

Estimation



Chapter Goals

After completing this chapter, you should be able to:

- Distinguish between a point estimate and a confidence interval estimate
- Construct and interpret a confidence interval estimate for a single population mean using both the Z and t distributions
- Form and interpret a confidence interval estimate for a single population proportion



Confidence Intervals

Content 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, $\hat{p}$ (large samples)

Troy University Resources

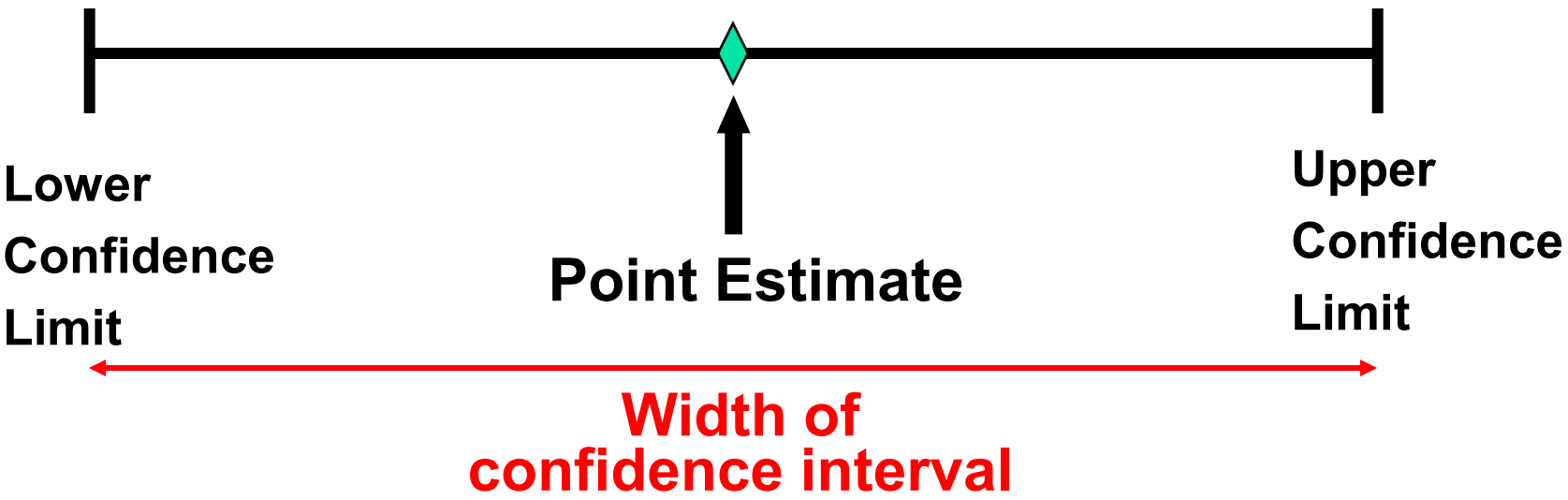


Definitions

- 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	$\hat{p}$



