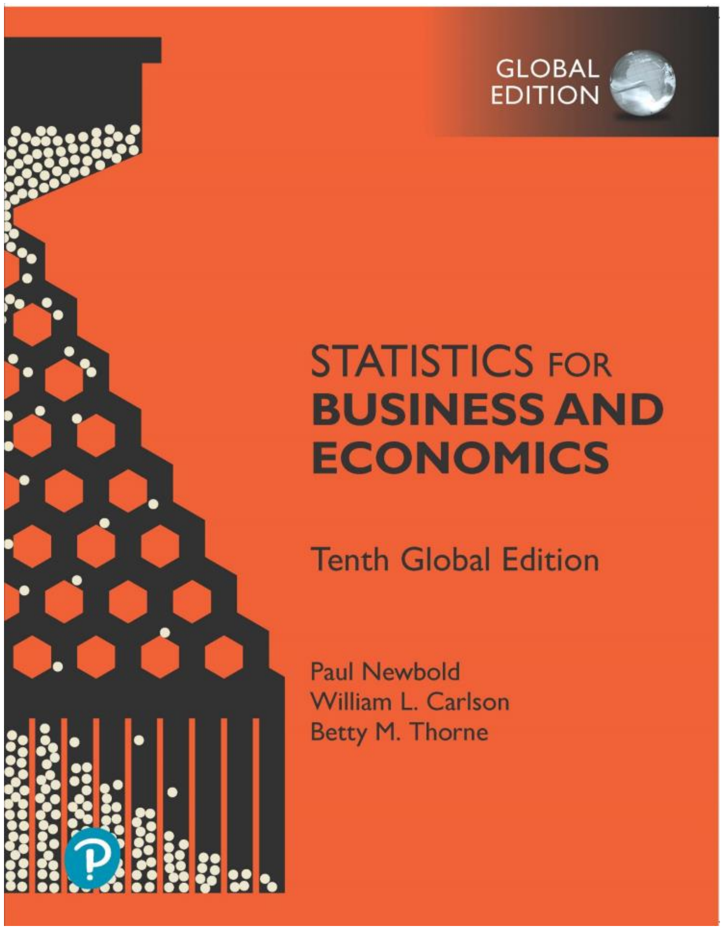
Statistics for Business and Economics

Tenth Edition, Global Edition



Chapter 7

Estimation: Single Population

Chapter Goals

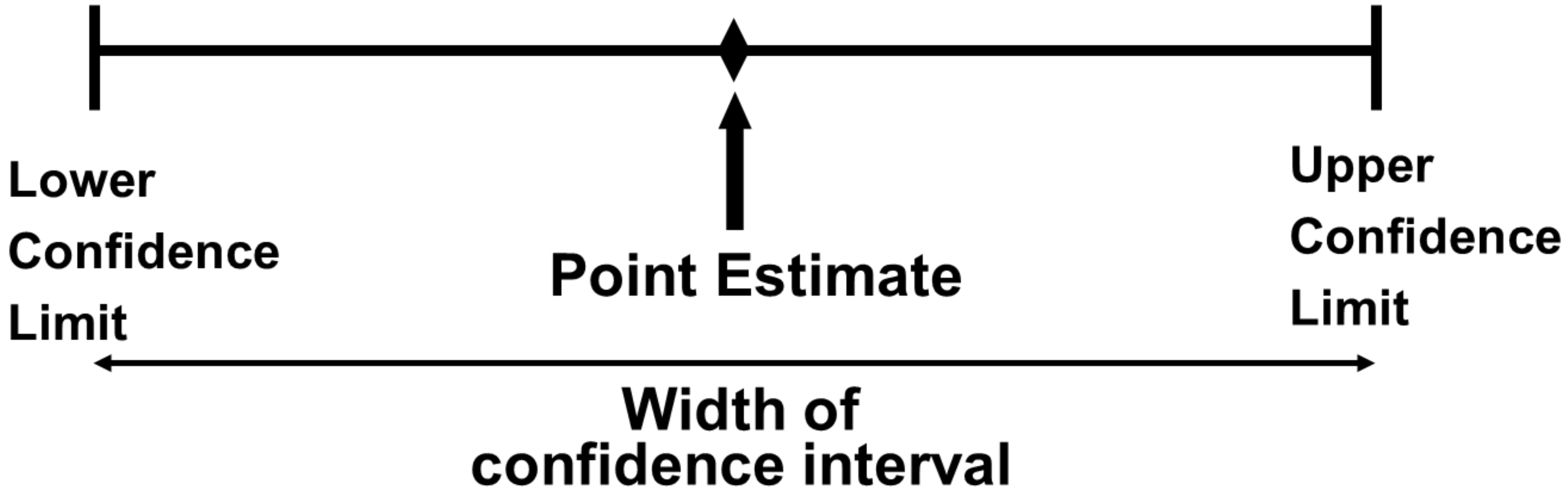
- Distinguish between a point estimate and a confidence interval estimate
- Construct and interpret a confidence interval estimate for
  - single population mean using both the Z and t distributions
  - single population proportion
  - variance of a normal population
- Determine the required sample size to estimate a mean or proportion within a specified margin of error

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	$\bar{x}$
Proportion	P	$\hat{p}$

Unbiasedness (1 of 2)

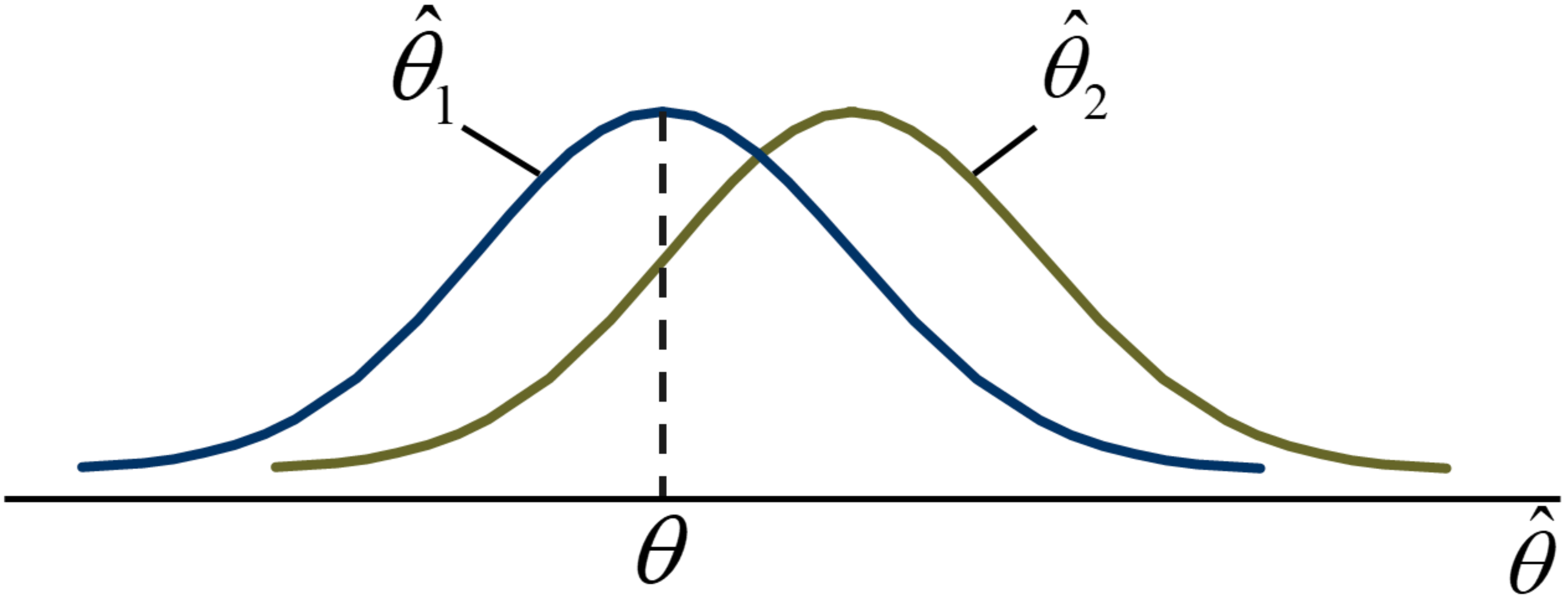
- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter $\boldsymbol{\theta}$ if its expected value is equal to that parameter:

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

- Examples:
  - The sample mean $\bar{x}$ 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 $\boldsymbol{\theta}$

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

- 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 $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$
  - 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

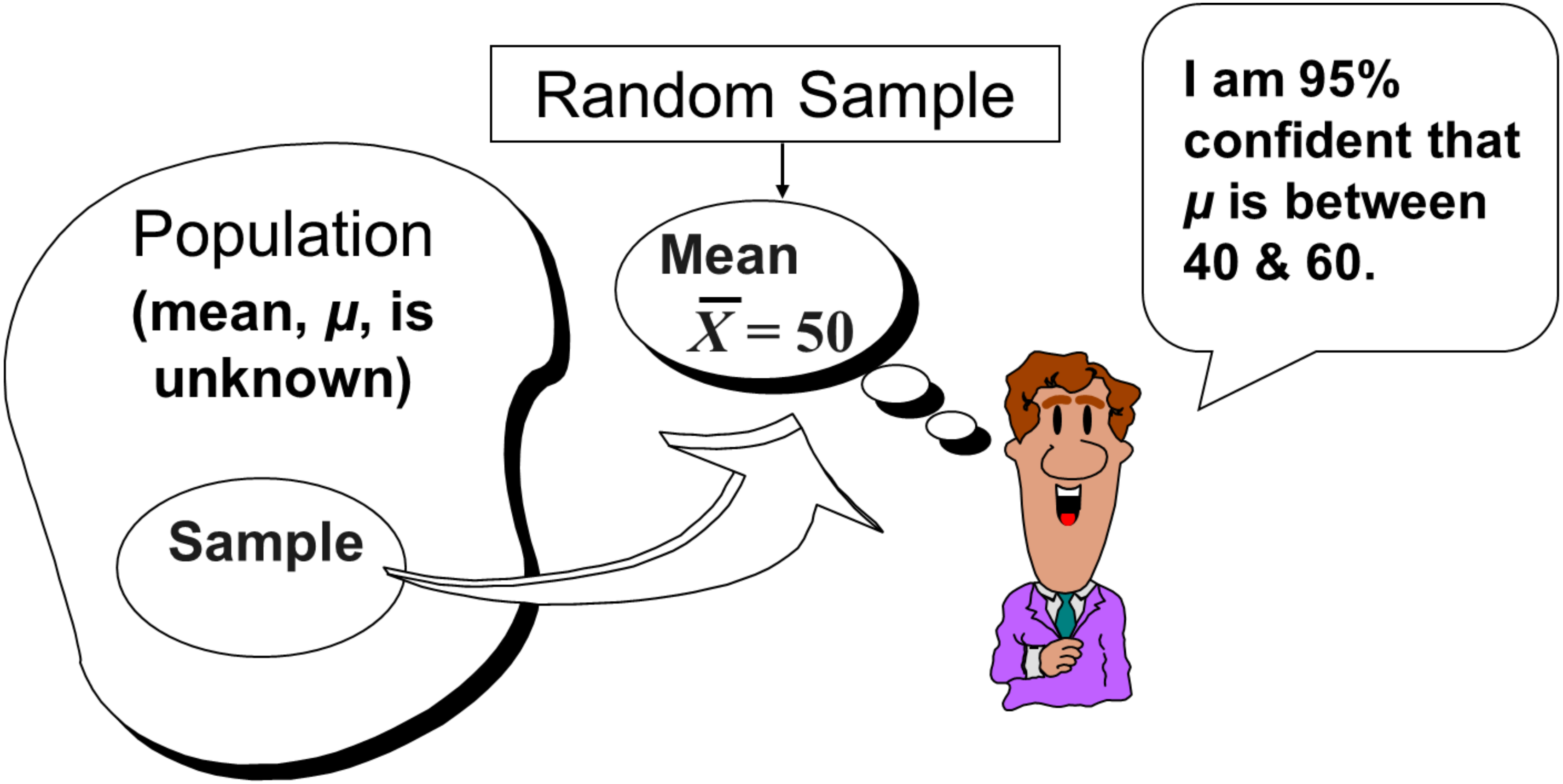
Confidence Interval Estimate

- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observation from one 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 $P(a < \boldsymbol{\theta} < b) = 1 - \boldsymbol{\alpha}$ then the interval from *a* to *b* is called a $100(1-\boldsymbol{\alpha})\%$ confidence interval of $\boldsymbol{\theta}$.
- The quantity $100(1-\boldsymbol{\alpha})\%$ is called the confidence level of the interval
  - $\boldsymbol{\alpha}$ is between 0 and 1
  - In repeated samples of the population, the true value of the parameter $\boldsymbol{\theta}$ would be contained in $100(1-\boldsymbol{\alpha})\%$ of intervals calculated this way.
  - The confidence interval calculated in this manner is written as $a < \boldsymbol{\theta} < b$ with $100(1-\boldsymbol{\alpha})\%$ confidence

