Confidence Intervals

Chapter 8 (8.1, 8.2, 8.4)

Learning Objectives

In this chapter, you learn:

- To construct and interpret confidence interval estimates for population means
- How to determine the sample size necessary to develop a confidence interval of a desired length

Chapter Outline

Content of this chapter

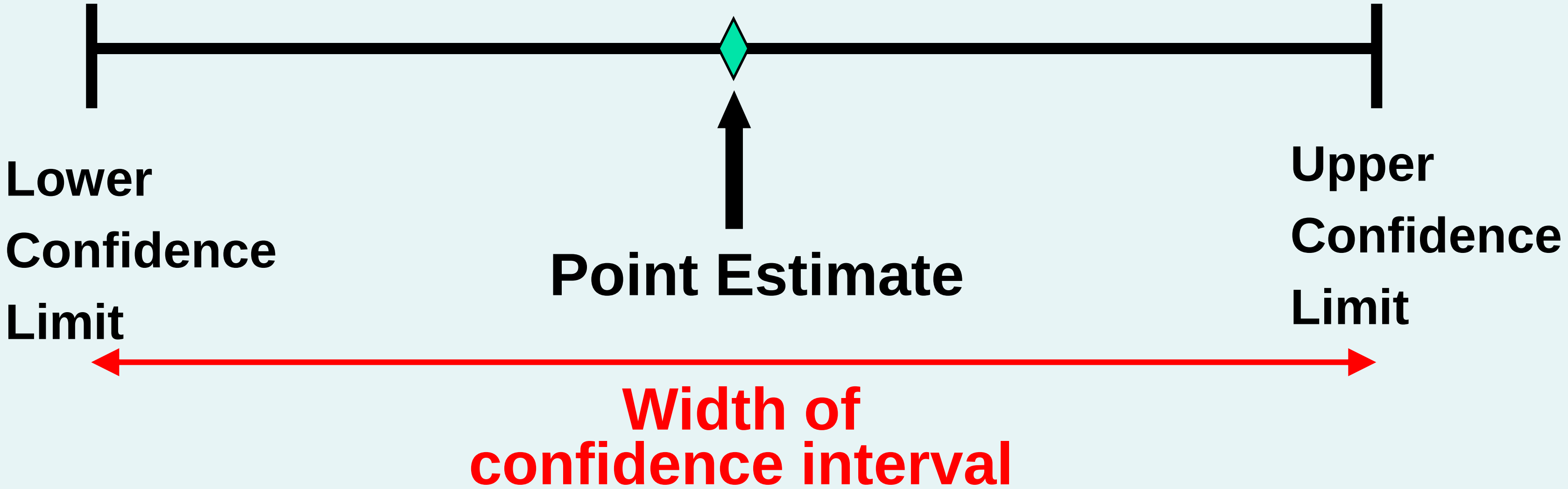
- Confidence Intervals for the Population Mean, μ
  - when Population Standard Deviation σ is Known
  - when Population Standard Deviation σ is Unknown
- Determining the Required Sample Size

Point and Interval Estimates

- A point estimate is a single number.

Examples: sample mean, sample variance, sample proportion.

- A confidence interval provides additional information about the variability of the estimate



Definition

[a,b] is a **(1-α)100% confidence interval** for a parameter, if it contains this parameter with probability (1-α).

$$P(a \leq \text{parameter} \leq b) = 1-\alpha$$

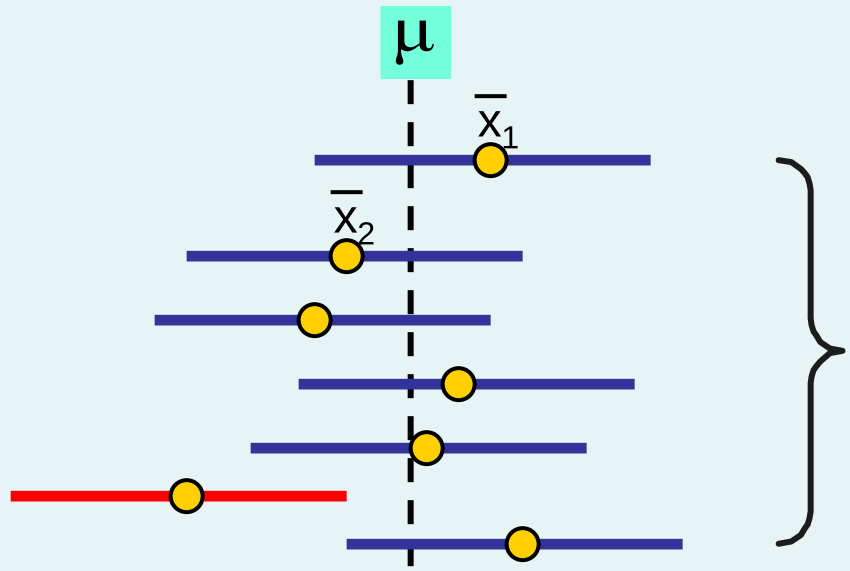
For example, **[a,b]** is a 95% confidence interval for the population mean μ if

$$P(a \leq \mu \leq b) = 0.95$$

Confidence level **(1-α)**

P(a ≤ parameter ≤ b) = 1-α

Parameter is not random! The interval [a,b] is:



Confidence Intervals

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

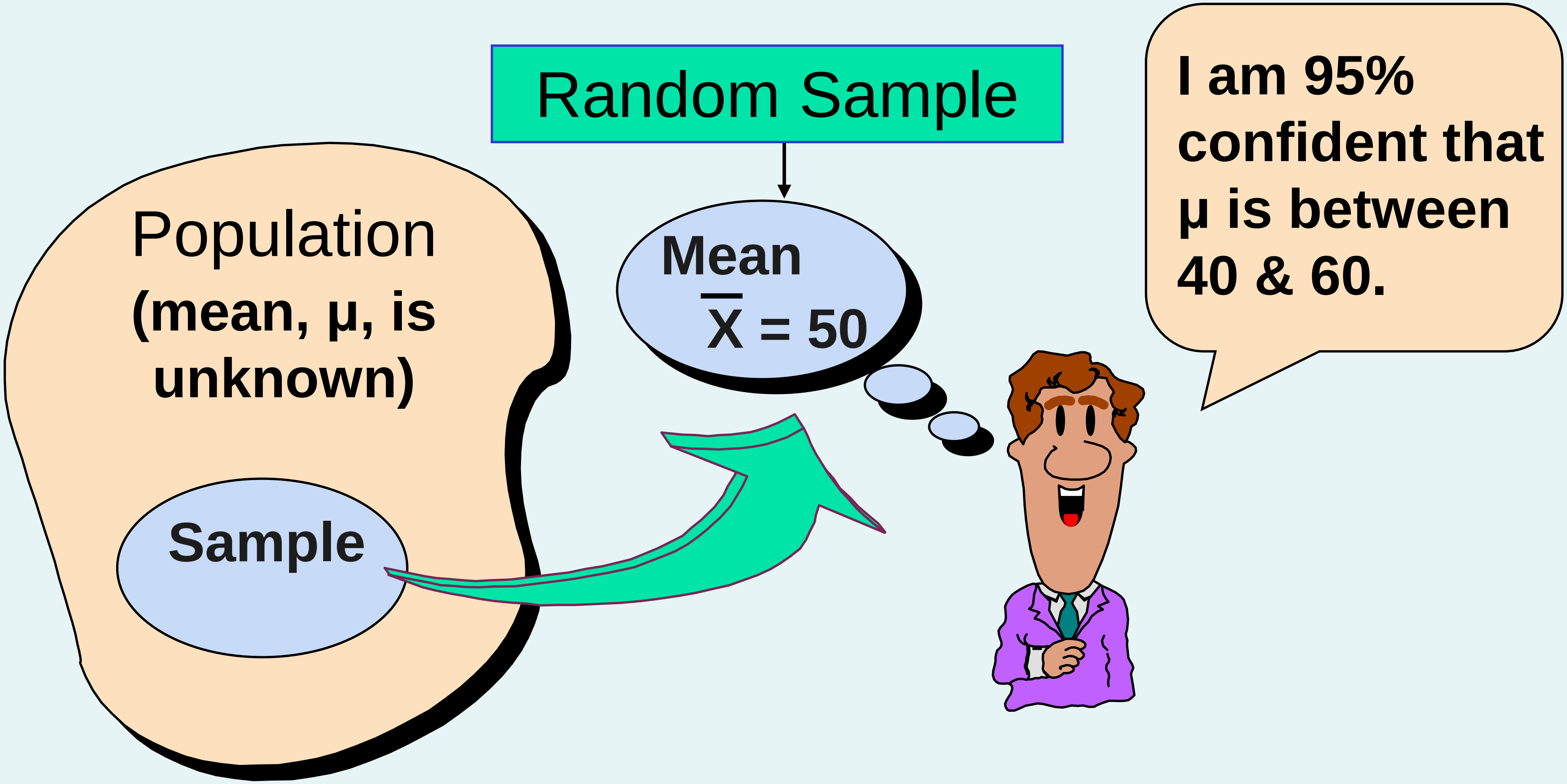
(α)100% do not.

