Statistics for Business and Economics

7th Edition

Chapter 7

Estimation: Single Population

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
- Create confidence interval estimates for the variance of a normal population

Confidence Intervals

Contents of this chapter:

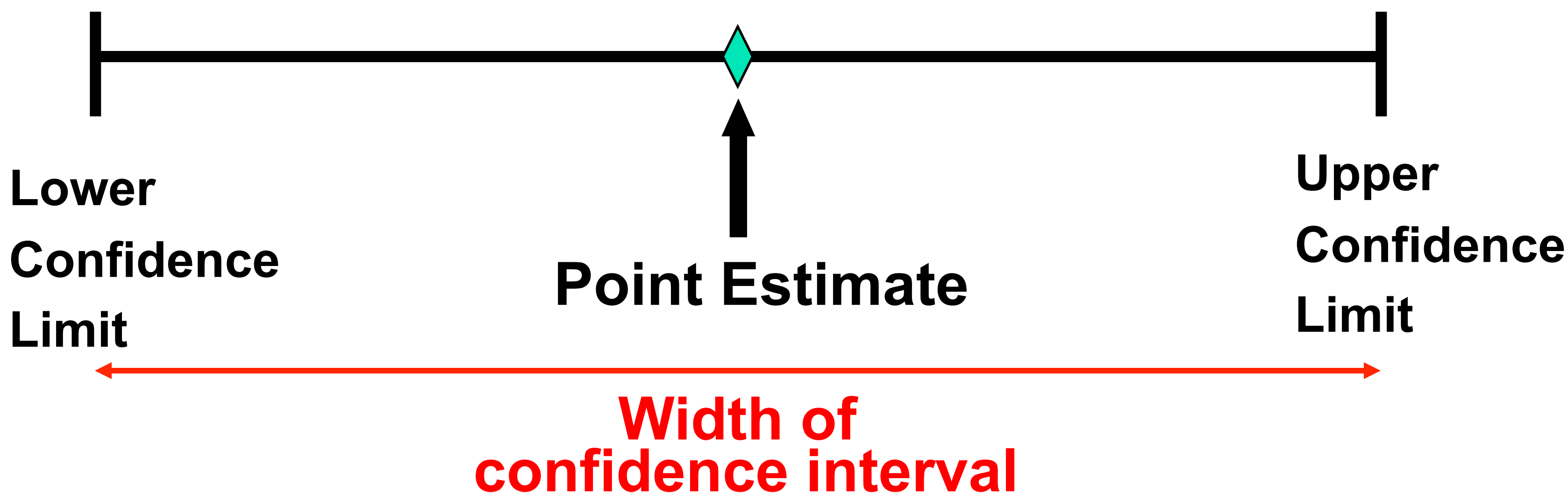
- Confidence Intervals for the Population Mean, μ
  - when Population Variance σ^2 is Known
  - when Population Variance σ^2 is Unknown
- Confidence Intervals for the Population Proportion, $\hat{p}$ (large samples)
- Confidence interval estimates for the variance of a normal population

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	$\hat{p}$

Unbiasedness

- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if the expected value, or mean, of the sampling distribution of $\hat{\theta}$ is θ,

