A Course In Business Statistics

4th Edition

Chapter 7

Estimating Population Values

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
- Determine the required sample size to estimate a single population mean within a specified margin of error
- Form and interpret a confidence interval estimate for a single population proportion

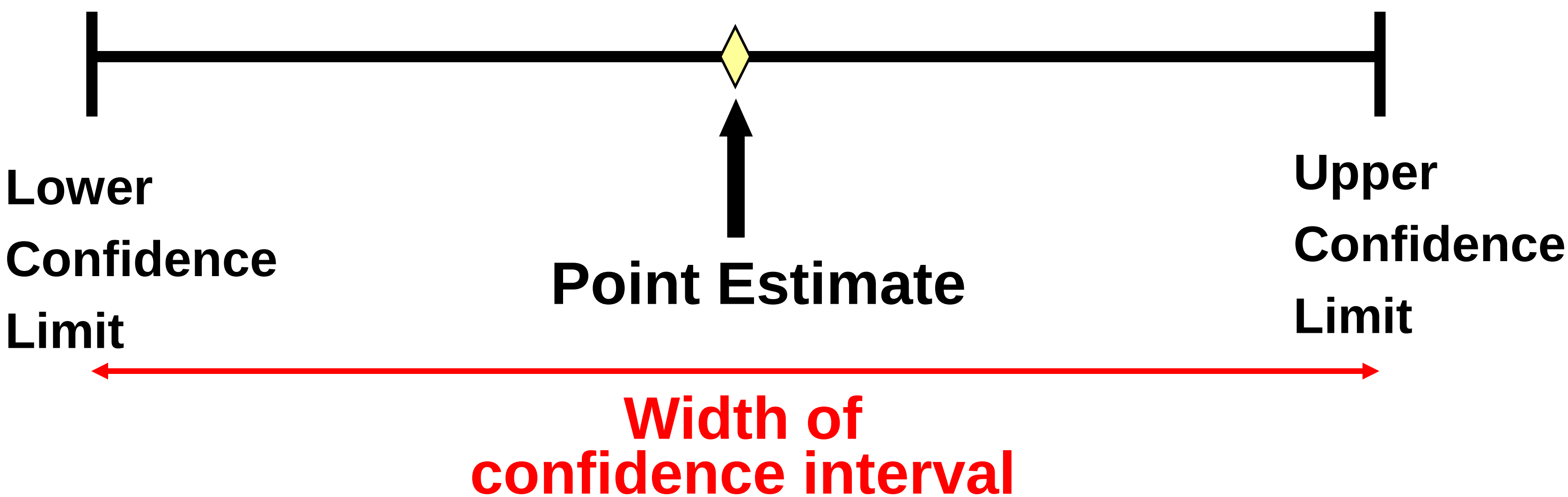
Confidence Intervals

Content of this chapter

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

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	$\bar{p}$

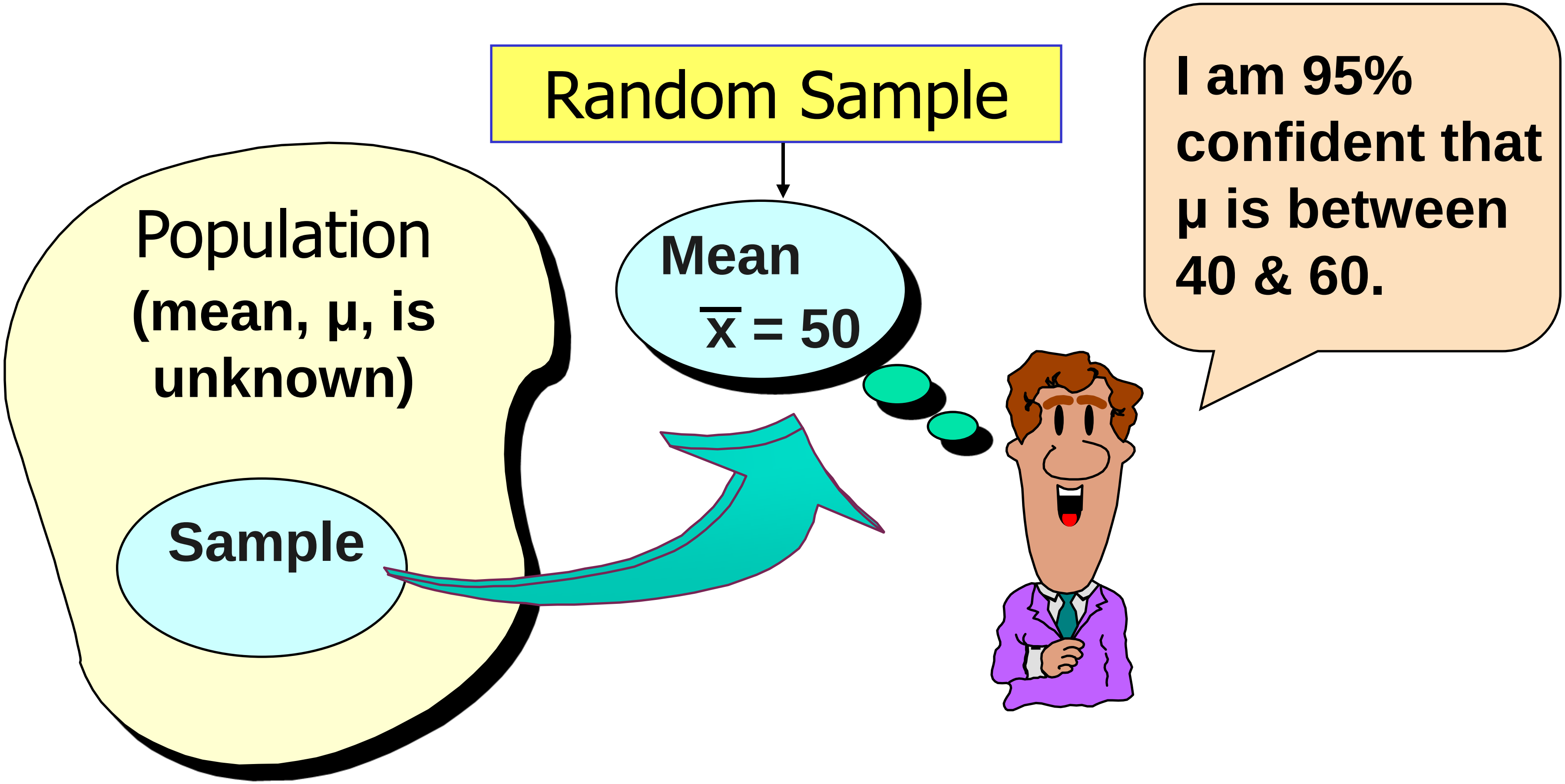
Confidence Intervals

- 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
    - Never 100% sure

Estimation Process



General Formula

- The general formula for all confidence intervals is:

Point Estimate ± (Critical Value)(Standard Error)

Confidence Level

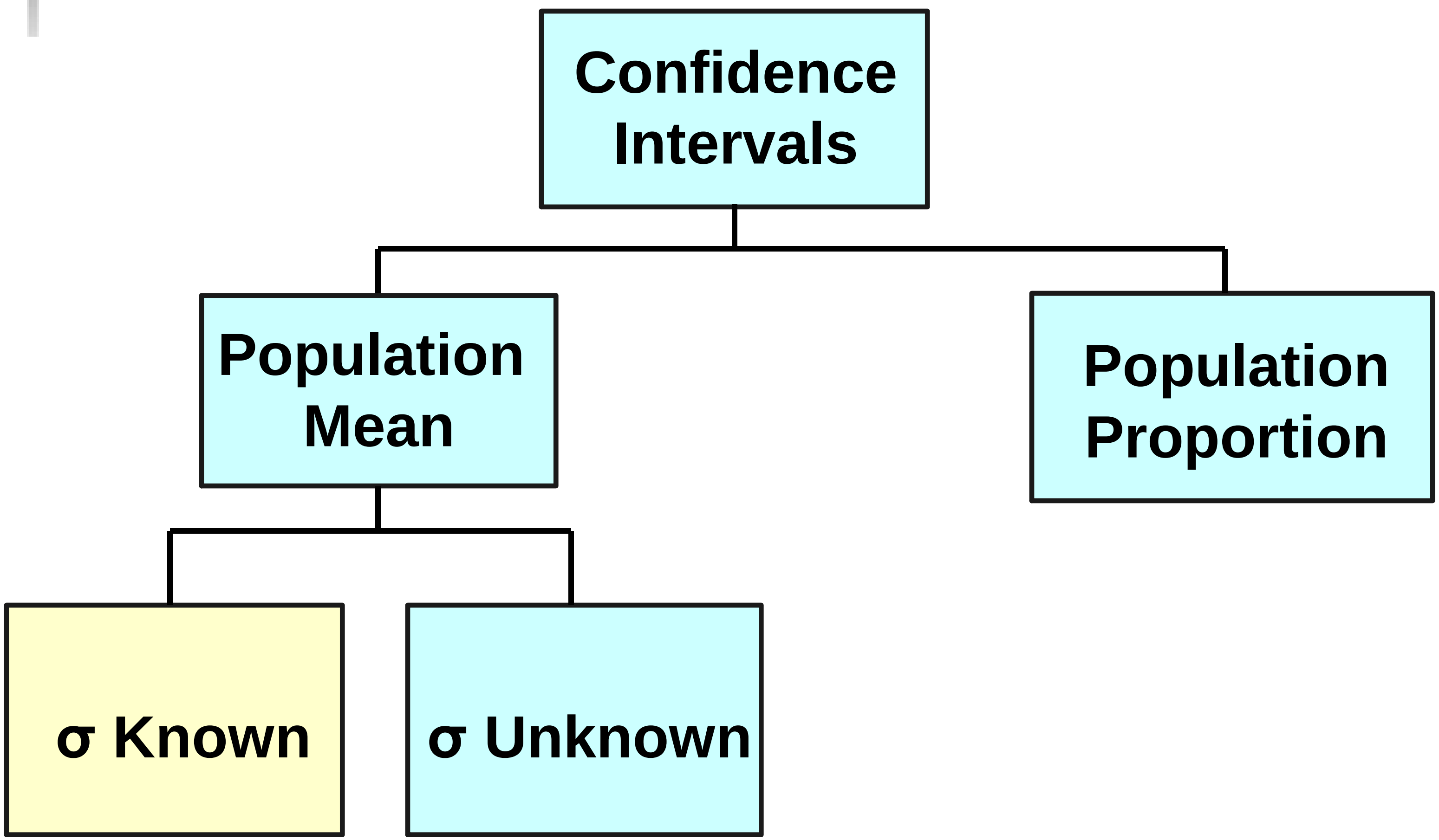
- Confidence Level
  - Confidence in which the interval will contain the unknown population parameter
- A percentage (less than 100%)

Confidence Level, $(1-\alpha)$

(continued)

- Suppose confidence level = 95%
- Also written $(1 - \alpha) = .95$
- A relative frequency interpretation:
  - In the long run, 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

