Statistics for Managers Using Microsoft® Excel 5th Edition

Chapter 6 (Textbook Ch8)

Confidence Interval Estimation

Chapter Goals

After completing this chapter, you should understand:

- Point estimate and a confidence interval estimate
- Confidence interval estimate for a single population mean
- Confidence interval estimate for a single population proportion
- Determining the required sample size to estimate a mean or proportion within a specified margin of error

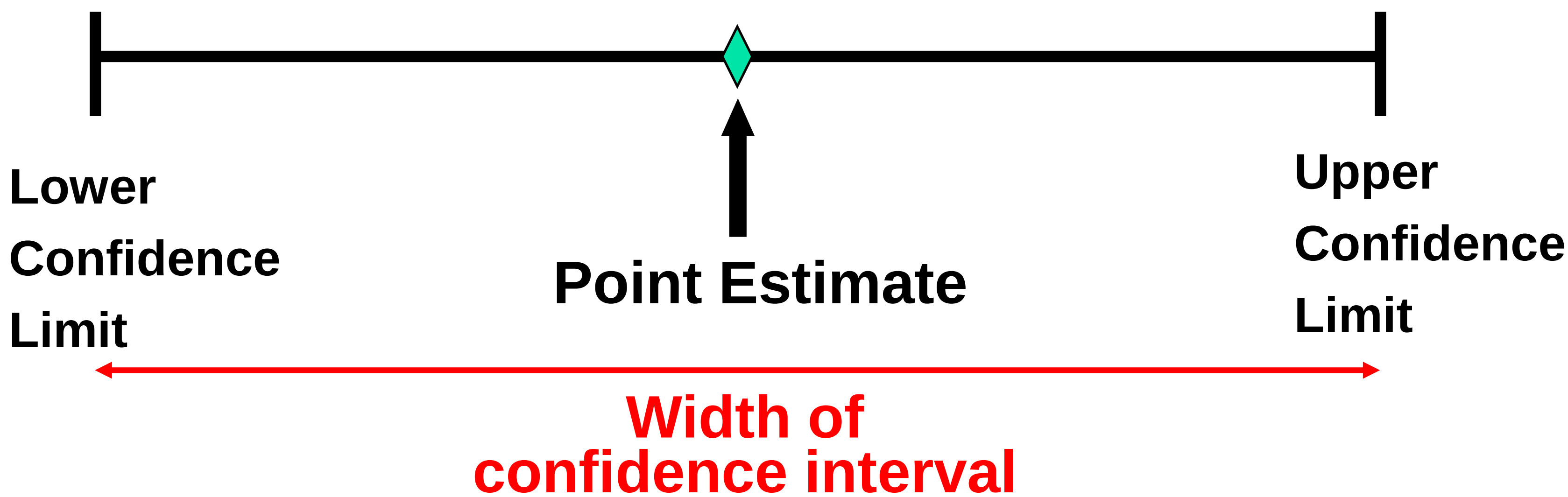
Confidence Intervals

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
- Confidence Intervals for the Population Proportion, p
- Determining the Required Sample Size