Estimation Process



- 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

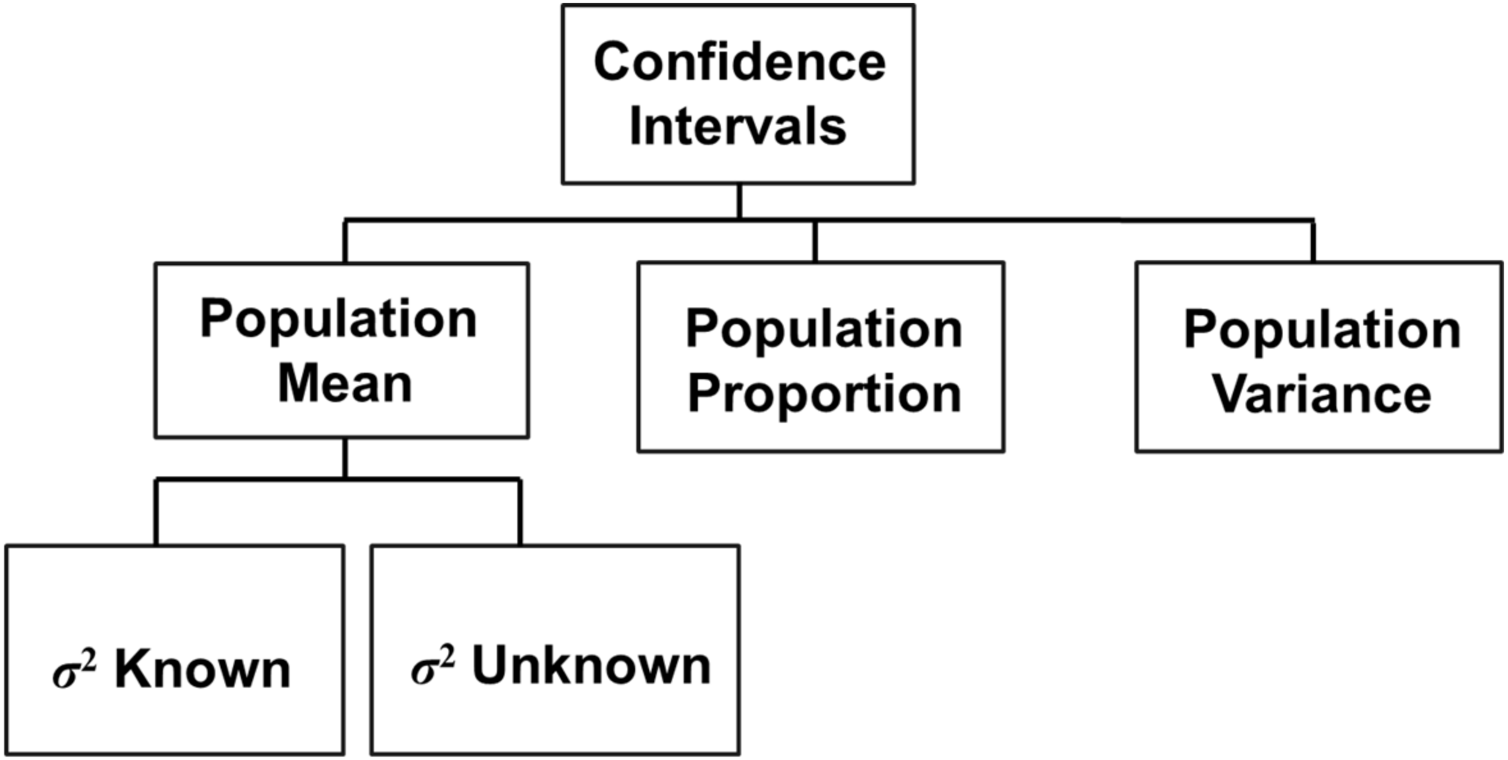
- The general form for all confidence intervals is:

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

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

- 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

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

- The endpoints of the interval are

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

$$\text{LCL} = \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$ 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

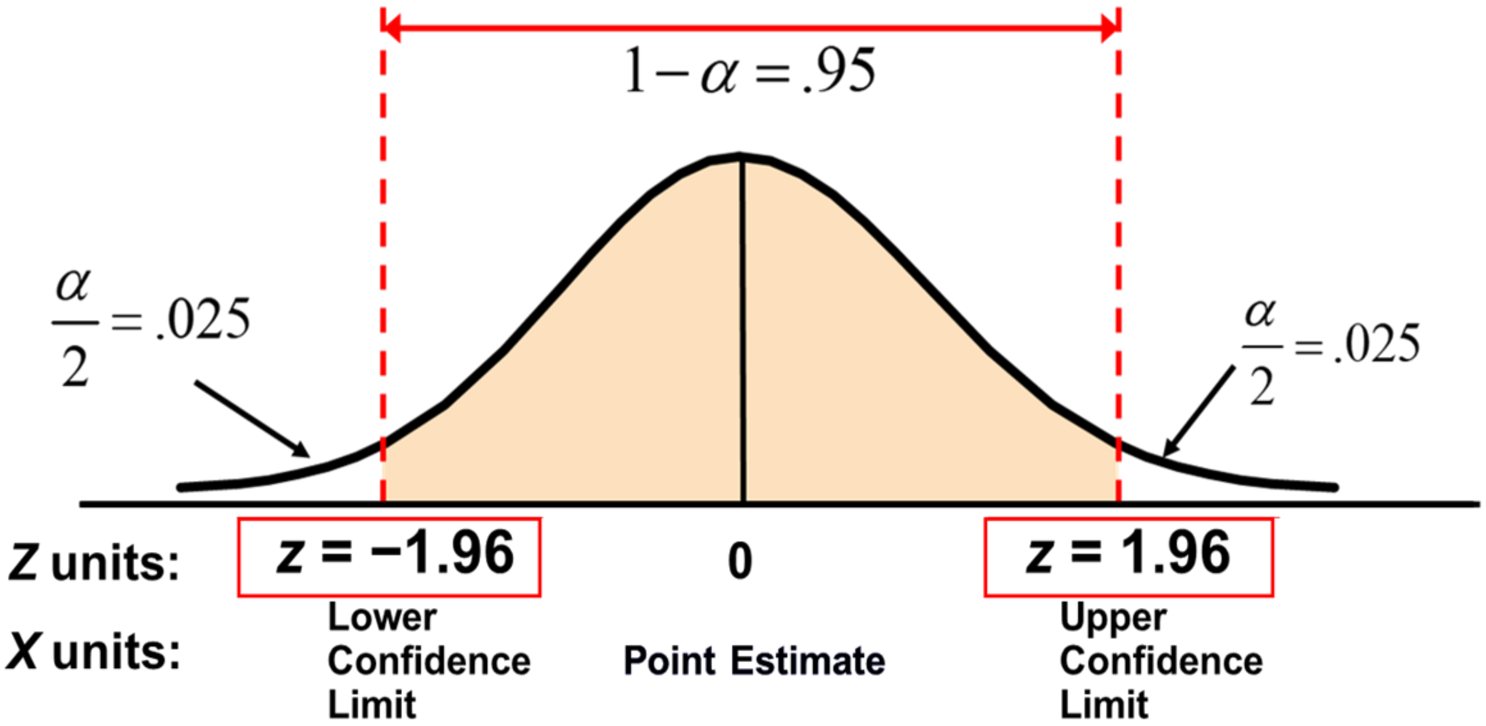
Reducing the Margin of Error

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

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, $(1-\alpha)\downarrow$

- Consider a 95% confidence interval:



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

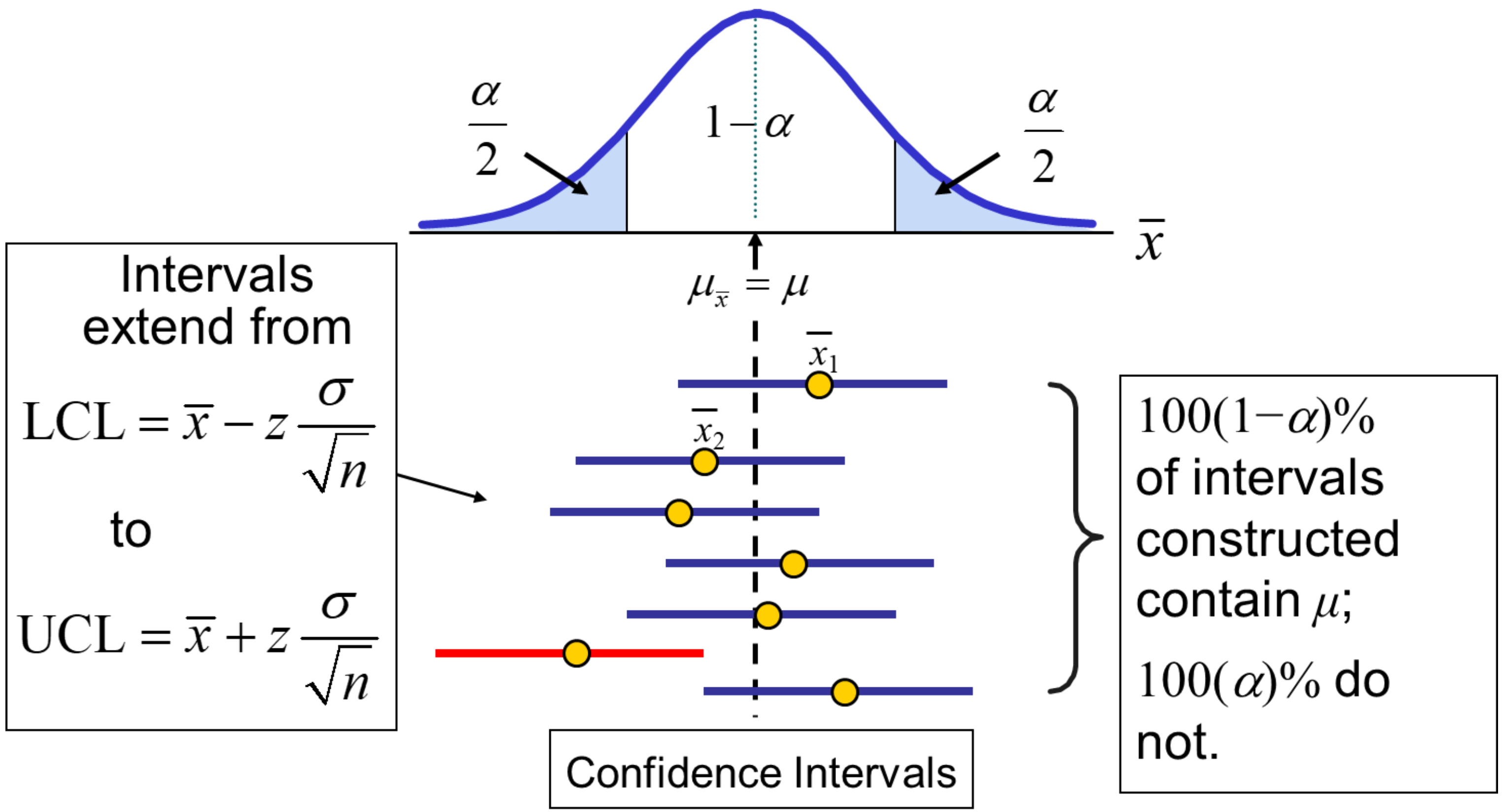
Common Levels of Confidence

- 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

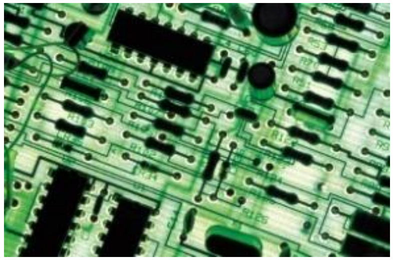
Intervals and Level of Confidence

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)