Confidence Interval Estimate

- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observations from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence

    Such as 95% confident, 99% confident

Estimation Process



General Formula

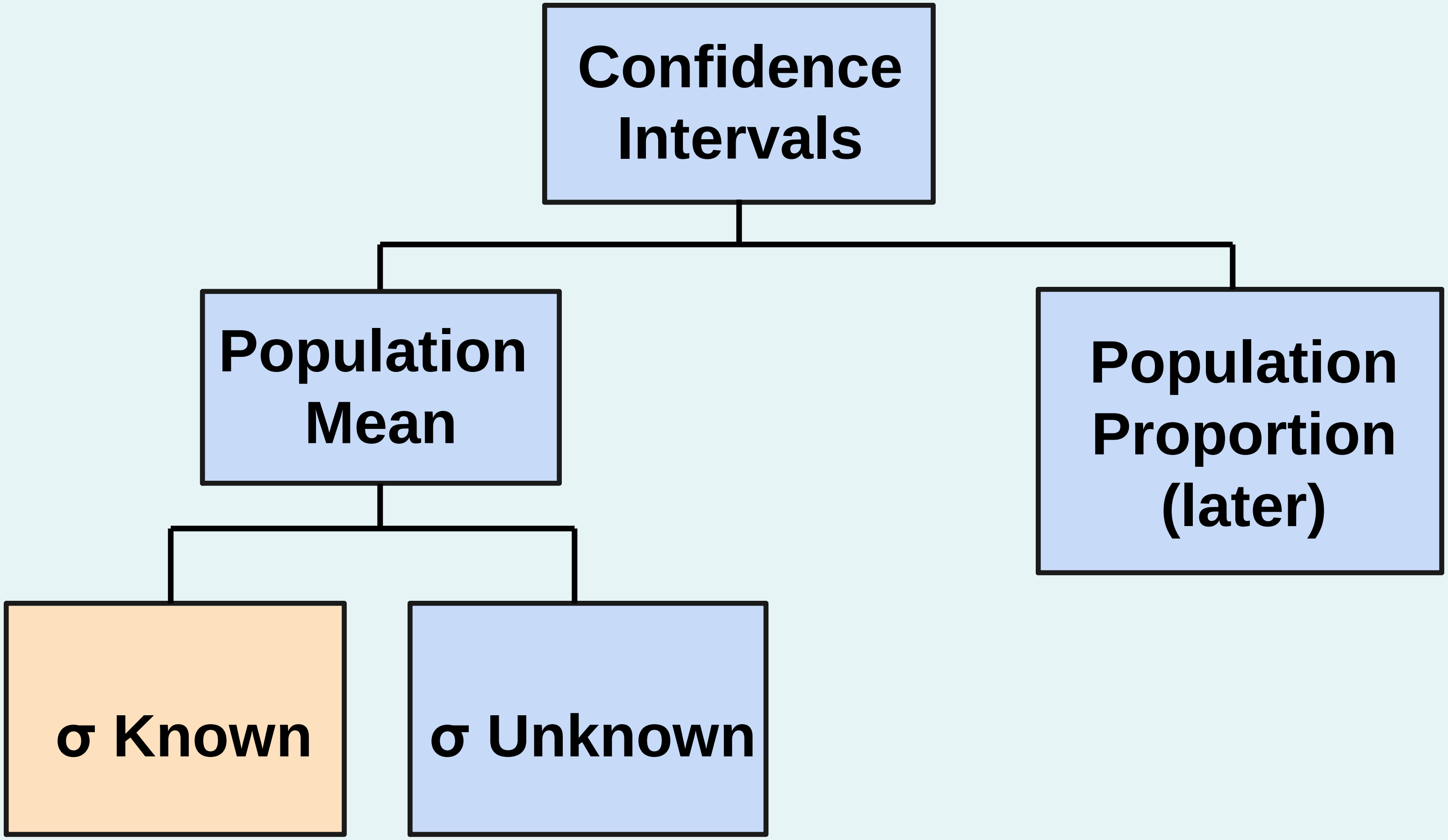
- The general formula for all symmetric confidence intervals is:

Point Estimate ± (Critical Value)(Standard Error)

Where:

- Point Estimate is the sample statistic estimating the population parameter of interest
- Critical Value is a table value based on the sampling distribution of the point estimate and the desired confidence level
- Standard Error is the standard deviation of the point estimate

Confidence Intervals



Confidence Interval for μ (σ Known)

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

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where $\bar{X}$ is the point estimate

$Z_{\alpha/2}$ is the normal distribution critical value for a probability of $\alpha/2$ in each tail

$\sigma/\sqrt{n}$ is the standard error

Example: $X_1, \ldots, X_n$ from Normal(μ, σ) with unknown μ, known σ

1. Estimate $\theta = \mu$ by its estimator $\bar{X} = \frac{1}{n}\sum X_i$.

2. Find its distribution: *Normal* with

$$E(\bar{X}) = \mu$$

$$\text{Var}(\bar{X}) = \frac{1}{n^2}\sum_{1}^{n}\text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Therefore,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ is Normal(0,1)}$$

3. Find *critical values* $\pm z_{\alpha/2}$ such that

$$P\left\{-z_{\alpha/2} < Z < z_{\alpha/2}\right\}$$

for $Z \sim$ Normal(0,1).

