Econ 325: Introduction to Empirical Economics

Lecture 7

Estimation: Single Population

Parameters

- A parameter is some constant that summarizes the feature of population distribution.
- Examples
  - μ population mean
  - σ^2 population variance
  - p population fraction
- We often use θ (``theta'') to denote a parameter

Estimation problem

- Given a sample, we would like to make our best guess about a parameter of interest.

- Examples:
  - Sample mean $\bar{X}$ is our guess of population mean μ
  - Sample variance s^2 is our guess of population variance σ^2
  - Sample fraction $\hat{p}$ is our guess of population fraction p

7.1

Point Estimator

- A point estimator of a population parameter θ is a function of random sample:

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$$

- Example:

$$\bar{X} = \bar{X}(X_1, X_2, \ldots, X_n) \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$$

- A specific realized value of that random variable is called an point estimate.

Point Estimates

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{X}$
Variance	σ^2	s^2

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,

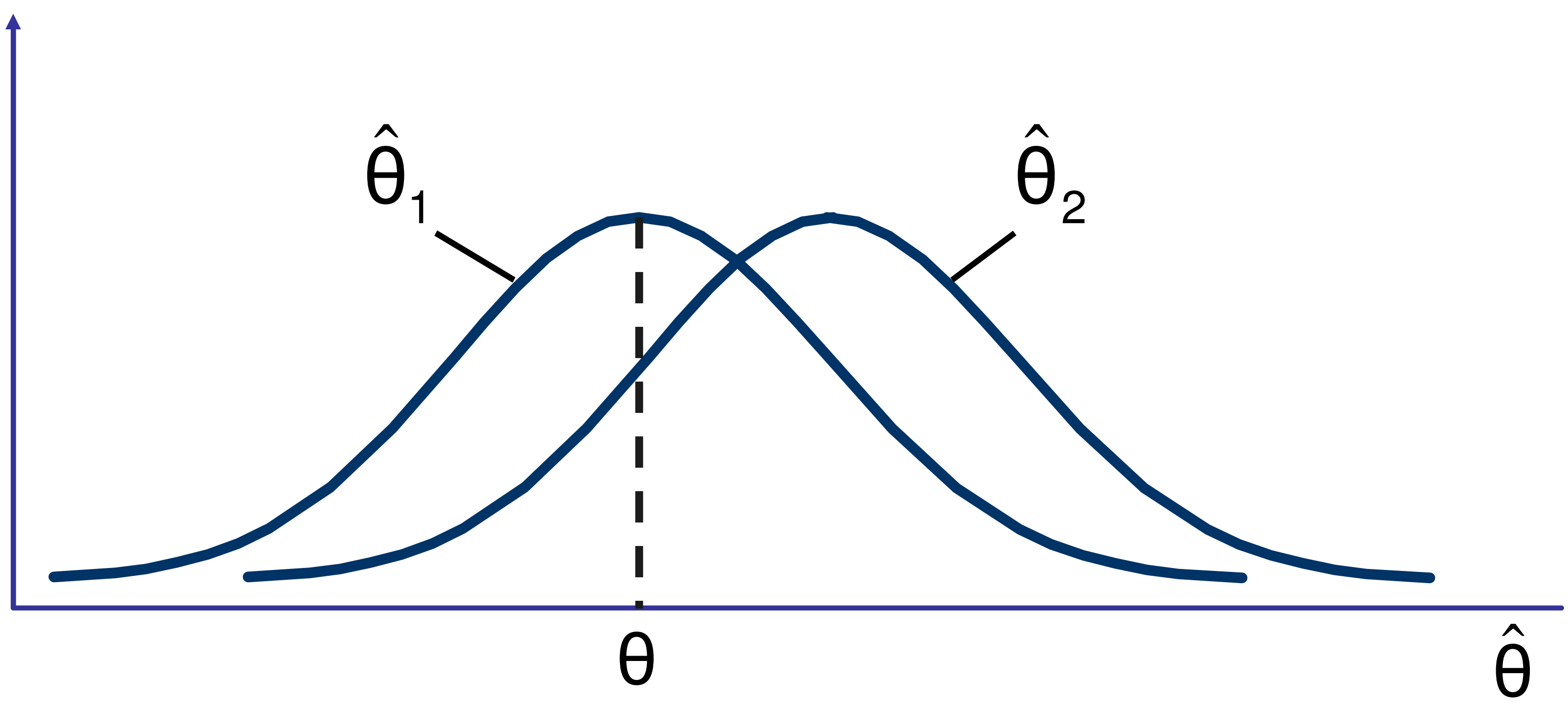
$$E(\hat{\theta}) = \theta$$

- 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

Clicker Question 7-1

- Given a random sample of n = 2, consider two estimators for μ:

$$\text{(i) } \bar{X} = \frac{1}{2}(X_1 + X_2) \text{ and } \text{(ii) } \hat{X} = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

A). (i) is unbiased but the bias of (ii) is not zero.