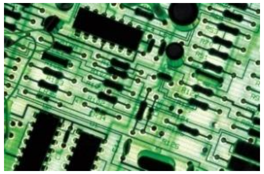
- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is .35 ohms.
- Solution:

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

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

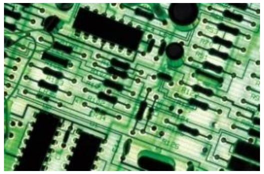
$$= 2.20 \pm .2068$$

$$1.9932 < \mu < 2.4068$$

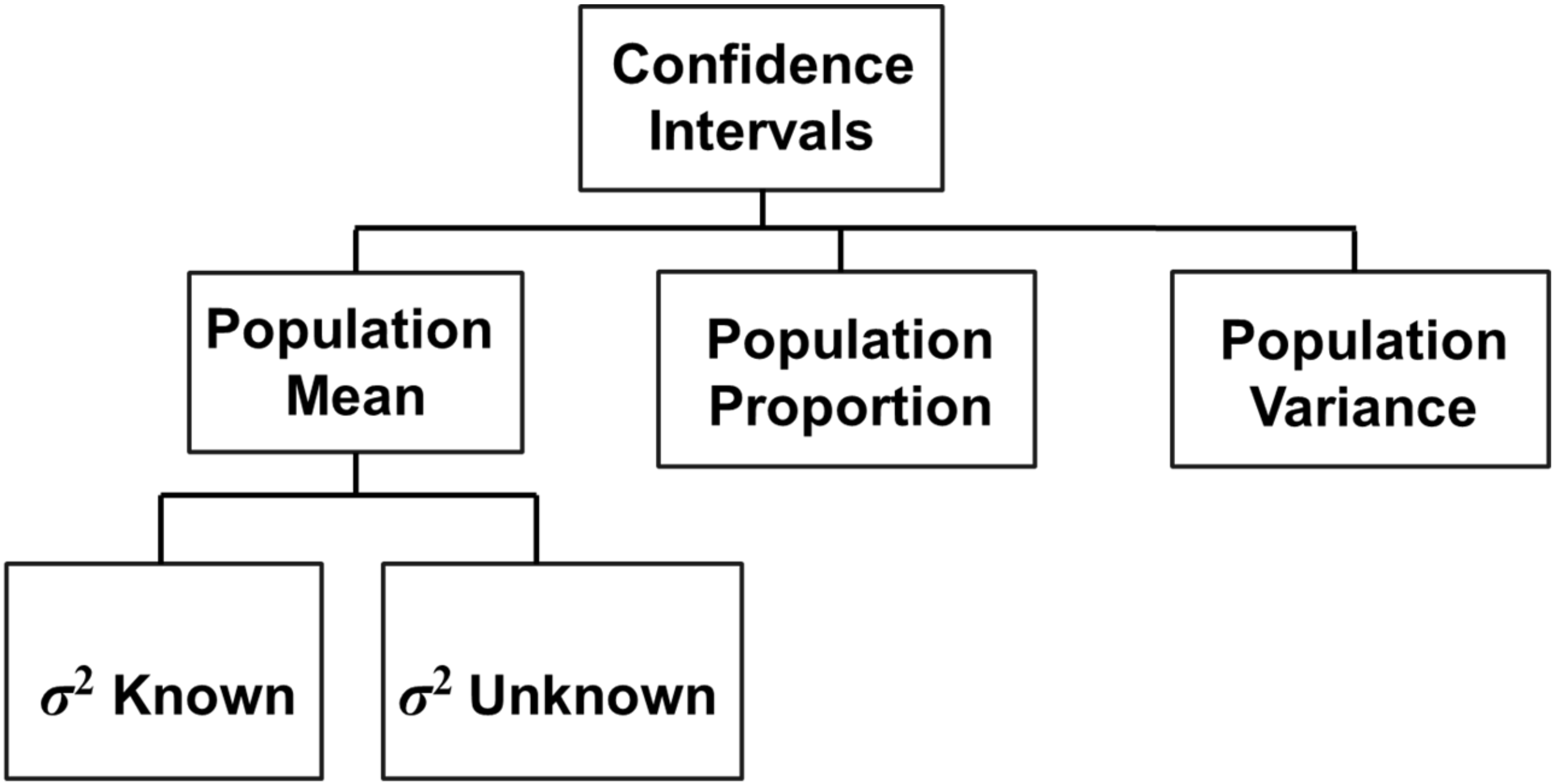


Interpretation (1 of 2)

- 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)



(From normally distributed populations)

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

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

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

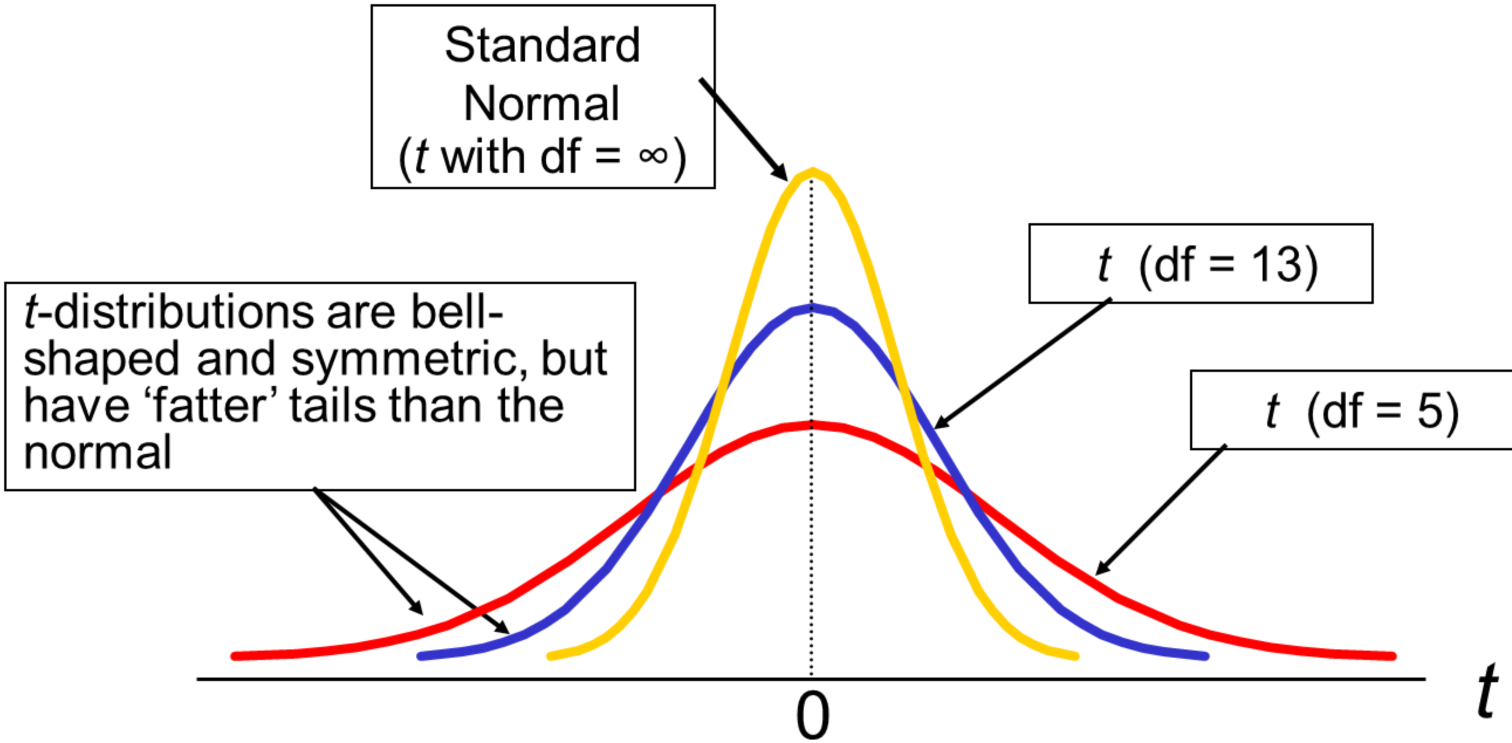
Student's *t* Distribution (2 of 3)

- 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

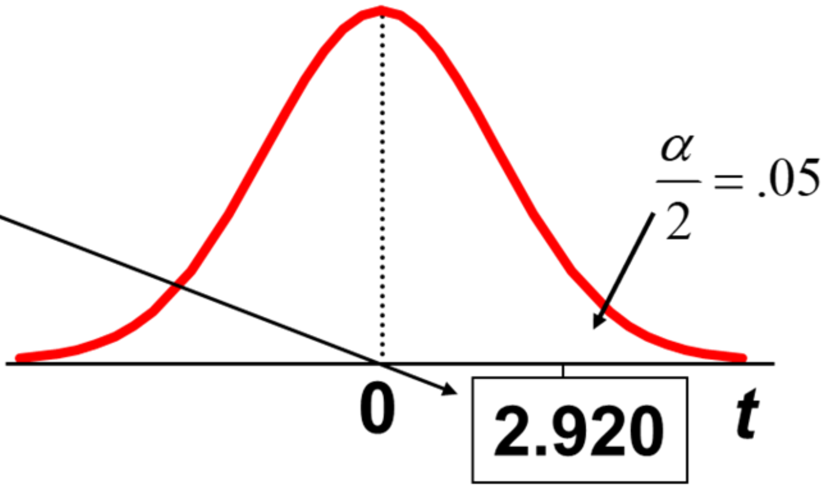


Student's *t* Table

	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 $t_{n-1,\frac{\alpha}{2}} = t_{24,.025} = 2.0639$