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	p_s

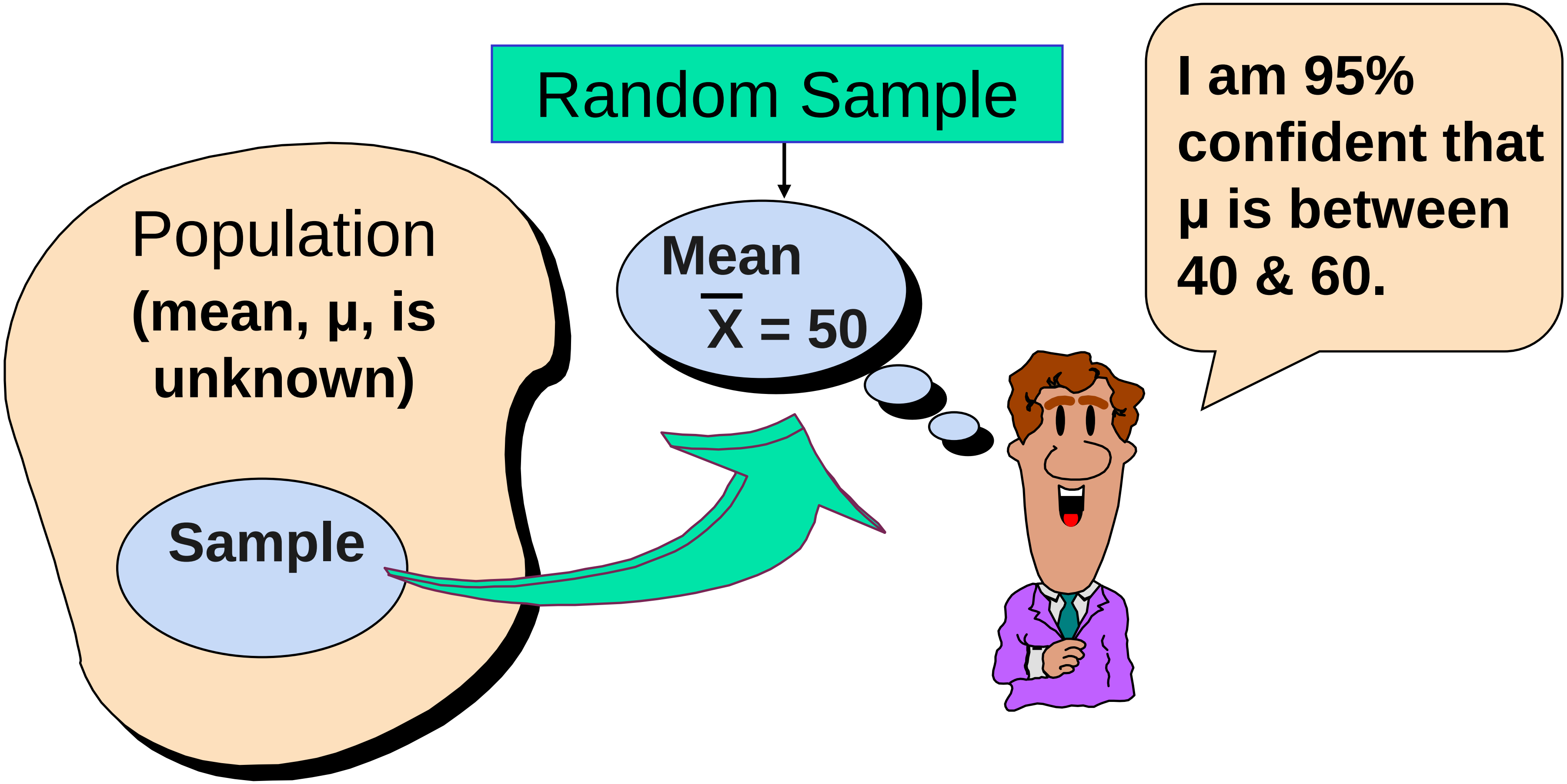
Confidence Intervals

- 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

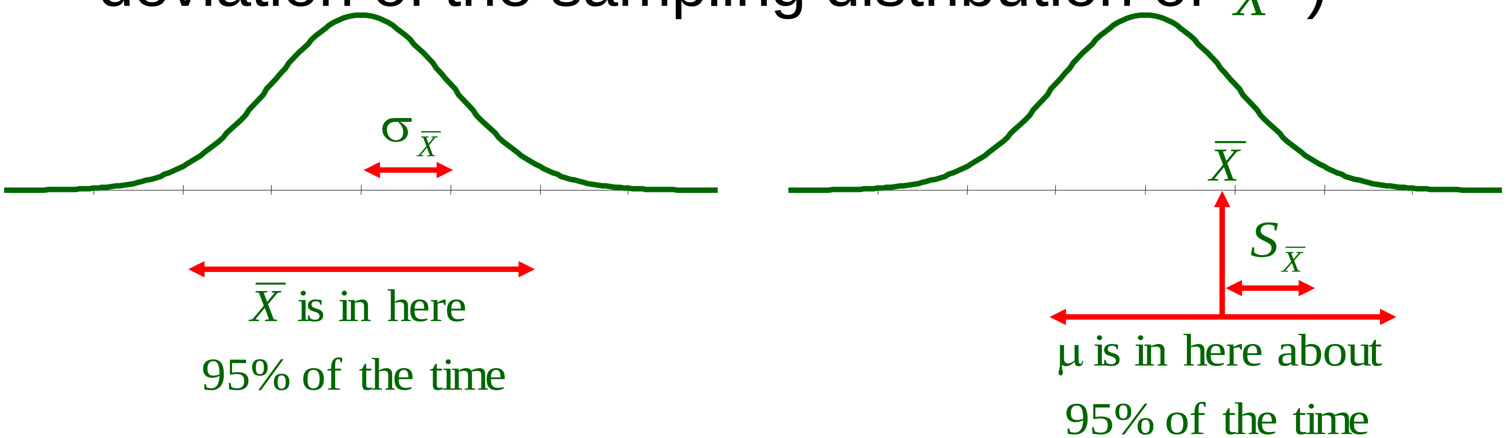
- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observation from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - Can never be 100% confident

Estimation Process



Why does it work?

- $\bar{X}$ is within $2S_{\bar{X}}$ of its mean μ about 95% of the time
  - Because $S_{\bar{X}}$ is an estimator of $\sigma_{\bar{X}}$ (the standard deviation of the sampling distribution of $\bar{X}$)



- This also says that μ is within $2S_{\bar{X}}$ of $\bar{X}$ about 95% of the time

Why does it work?

- $P(\mu - 2\sigma_{\bar{X}} < \bar{X} < \mu + 2\sigma_{\bar{X}}) \approx 0.95 \Leftrightarrow$

  $P(-2\sigma_{\bar{X}} < \bar{X} - \mu < +2\sigma_{\bar{X}}) \approx 0.95 \Leftrightarrow$

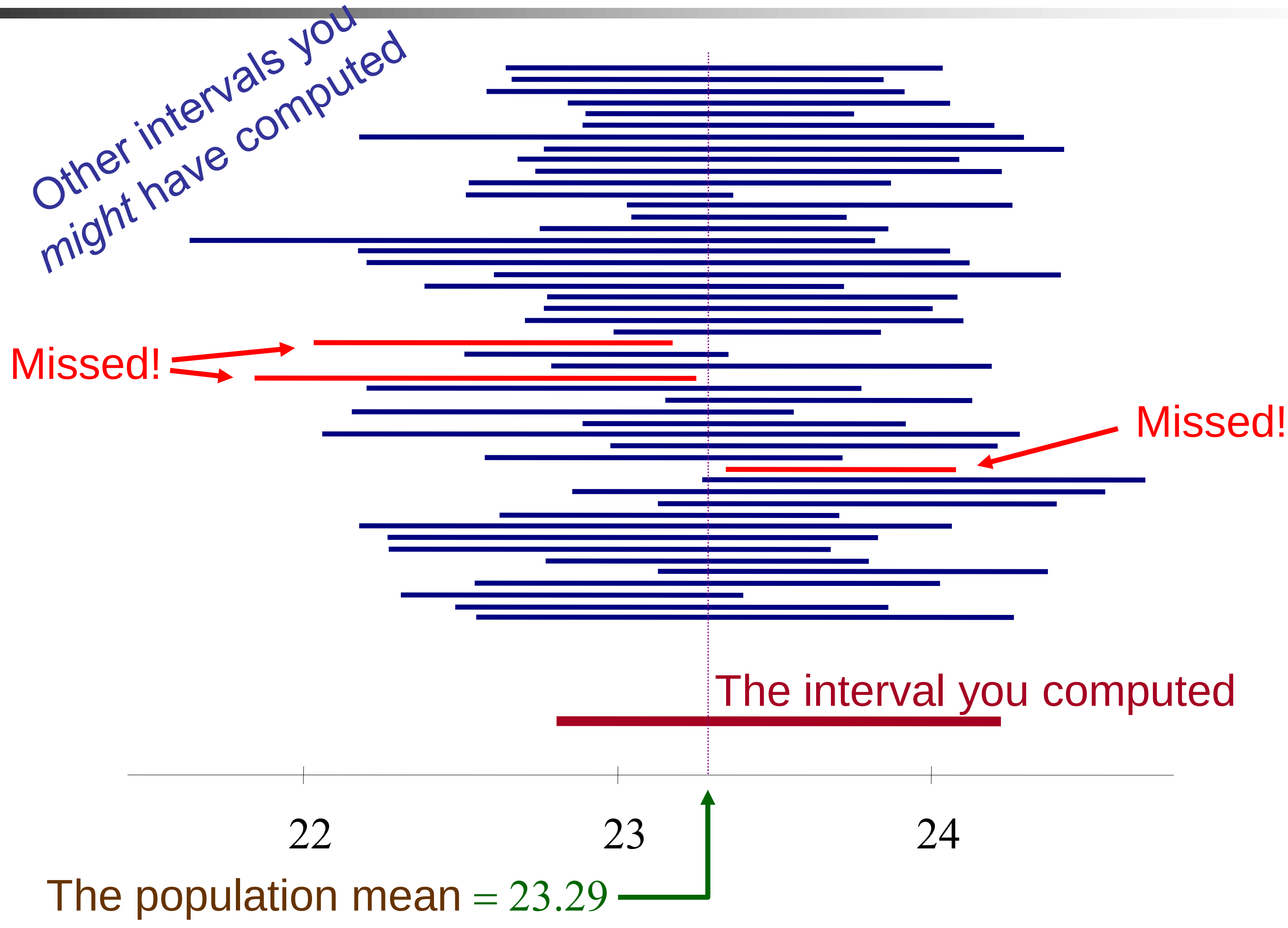
  $P(-\bar{X} - 2\sigma_{\bar{X}} < -\mu < -\bar{X} + 2\sigma_{\bar{X}}) \approx 0.95 \Leftrightarrow$

  $P(\bar{X} - 2\sigma_{\bar{X}} < \mu < \bar{X} + 2\sigma_{\bar{X}}) \approx 0.95 \Leftrightarrow$