Confidence Intervals

- 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 intervals

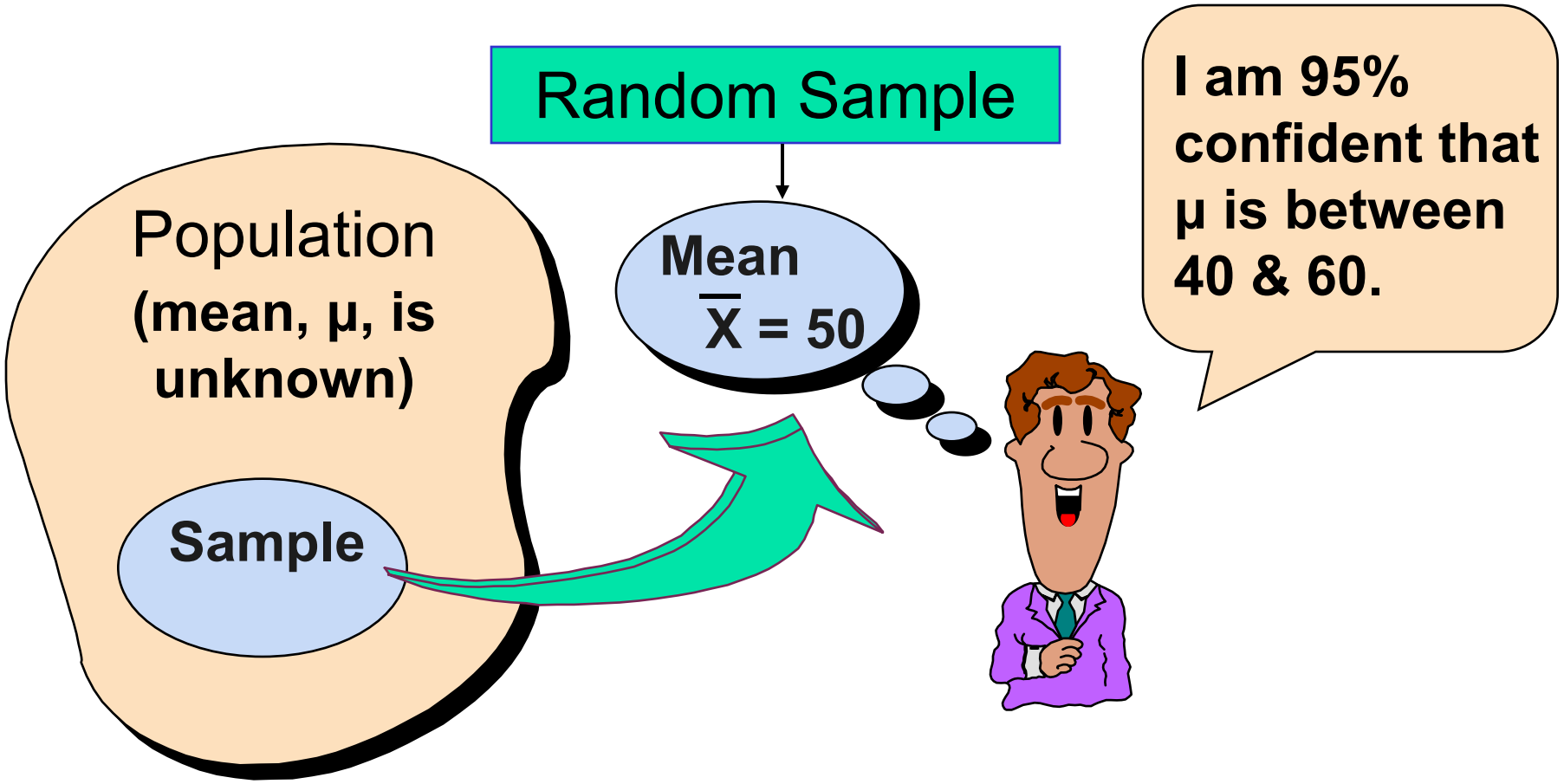
Confidence Interval Estimate

- 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 $(1 - \alpha)$ is called the confidence level of the interval (α 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, (1-α)

(continued)

- Suppose confidence level = 95%
- Also written (1 - α) = 0.95
- A relative frequency interpretation:
  - From repeated samples, 95% of all the confidence intervals that can be constructed 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