4. Then we have

$$P\left\{-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right\} = 1-\alpha$$

Solve for μ:

$$P\left\{\bar{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

5. Hence,

$$\bar{X} \pm \frac{z_{\alpha/2}\sigma}{\sqrt{n}} = \left[\bar{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\,,\ \bar{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right]$$

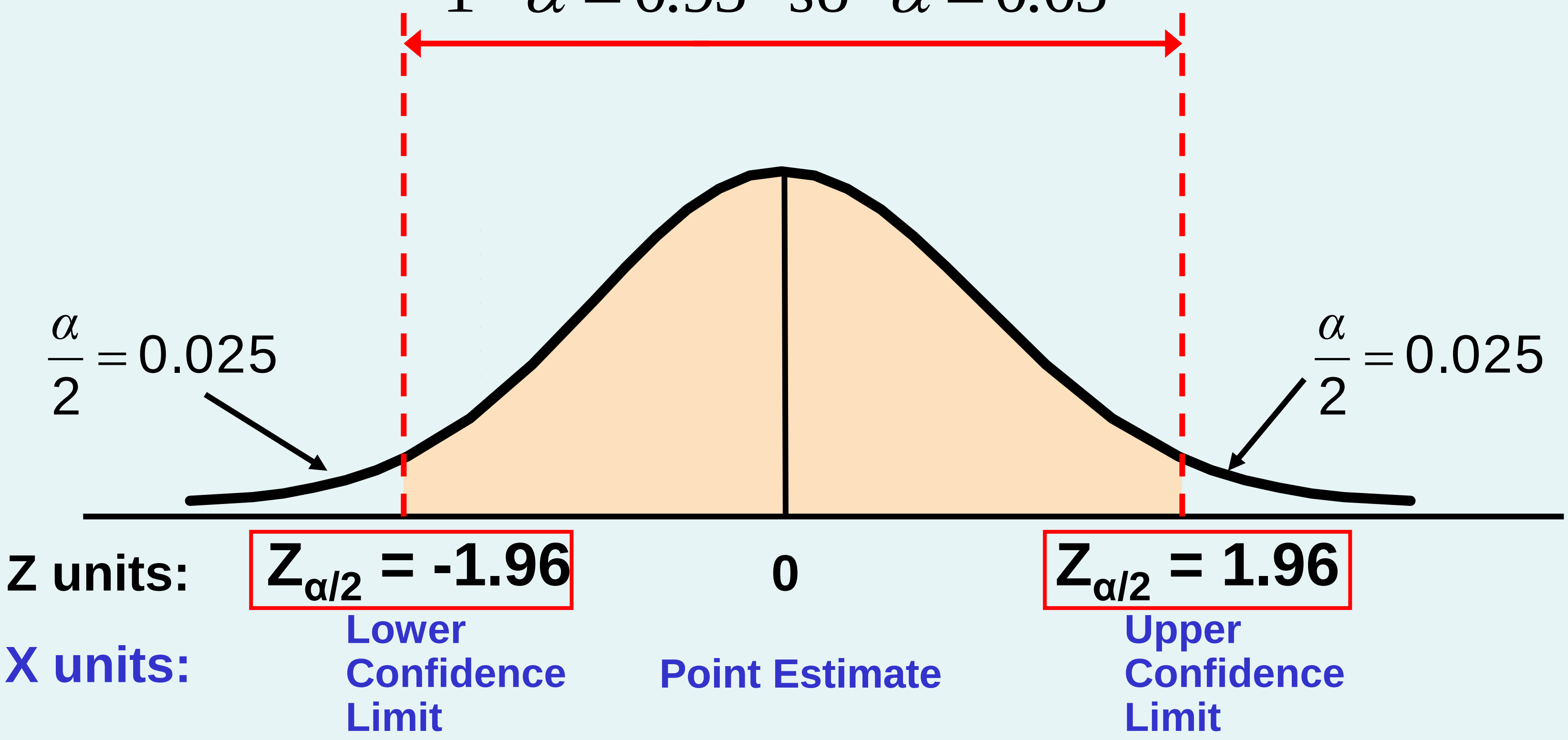
is a $(1-\alpha)100\%$ confidence interval for μ.

Finding the Critical Value, $Z_{\alpha/2}$

- For a 95% confidence interval:

$$Z_{\alpha/2} = \pm 1.96$$

$$1-\alpha = 0.95 \quad \text{so} \quad \alpha = 0.05$$



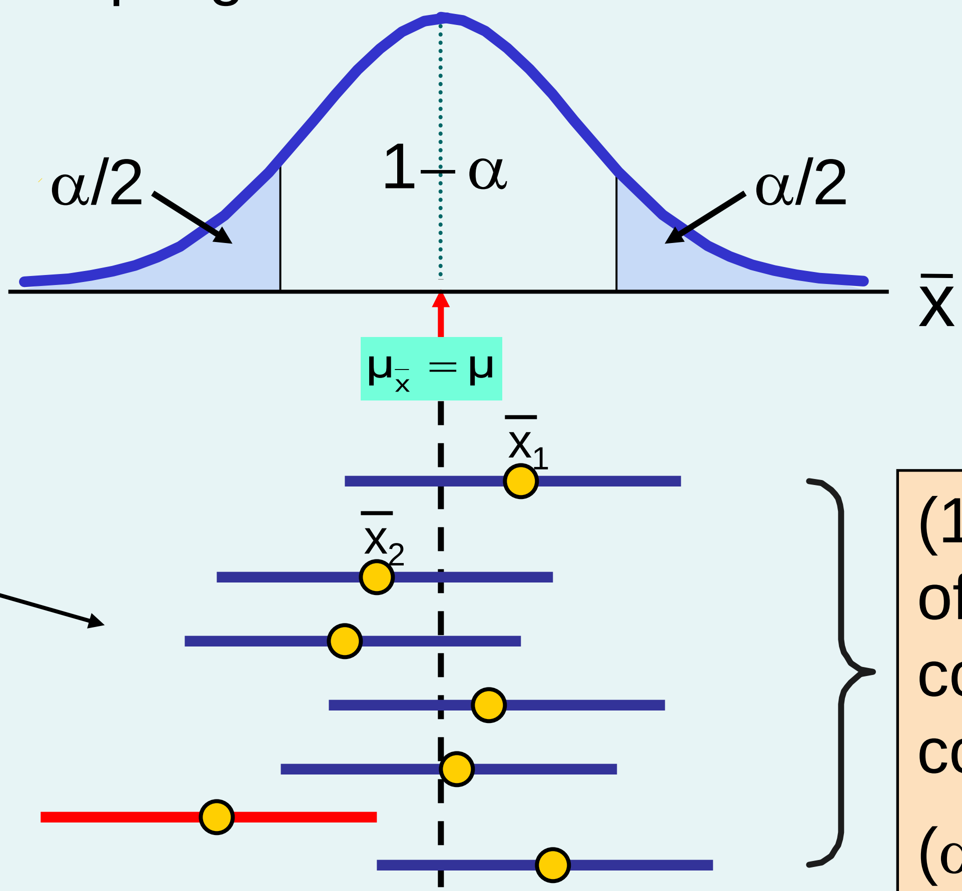
Common Levels of Confidence

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

Confidence Level	***Confidence Coefficient,*** $1-\alpha$	$Z_{\alpha/2}$ ***value***
80%	**0.80**	**1.28**
90%	**0.90**	**1.645**
95%	**0.95**	**1.96**
98%	**0.98**	**2.33**
99%	**0.99**	**2.58**
99.8%	**0.998**	**3.08**
99.9%	**0.999**	**3.27**

Intervals and Level of Confidence

Sampling Distribution of the Mean



Confidence Intervals

Intervals extend from

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

to

$$\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

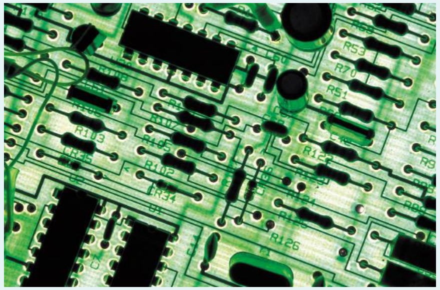
(1-α)100% of intervals constructed contain μ;

(α)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 0.35 ohms.

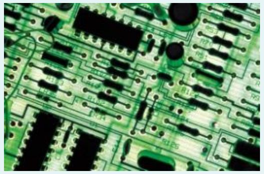
- Solution:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96\,(0.35/\sqrt{11})$$