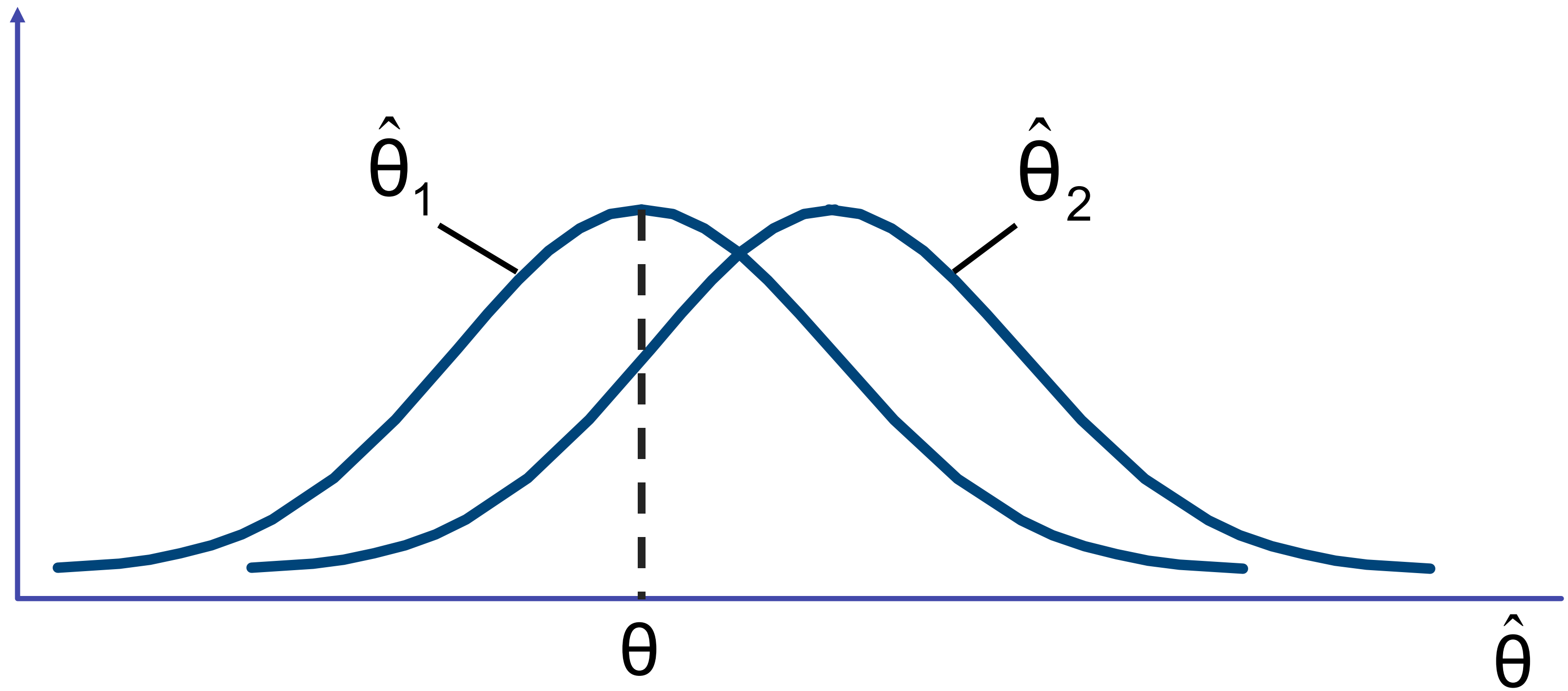
$$E(\hat{\theta}) = \theta$$

- Examples:
  - The sample mean $\bar{x}$ is an unbiased estimator of μ
  - The sample variance s^2 is an unbiased estimator of σ^2
  - The sample proportion $\hat{p}$ is an unbiased estimator of P

Unbiasedness

(continued)

- $\hat{\theta}_1$ is an unbiased estimator, $\hat{\theta}_2$ is biased:



Bias

- Let $\hat{\theta}$ be an estimator of θ

- The bias in $\hat{\theta}$ is defined as the difference between its mean and θ

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

- The bias of an unbiased estimator is 0

Most Efficient Estimator

- Suppose there are several unbiased estimators of θ
- The most efficient estimator or the minimum variance unbiased estimator of θ is the unbiased estimator with the smallest variance
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ, based on the same number of sample observations. Then,
  - $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$
  - The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is the ratio of their variances:

$$\text{Relative Efficiency} = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}$$

7.2

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

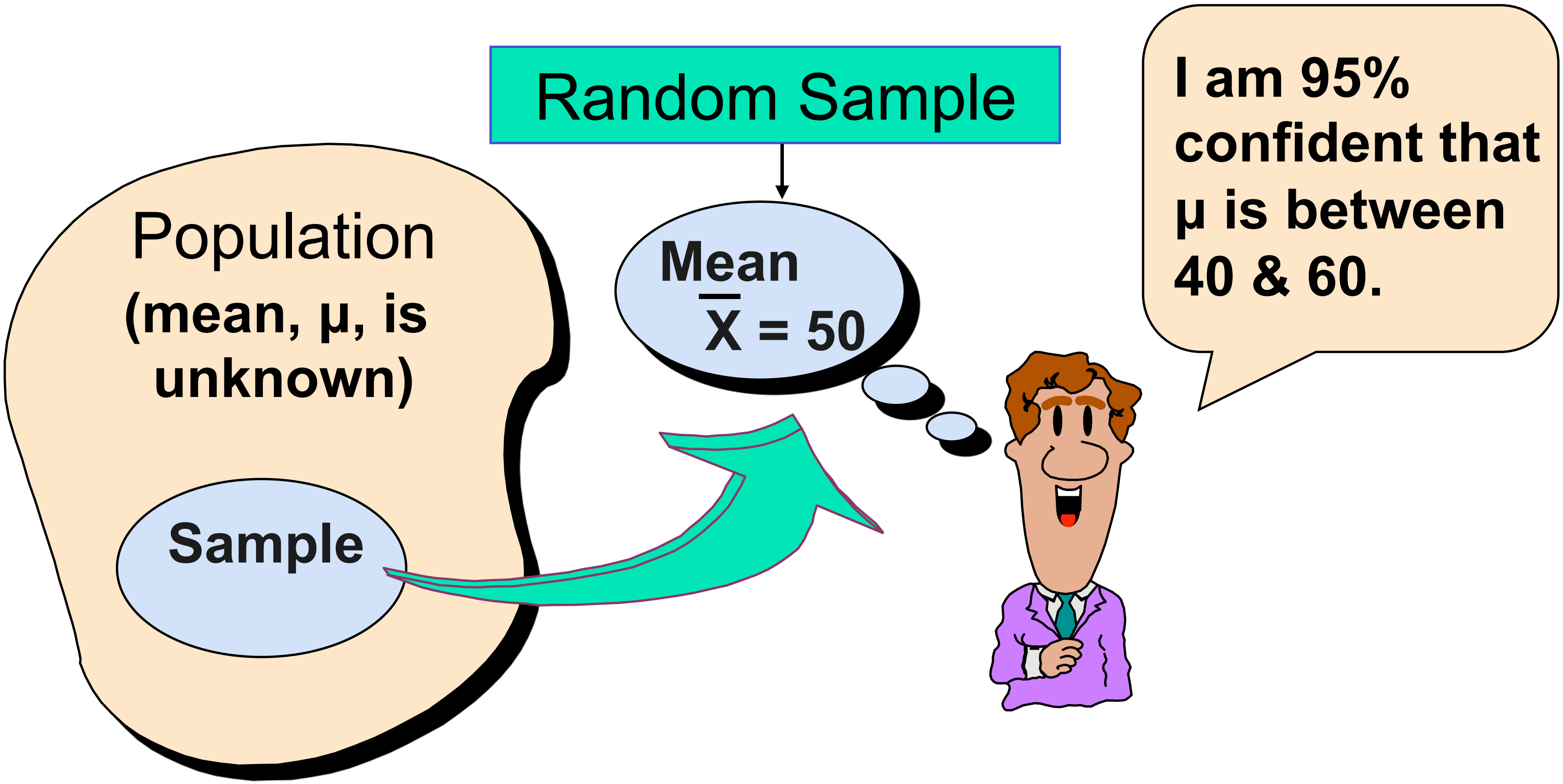
Confidence Interval Estimate

- 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, $(1-\alpha)$

(continued)

