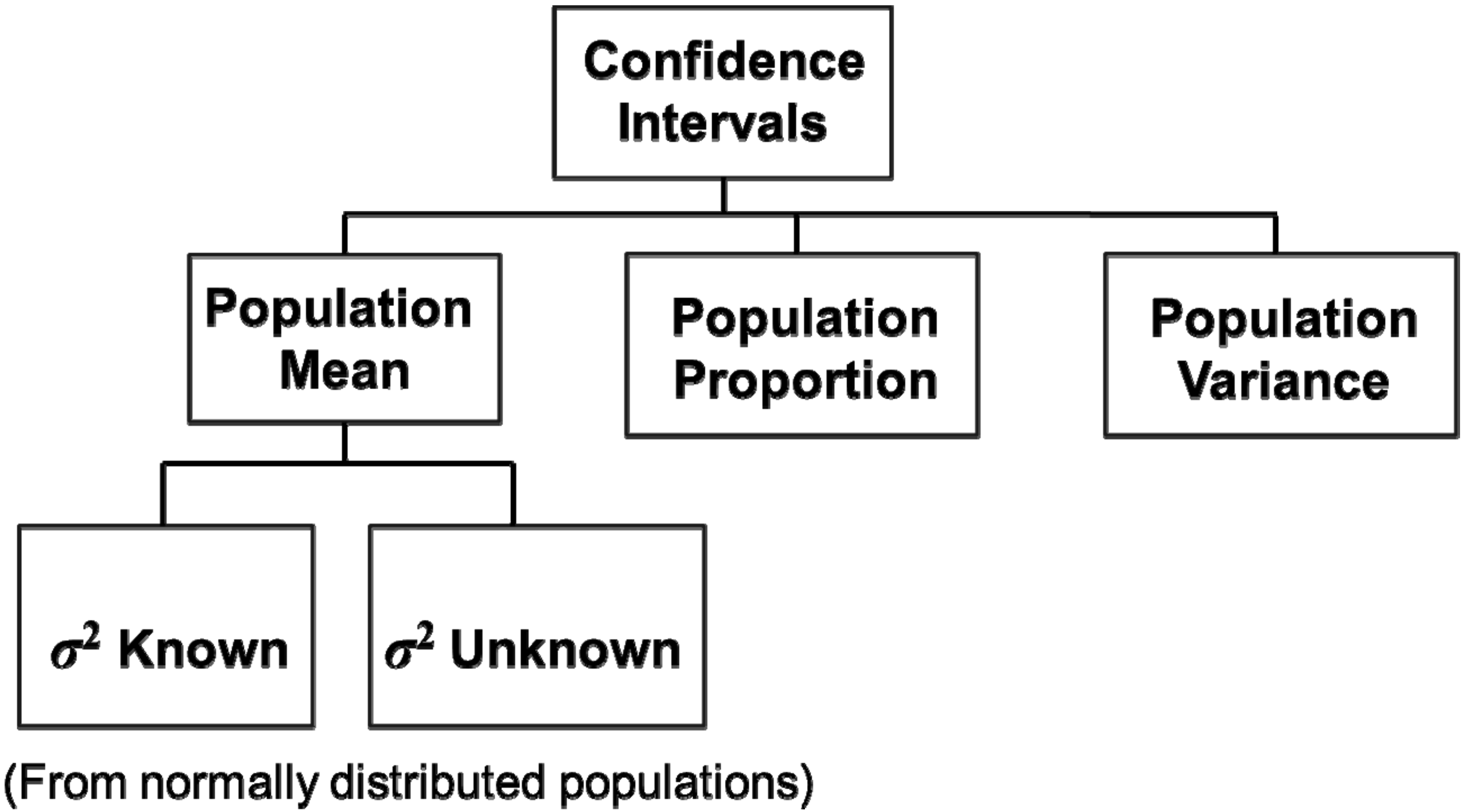


Interpretation (1 of 2)