Estimate $\sigma_{\bar{X}}$ by $S_{\bar{X}}$, then

$$P(\bar{X} - 2S_{\bar{X}} < \mu < \bar{X} + 2S_{\bar{X}}) \approx 0.95.$$

Imagine Many Samples



General Formula

- The general formula for all confidence intervals is:

Point Estimate ± (Critical Value)(Standard Error)

Confidence Level

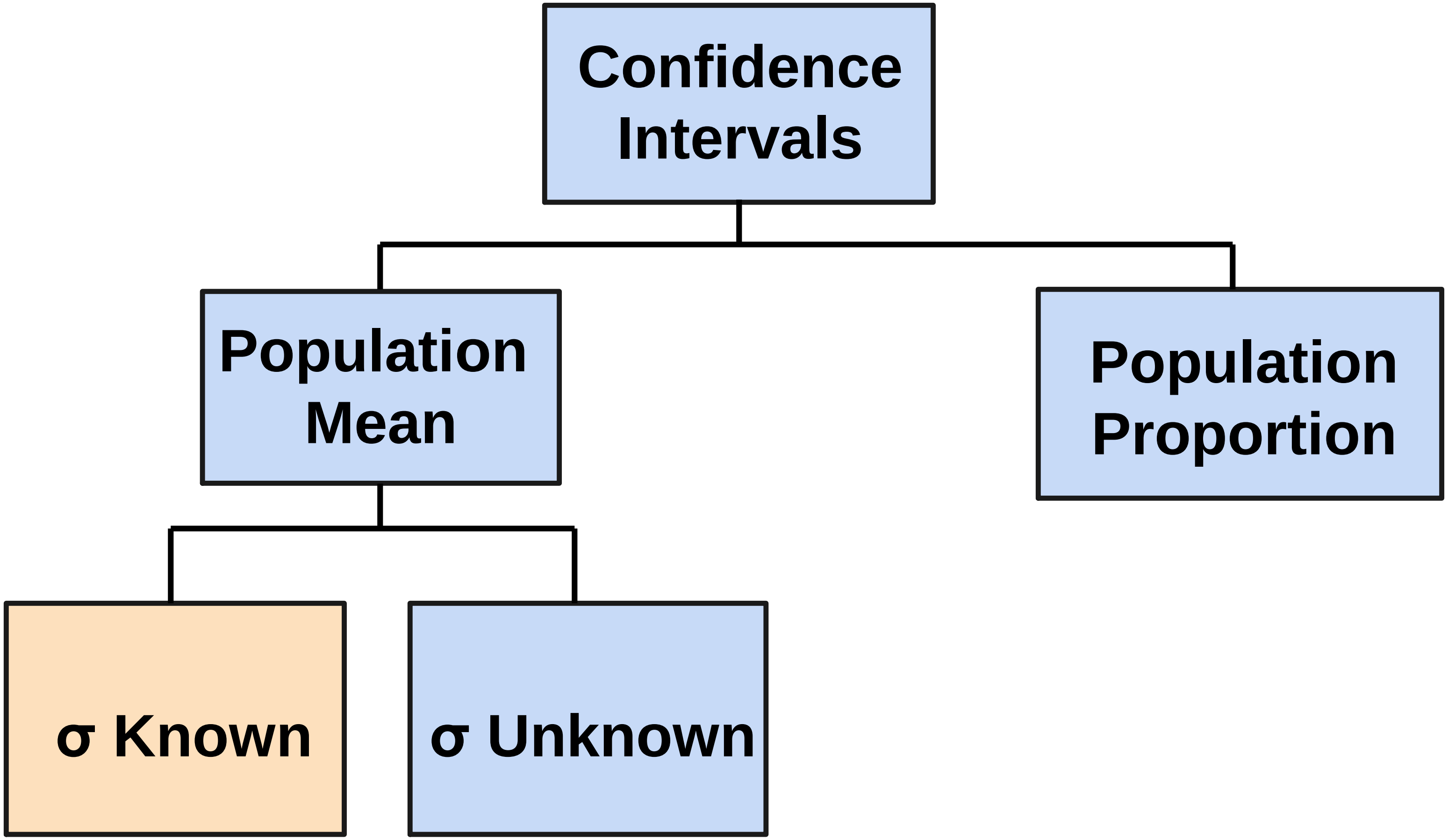
- Confidence Level
  - Confidence in which the interval will contain the unknown population parameter
- A percentage (less than 100%)

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)

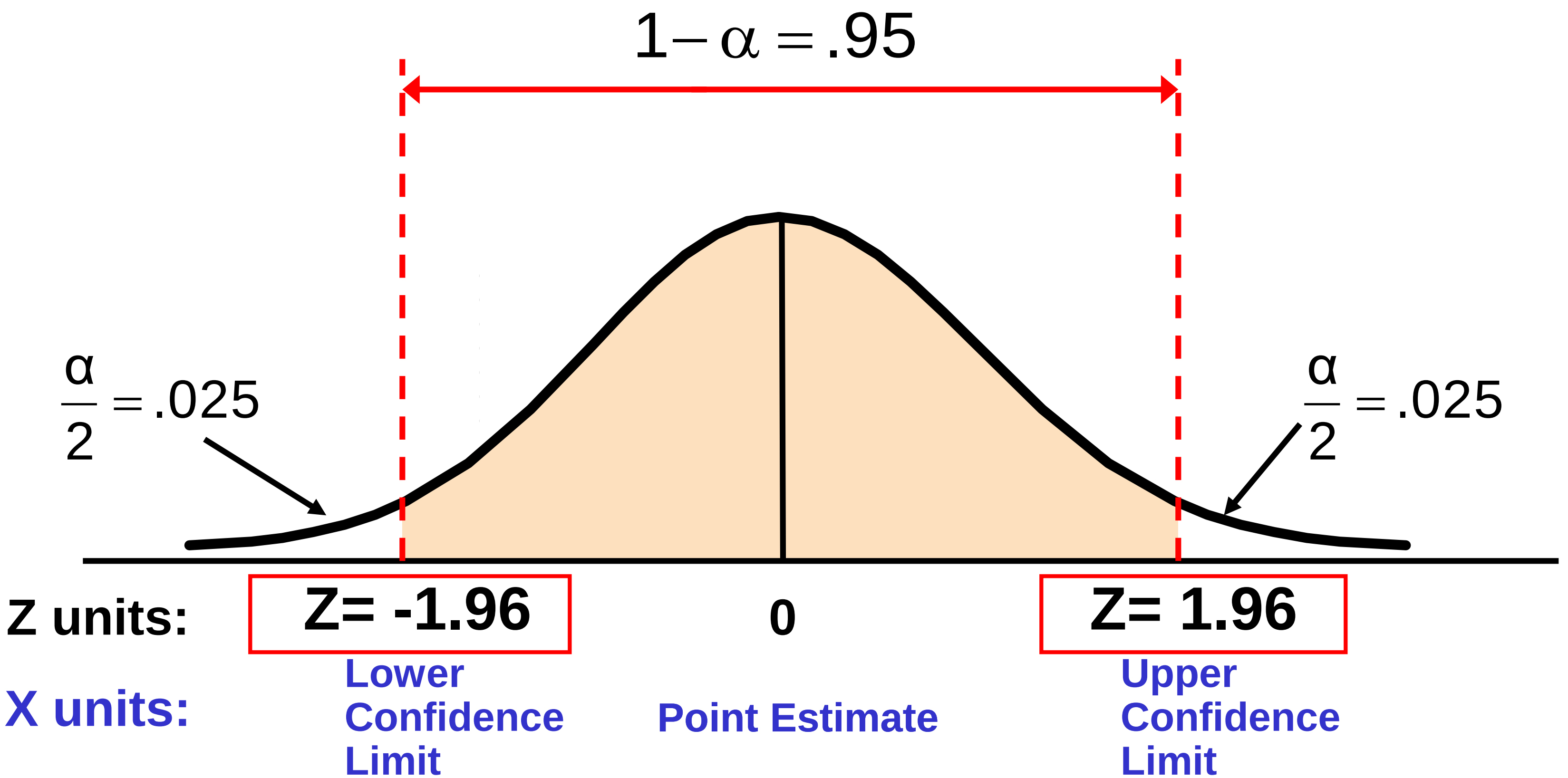
- Assumptions
  - Population standard deviation σ is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