- 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 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

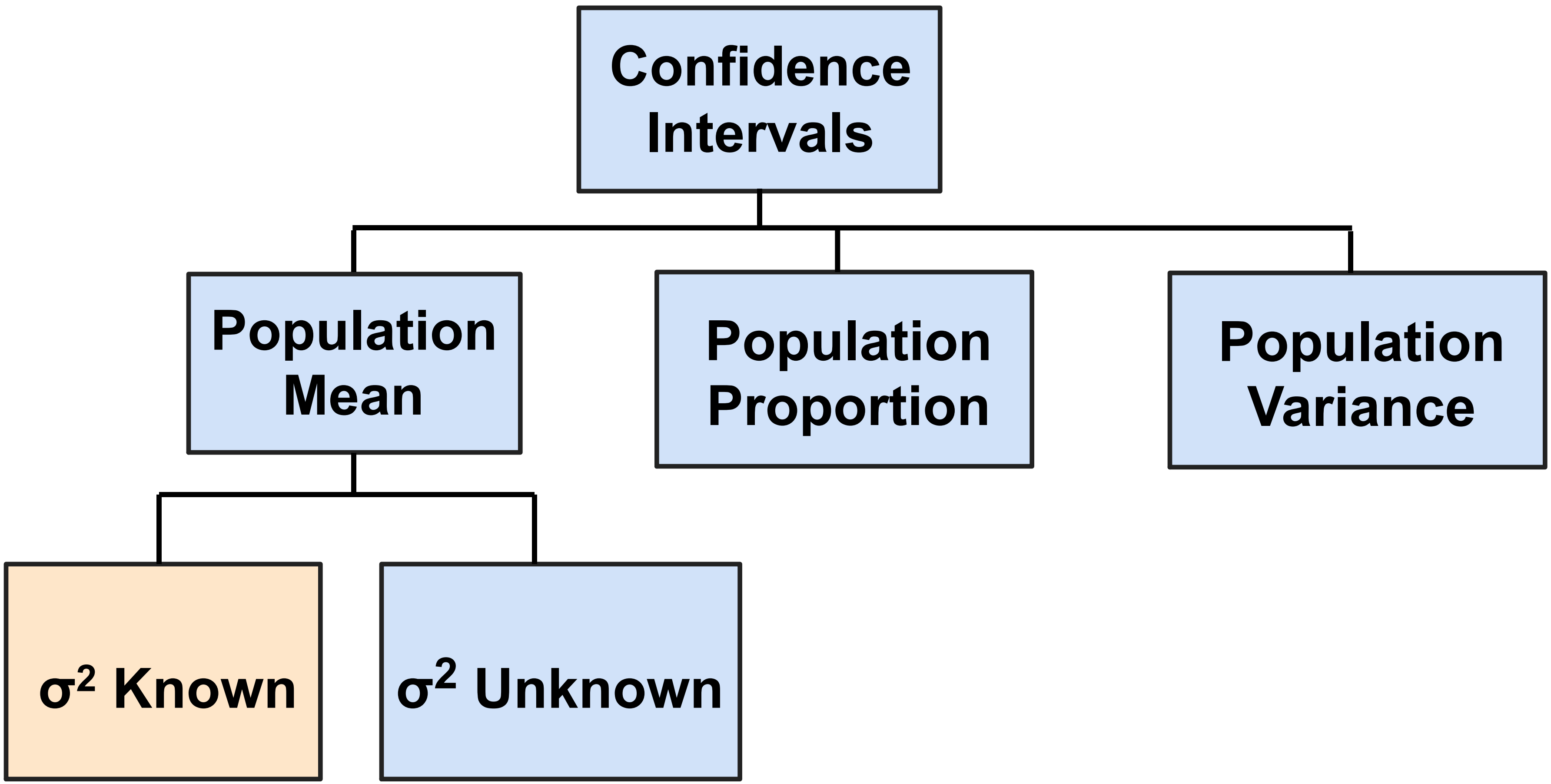
General Formula

- 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



7.2

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

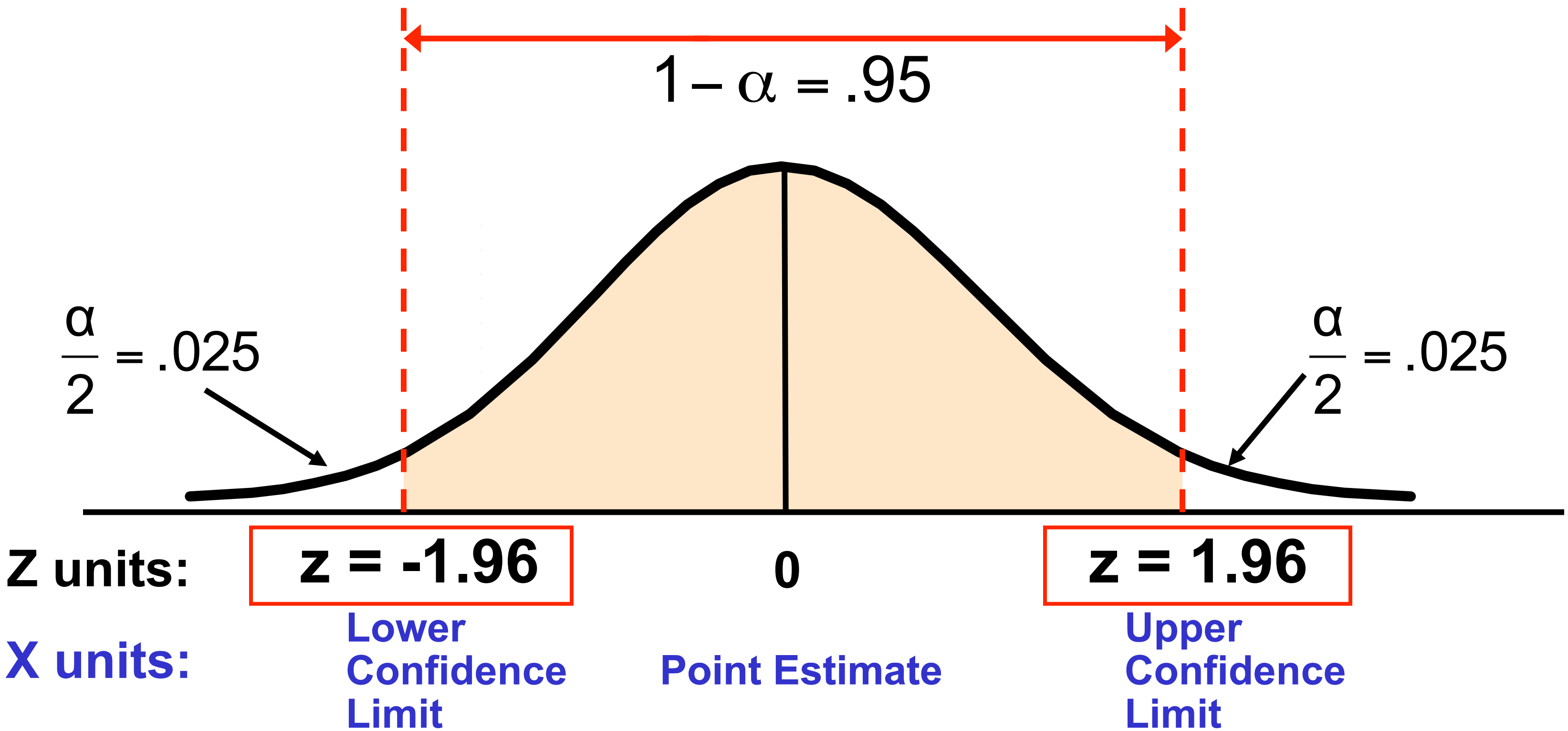
$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

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

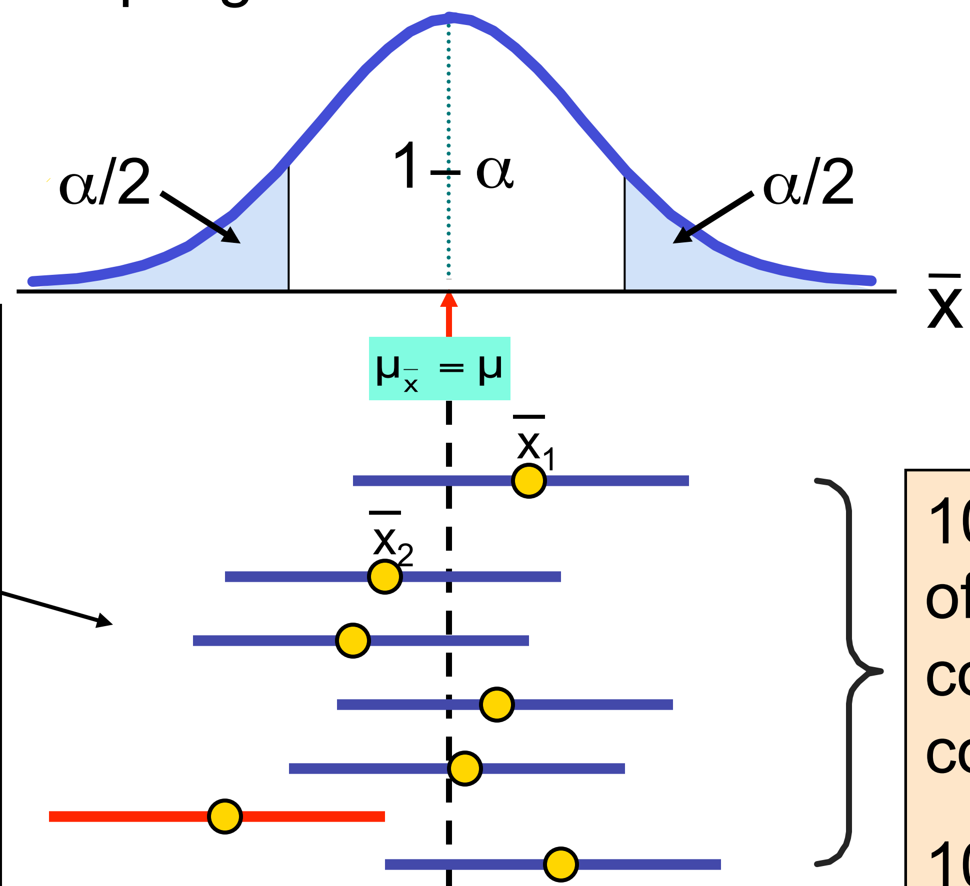
Common Levels of Confidence

- 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	2.58
99.8%	.998	3.08
99.9%	.999	3.27

Intervals and Level of Confidence

Sampling Distribution of the Mean



Confidence Intervals

Intervals extend from

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

to

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

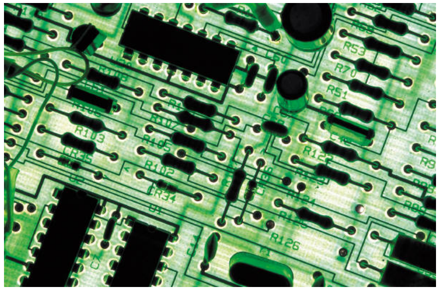
$100(1-\alpha)\%$ of intervals constructed contain μ;

$100(\alpha)\%$ 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.