$$= 2.20 \pm 0.2068$$

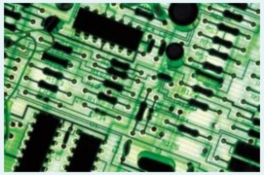
$$\boxed{1.9932 \le \mu \le 2.4068}$$



Interpretation

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms

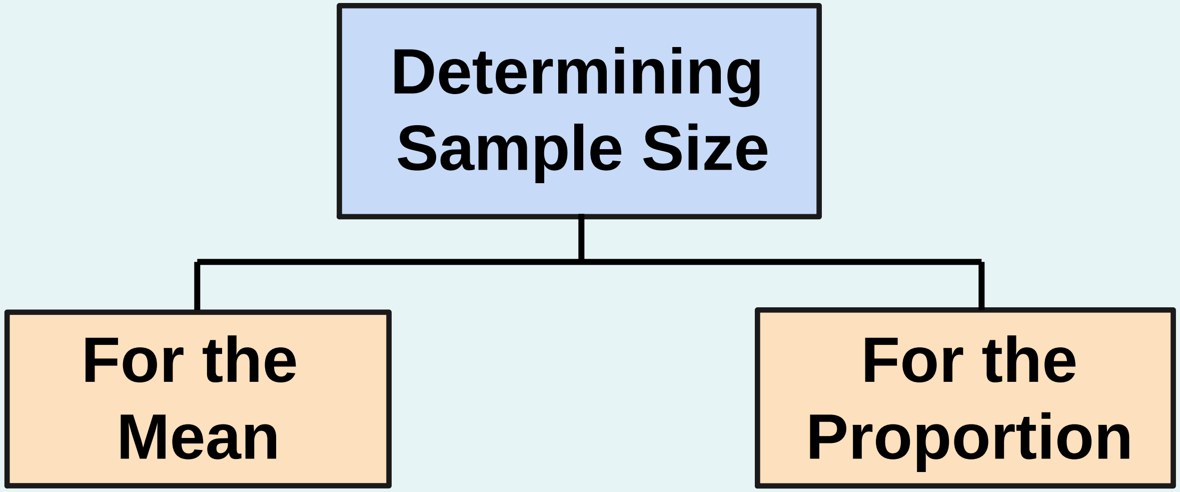
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean



Example

- Page 276, # 8.9

Determining Sample Size



Sampling Error

- The required sample size can be found to reach a desired margin of error (e) with a specified level of confidence ($1 - \alpha$)

- The margin of error is also called sampling error
  - amount of imprecision in the estimate of the population parameter, due to using a **random sample**
  - amount added and subtracted to the point estimate to form the confidence interval

Determining Sample Size

Determining Sample Size

For the Mean

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Sampling error (margin of error)

$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Determining Sample Size

(continued)

Determining Sample Size

For the Mean

$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Now solve for n to get

$$n = \frac{Z_{\alpha/2}^2 \sigma^2}{e^2}$$

Required Sample Size Example

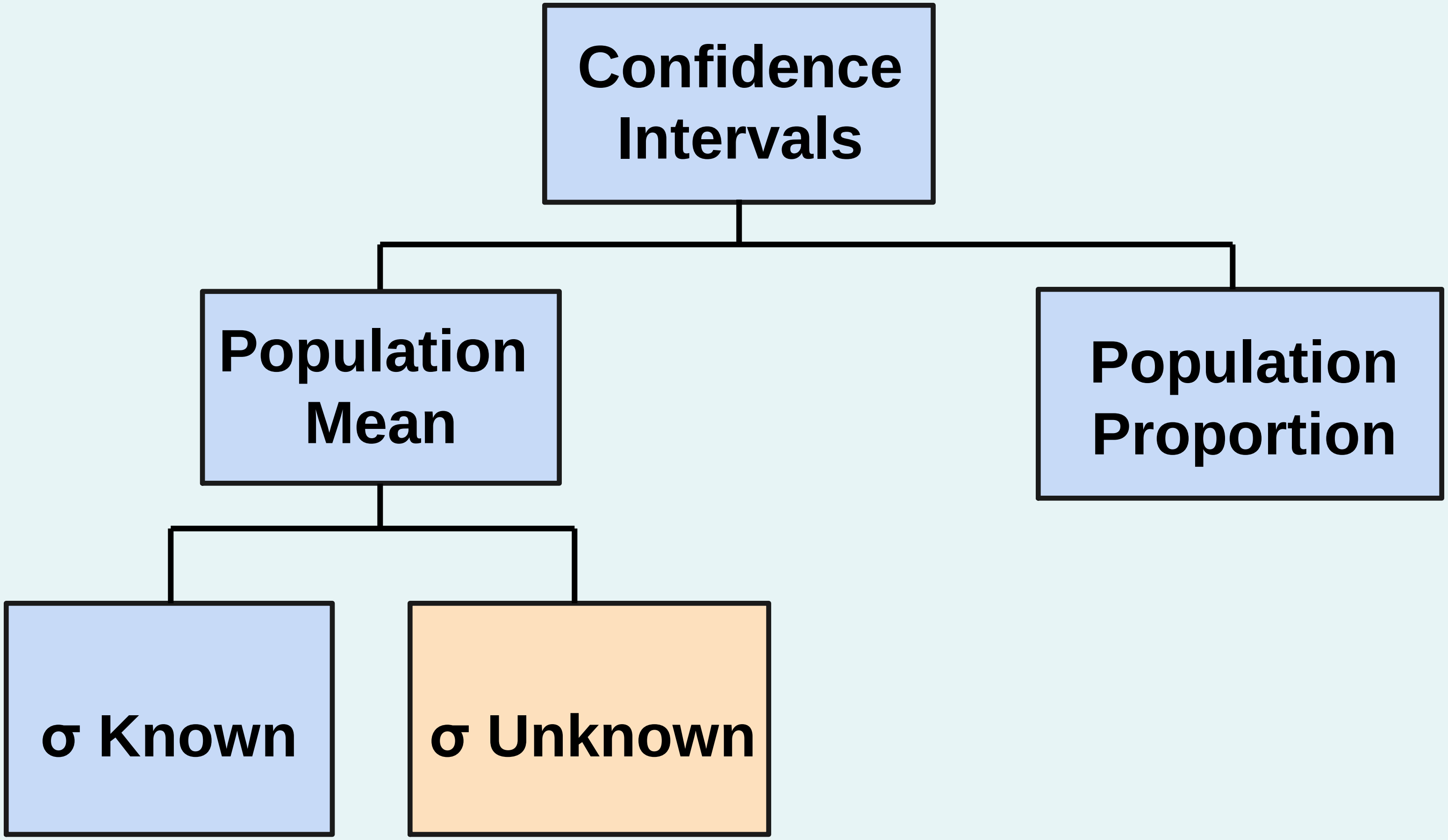
If σ = 45, what sample size is needed to estimate the mean within ± 5 with 90% confidence?