(where Z is the normal distribution critical value for a probability of α/2 in each tail)

Finding the Critical Value, Z

- Consider a 95% confidence interval: $Z = \pm 1.96$



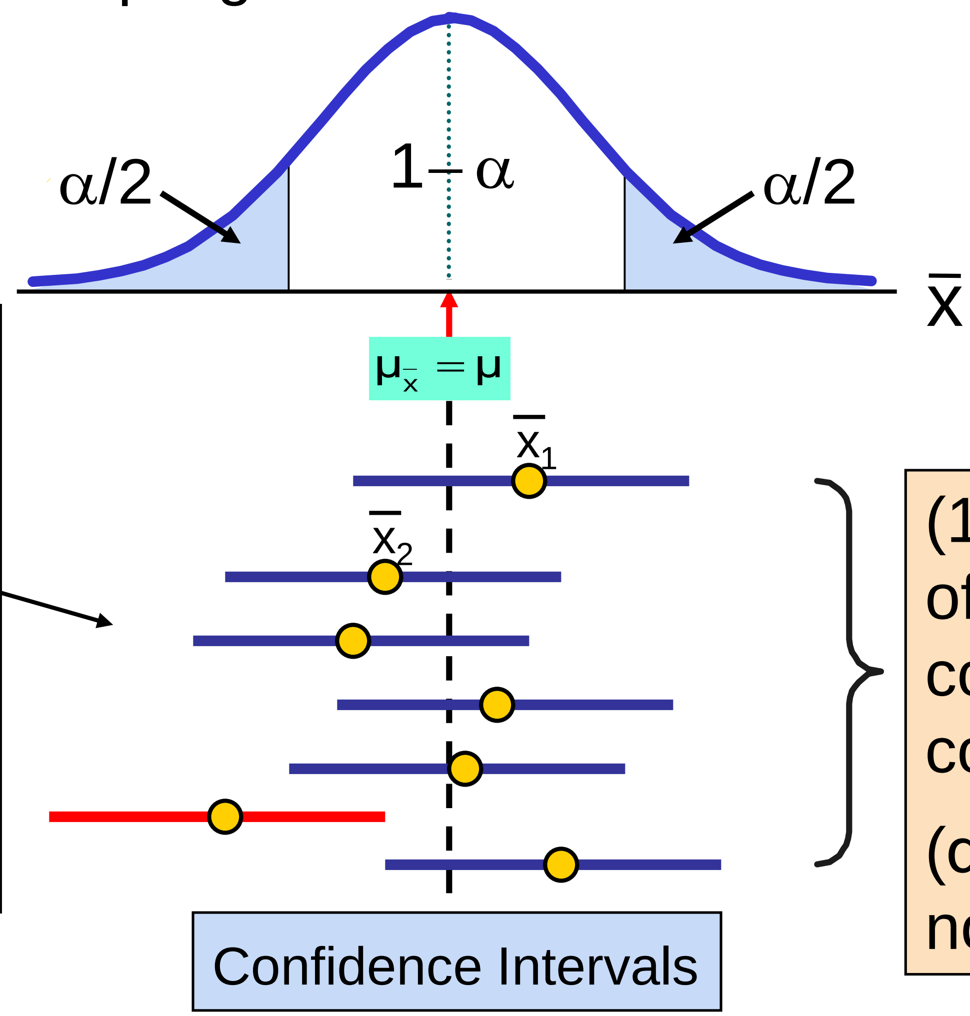
Common Levels of Confidence

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

Confidence Level	***Confidence Coefficient,*** $1-\alpha$	***Z 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



Intervals extend from

$$\bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

to

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}}$$

(1-α)x100% of intervals constructed contain μ;

(α)x100% 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 .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:

$$\overline{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

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

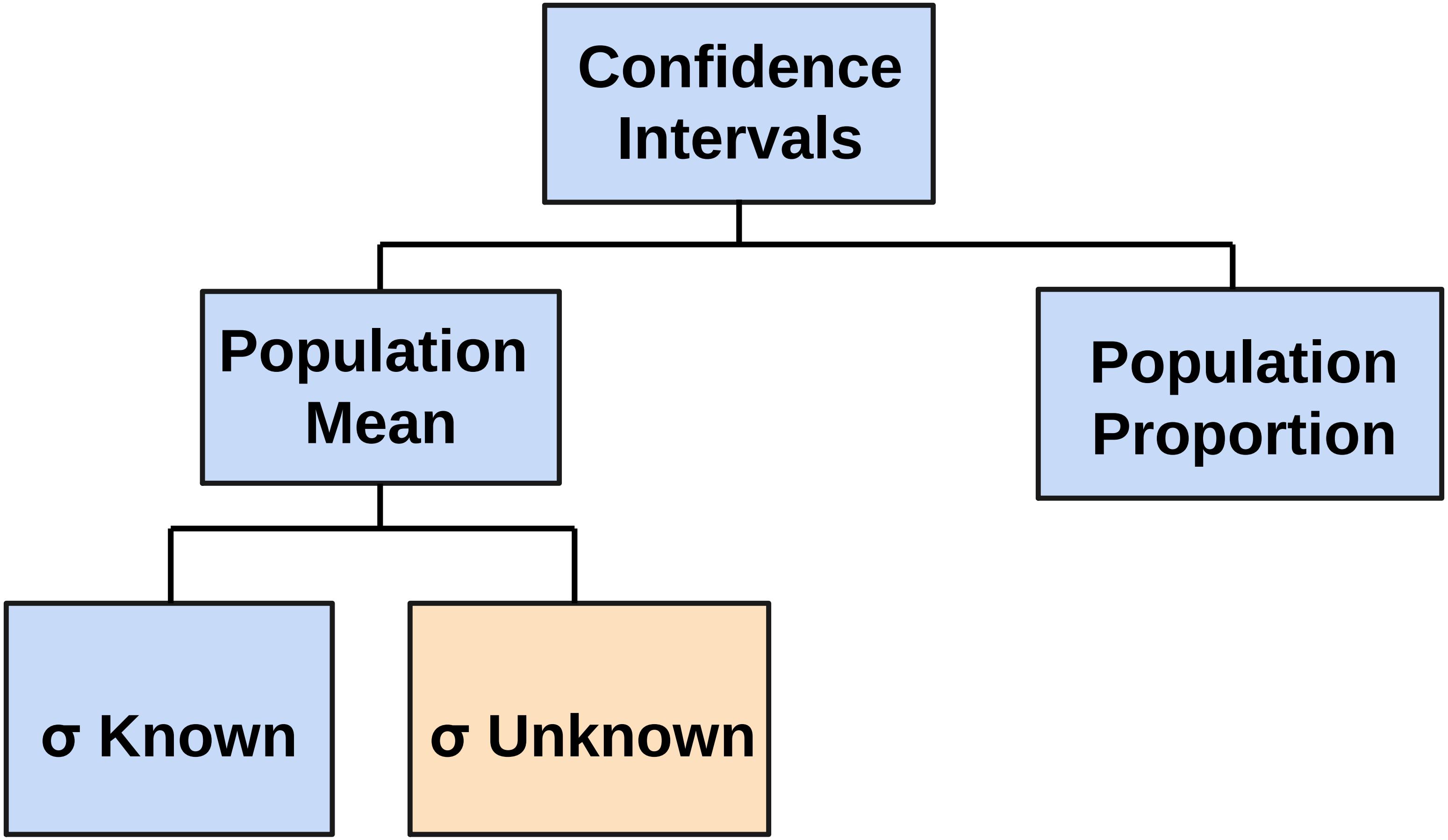
$$= 2.20 \pm .2068$$

$$\boxed{(1.9932\ ,\ 2.4068)}$$

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



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:

$$\boxed{\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}}}$$

(where t is the critical value of the t distribution with n-1 d.f. and an area of α/2 in each tail)

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

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

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?

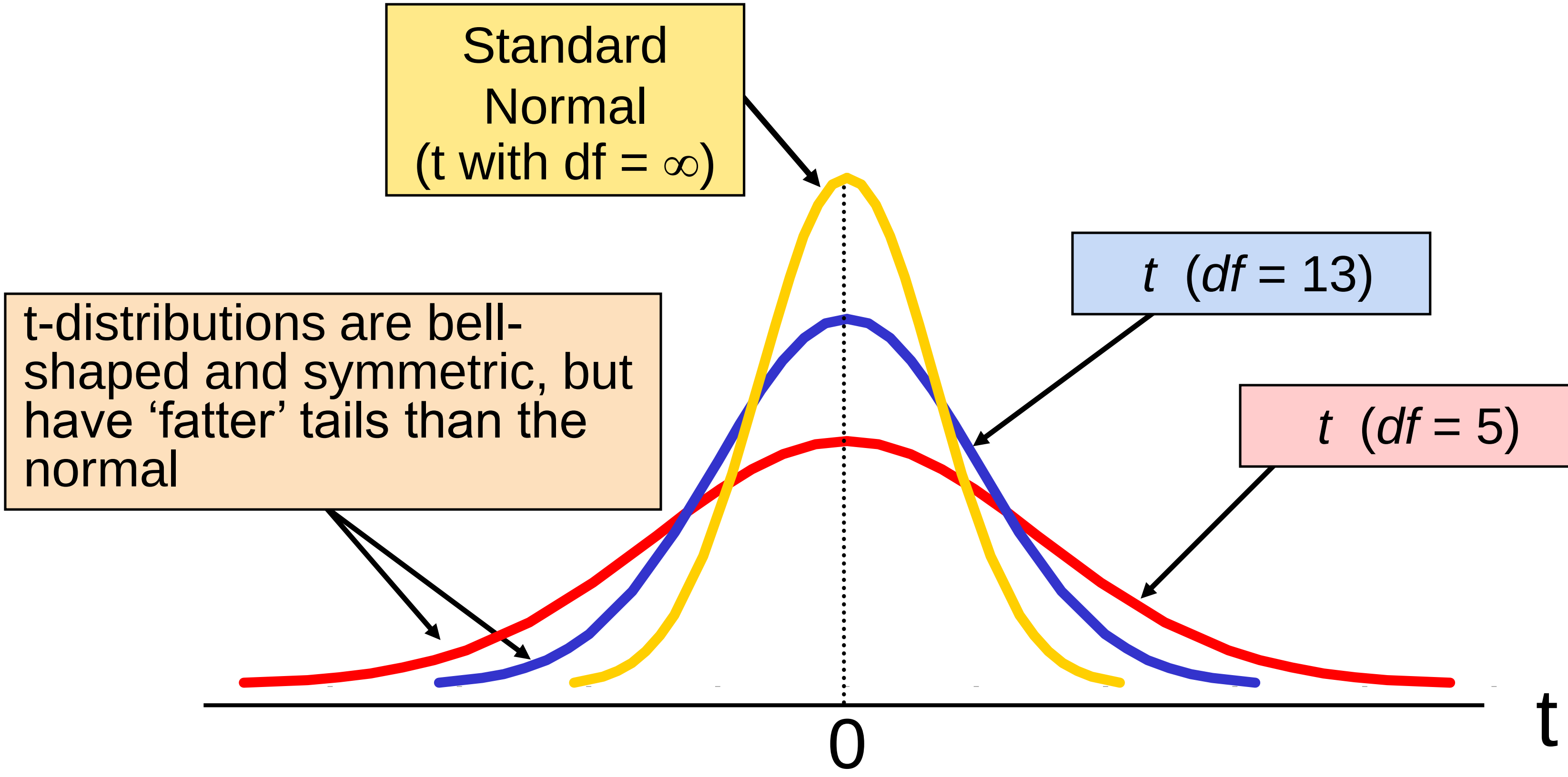
If the mean of these three values is 8.0,
then X_3 must be 9
(i.e., X_3 is not free to vary)