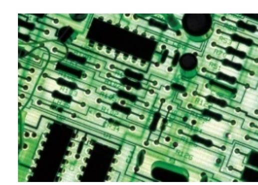
- Solution:

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

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

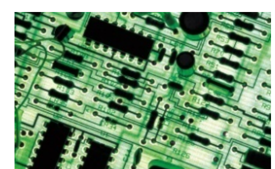
$$= 2.20 \pm .2068$$

$$1.9932 < \mu < 2.4068$$

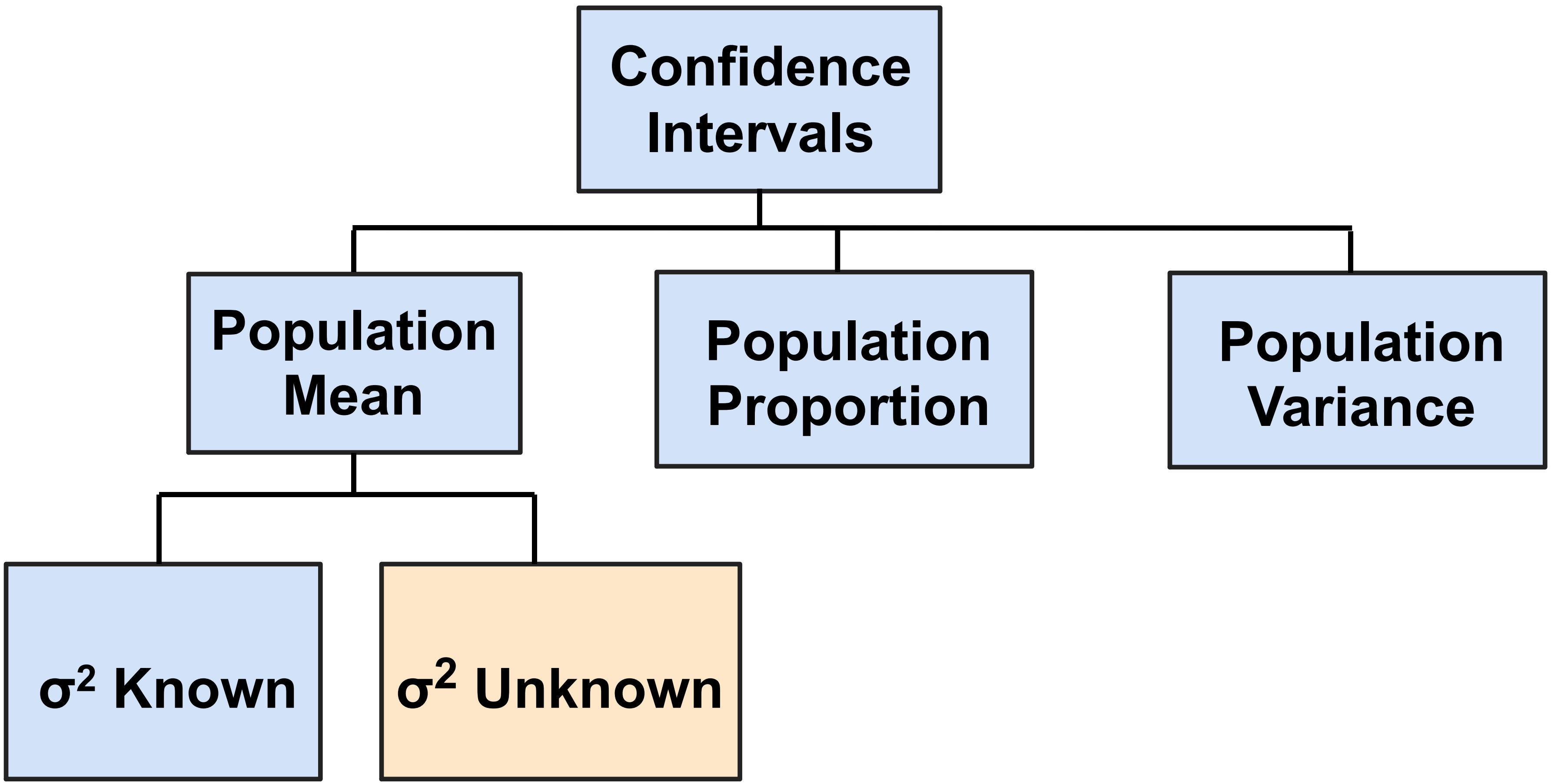


Interpretation

- 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

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

Margin of Error

- The confidence interval,

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

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

where ME is called the margin of error:

$$ME = t_{n\text{-}1,\alpha/2}\frac{\sigma}{\sqrt{n}}$$

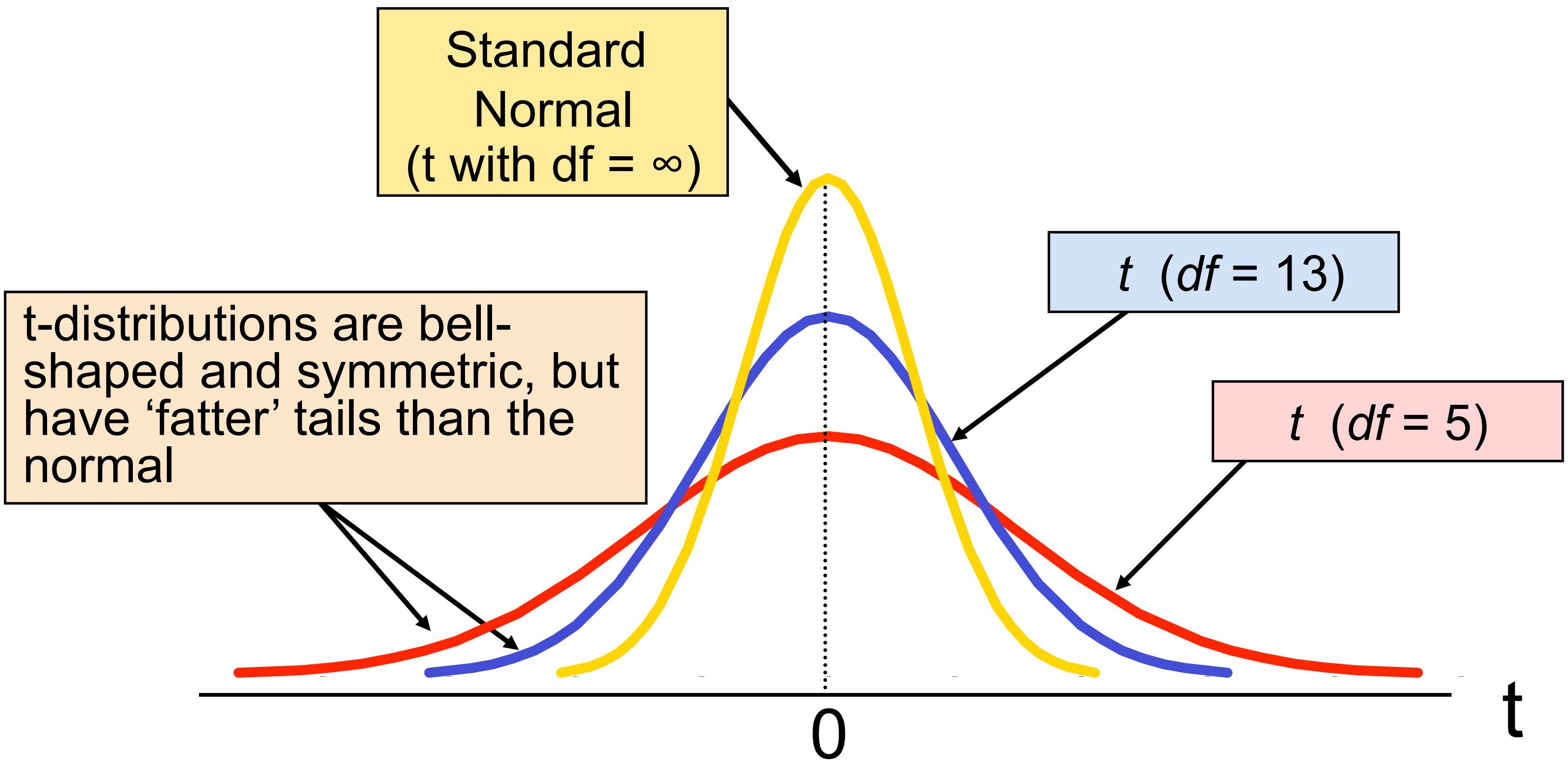
Student’s t Distribution

- 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

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

Student's t Distribution

Note: $t \longrightarrow Z$ as n increases

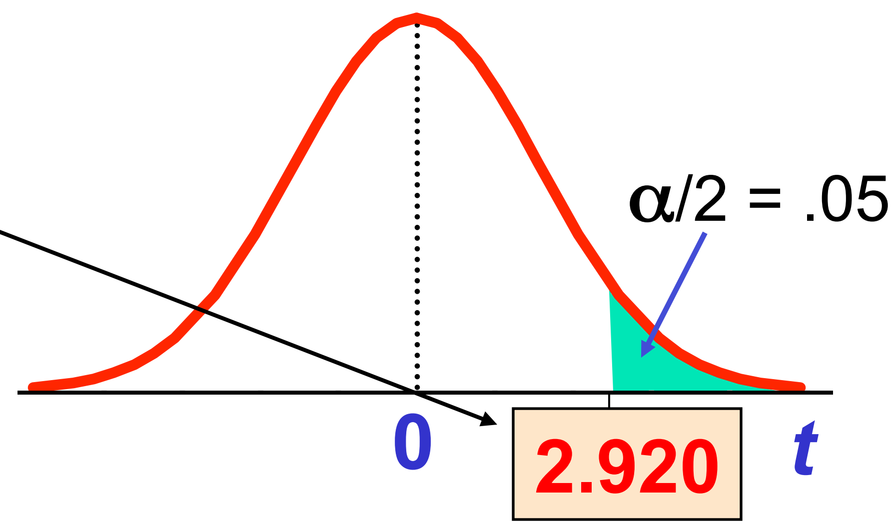


Student's t Table

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 = .10
$\alpha/2$ = .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