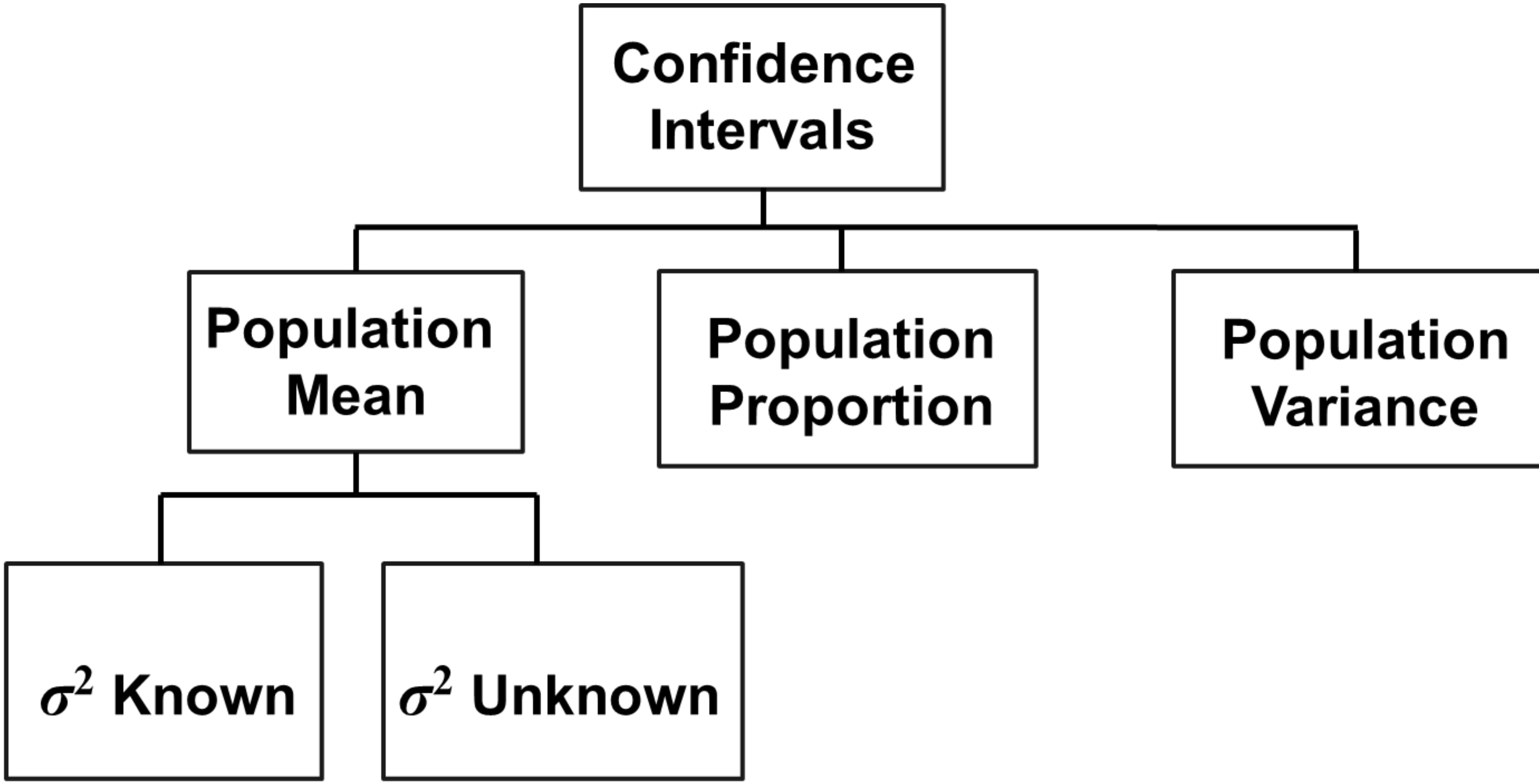
The confidence interval is

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

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

$$46.698 < \mu < 53.302$$

Section 7.4 Confidence Interval Estimation for Population Proportion (1 of 2)



Section 7.4 Confidence Interval Estimation for Population Proportion (2 of 2)

- An interval estimate for the population proportion (P) can be calculated by adding an allowance for uncertainty to the sample proportion $(\hat{p})$

Confidence Intervals for the Population Proportion

- Recall that the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

$$\sigma_P = \sqrt{\frac{P(1-P)}{n}}$$

- We will estimate this with sample data:

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

Confidence Interval Endpoints

- The confidence interval for the population proportion is given by

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

- where
  - $z_{\frac{\alpha}{2}}$ is the standard normal value for the level of confidence desired
  - $\hat{p}$ is the sample proportion
  - *n* is the sample size
  - $nP(1-P) > 5$

Example 3 (1 of 2)

- A random sample of 100 people shows that 25 are left-handed.
- Form a 95% confidence interval for the true proportion of left-handers

Example 3 (2 of 2)

- A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

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

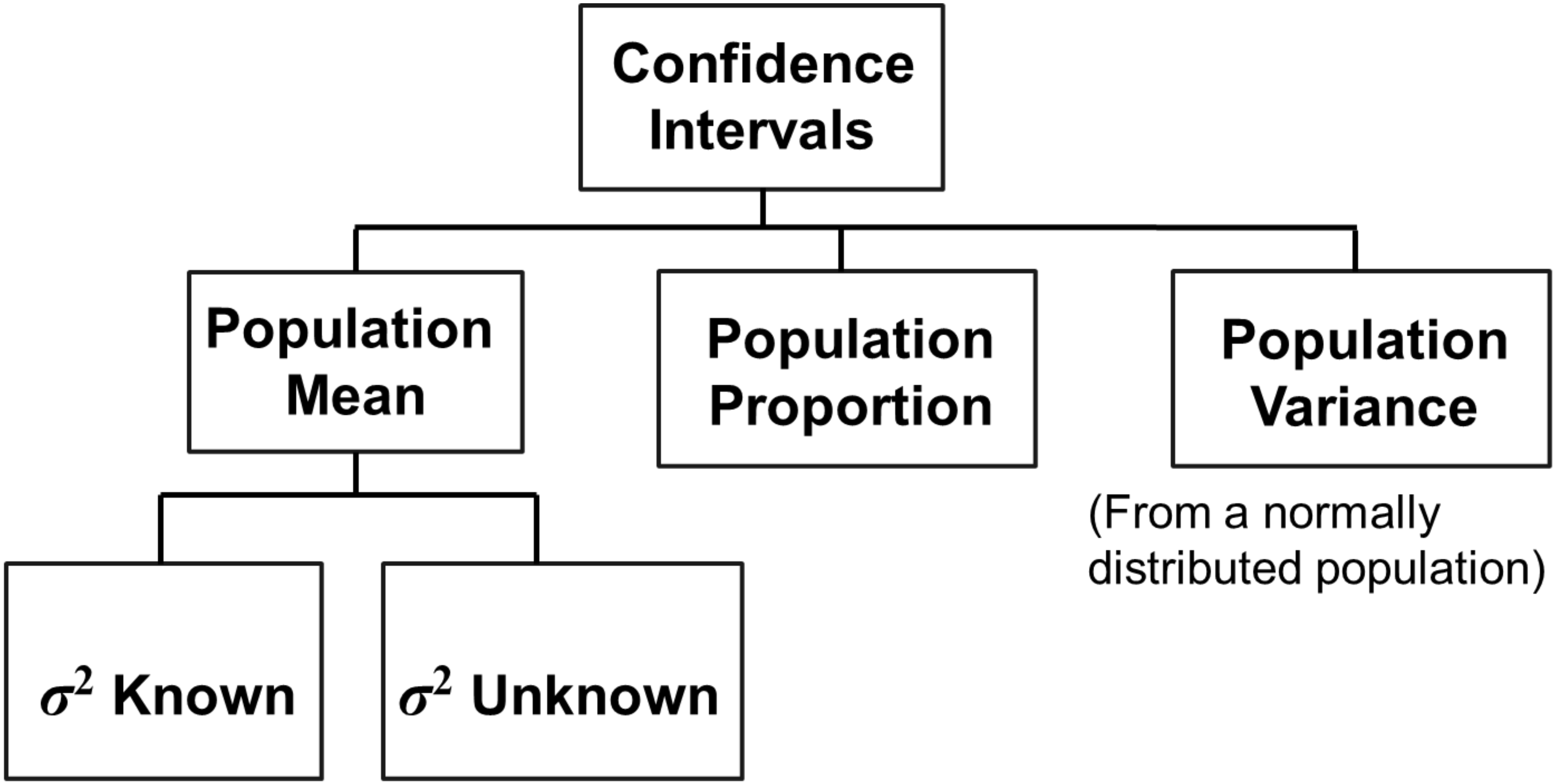
$$\frac{25}{100} \pm 1.96 \sqrt{\frac{.25(.75)}{100}}$$

$$0.1651 < P < 0.3349$$

Interpretation (2 of 2)

- We are 95% confident that the true proportion of left-handers in the population is between 16.51% and 33.49%.
- Although the interval from 0.1651 to 0.3349 may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

Section 7.5 Confidence Interval Estimation for the Variance



Confidence Intervals for the Population Variance (1 of 3)

- Goal: Form a confidence interval for the population variance, σ^2
  - The confidence interval is based on the sample variance, s^2
  - Assumed: the population is normally distributed

Confidence Intervals for the Population Variance (2 of 3)

The random variable

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

follows a chi-square distribution with (n − 1) degrees of freedom

Where the chi-square value $\chi^2_{n-1,\alpha}$ denotes the number for which

$$P\left(\chi^2_{n-1} > \chi^2_{n-1,\alpha}\right) = \alpha$$

Confidence Intervals for the Population Variance (3 of 3)

The $100(1-\alpha)\%$ confidence interval for the population variance is given by

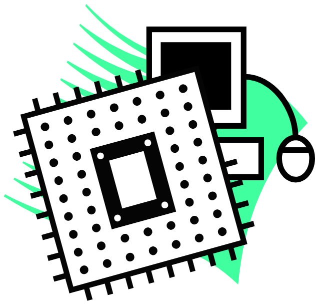
$$\text{LCL} = \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}$$

$$\text{UCL} = \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}$$

Example 4

You are testing the speed of a batch of computer processors. You collect the following data (in Mhz):

Sample size	**17**
Sample mean	**3004**
Sample std dev	**74**



Assume the population is normal.

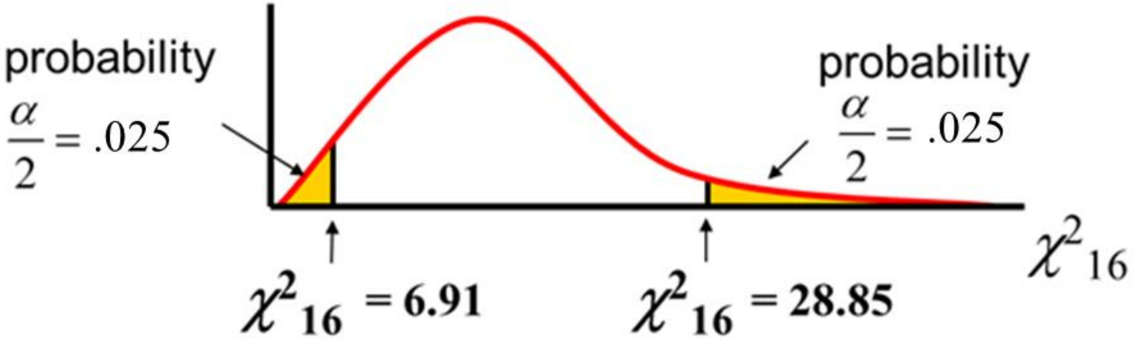
Determine the 95% confidence interval for σ_x^2

Finding the Chi-Square Values

- n = 17 so the chi-square distribution has (n − 1) = 16 degrees of freedom
- $\alpha = 0.05,$ so use the the chi-square values with area 0.025 in each tail:

$$\chi^2_{n-1,\frac{\alpha}{2}} = \chi^2_{16,0.025} = 28.85$$

$$\chi^2_{n-1,1-\frac{\alpha}{2}} = \chi^2_{16,0.975} = 6.91$$



Calculating the Confidence Limits

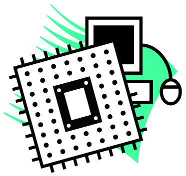
- The 95% confidence interval is

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

$$\frac{(17-1)(74)^2}{28.85} < \sigma^2 < \frac{(17-1)(74)^2}{6.91}$$

$$3037 < \sigma^2 < 12680$$

Converting to standard deviation, we are 95% confident that the population standard deviation of CPU speed is between 55.1 and 112.6 Mhz