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

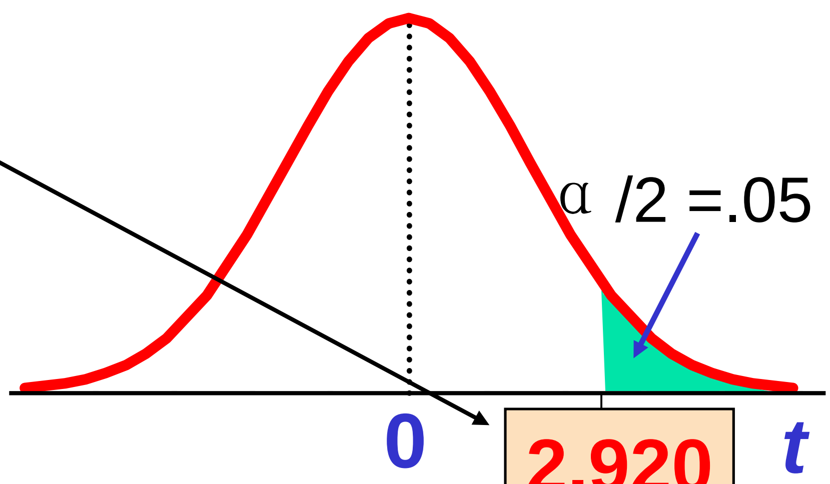


Student’s t Table

	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

Let: n = 3
df = *n* - 1 = 2
α = .10
α /2 =.05

The body of the table contains t values, not probabilities



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	2.58

Note: $t \rightarrow Z$ as n increases

Example

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

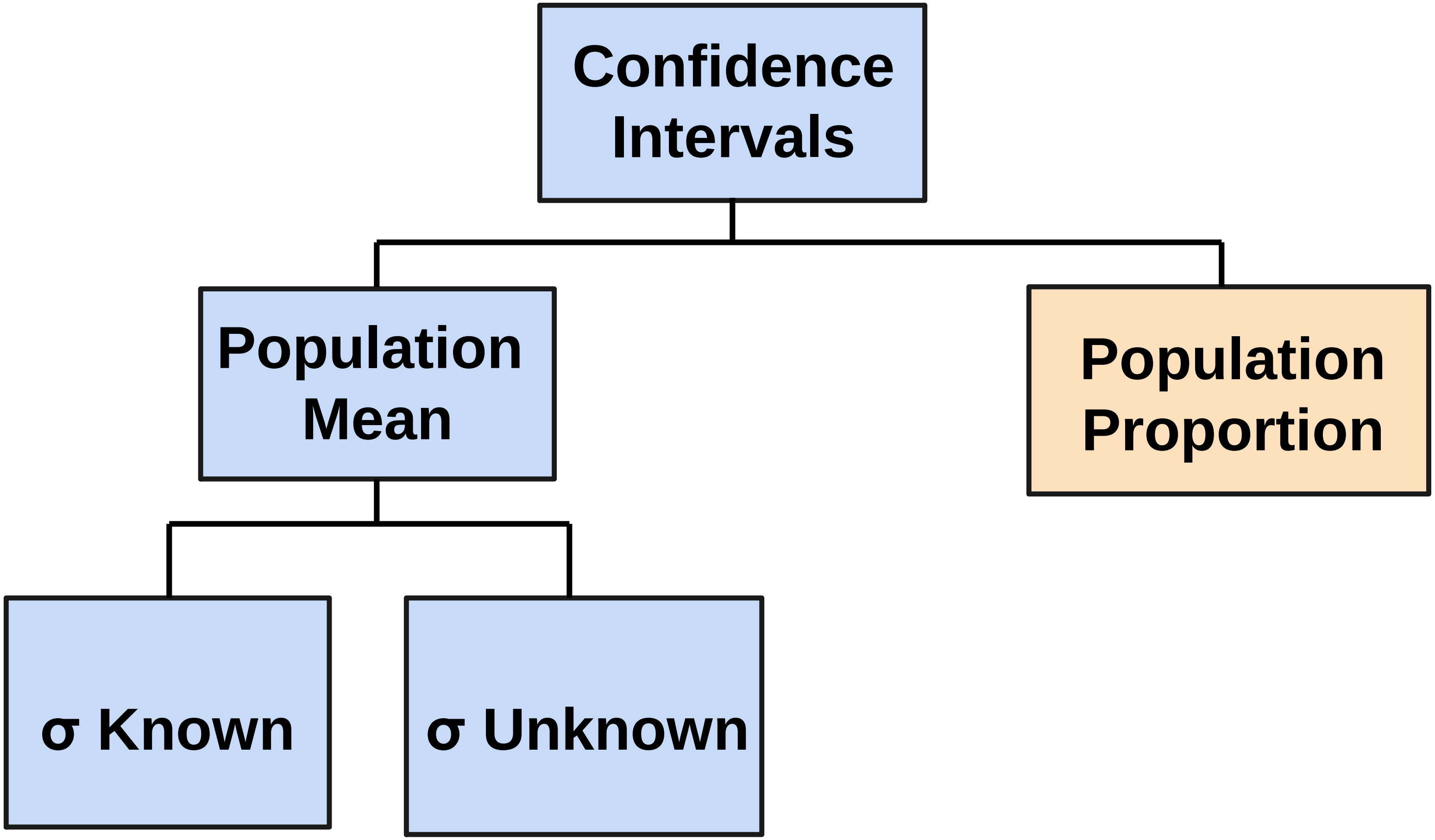
- d.f. = n – 1 = 24, so $t_{\alpha/2\,,\,n-1} = t_{.025,24} = 2.0639$

The confidence interval is

$$\bar{X} \pm t_{\alpha/2,\,n\text{-}1} \frac{S}{\sqrt{n}} \;=\; 50 \pm (2.0639)\frac{8}{\sqrt{25}}$$

(46.698 , 53.302)

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 (p_s)

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

$$\boxed{\sigma_p = \sqrt{\frac{p(1-p)}{n}}}$$

- We will estimate this with sample data:

$$\boxed{\sqrt{\frac{p_s(1-p_s)}{n}}}$$

Confidence Interval Endpoints

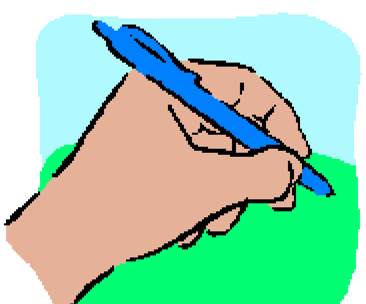
- Upper and lower confidence limits for the population proportion are calculated with the formula

$$p_s \pm Z\sqrt{\frac{p_s(1-p_s)}{n}}$$

- where
  - Z is the standard normal value for the level of confidence desired
  - p_s 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.

$$p_s \pm Z\sqrt{p_s(1-p_s)/n}$$

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

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

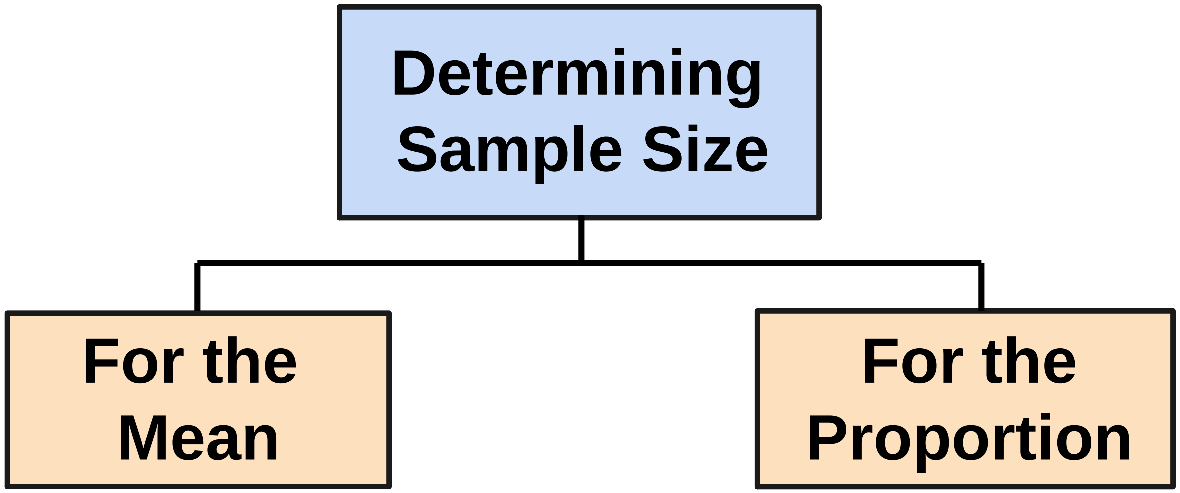
$$\boxed{(0.1651\,,\ 0.3349)}$$

Interpretation

- We are 95% confident that the true percentage of left-handers in the population is between 16.51% and 33.49%.