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.645)^2 (45)^2}{5^2} = 219.19$$

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

(Always round **up**)

Confidence Intervals



Do You Ever Truly Know σ?

- Probably not!
- In virtually all real world business situations, σ is not known.
- If there is a situation where σ is known then μ is often also known (since to calculate σ you need to know μ.)
- If you truly know μ there would be no need to gather a sample to estimate it.

Confidence Interval for μ (σ 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 Student's t distribution instead of the normal distribution

When σ is unknown

Data $X_1, ..., X_n$ from Normal(μ, σ) with unknown μ, **unknown** σ

1. Estimate σ by $S = \sqrt{\frac{1}{n-1} \sum_1^n (X_i - \bar{X})^2}$

2. Use t-distribution with $(n-1)$ degrees of freedom instead of Normal.

For large n, use Normal approximation.

Result:

$$\bar{X} \pm \frac{t_{\alpha/2, n-1} S}{\sqrt{n}}$$

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:

$$\boxed{\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}}$$

(where $t_{\alpha/2}$ is the critical value of the t distribution with n -1 degrees of freedom and an area of α/2 in each tail)

Student's t Distribution

- The t is a family of distributions
- The $t_{\alpha/2}$ value depends on degrees of freedom (d.f.)
  - Number of observations that are free to vary after sample mean has been calculated

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

Degrees of Freedom (df)

What are degrees of freedom?

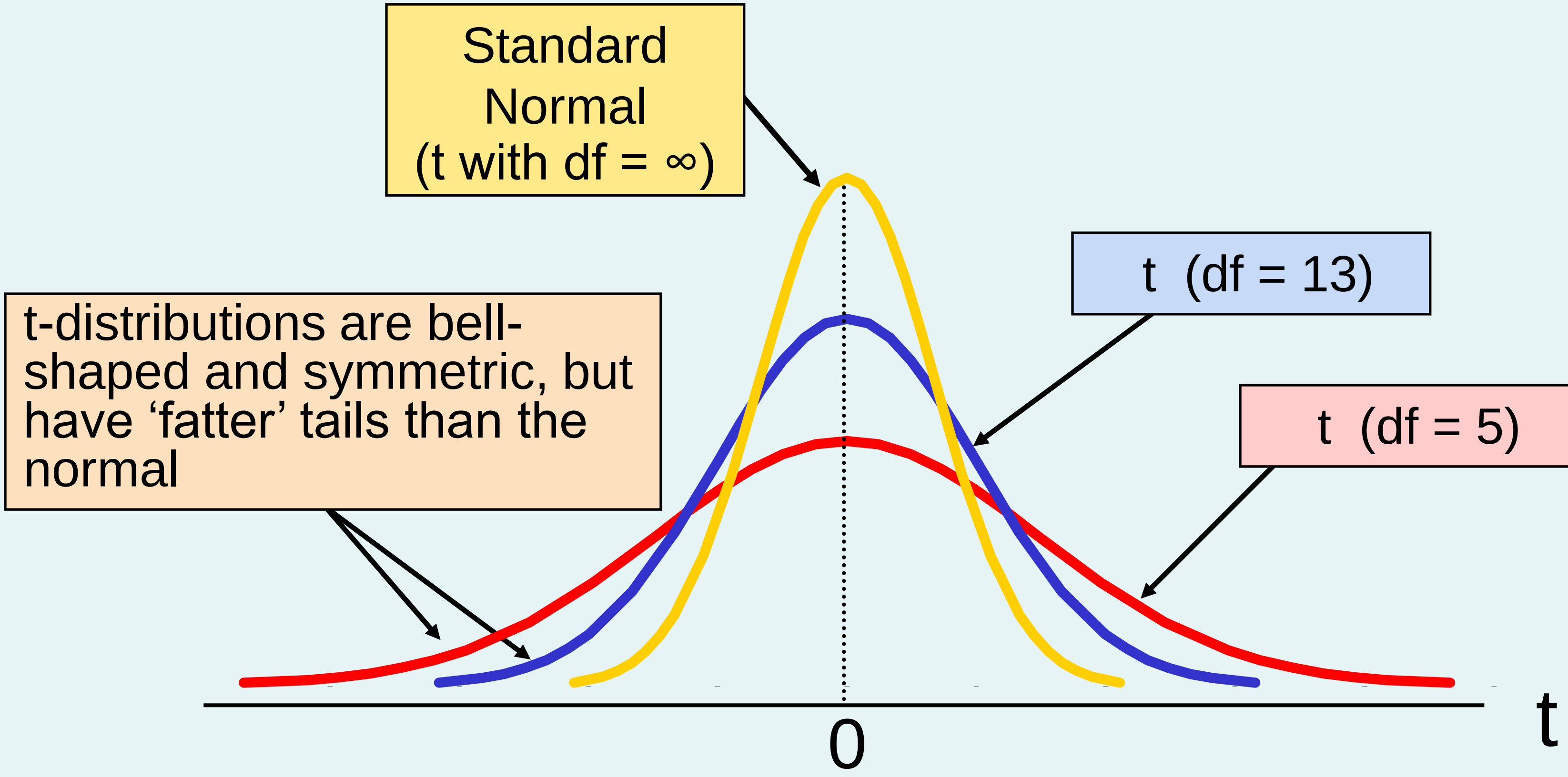
This is dimension of the space, the number of freely varying quantities used to estimate the standard deviation.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