Confidence Intervals



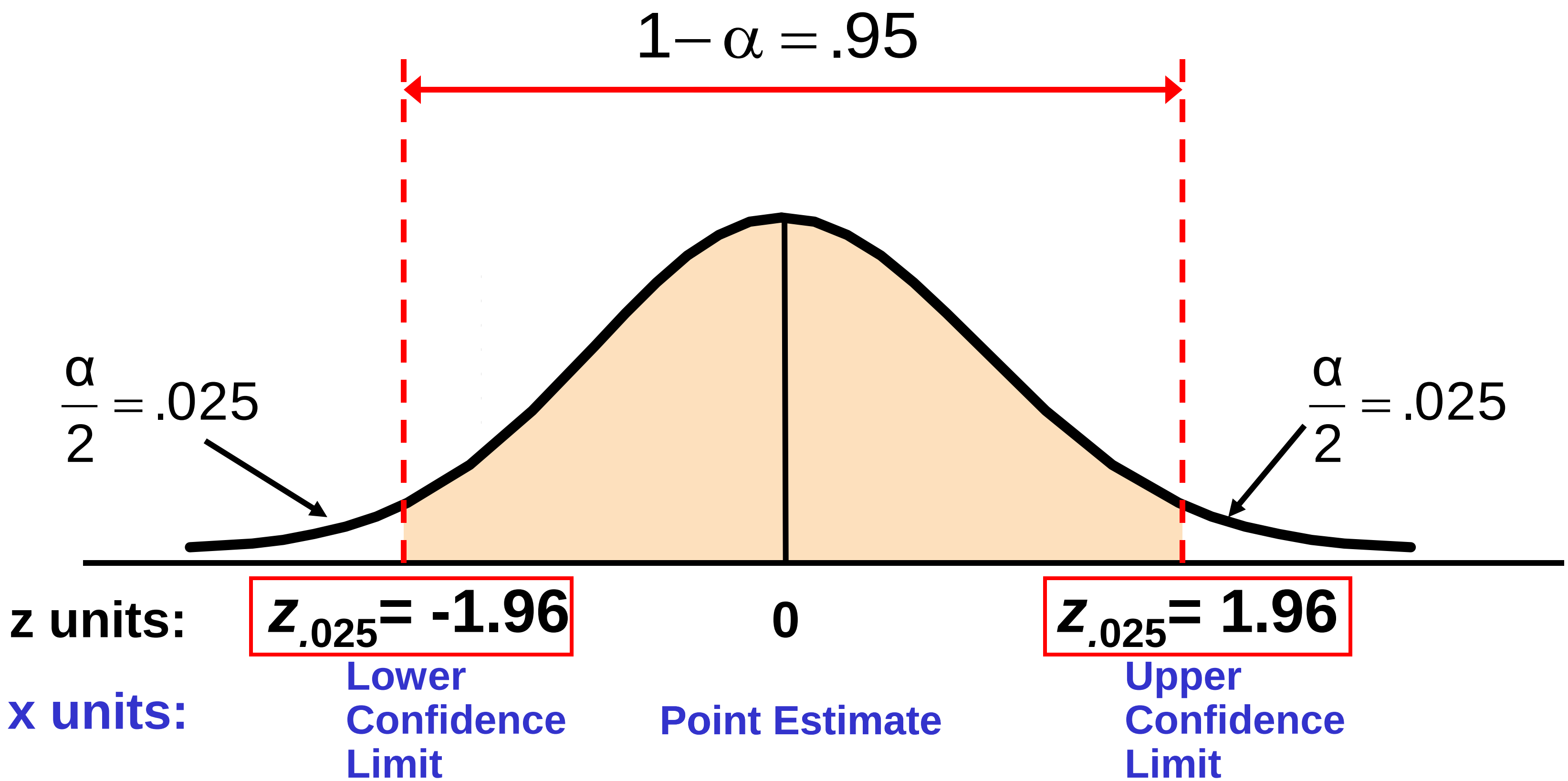
Confidence Interval for μ (σ Known)

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

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

Finding the Critical Value

- Consider a 95% confidence interval: $\boxed{z_{\alpha/2} = \pm 1.96}$



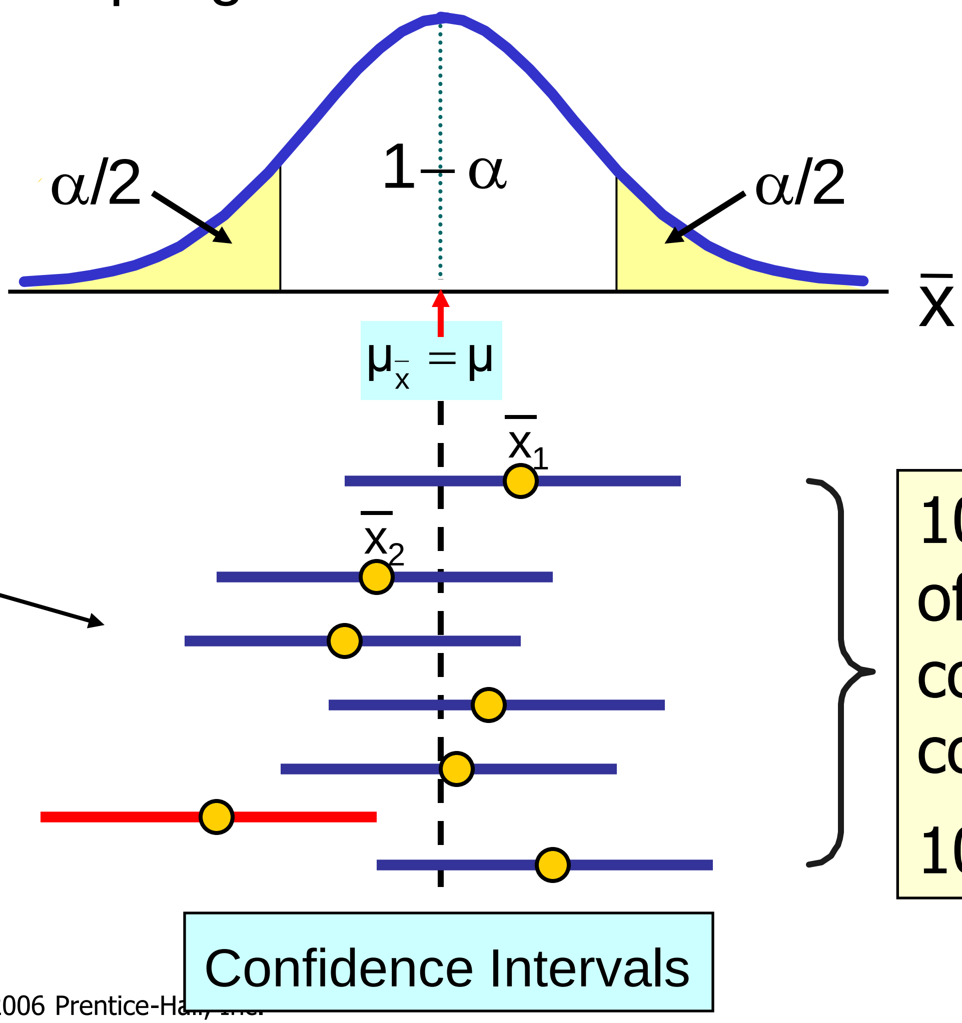
Common Levels of Confidence

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

Confidence Level	*Confidence Coefficient,* $1-\alpha$	*z value,* $Z_{\alpha/2}$
80%	**.80**	**1.28**
90%	**.90**	**1.645**
95%	**.95**	**1.96**
98%	**.98**	**2.33**
99%	**.99**	**2.57**
99.8%	**.998**	**3.08**
99.9%	**.999**	**3.27**

Interval and Level of Confidence

Sampling Distribution of the Mean



Intervals extend from

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

to

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

$100(1-\alpha)\%$ of intervals constructed contain μ;

$100\alpha\%$ do not.