Section 7.6 Confidence Interval Estimation: Finite Populations

- If the sample size is more than 5% of the population size (and sampling is without replacement) then a finite population correction factor must be used when calculating the standard error

Finite Population Correction Factor

- Suppose sampling is without replacement and the sample size is large relative to the population size
- Assume the population size is large enough to apply the central limit theorem
- Apply the finite population correction factor when estimating the population variance

$$\text{finite population correction factor} = \frac{N-n}{N-1}$$

Estimating the Population Mean

- Let a simple random sample of size *n* be taken from a population of *N* members with mean μ
- The sample mean is an unbiased estimator of the population mean μ
- The point estimate is:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Finite Populations: Mean

- If the sample size is more than 5% of the population size, an unbiased estimator for the variance of the sample mean is

$$\hat{\sigma}_{\bar{x}}^2 = \frac{s^2}{n}\left(\frac{N-n}{N-1}\right)$$

- So the $100(1-\alpha)\%$ confidence interval for the population mean is

$$\bar{x} \pm t_{n-1,\frac{\alpha}{2}}\hat{\sigma}_{\bar{x}}$$

Estimating the Population Total (1 of 2)

- Consider a simple random sample of size *n* from a population of size *N*
- The quantity to be estimated is the population total $N\mu$
- An unbiased estimation procedure for the population total $N\mu$ yields the point estimate $N\bar{x}$

Estimating the Population Total (2 of 2)

- An unbiased estimator of the variance of the population total is

$$N^2 \hat{\sigma}_{\bar{x}}^2 = N^2 \frac{s^2}{n} \left(\frac{N-n}{N-1} \right)$$

- A $100(1-\alpha)\%$ confidence interval for the population total is

$$N\bar{x} \pm t_{n-1,\frac{\alpha}{2}} N \hat{\sigma}_{\bar{x}}$$

Confidence Interval for Population Total: Example

A firm has a population of 1000 accounts and wishes to estimate the value of the total population balance

A sample of 80 accounts is selected with average balance of $87.60 and standard deviation of $22.30

Find the 95% confidence interval estimate of the total balance

Example Solution

$$N = 1000, \quad n = 80, \quad \bar{x} = 87.6, \quad s = 22.3$$

$$N^2 \hat{\sigma}_{\bar{x}}^2 = N^2 \frac{s^2}{n} \frac{(N-n)}{N-1} = (1000)^2 \frac{(22.3)^2}{80} \frac{920}{999} = 5724559.6$$

$$N \hat{\sigma}_{\bar{x}} = \sqrt{5724559.6} = 2392.6$$

$$N\bar{x} \pm t_{79,0.025} N\hat{\sigma}_{\bar{x}} = (1000)(87.6) \pm (1.9905)(2392.6)$$

$$82837.53 < N\mu < 92362.47$$

The 95% confidence interval for the population total balance is $82,837.53 to $92,362.47

Estimating the Population Proportion: Finite Population

- Let the true population proportion be *P*
- Let $\hat{p}$ be the sample proportion from *n* observations from a simple random sample
- The sample proportion, $\hat{p}$, is an unbiased estimator of the population proportion, *P*

Confidence Intervals for Population Proportion: Finite Population

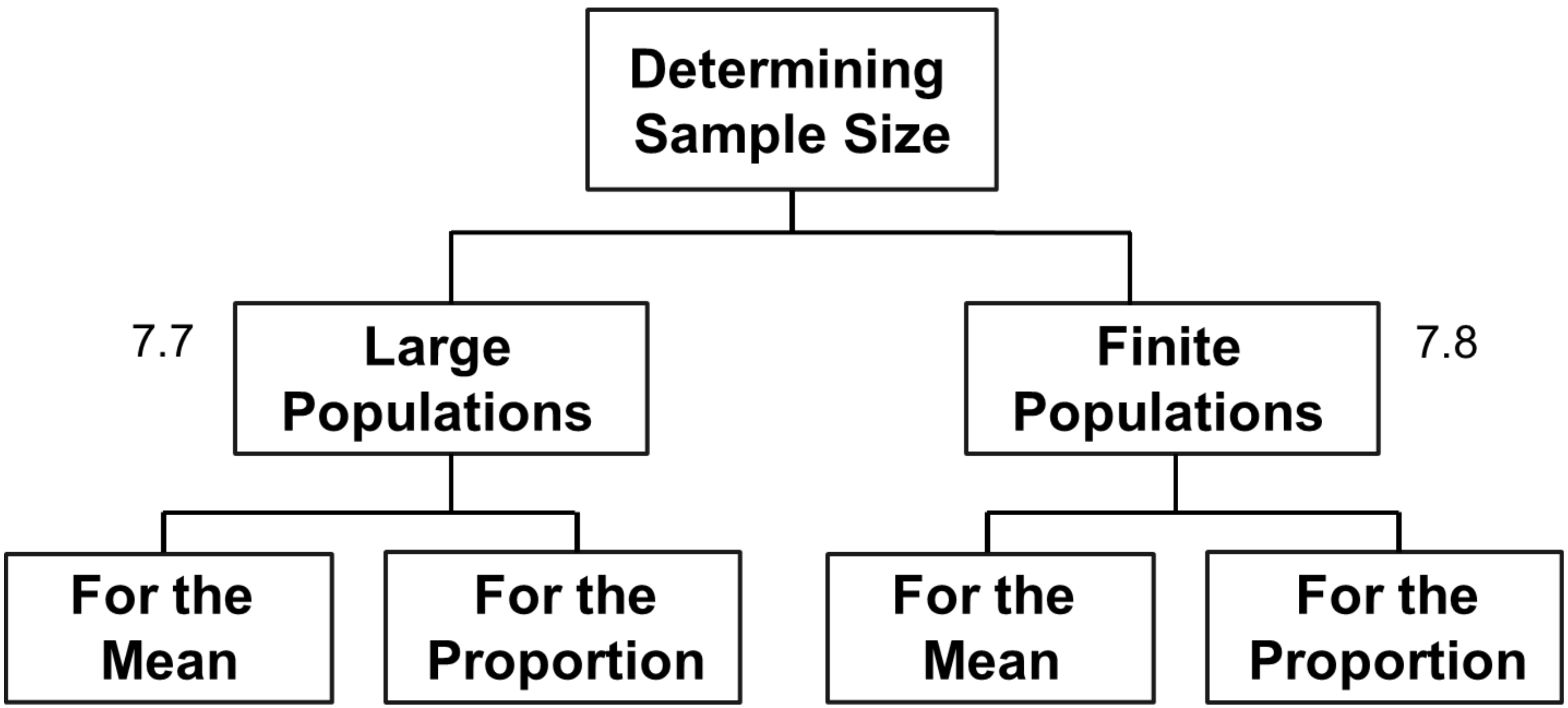
- If the sample size is more than 5% of the population size, an unbiased estimator for the variance of the population proportion is

$$\hat{\sigma}_{\hat{p}}^{2} = \frac{\hat{p}(1-\hat{p})}{n}\left(\frac{N-n}{N-1}\right)$$

- So the $100(1-\alpha)\%$ confidence interval for the population proportion is

$$\hat{p} \pm z_{\frac{\alpha}{2}}\hat{\sigma}_{\hat{p}}$$

Sample-Size Determination (1 of 2)



Section 7.7 Sample-Size Determination: Large Populations (1 of 2)

Large Populations

For the Mean

(Known population variance)

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

Margin of Error (sampling error)

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

Section 7.7 Sample-Size Determination: Large Populations (2 of 2)

Large Populations

For the Mean

(Known population variance)

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

⇨ Now solve for *n* to get ⇨

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

Sample-Size Determination (2 of 2)

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence $(1-\alpha)$, which determines the $z_{\frac{\alpha}{2}}$ value
  - The acceptable margin of error (sampling error), *ME*
  - The population standard deviation, σ

Required Sample Size Example

If $\sigma = 45$, what sample size is needed to estimate the mean within ± 5 with 90% confidence?

$$n = \frac{z_{\frac{\alpha}{2}}^2 \sigma^2}{ME^2} = \frac{(1.645)^2 (45)^2}{5^2} = 219.19$$

So the required sample size is **n = 220**

(Always round up)

Sample Size Determination: Population Proportion (1 of 3)

Large Populations

For the Proportion

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

$$ME = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Margin of Error (sampling error)

Sample Size Determination: Population Proportion (2 of 3)

Large Populations

For the Proportion

$$ME = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}(1-\hat{p})$ cannot be larger than 0.25, when $\hat{p} = 0.5$

⇨

Substitute 0.25 for $\hat{p}(1-\hat{p})$ and solve for *n* to get

⇨

$$n = \frac{0.25 \; z_{\frac{\alpha}{2}}^2}{ME^2}$$

Sample Size Determination: Population Proportion (3 of 3)

- The sample and population proportions, $\hat{p}$ and *P*, are generally not known (since no sample has been taken yet)
- $P(1-P) = 0.25$ generates the largest possible margin of error (so guarantees that the resulting sample size will meet the desired level of confidence)
- To determine the required sample size for the proportion, you must know:
  - The desired level of confidence $(1-\alpha)$, which determines the critical $z_{\frac{\alpha}{2}}$ value
  - The acceptable sampling error (margin of error), *ME*
  - Estimate $P(1-P) = 0.25$

Required Sample Size Example: Population Proportion (1 of 2)

How large a sample would be necessary to estimate the true proportion defective in a large population within $\pm 3\%$, with 95% confidence?

Required Sample Size Example: Population Proportion (2 of 2)

Solution:

For 95% confidence, use $z_{0.025} = 1.96$

ME = 0.03

Estimate $P(1-P) = 0.25$

$$n = \frac{0.25 \; z_{\frac{\alpha}{2}}^{2}}{ME^{2}} = \frac{(0.25)(1.96)^{2}}{(0.03)^{2}} = 1067.11$$

So use *n* = 1068

Chapter Summary (1 of 2)

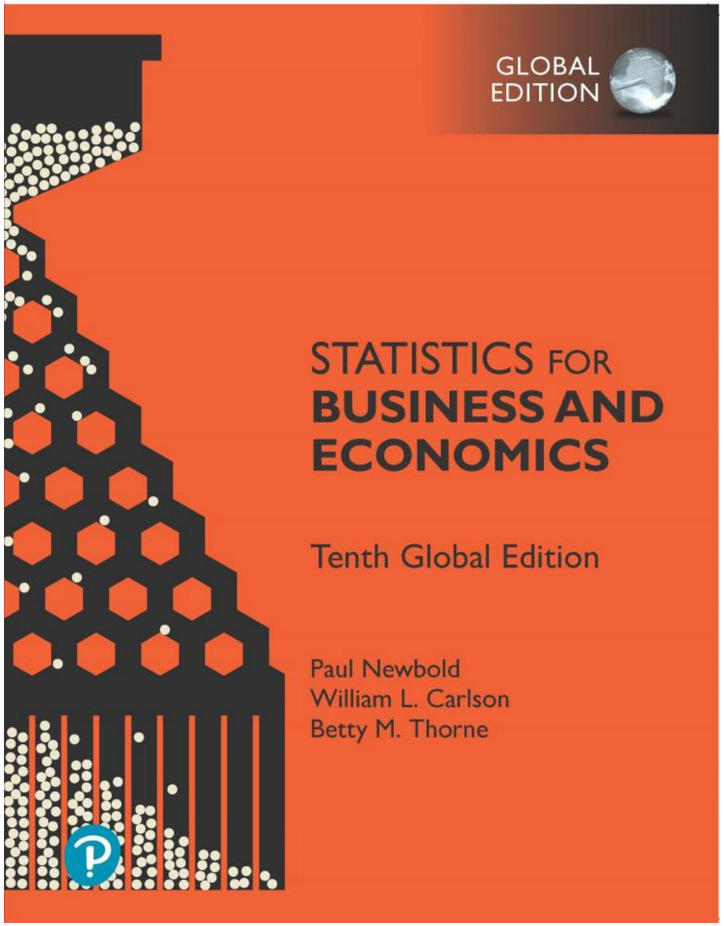
- Introduced the concept of confidence intervals
- Discussed point estimates
- Developed confidence interval estimates
- Created confidence interval estimates for the mean (σ^2 known)
- Introduced the Student's *t* distribution
- Determined confidence interval estimates for the mean (σ^2 unknown)