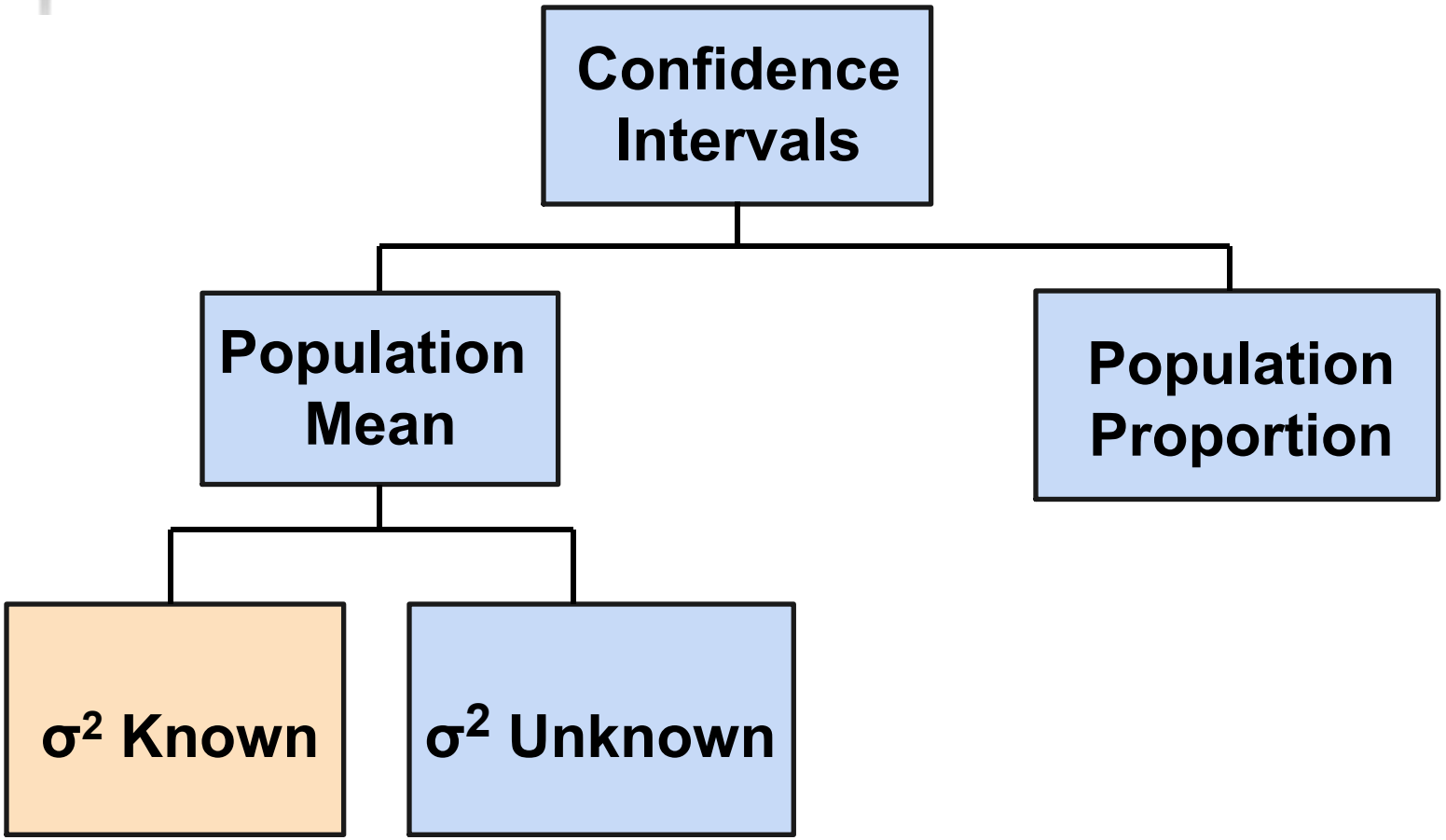
General Formula

- The general formula for all confidence intervals is:

Point Estimate ± (Reliability Factor)(Standard Error)

- The value of the reliability factor depends on the desired level of confidence

Confidence Intervals



Confidence Interval for μ (σ^2 Known)