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

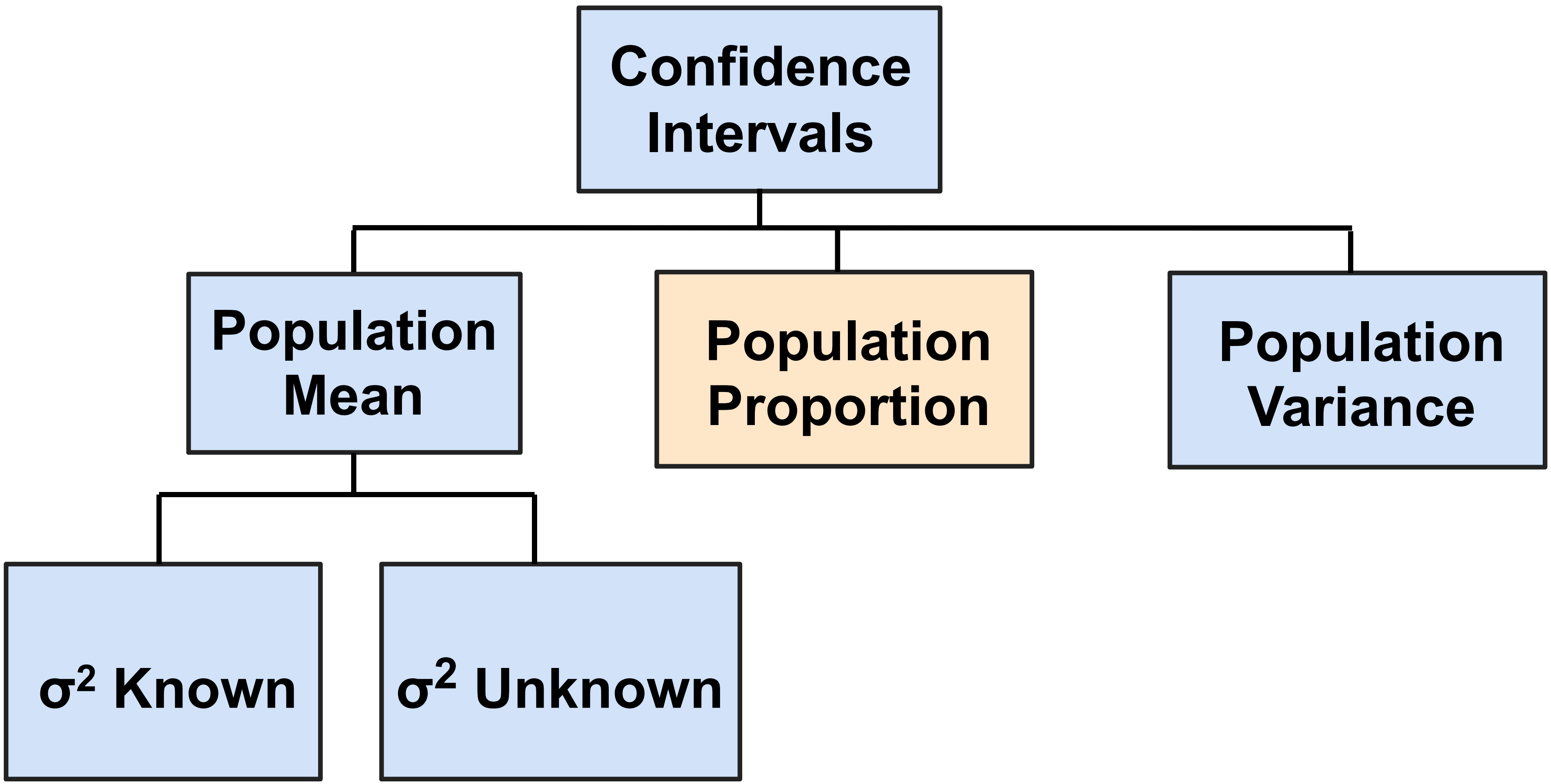
$$\bar{x} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

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

$$46.698 < \mu < 53.302$$

7.4

Confidence Intervals



Confidence Intervals for the Population Proportion

- 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
  - $nP(1-P) > 5$

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.