Chapter Summary (2 of 2)

- Created confidence interval estimates for the proportion
- Created confidence interval estimates for the variance of a normal population
- Applied the finite population correction factor to form confidence intervals when the sample size is not small relative to the population size
- Determined required sample size to meet confidence and margin of error requirements

Statistics for Business and Economics

Tenth Edition, Global Edition



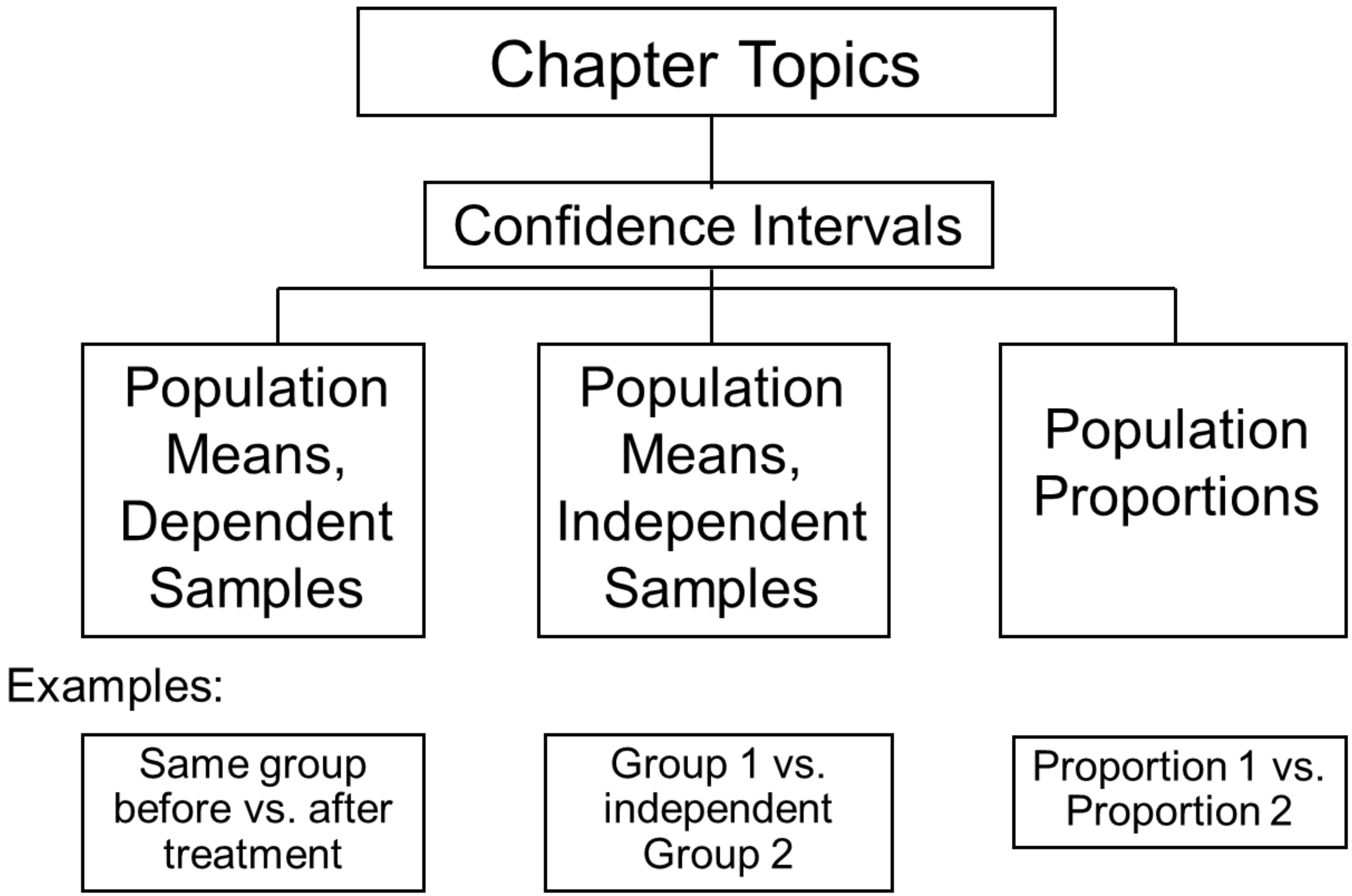
Chapter 8

Estimation: Additional Topics

Chapter Goals

- Form confidence intervals for the difference between two means from dependent samples
- Form confidence intervals for the difference between two independent population means (standard deviations known or unknown)
- Compute confidence interval limits for the difference between two independent population proportions

Estimation: Additional Topics



Pearson

Section 8.1 Dependent Samples

Dependent samples

Confidence Interval Estimation of the Difference Between Two Normal Population Means: Dependent Samples

Tests Means of 2 Related Populations

- Paired or matched samples
- Repeated measures (before/after)
- Use difference between paired values:

$$d_i = x_i - y_i$$

- Eliminates Variation Among Subjects
- Assumptions:
  - Both Populations Are Normally Distributed

Mean Difference

Dependent samples

The i^{th} paired difference is d_i, where

$$d_i = x_i - y_i$$

The point estimate for the population mean paired difference is $\bar{d}$:

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

The sample standard deviation is:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2}{n-1}}$$

n is the number of matched pairs in the sample

Confidence Interval for Mean Difference (1 of 2)

Dependent samples

The confidence interval for the difference between two population means, μ_d, is

$$\bar{d} \pm t_{n-1,\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}}$$

Where

n = the sample size

(number of matched pairs in the paired sample)

Confidence Interval for Mean Difference (2 of 2)

Dependent samples

- The margin of error is

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

- $t_{n-1,\frac{\alpha}{2}}$ is the value from the Student's *t* distribution with $(n-1)$ degrees of freedom for which

$$p\left(t_{n-1} > t_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

Paired Samples Example (1 of 2)

Dependent samples

- Six people sign up for a weight loss program. You collect the following data:

	Weight:		
Person	Before (x)	After (y)	Difference, d_i
1	136	125	11
2	205	195	10
3	157	150	7
4	138	140	− 2
5	175	165	10
6	166	160	6
			42

$$\bar{d} = \frac{\sum d_i}{n} = 7.0$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = 4.82$$

Paired Samples Example (2 of 2)

Dependent samples

- For a 95% confidence level, the appropriate *t* value is

$$t_{n-1,\frac{\alpha}{2}} = t_{5,.025} = 2.571$$

- The 95% confidence interval for the difference between means, μ_d, is

$$\bar{d} \pm t_{n-1,\frac{\alpha}{2}} \frac{S_d}{\sqrt{n}}$$

$$7 \pm (2.571)\frac{4.82}{\sqrt{6}}$$

$$-1.94 < \mu_d < 12.06$$

Since this interval contains zero, we cannot be 95% confident, given this limited data, that the weight loss program helps people lose weight

Section 8.2 Difference Between Two Means: Independent Samples

Population means, independent samples

Confidence Interval Estimation of the Difference Between Two Normal Population Means: Independent Samples

Goal: Form a confidence interval for the difference between two population means, $\mu_x - \mu_y$

Difference Between Two Means: Independent Samples (1 of 2)

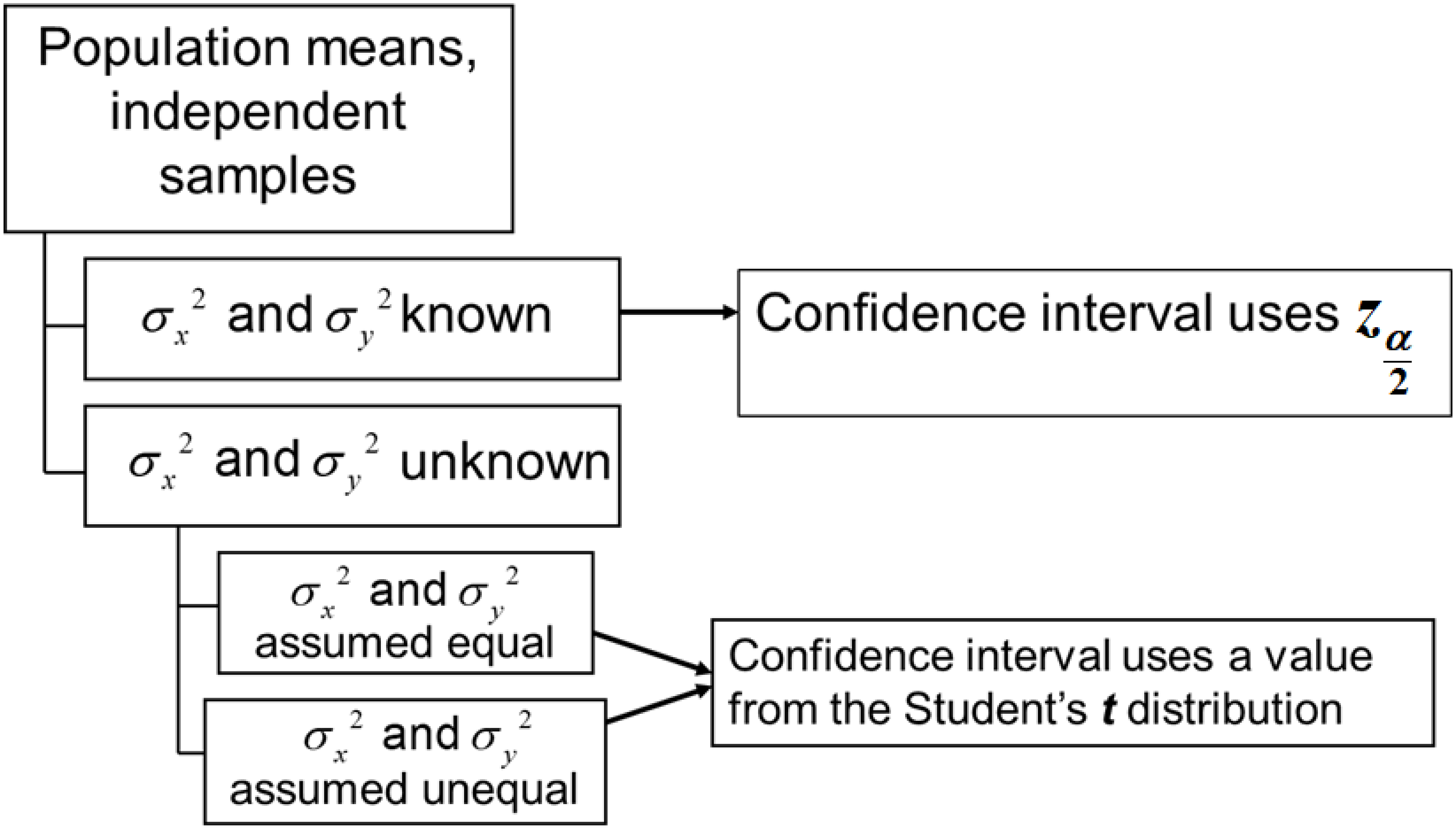
Population means, independent samples

Goal: Form a confidence interval for the difference between two population means, $\mu_x - \mu_y$

- Different data sources
  - Unrelated
  - Independent
    - Sample selected from one population has no effect on the sample selected from the other population
- The point estimate is the difference between the two sample means:

$$\bar{x} - \bar{y}$$

Difference Between Two Means: Independent Samples (2 of 2)