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

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

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

Margin of Error

- The confidence interval,

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

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

  where ME is called the margin of error

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

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

Reducing the Margin of Error

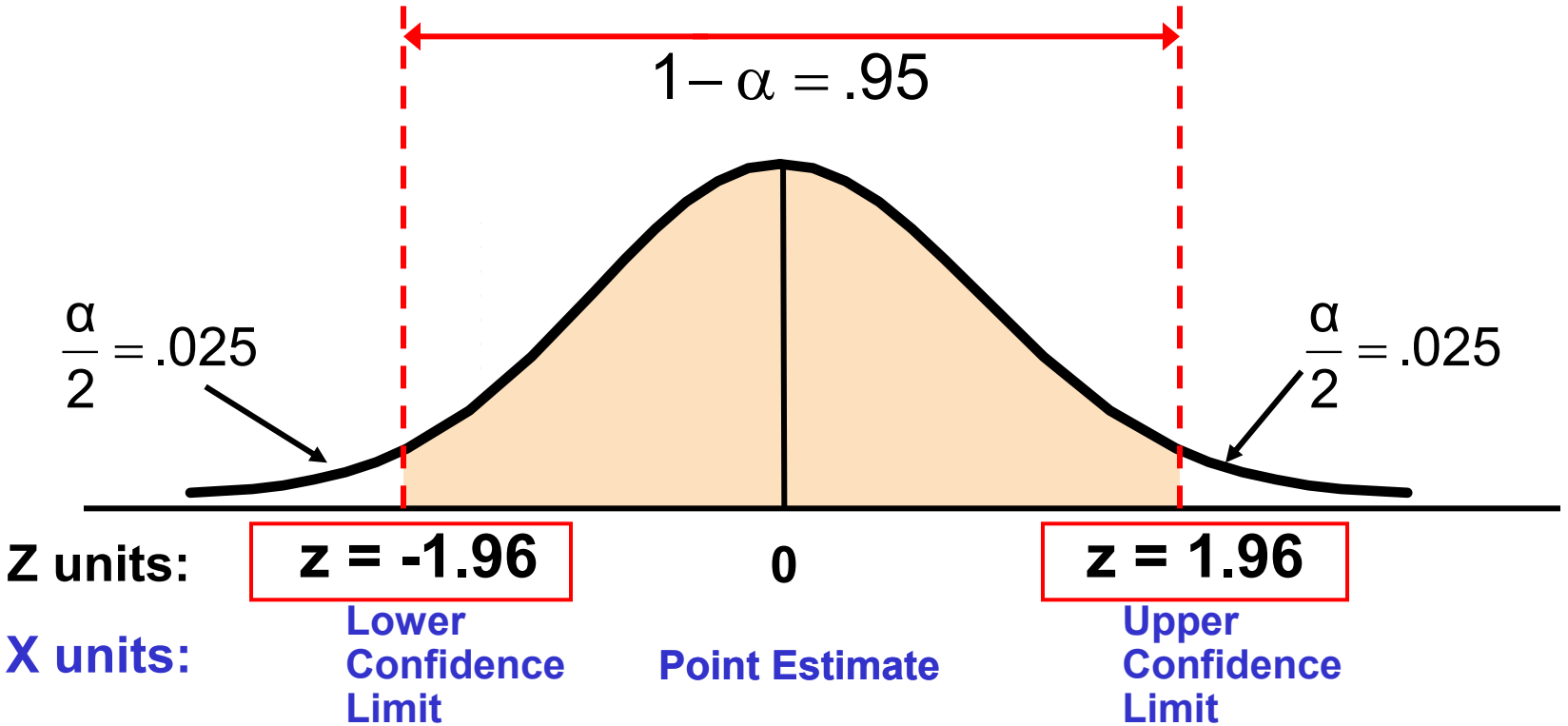
$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

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 the Reliability Factor, $z_{\alpha/2}$

- Consider a 95% confidence interval:



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

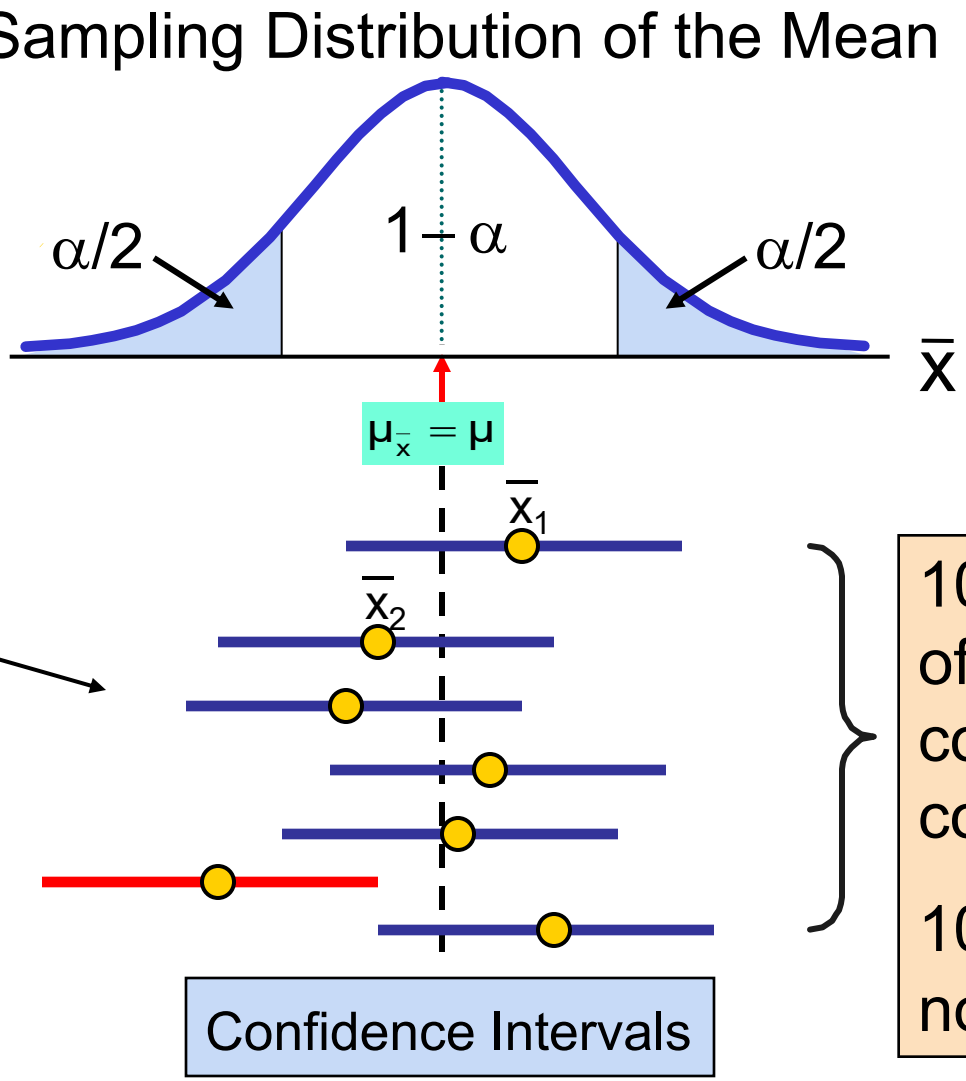
Troy University Resources

Common Levels of Confidence

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

Confidence Level	*Confidence Coefficient,* $1-\alpha$	$Z_{\alpha/2}$ *value*
80%	.80	1.28
90%	.90	1.645
95%	.95	1.96
98%	.98	2.33
99%	.99	2.575
99.8%	.998	3.08
99.9%	.999	3.27

Intervals and Level of Confidence



Intervals extend from

$$\bar{x} - z\frac{\sigma}{\sqrt{n}}$$

to

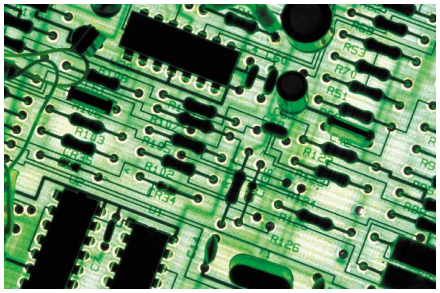
$$\bar{x} + z\frac{\sigma}{\sqrt{n}}$$

100(1-□)% of intervals constructed contain μ;

100(□)% do not.

Example

- 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

(continued)

- 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.

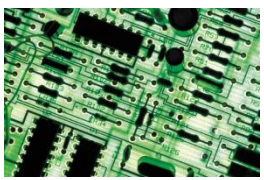
- Solution:

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96\,(.35/\sqrt{11})$$

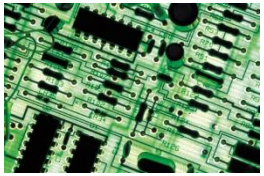
$$= 2.20 \pm .2068$$

$$1.9932 < \mu < 2.4068$$

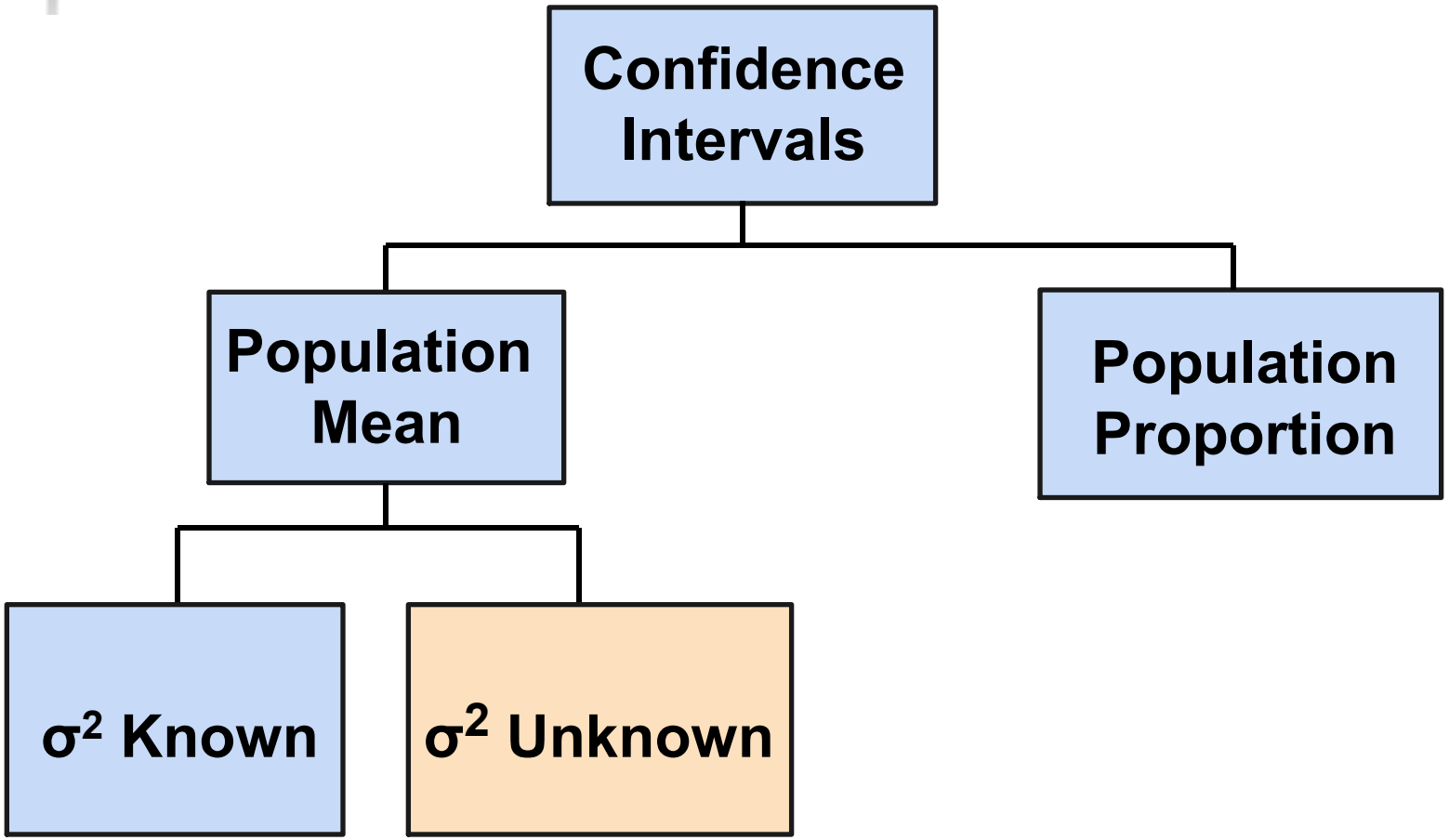


Interpretation

- 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



Confidence Intervals



Student's t Distribution

- 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

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

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

Confidence Interval for μ (σ^2 Unknown)

- 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

Troy University Resources

Confidence Interval for μ (σ Unknown)

(continued)

- 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:

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

where $t_{n-1,\alpha/2}$ is the critical value of the t distribution with n-1 d.f. and an area of α/2 in each tail