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

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

$$0.1651 < P < 0.3349$$

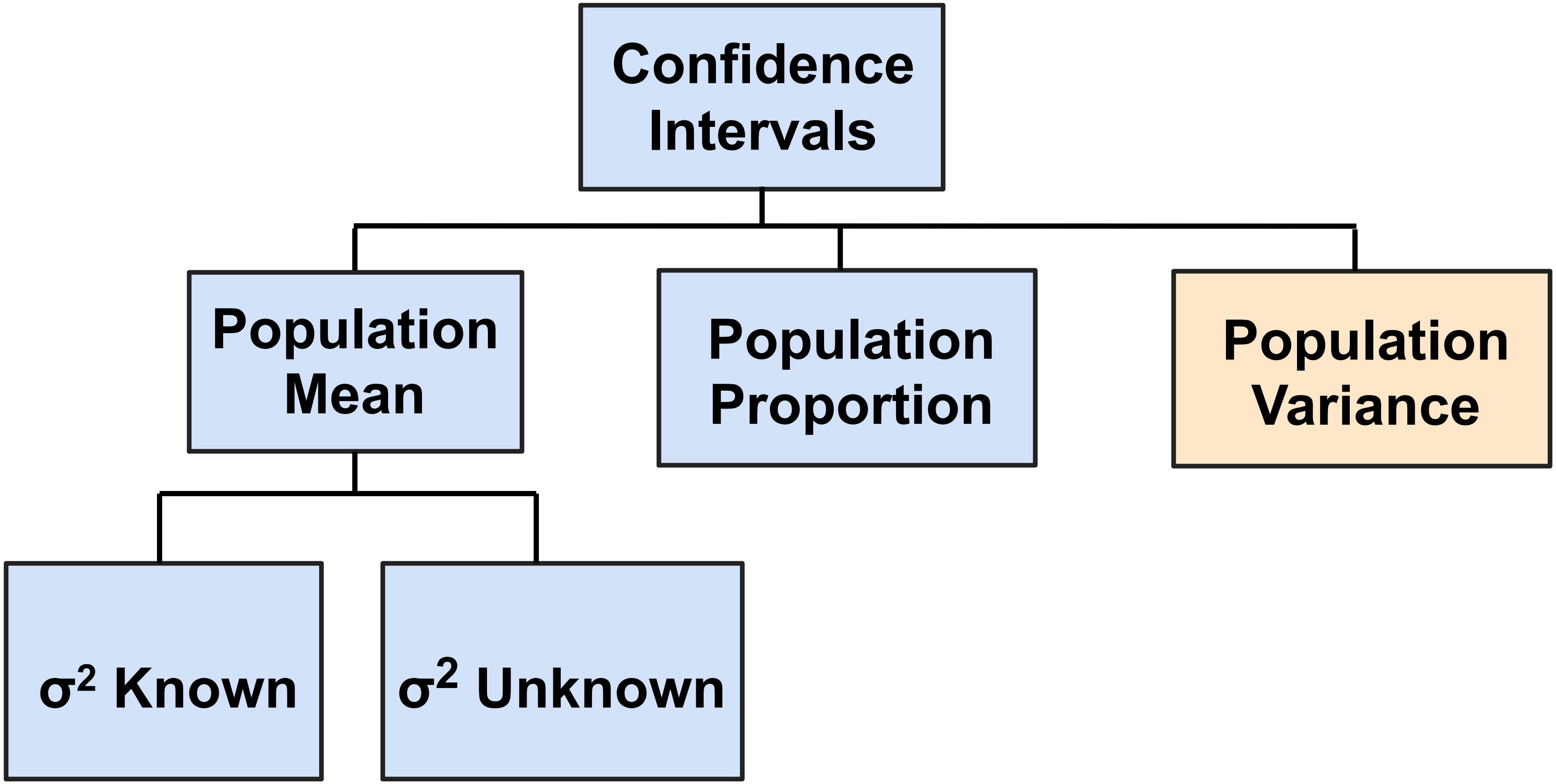
Interpretation

- 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.

Confidence Intervals



Confidence Intervals for the Population Variance

- 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

(continued)

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

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

Confidence Intervals for the Population Variance

(continued)

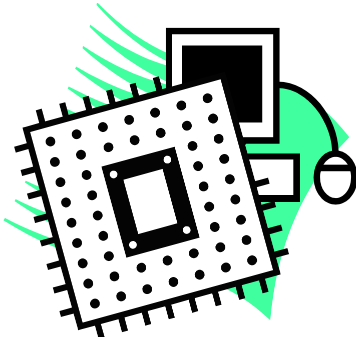
The (1 - α)% confidence interval for the population variance is

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

Example

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**



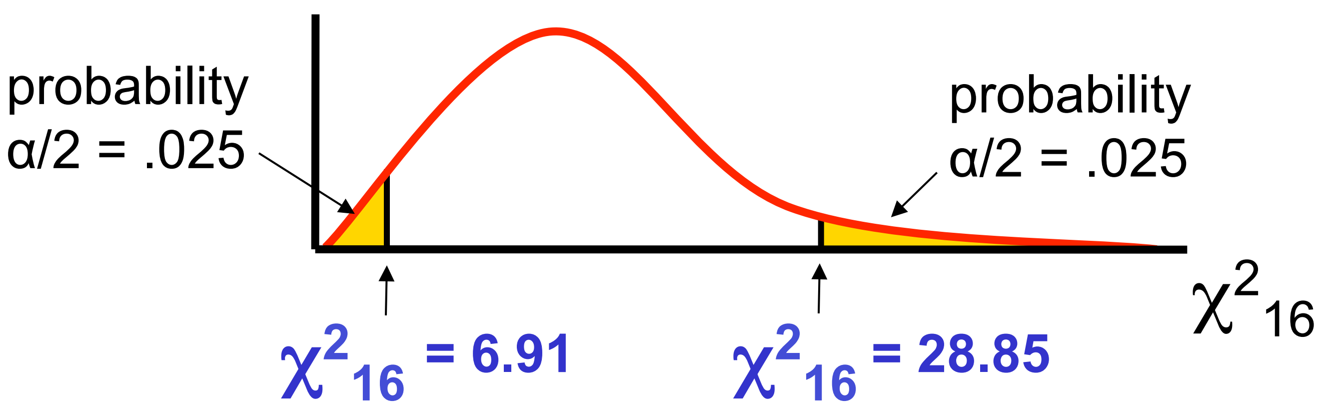
Assume the population is normal.
Determine the 95% confidence interval for σ_x^2

Finding the Chi-square Values

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

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

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



Calculating the Confidence Limits

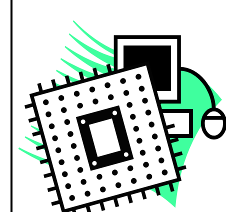
- The 95% confidence interval is

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

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

$$3037 < \sigma^2 < 12683$$

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



7.6

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-α)% confidence interval for the population mean is

$$\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$$

Estimating the Population Total

- 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

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

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

- A 100(1 - α)% confidence interval for the population total is

$$N\bar{x} - t_{n-1,\alpha/2} N\hat{\sigma}_{\bar{x}} < N\mu < N\bar{x} + t_{n-1,\alpha/2} N\hat{\sigma}_{\bar{x}}$$

Confidence Interval for Population Total: Example

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

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

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

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

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

$$82837.53 < N\mu < 92362.47$$

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

Estimating the Population Proportion

- 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

Finite Populations: Proportion

- 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-α)% confidence interval for the population proportion is

$$\hat{p} - z_{\alpha/2}\hat{\sigma}_{\hat{p}} < P < \hat{p} + z_{\alpha/2}\hat{\sigma}_{\hat{p}}$$

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)

Chapter Summary

(continued)

- 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