Sigmas Known (1 of 2)

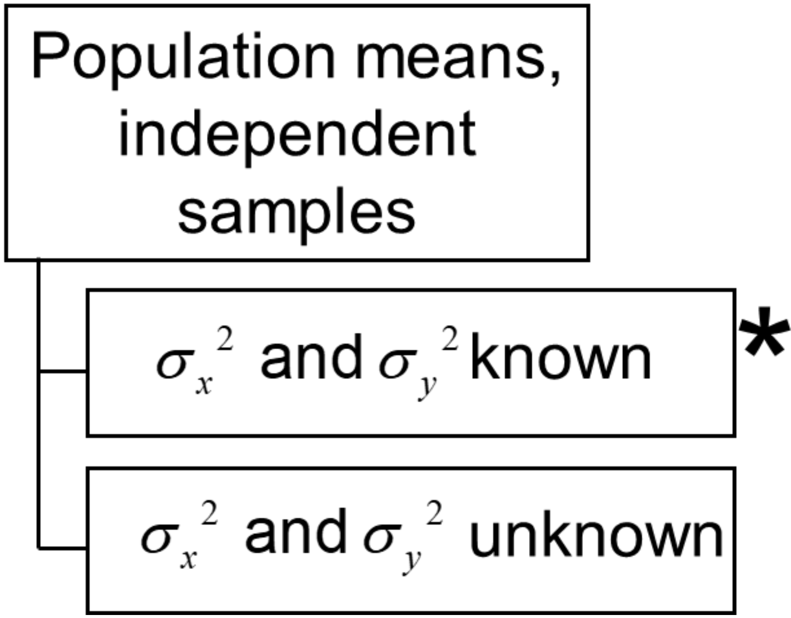
Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

Assumptions:

- Samples are randomly and independently drawn
- both population distributions are normal
- Population variances are known

Sigmas Known (2 of 2)



When σ_x and σ_y are known and both populations are normal, the variance of $\bar{X} - \bar{Y}$ is

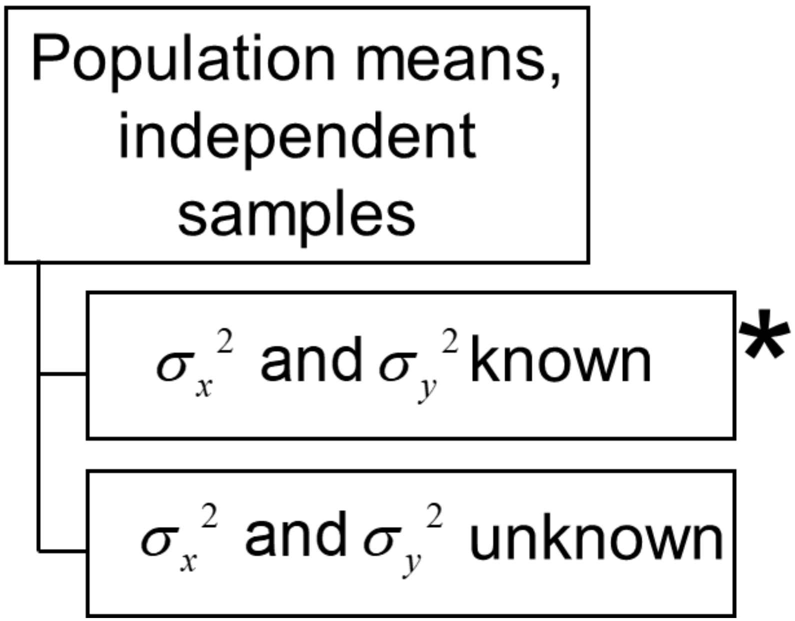
$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{\sigma_x^{\;2}}{n_x} + \frac{\sigma_y^{\;2}}{n_y}$$

…and the random variable

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_x^2}{n_X} + \dfrac{\sigma_y^2}{n_Y}}}$$

has a standard normal distribution

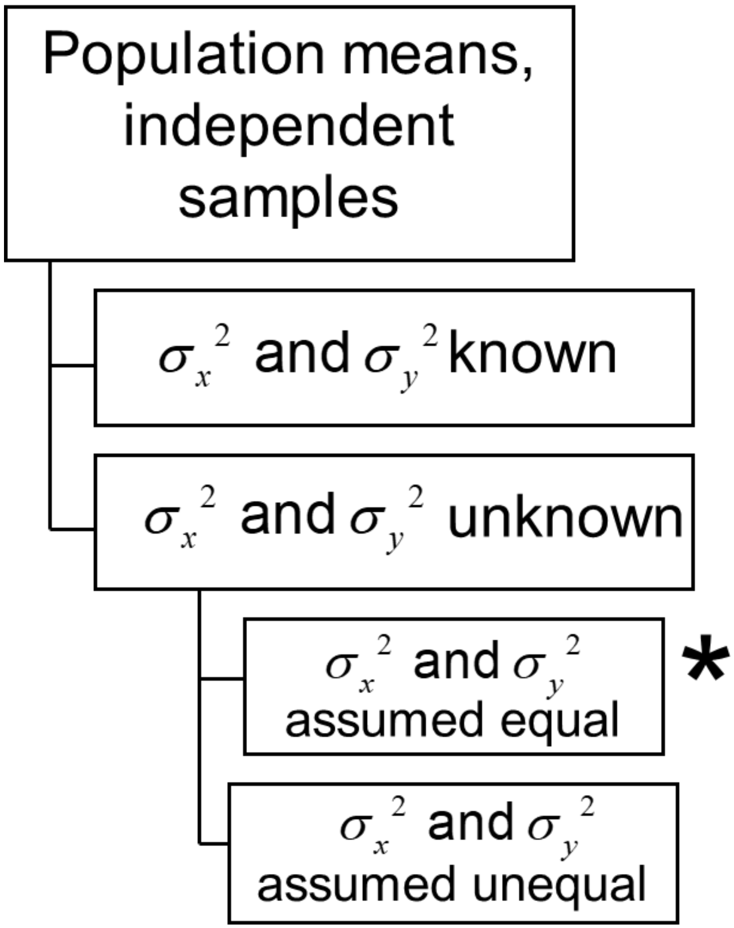
Confidence Interval, Sigmas Known



The confidence interval for $\mu_x - \mu_y$ is:

$$\left(\bar{x} - \bar{y}\right) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

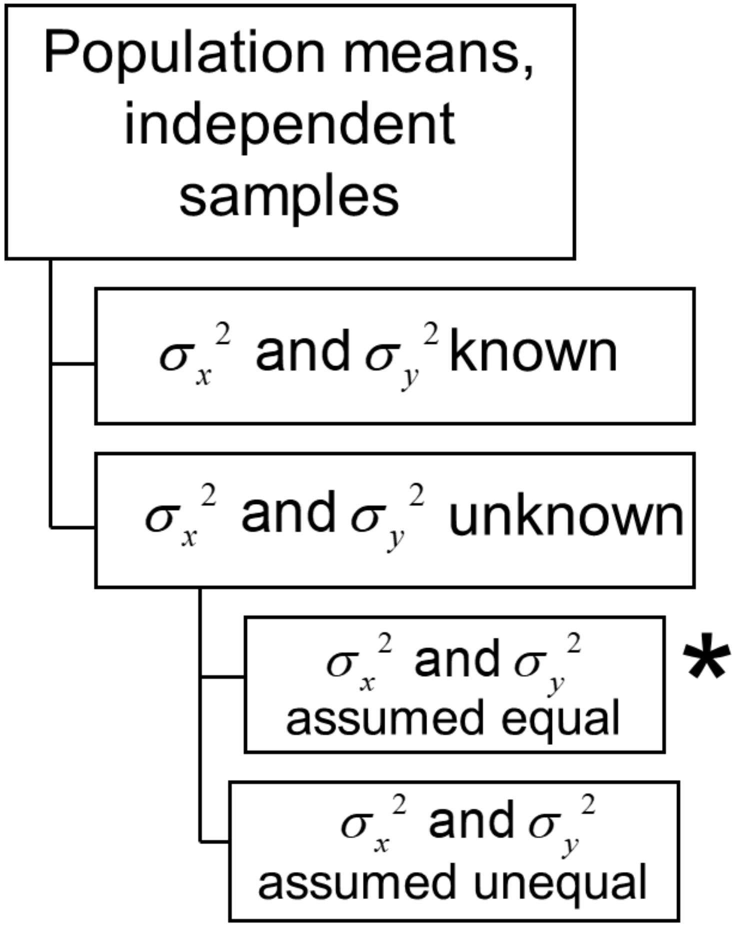
Sigmas Unknown, Assumed Equal (1 of 3)



Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown but assumed equal

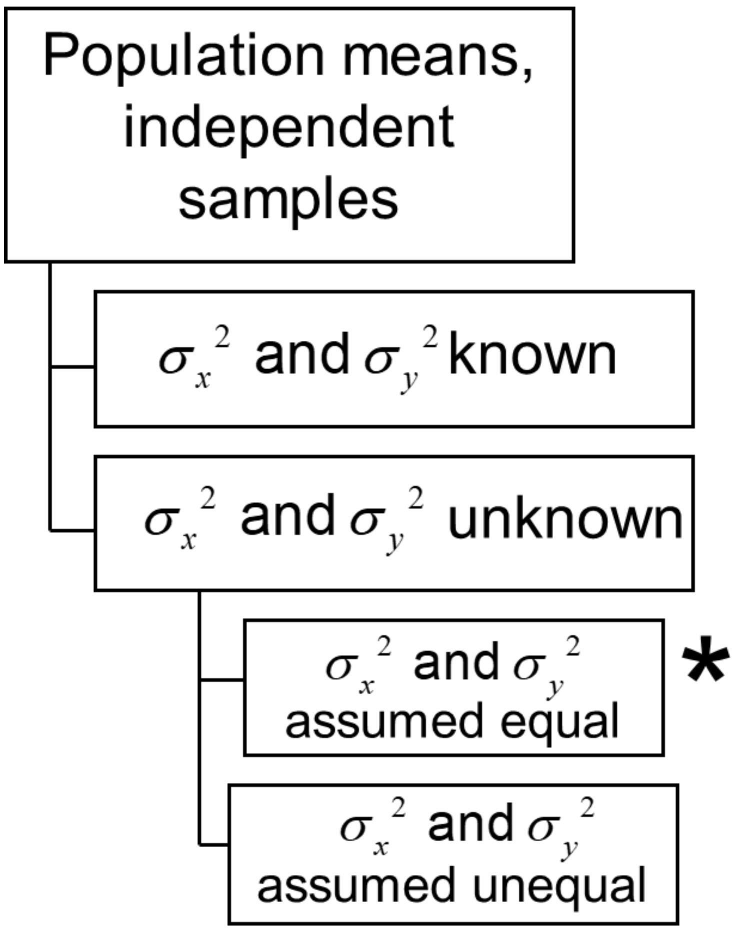
Sigmas Unknown, Assumed Equal (2 of 3)



Forming interval estimates:

- The population variances are assumed equal, so use the two sample standard deviations and pool them to estimate σ
- use a *t* value with $(n_x + n_y - 2)$ degrees of freedom

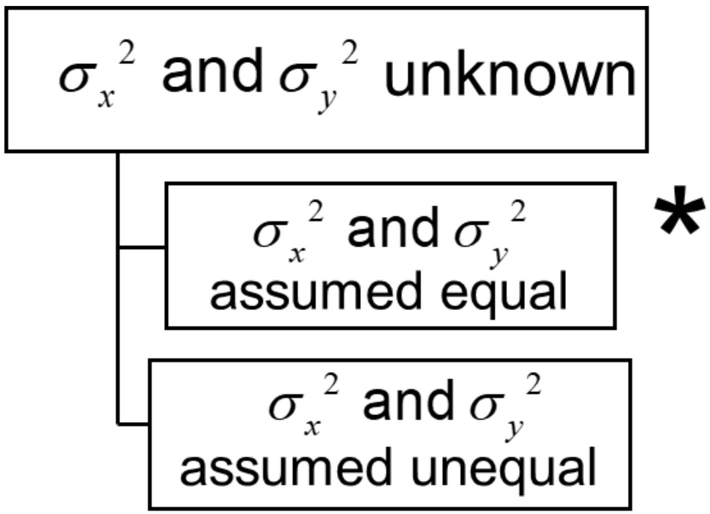
Sigmas Unknown, Assumed Equal (3 of 3)