$$P(t_{n-1} > t_{n-1,\alpha/2}) = \alpha/2$$

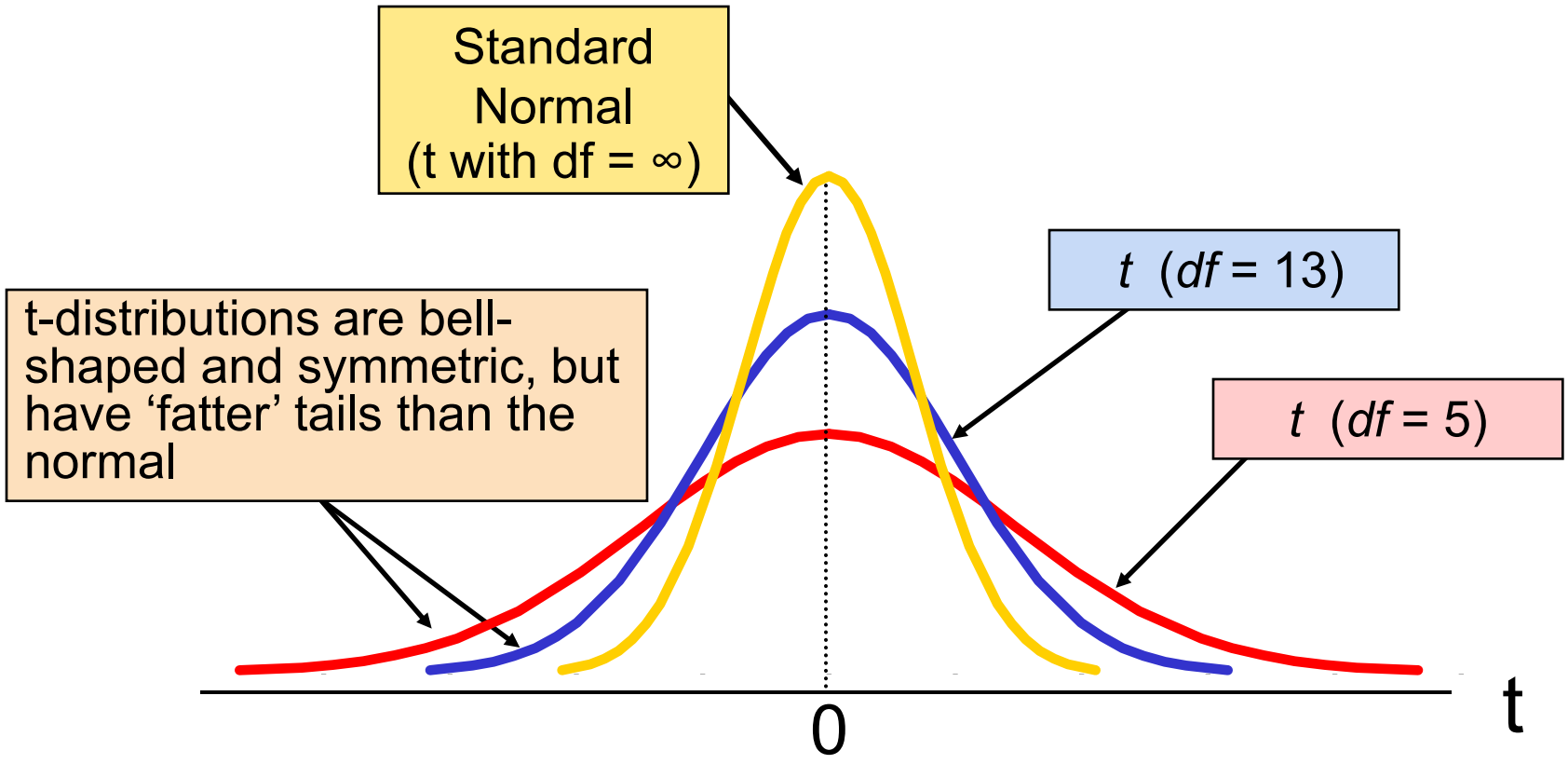
Student's t Distribution

- 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

$$d.f. = n - 1$$

Student's t Distribution

Note: t ⟶ Z as n increases

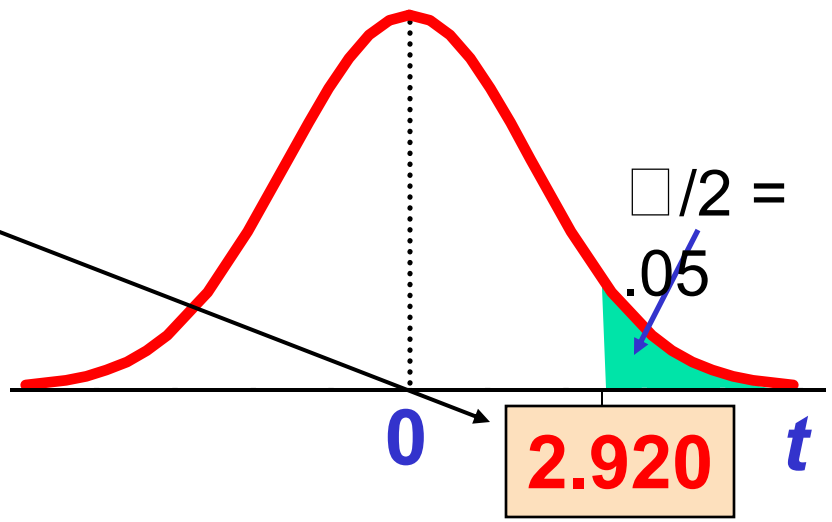


Student's t Table

	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
□ = .10
□/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 → Z as n increases