Margin of Error

- Margin of Error (e): the amount added and subtracted to the point estimate to form the confidence interval

Example: Margin of error for estimating μ, σ known:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad\qquad e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Factors Affecting Margin of Error

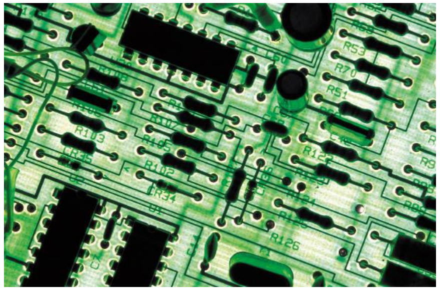
$$e = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Data variation, σ : e ↓ as σ ↓
- Sample size, n : e ↓ as n ↑
- Level of confidence, 1 - α : e ↓ if 1 - α ↓

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 .35 ohms.

- Determine a 95% confidence interval for the true mean resistance of the population.



Example

(continued)

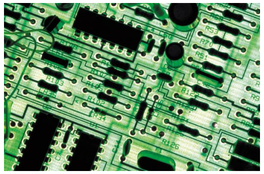
- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is .35 ohms.

- Solution: $\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

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

$$= 2.20 \pm .2068$$

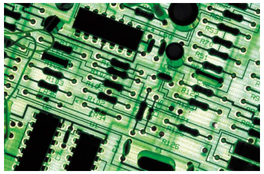
$$\boxed{1.9932 \; \ldots\ldots\ldots\; \ldots\ldots \; 2.4068}$$



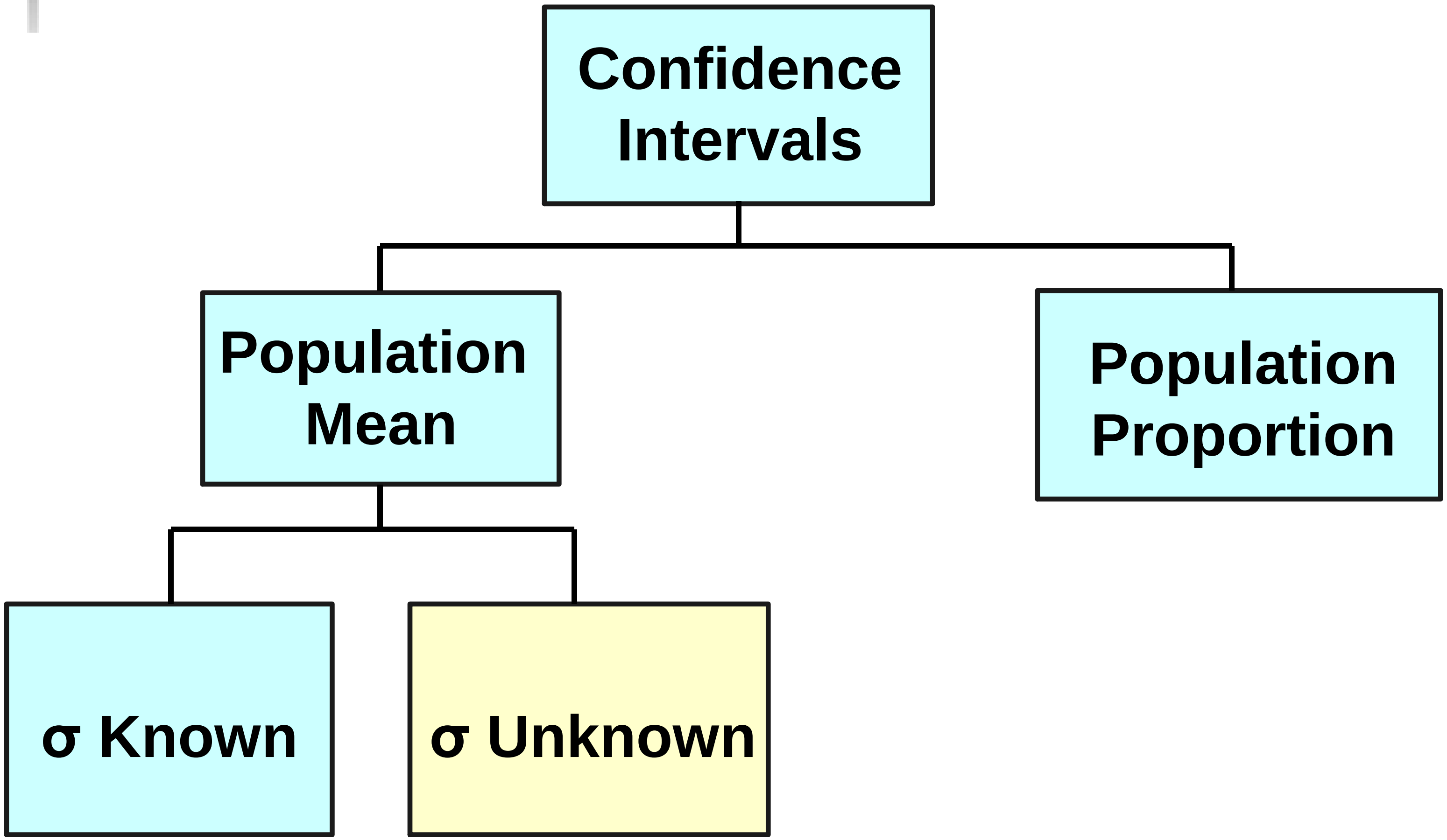
Interpretation

- We are 98% 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, 98% of intervals formed in this manner will contain the true mean
- An **incorrect** interpretation is that there is 98% probability that this interval contains the true population mean.