- Although the interval from .1651 to .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.

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 - α)

- The margin of error is also called sampling error
  - the amount of imprecision in the estimate of the population parameter
  - the amount added and subtracted to the point estimate to form the confidence interval

Determining Sample Size

Determining Sample Size

For the Mean

Sampling error (margin of error)

$$\bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$$

$$e = Z\frac{\sigma}{\sqrt{n}}$$

Determining Sample Size

(continued)

Determining Sample Size

For the Mean

$$e = Z\frac{\sigma}{\sqrt{n}}$$

Now solve for n to get

$$\boxed{n = \frac{Z^2\,\sigma^2}{e^2}}$$

Determining Sample Size

(continued)

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence ($1 - \alpha$), which determines the critical Z value
  - The acceptable sampling error (margin of error), e
  - The standard deviation, σ

Required Sample Size Example

If σ = 45, what sample size is needed to estimate the mean within $\pm$ 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)

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

Determining Sample Size

Determining Sample Size

For the Proportion

$$p_s \pm Z\sqrt{\frac{p_s(1-p_s)}{n}}$$

$$e = Z\sqrt{\frac{p(1-p)}{n}}$$

Sampling error (margin of error)

Determining Sample Size

(continued)

Determining Sample Size

For the Proportion

$$e = Z\sqrt{\frac{p(1-p)}{n}}$$

Now solve for n to get

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

Determining Sample Size

(continued)

- To determine the required sample size for the proportion, you must know:
  - The desired level of confidence (1 - α), which determines the critical Z value
  - The acceptable sampling error (margin of error), e
  - The true proportion of "successes", p
    - p can be estimated with a pilot sample, if necessary (or conservatively use p = .50)

Required Sample Size Example

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 $p_s = .12$)

Required Sample Size Example

(continued)

Solution:

For 95% confidence, use $Z = 1.96$

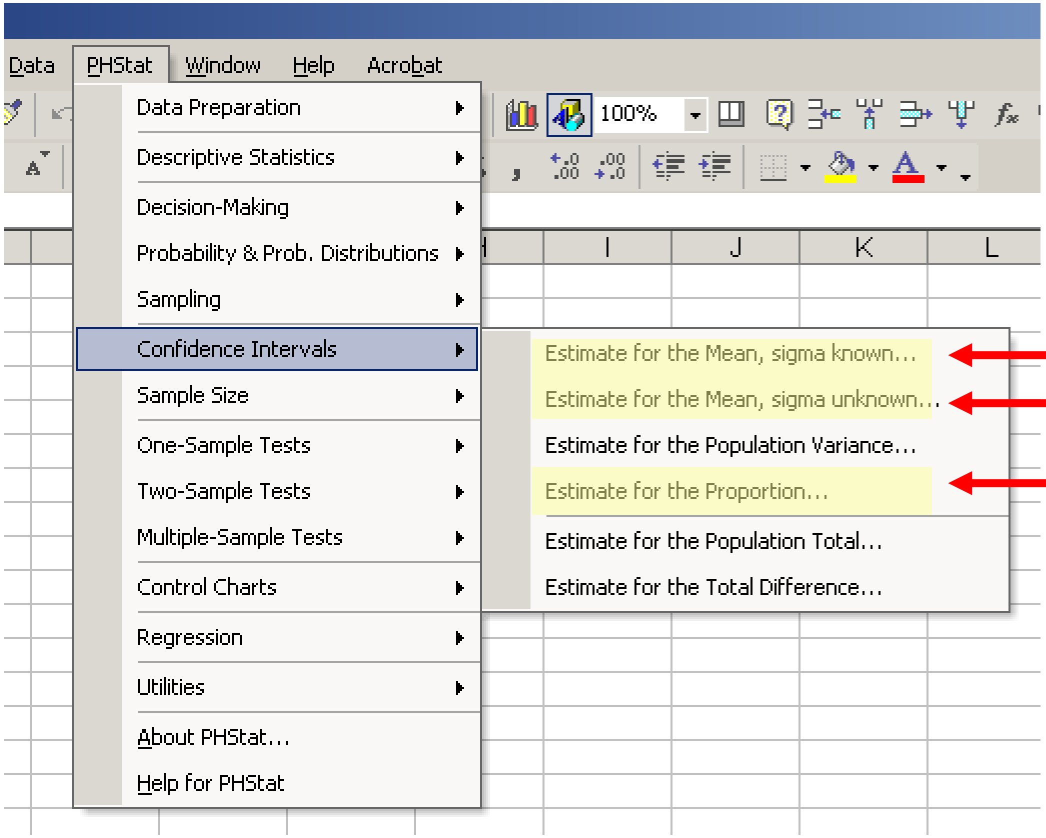
$e = .03$

$p_s = .12$, so use this to estimate p

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

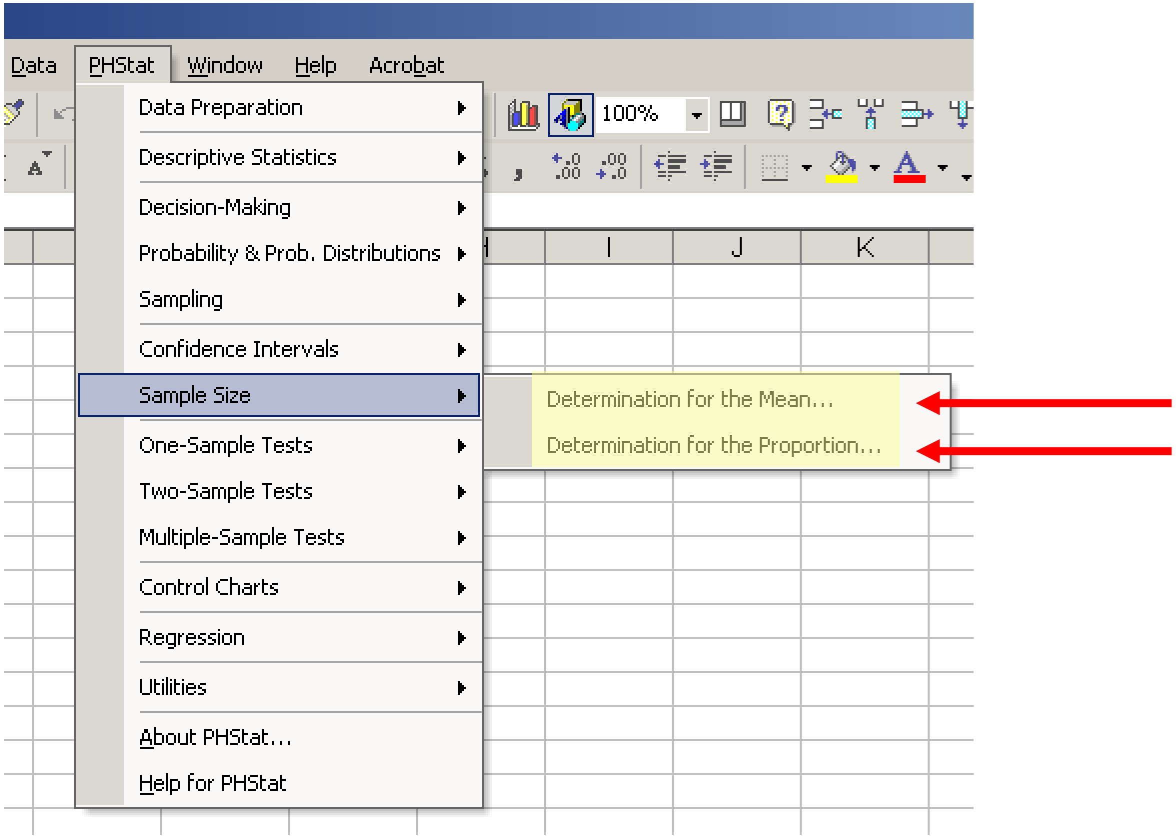
So use $n = 451$

PHStat Interval Options



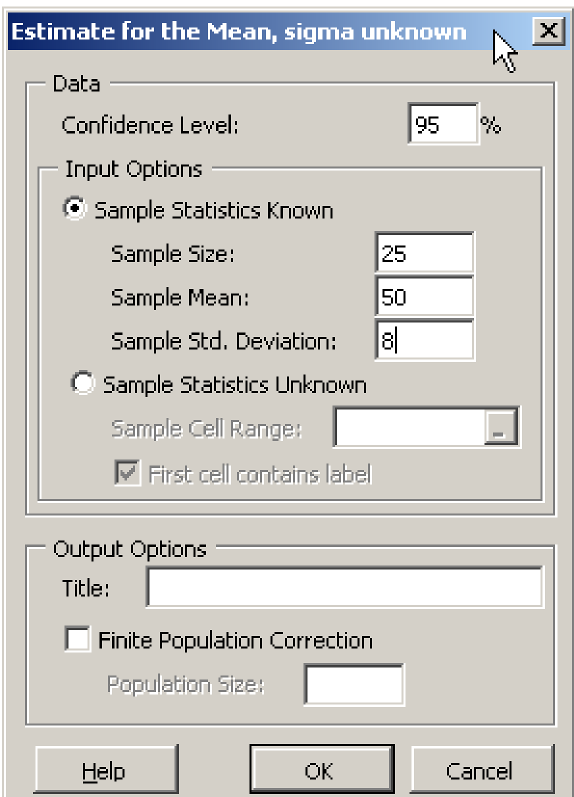
options

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

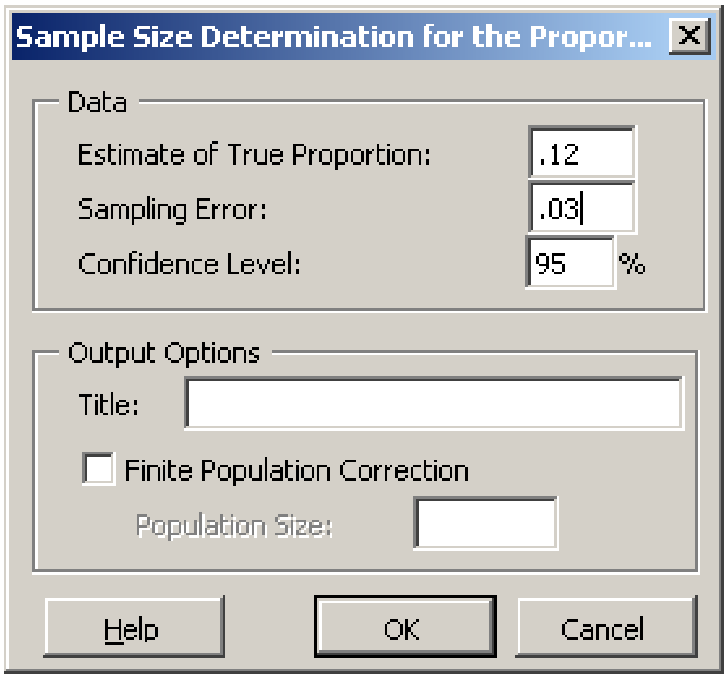


	A	B
1	**Confidence Interval Estimate for the Mean**	
2		
3	**Data**	
4	**Sample Standard Deviation**	**8**
5	**Sample Mean**	**50**
6	**Sample Size**	**25**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	1.6
11	Degrees of Freedom	24
12	*t* Value	2.063898137
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 $p_s = .12$)



	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**

Confidence Interval for Population Total Amount

- Point estimate:

$$\boxed{\text{Population total} = N\bar{X}}$$

- Confidence interval estimate:

$$\boxed{N\bar{X} \pm N(t_{n-1})\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

(This is sampling without replacement, so use the finite population correction in the confidence interval formula)

Confidence Interval for Population Total: Example

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

A sample of 80 accounts is selected with average balance of $87.6 and standard deviation of $22.3.

Find the 95% confidence interval estimate of the total balance.

Example Solution

$$N = 1000, \quad n = 80, \quad \bar{X} = 87.6, \quad S = 22.3$$

$$N\bar{X} \pm N(t_{n-1})\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$= (1000)(87.6) \pm (1000)(1.9905)\frac{22.3}{\sqrt{80}}\sqrt{\frac{1000-80}{1000-1}}$$

$$= 87{,}600 \pm 4{,}762.48$$

The 95% confidence interval for the population total balance is \$82,837.52 to \$92,362.48

Confidence Interval for Total Difference

- Point estimate:

$$\text{Total Difference} = N\bar{D}$$

- Where the average difference, $\bar{D}$, is:

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$$

$$\text{where } D_i = \text{audited value - original value}$$

Confidence Interval for Total Difference

(continued)

- Confidence interval estimate:

$$N\bar{D} \pm N(t_{n-1})\frac{S_D}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

where

$$S_D = \frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}$$

One Sided Confidence Intervals

- Application: find the upper bound for the proportion of items that do not conform with internal controls

$$\text{Upper bound} = p_s + Z\sqrt{\frac{p_s(1-p_s)}{n}}\sqrt{\frac{N-n}{N-1}}$$

- where
  - Z is the standard normal value for the level of confidence desired
  - p_s is the sample proportion of items that do not conform
  - n is the sample size
  - N is the population size