Example

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

The confidence interval is

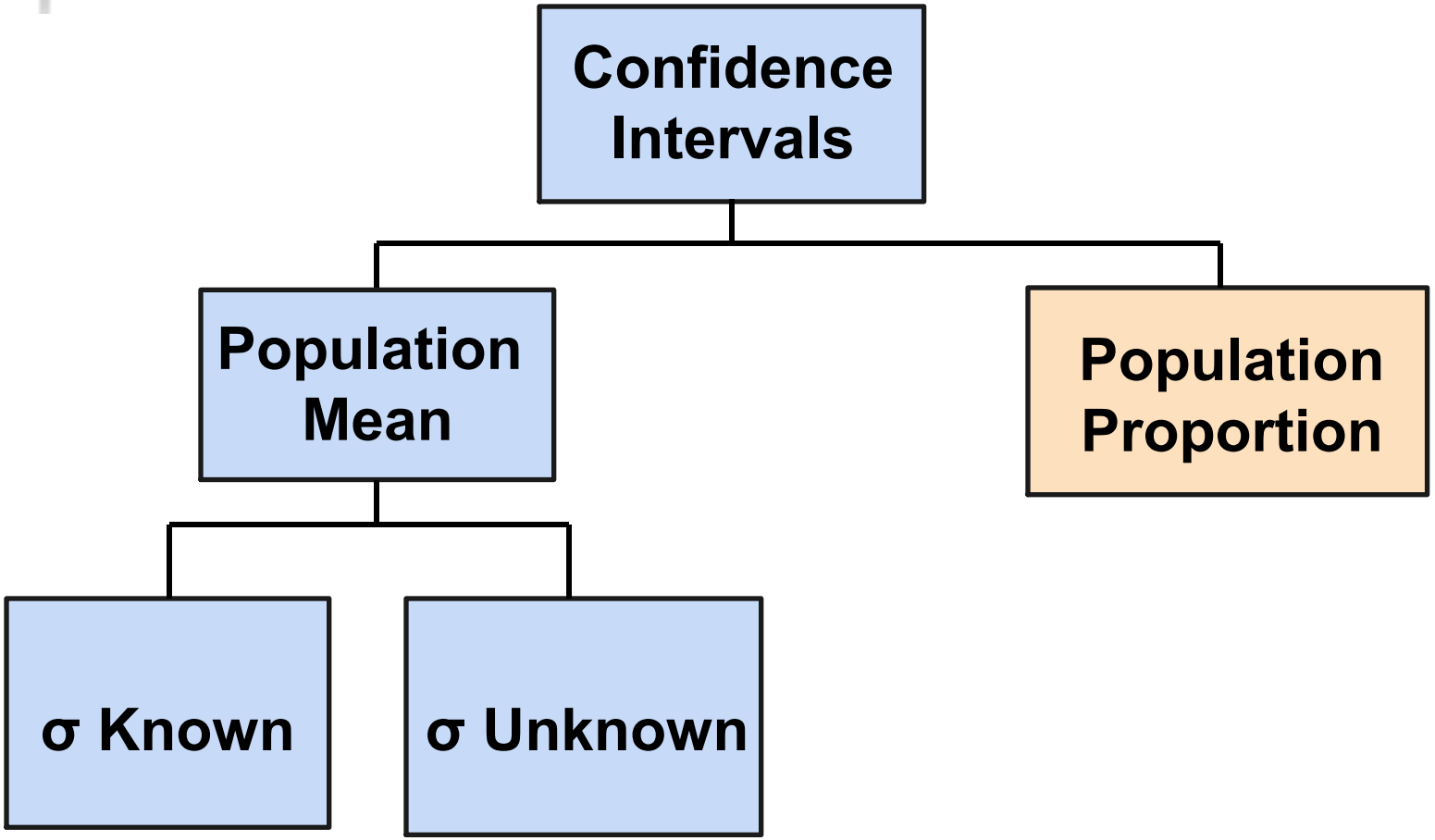
$$\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

$$50 - (2.0639)\frac{8}{\sqrt{25}} < \mu < 50 + (2.0639)\frac{8}{\sqrt{25}}$$

$$46.698 < \mu < 53.302$$

Troy University Resources

Confidence Intervals



Confidence Intervals for the Population Proportion, p

- 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, p

(continued)

- 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

- Upper and lower confidence limits for the population proportion are calculated with the formula

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

- where
  - $z_{\alpha/2}$ is the standard normal value for the level of confidence desired
  - $\hat{p}$ is the sample proportion
  - n is the sample size

Example

- 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

(continued)

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

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

$$\frac{25}{100} - 1.96\sqrt{\frac{.25(.75)}{100}} < P < \frac{25}{100} + 1.96\sqrt{\frac{.25(.75)}{100}}$$

$$0.1651 < P < 0.3349$$



Troy University Resources

Interpretation

- We are 95% confident that the true percentage 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.





Chapter Summary

- 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)
- Created confidence interval estimates for the proportion