  (This interval either does or does not contain the true mean, there is no probability for a single interval)



Confidence Intervals



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 t distribution instead of the normal distribution

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

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

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

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

Degrees of Freedom (df)

Idea: Number of observations that are free to vary after sample mean has been calculated

Example: Suppose the mean of 3 numbers is 8.0

Let $x_1 = 7$
Let $x_2 = 8$
What is x_3?

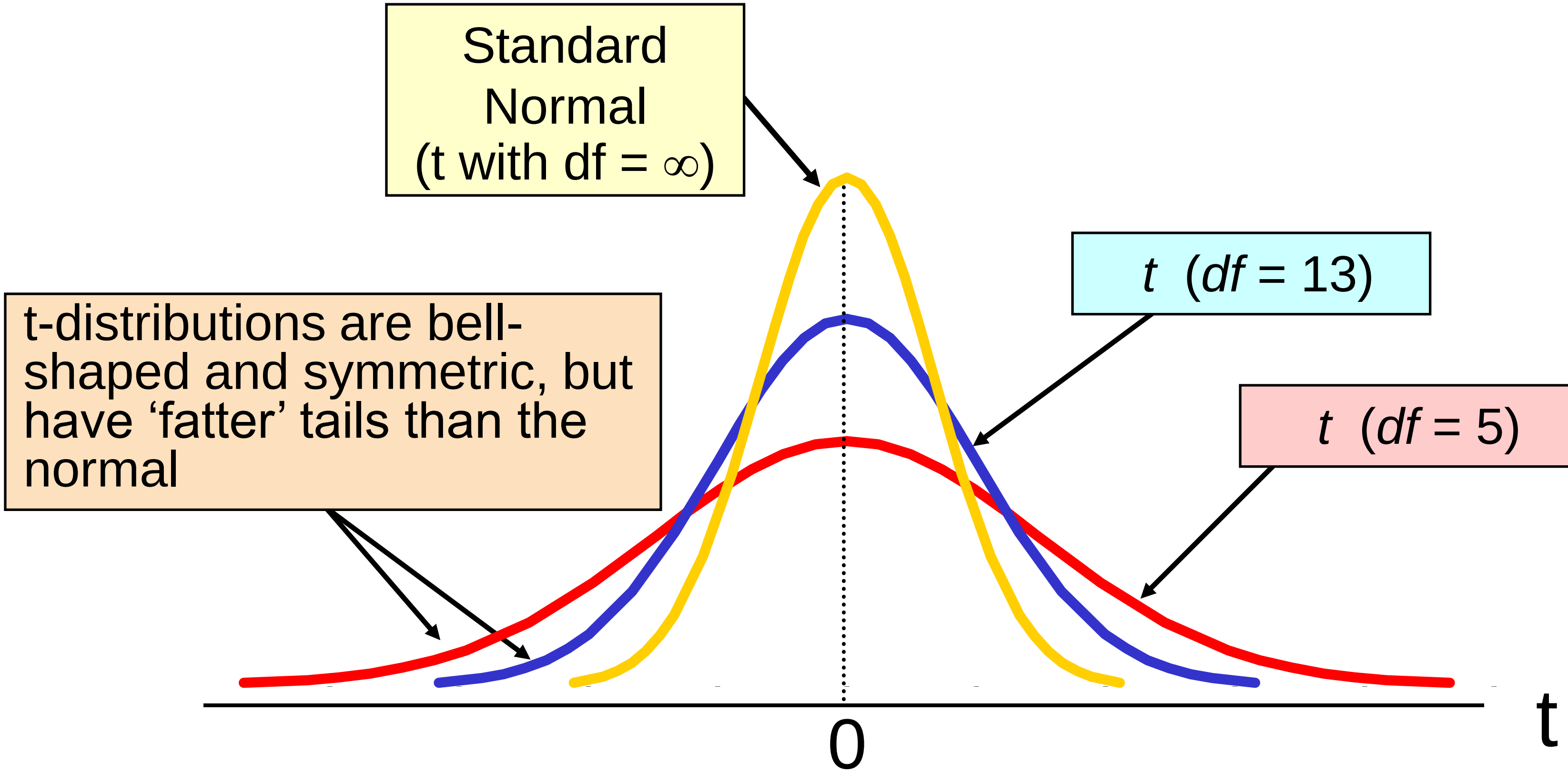
If the mean of these three values is 8.0,
then x_3 must be 9
(i.e., x_3 is not free to vary)

Here, n = 3, so degrees of freedom $= n - 1 = 3 - 1 = 2$

(2 values can be any numbers, but the third is not free to vary for a given mean)

Student’s t Distribution

Note: $t \longrightarrow z$ as n increases

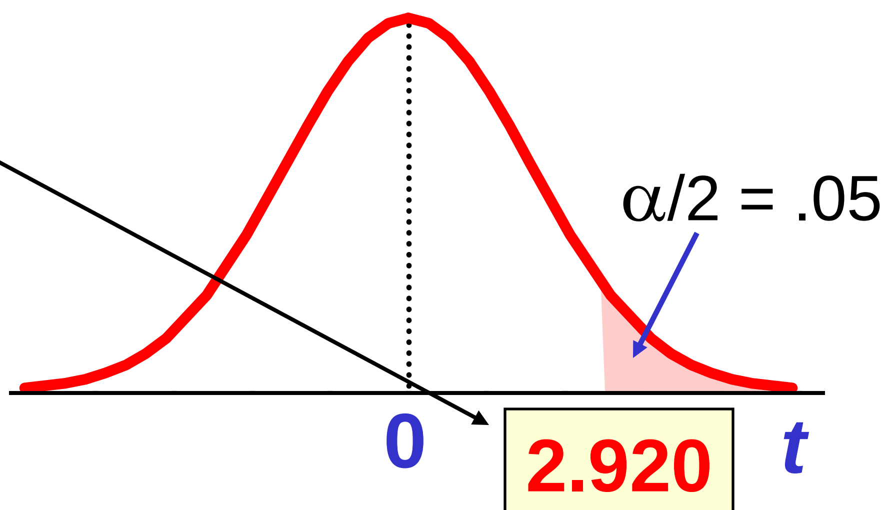


Student's t Table

	Upper Tail Area		
df	.25	.10	**.05**
1	1.000	3.078	6.314
2	0.817	1.886	**2.920**
3	0.765	1.638	2.353