The pooled variance is

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

Confidence Interval, Sigmas Unknown, Equal



The confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{x} - \bar{y}) \pm t_{n_x + n_y - 2, \frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}$$

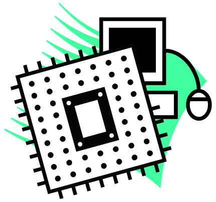
Where $s_p^2 = \dfrac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$

Pooled Variance Example

You are testing two computer processors for speed. Form a confidence interval for the difference in CPU speed. You collect the following speed data (in Mhz):

	CPU_x	CPU_y
Number Tested	**17**	**14**
Sample mean	**3004**	**2538**
Sample std dev	**74**	**56**

Assume both populations are normal with equal variances, and use 95% confidence



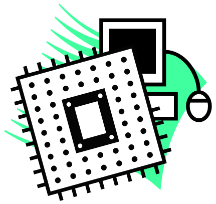
Calculating the Pooled Variance

The pooled variance is:

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{(n_x - 1) + (n_y - 1)} = \frac{(17-1)74^2 + (14-1)56^2}{(17-1)+(14-1)} = 4427.03$$

The *t* value for a 95% confidence interval is:

$$t_{n_x + n_y - 2, \frac{\alpha}{2}} = t_{29, 0.025} = 2.045$$



Calculating the Confidence Limits

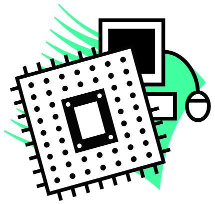
- The 95% confidence interval is

$$(\bar{x} - \bar{y}) \pm t_{n_x + n_y - 2, \frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}$$

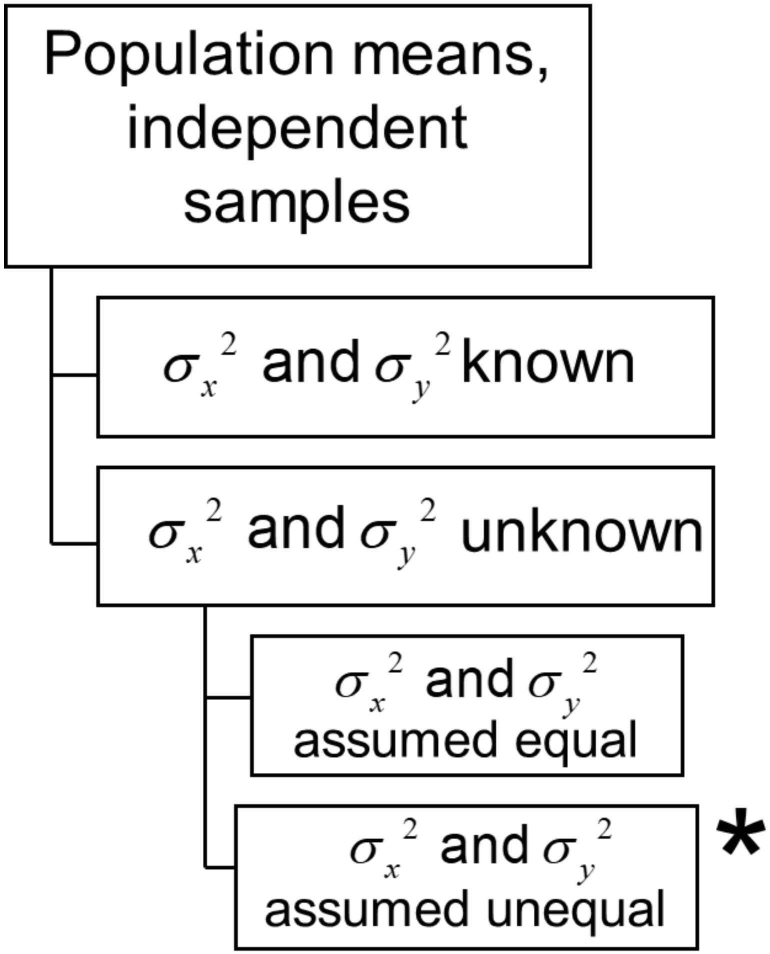
$$(3004 - 2538) \pm (2.054)\sqrt{\frac{4427.03}{17} + \frac{4427.03}{14}}$$

$$416.69 < \mu_X - \mu_Y < 515.31$$

We are 95% confident that the mean difference in CPU speed is between 416.69 and 515.31 Mhz.



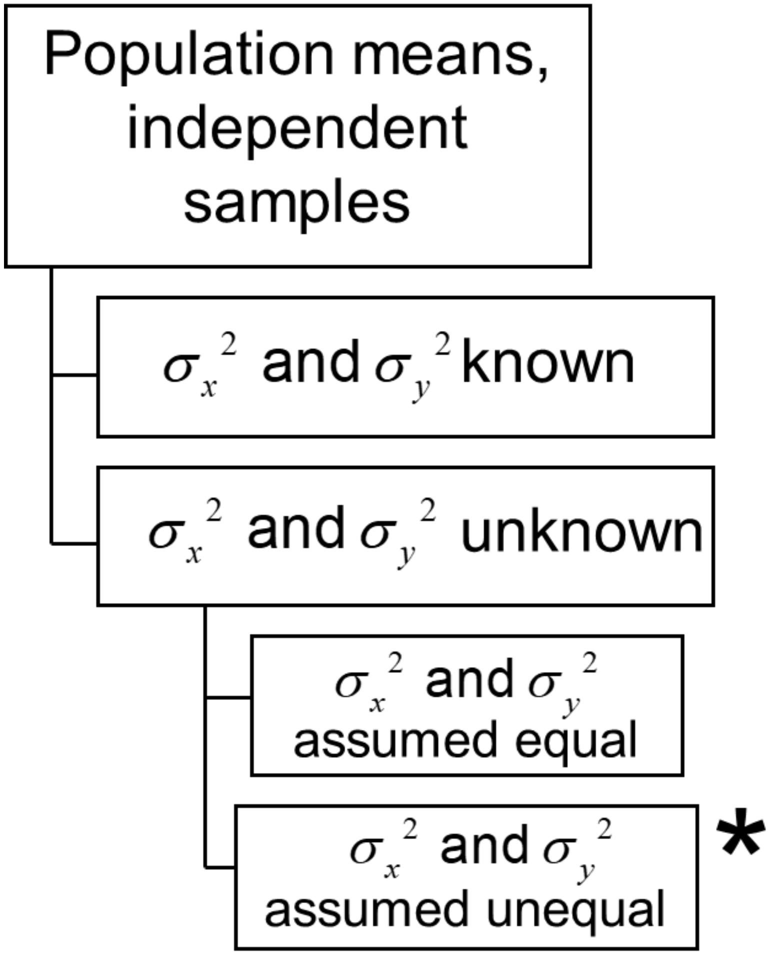
Sigmas Unknown, Assumed Unequal (1 of 2)



Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown and assumed unequal

Sigmas Unknown, Assumed Unequal (2 of 2)

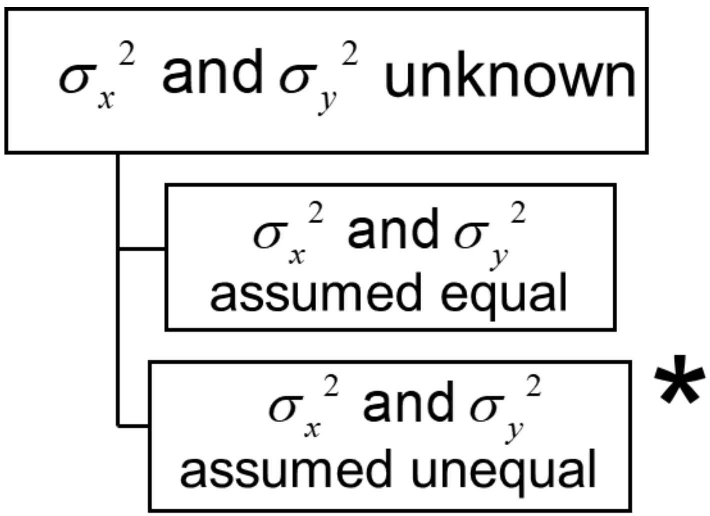


Forming interval estimates:

- The population variances are assumed unequal, so a pooled variance is not appropriate
- use a *t* value with ν degrees of freedom, where

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{(n_x - 1)} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{(n_y - 1)}}$$

Confidence Interval, Sigmas Unknown, Assumed Unequal



The confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{x} - \bar{y}) \pm t_{\nu, \frac{\alpha}{2}} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

Where

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{(n_x - 1)} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{(n_y - 1)}}$$

Section 8.3 Two Population Proportions (1 of 2)

Population proportions

Confidence Interval Estimation of the Difference Between Two Population Proportions (Large Samples)

Goal: Form a confidence interval for the difference between two population proportions, $P_x - P_y$

Section 8.3 Two Population Proportions (2 of 2)

Population proportions

Goal: Form a confidence interval for the difference between two population proportions, $P_x - P_y$

Assumptions:

Both sample sizes are large (generally at least 40 observations in each sample)

The point estimate for the difference is $\hat{p}_x - \hat{p}_y$

Two Population Proportions

Population proportions

- The random variable

$$Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}}$$

is approximately normally distributed

Confidence Interval for Two Population Proportions

Population proportions

The confidence limits for

$P_x - P_y$ are :

$$\left(\hat{p}_x - \hat{p}_y\right) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_x\left(1-\hat{p}_x\right)}{n_x} + \frac{\hat{p}_y\left(1-\hat{p}_y\right)}{n_y}}$$

Example: Two Population Proportions (1 of 3)

Form a 90% confidence interval for the difference between the proportion of men and the proportion of women who have college degrees.

- In a random sample, 26 of 50 men and 28 of 40 women had an earned college degree

Example: Two Population Proportions (2 of 3)

Men: $\hat{p}_x = \frac{26}{50} = 0.52$

Women: $\hat{p}_y = \frac{28}{40} = 0.70$

$$\sqrt{\frac{\hat{p}_x\left(1-\hat{p}_x\right)}{n_x} + \frac{\hat{p}_y\left(1-\hat{p}_y\right)}{n_y}} = \sqrt{\frac{0.52\left(0.48\right)}{50} + \frac{0.70\left(0.30\right)}{40}} = 0.1012$$

For 90% confidence, $Z_{\frac{\alpha}{2}} = 1.645$

Example: Two Population Proportions (3 of 3)

The confidence limits are:

$$\left(\hat{p}_x - \hat{p}_y\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_x\left(1-\hat{p}_x\right)}{n_x} + \frac{\hat{p}_y\left(1-\hat{p}_y\right)}{n_y}}$$

$$= \left(.52 - .70\right) \pm 1.645\left(0.1012\right)$$

so the confidence interval is

$$-0.3465 < P_x - P_y < -0.0135$$

Since this interval does not contain zero we are 90% confident that the two proportions are not equal

Chapter Summary

- Compared two dependent samples (paired samples)
  - Formed confidence intervals for the paired difference
- Compared two independent samples
  - Formed confidence intervals for the difference between two means, population variance known, using z
  - Formed confidence intervals for the differences between two means, population variance unknown, using t
- Formed confidence intervals for the differences between two population proportions