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean



Section 7.3 Confidence Interval Estimation for the Mean (Sigma Squared Unknown)



Student's *t* Distribution (1 of 3)

- Consider a random sample of *n* observations
  - with mean $\bar{x}$ and standard deviation *s*
  - from a normally distributed population with mean μ
- Then the variable

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

  follows the Student's *t* distribution with (*n* − 1) degrees of freedom

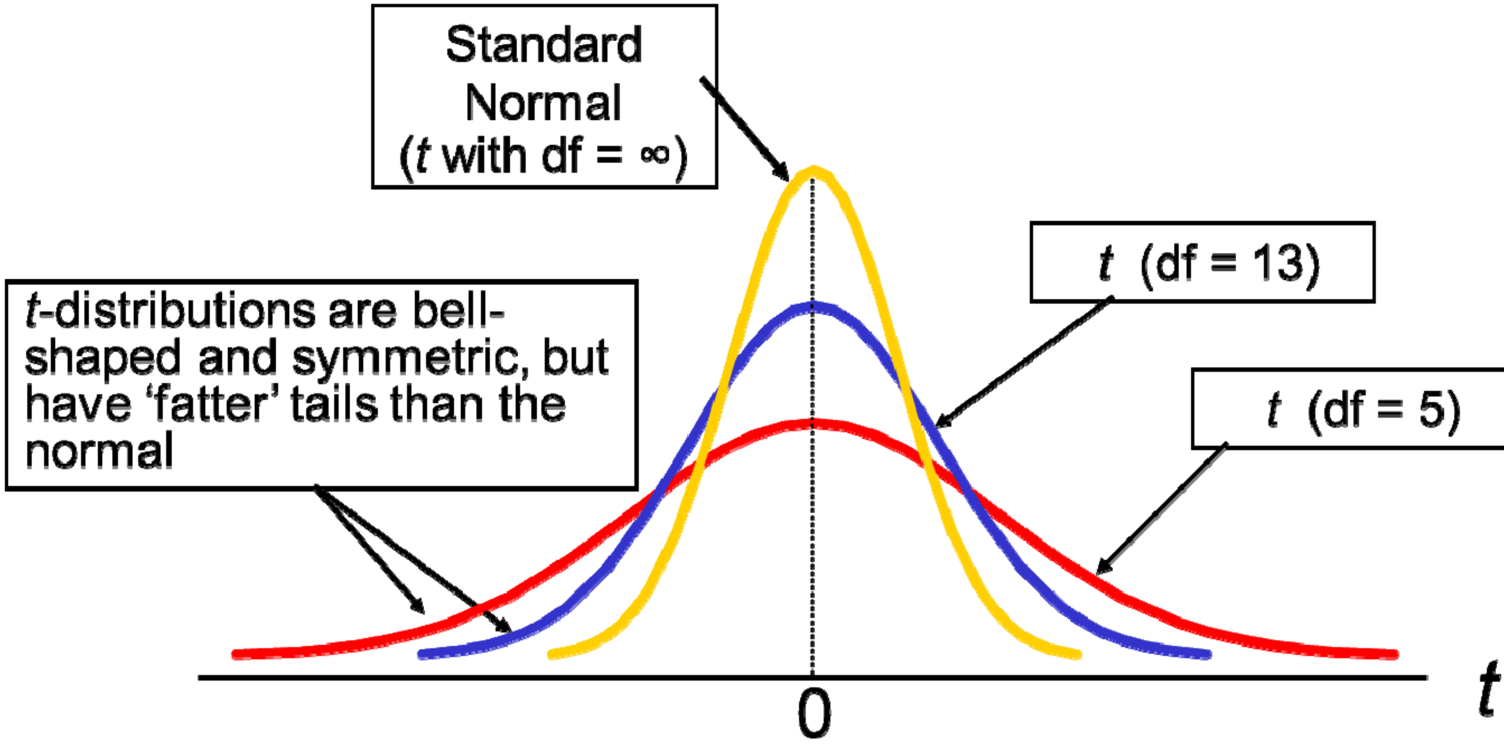
Student's *t* Distribution (2 of 3)