The body of the table contains t values, not probabilities

Let: n = 3
df = n - 1 = 2
α = .10
α/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	1.28
.90	1.812	1.725	1.697	1.64
.95	2.228	2.086	2.042	1.96
.99	3.169	2.845	2.750	2.57

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

The confidence interval is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

46.698 ………………. 53.302

Approximation for Large Samples

- Since t approaches z as the sample size increases, an approximation is <u>sometimes</u> used when $n \geq 30$:

Technically correct	Approximation for large n
$\bar{x} \pm t_{\alpha/2} \dfrac{s}{\sqrt{n}}$	$\bar{x} \pm z_{\alpha/2} \dfrac{s}{\sqrt{n}}$

Determining Sample Size

- The required sample size can be found to reach a desired margin of error (e) and level of confidence (1 - α)
  - Required sample size, σ known:

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

Required Sample Size Example

If σ = 45, what sample size is needed to be 90% confident of being correct within ± 5?

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{e} \right)^2 = \left(\frac{1.645(45)}{5} \right)^2 = 219.19$$

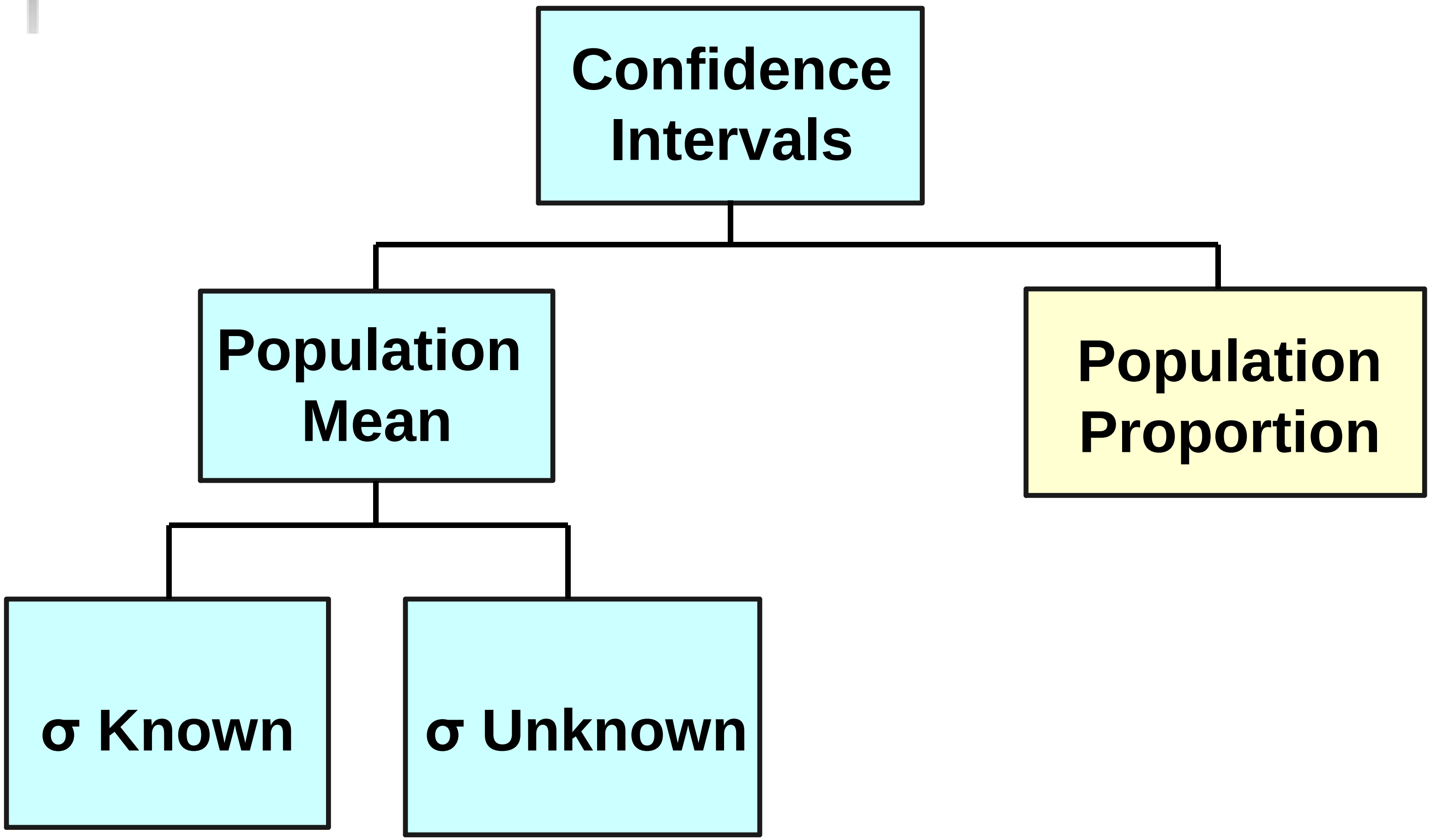
So the required sample size is **n = 220**

(Always round up)

If 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

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 ($\bar{p}$)

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:

$$s_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Confidence interval endpoints

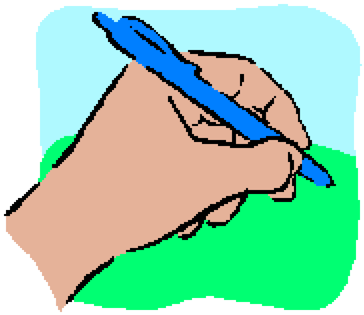
- Upper and lower confidence limits for the population proportion are calculated with the formula

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

- where
  - z is the standard normal value for the level of confidence desired
  - $\bar{p}$ 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.

1. $\bar{p} = 25/100 = .25$

2. $S_{\bar{p}} = \sqrt{\bar{p}(1-\bar{p})/n} = \sqrt{.25(.75)/n} = .0433$

3. $.25 \pm 1.96\,(.0433)$

$$\boxed{0.1651\ \ldots\ldots\ 0.3349}$$

Interpretation

- We are 95% confident that the true percentage of left-handers in the population is between

  16.51% and 33.49%.