is based on *n* quantities, but only *(n-1)* vary freely.

Student’s t Distribution

t converges to Z as n increases

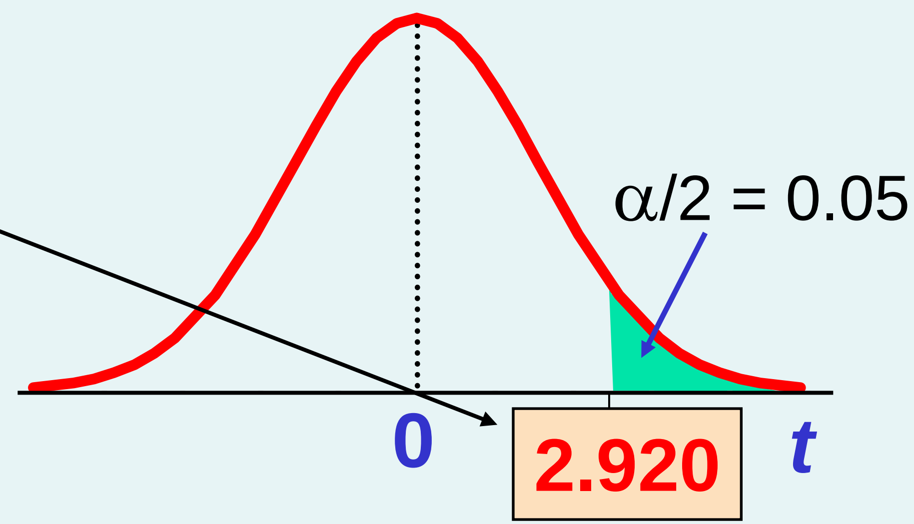


Student's t Table

df	Upper Tail Area		
	.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
α = 0.10
$\alpha/2$ = 0.05



Selected t distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z (∞ d.f.)
0.80	1.372	1.325	1.310	1.28
0.90	1.812	1.725	1.697	1.645
0.95	2.228	2.086	2.042	1.96
0.99	3.169	2.845	2.750	2.58

Note: $t \longrightarrow Z$ as n increases

t-values are higher than Z-values (price for unknown σ)

Example of a confidence interval (case of unknown variance)

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

- d.f. = n – 1 = 24, so $t_{\alpha/2} = t_{0.025} = 2.0639$

The confidence interval is

$$\overline{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

$46.698 \leq \mu \leq 53.302$

Excel commands

= tinv (α, degrees of freedom)

Returns the two-tailed inverse of the Student's t-distribution.

= confidence.t (α, standard deviation, sample size)

gives the margin of a (1-α)100% confidence interval.

Examples

- Page 283, #8.16 – manually
- Page 283, #8.20 – use Excel

How to determine the sample size when σ is unknown?

- If unknown, σ can be estimated when using the required sample size formula
  - Use a value for σ that is expected to be at least as large as the true σ
  - Select a pilot sample and estimate σ with the sample standard deviation, S

Ethical Issues (usual practice)

- A confidence interval estimate (reflecting sampling error) should always be included when reporting a point estimate
- The level of confidence should always be reported
- The sample size should be reported
- An interpretation of the confidence interval estimate should also be provided
  - *Examples: USA Today survey polls*

Examples

- Page 292, # 8.43ab
- Page 292, # 8.45abc

Chapter Summary

In this chapter we discussed

- The concept of confidence intervals
- Point estimates & confidence interval estimates
- Confidence interval for the mean (σ known)
- Confidence interval for the mean (σ unknown)
- Determining required sample size
- Ethical issues in confidence interval estimation