- The t is a family of distributions
- The t value depends on degrees of freedom (d.f.)
  - Number of observations that are free to vary after sample mean has been calculated

$$\text{d.f.} = n - 1$$

Student's *t* Distribution (3 of 3)

Note: $t \rightarrow Z$ as *n* increases



Student's *t* Table

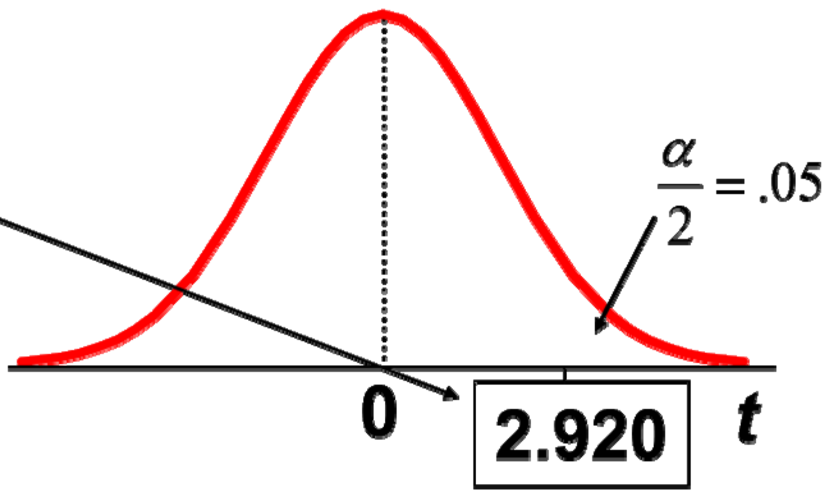
	Upper Tail Area		
df	.10	**.05**	.025
1	3.078	6.314	12.706
2	1.886	**2.920**	4.303
3	1.638	2.353	3.182

The body of the table contains *t* values, not probabilities

Let: $n = 3$
df $= n - 1 = 2$

$\alpha = .10$

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



t Distribution Values

With comparison to the *Z* value

Confidence Level	*t* (10 d.f.)	*t* (20 d.f.)	*t* (30 d.f.)	*z*
.80	1.372	1.325	1.310	1.282
.90	1.812	1.725	1.697	1.645
.95	2.228	2.086	2.042	1.960
.99	3.169	2.845	2.750	2.576

Note: $t \rightarrow Z$ as *n* increases

Confidence Interval Estimation for the Mean (Sigma Squared Unknown) (1 of 2)

- If the population standard deviation σ is unknown, we can substitute the sample standard deviation, *s*
- This introduces extra uncertainty, since *s* is variable from sample to sample
- So we use the *t* distribution instead of the normal distribution

Confidence Interval Estimation for the Mean (Sigma Squared Unknown) (2 of 2)

- Assumptions
  - Population standard deviation is unknown
  - Population is normally distributed
  - If population is not normal, use large sample
- Use Student's *t* Distribution
- Confidence Interval Estimate:

$$\bar{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

where $t_{n-1,\frac{\alpha}{2}}$ is the critical value of the *t* distribution with *n* − 1 d.f.

and an area of $\frac{\alpha}{2}$ in each tail: $P\left(t_{n-1} > t_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$

Margin of Error (2 of 2)

- The confidence interval,

$$\bar{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

- Can also be written as $\bar{x} \pm ME$

where *ME* is called the margin of error:

$$ME = t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Example 2

A random sample of n = 25 has $\bar{x} = 50$ and s = 8. Form a 95% confidence interval for μ

– d.f. = n − 1 = 24, so $t_{n-1,\frac{\alpha}{2}} = t_{24,.025} = 2.0639$

The confidence interval is

$$\bar{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

$$46.698 < \mu < 53.302$$