- Although this range 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.

Changing the sample size

- Increases in the sample size reduce the width of the confidence interval.

Example:

- If the sample size in the above example is doubled to 200, and if 50 are left-handed in the sample, then the interval is still centered at .25, but the width shrinks to

.19 …… .31

Finding the Required Sample Size for proportion problems

Define the margin of error:

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

Solve for n:

$$n = \frac{z_{\alpha/2}^{2}\, p(1-p)}{e^2}$$

p can be estimated with a pilot sample, if necessary (or conservatively use p = .50)

What sample size...?

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

(Assume a pilot sample yields $\bar{p} = .12$)

What sample size...?

(continued)

Solution:

For 95% confidence, use Z = 1.96

E = .03

$\bar{p}$ = .12, so use this to estimate p

$$n = \frac{z_{\alpha/2}^2\, p(1-p)}{e^2} = \frac{(1.96)^2(.12)(1-.12)}{(.03)^2} = 450.74$$

So use n = 451

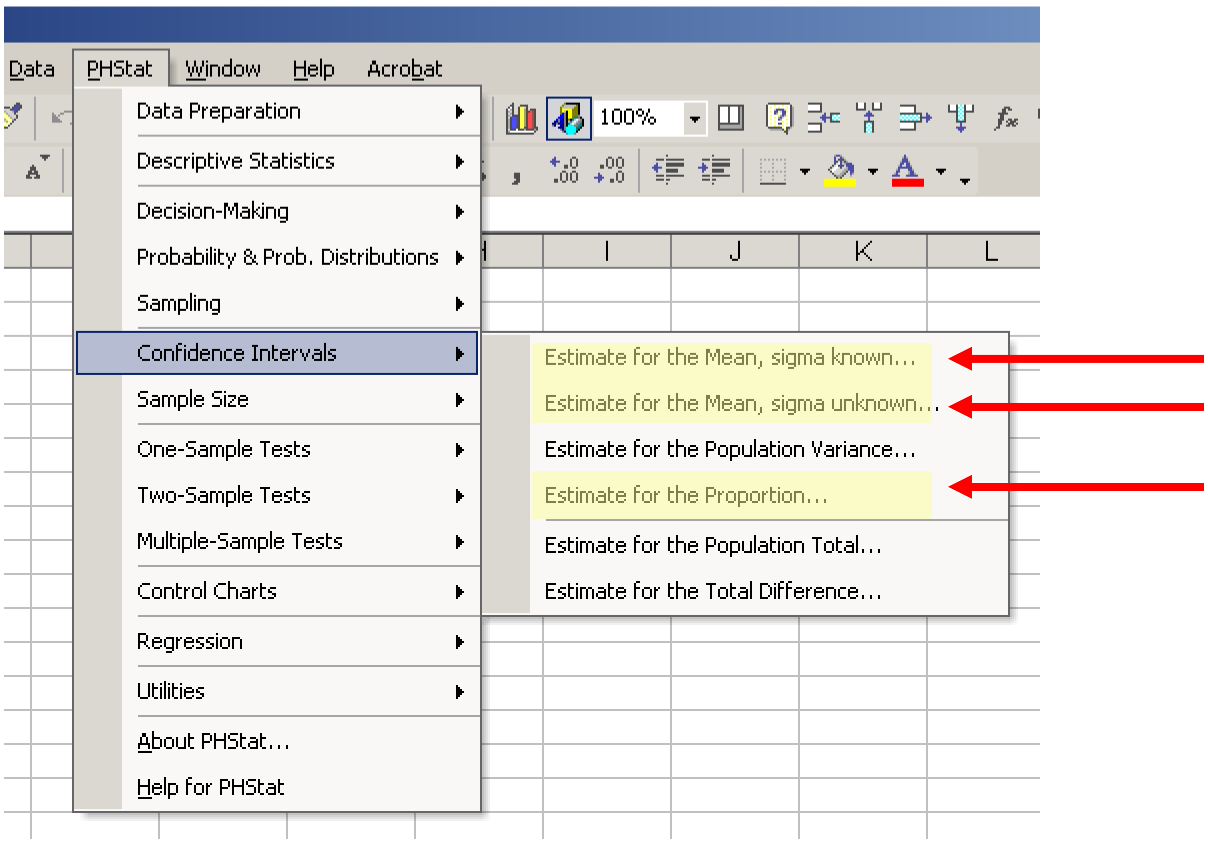
Using PHStat

- PHStat can be used for confidence intervals for the mean or proportion

- two options for the mean: known and unknown population standard deviation

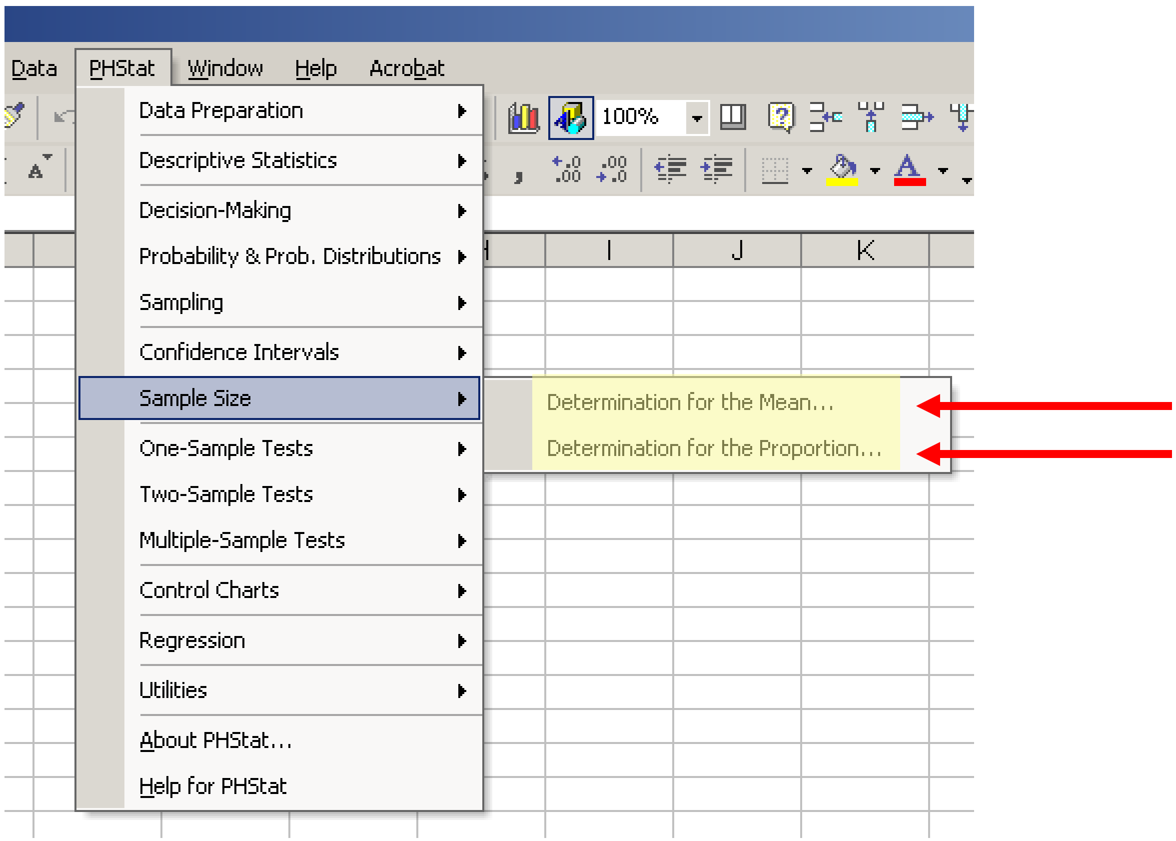
- required sample size can also be found

PHStat Interval Options



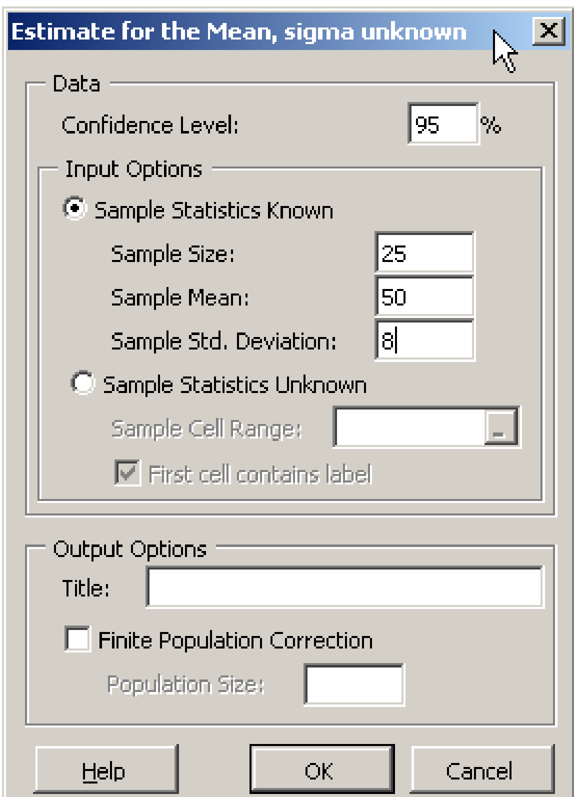
options

PHStat Sample Size Options



Using PHStat (for μ, σ unknown)

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

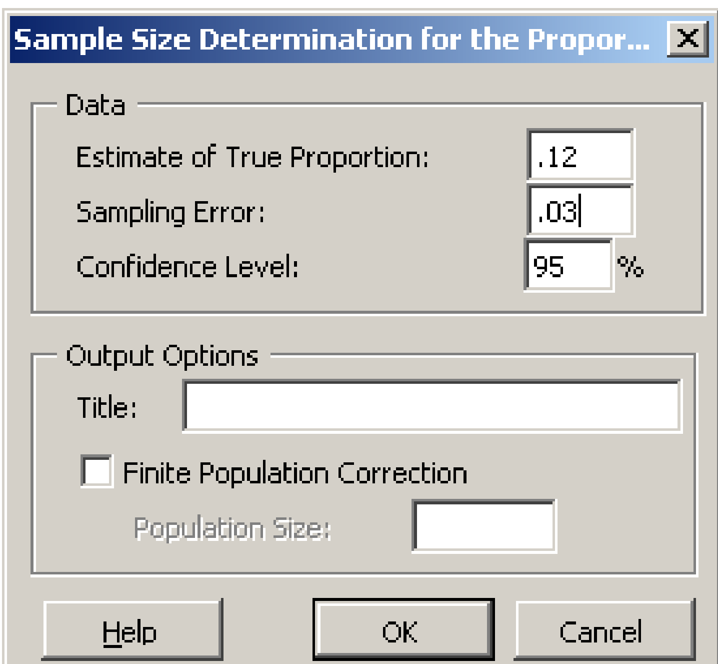


	A	B
1	**Confidence Interval Estimate for the Mean**	
2		
3	**Data**	
4	**Sample Standard Deviation**	**8**
5	**Sample Mean**	**50**
6	**Sample Size**	**25**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	1.6
11	Degrees of Freedom	24
12	*t* Value	2.063898137
13	Interval Half Width	3.302237019
14		
15	**Confidence Interval**	
16	**Interval Lower Limit**	**46.70**
17	**Interval Upper Limit**	**53.30**

Using PHStat (sample size for proportion)

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

(Assume a pilot sample yields p = .12)



	A	B
1	**Sample Size Determination**	
2		
3	**Data**	
4	**Estimate of True Proportion**	**0.12**
5	**Sampling Error**	**0.03**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Z Value	-1.95996279
10	Calculated Sample Size	450.7306177
11		
12	**Result**	
13	**Sample Size Needed**	**451**

Chapter Summary

- Illustrated estimation process
- Discussed point estimates
- Introduced interval estimates
- Discussed confidence interval estimation for the mean (σ known)
- Addressed determining sample size
- Discussed confidence interval estimation for the mean (σ unknown)
- Discussed confidence interval estimation for the proportion