B). The bias of (i) not zero but (ii) is unbiased.

C). Both are unbiased.

Efficiency

- We prefer the estimator with the smaller variance.
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ.
- Then,

  $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if

$$\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$$

The most efficient unbiased estimator of θ is the unbiased estimator with the smallest variance.

Clicker Question 7-2

- Given a random sample of n = 2, consider two estimators for μ:

$$\bar{X} = \frac{1}{2}(X_1 + X_2) \text{ and } \hat{X} = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

Which estimator is more efficient?

A). $\bar{X} = \frac{1}{2}(X_1 + X_2)$

B). $\hat{X} = \frac{1}{3}X_1 + \frac{2}{3}X_2$

C). Both are equally efficient.

Consistency

- A point estimator $\hat{\theta}$ is said to be a **consistent** estimator of θ if $\hat{\theta}$ converges in probability to θ, i.e.,

$$\hat{\theta} \xrightarrow{p} \theta$$

- By the Law of Large Numbers, the sample mean $\bar{X}_n$ is a consistent estimator of μ because $\bar{X}_n \xrightarrow{p} \mu$.

Clicker Question 7-2

- Consider the following estimator of $\mu = E[X]$:

$$\hat{X} = \frac{1}{n-1}\sum_{i=1}^{n} X_i$$

A). $\hat{X}$ is a **consistent estimator** of μ

B). $\hat{X}$ is **not a consistent estimator** of μ

Clicker Question 7-3

- Consider the following estimator of $\mu = E[X]$:

$$\hat{X} = \frac{1}{n-1}\sum_{i=1}^{n} X_i$$

A). $\hat{X}$ is an **unbiased estimator** of μ

B). $\hat{X}$ is **not an unbiased estimator** of μ

Unbiasedness and Consistency

- **Consistency** is a property of an estimator **when $n \to \infty$.** Consistency is the result of the Law of Large Numbers.

- **Unbiasedness** is a property of an estimator **when n is fixed.** It is nothing to do with the Law of Large Numbers.

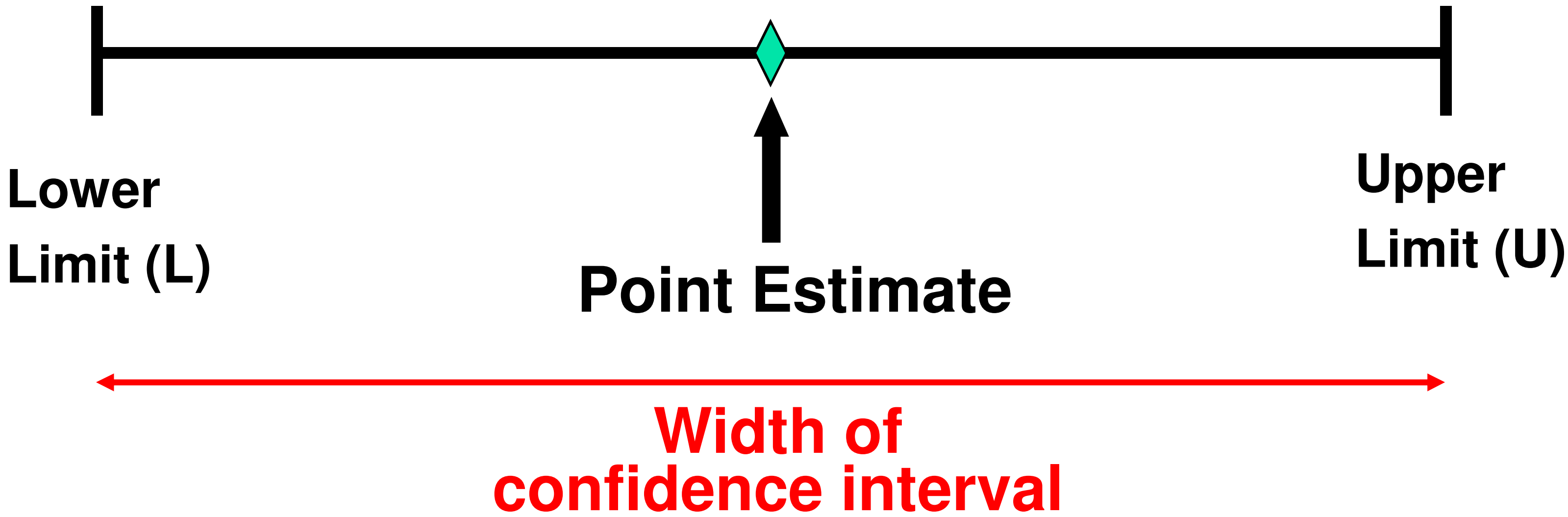
7.2

Confidence Intervals

- An interval estimate provides more information about a population characteristic than does a point estimate
- Such interval estimates are called confidence intervals

Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about variability



Confidence Interval Estimate

- An interval gives a range of values
- Based on observation from 1 sample
- The lower limit (L) and upper limit (U) are functions of the sample, e.g.,

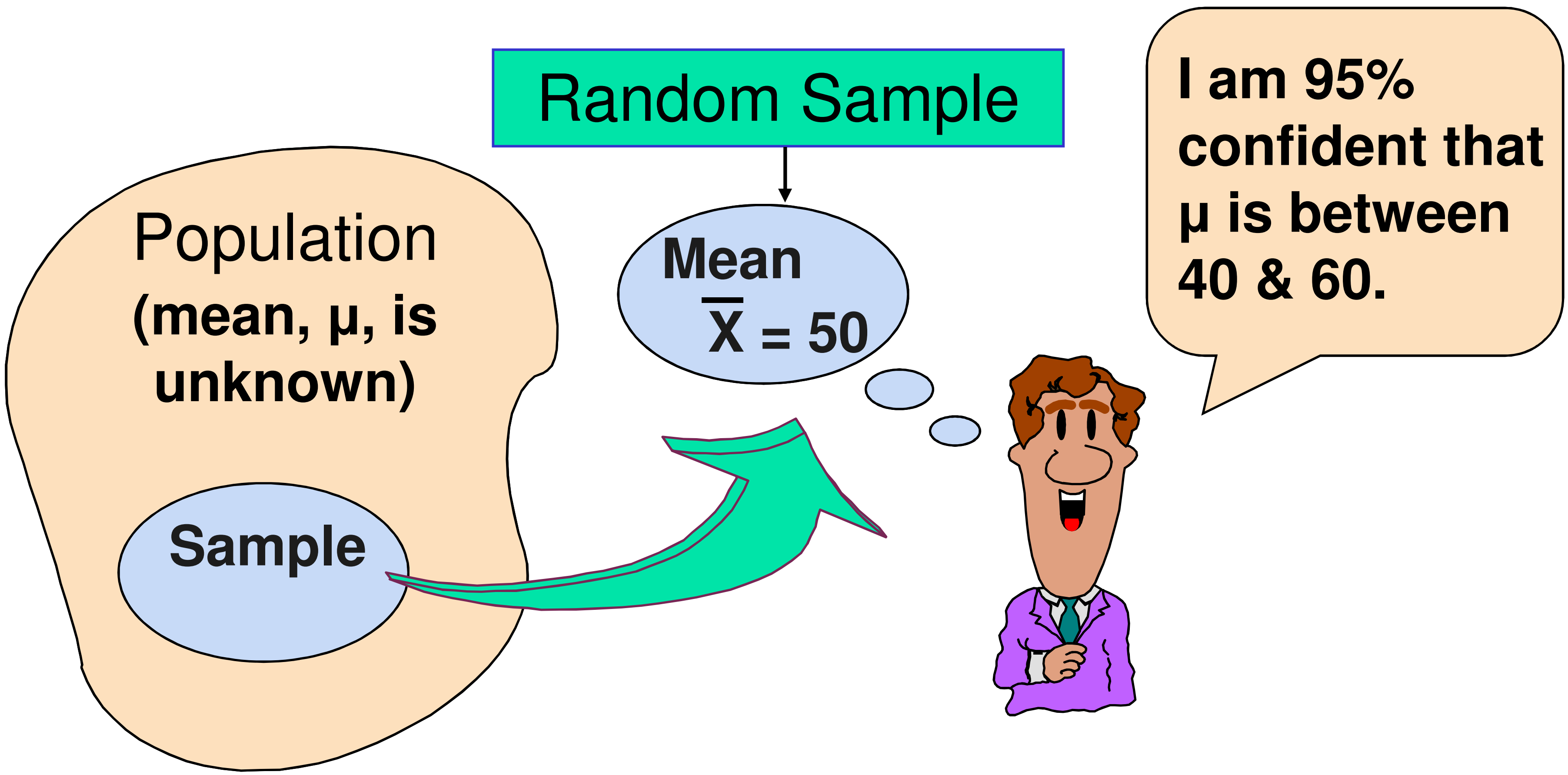
$$P(L(X_1, X_2, \dots, X_n) < \theta < U(X_1, X_2, \dots, X_n)) = 0.95$$

Confidence Interval and Confidence Level

- If $P(L < \theta < U) = 1 - \alpha$ then the interval from L to U 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)

Estimation Process



Confidence Level, $(1-\alpha)$

(continued)

- If confidence level = $(1 - \alpha) = 0.95$
- From repeated samples, 95% of all the confidence intervals will contain the true parameter
- A specific interval either will contain or will not contain the true parameter

General Formula

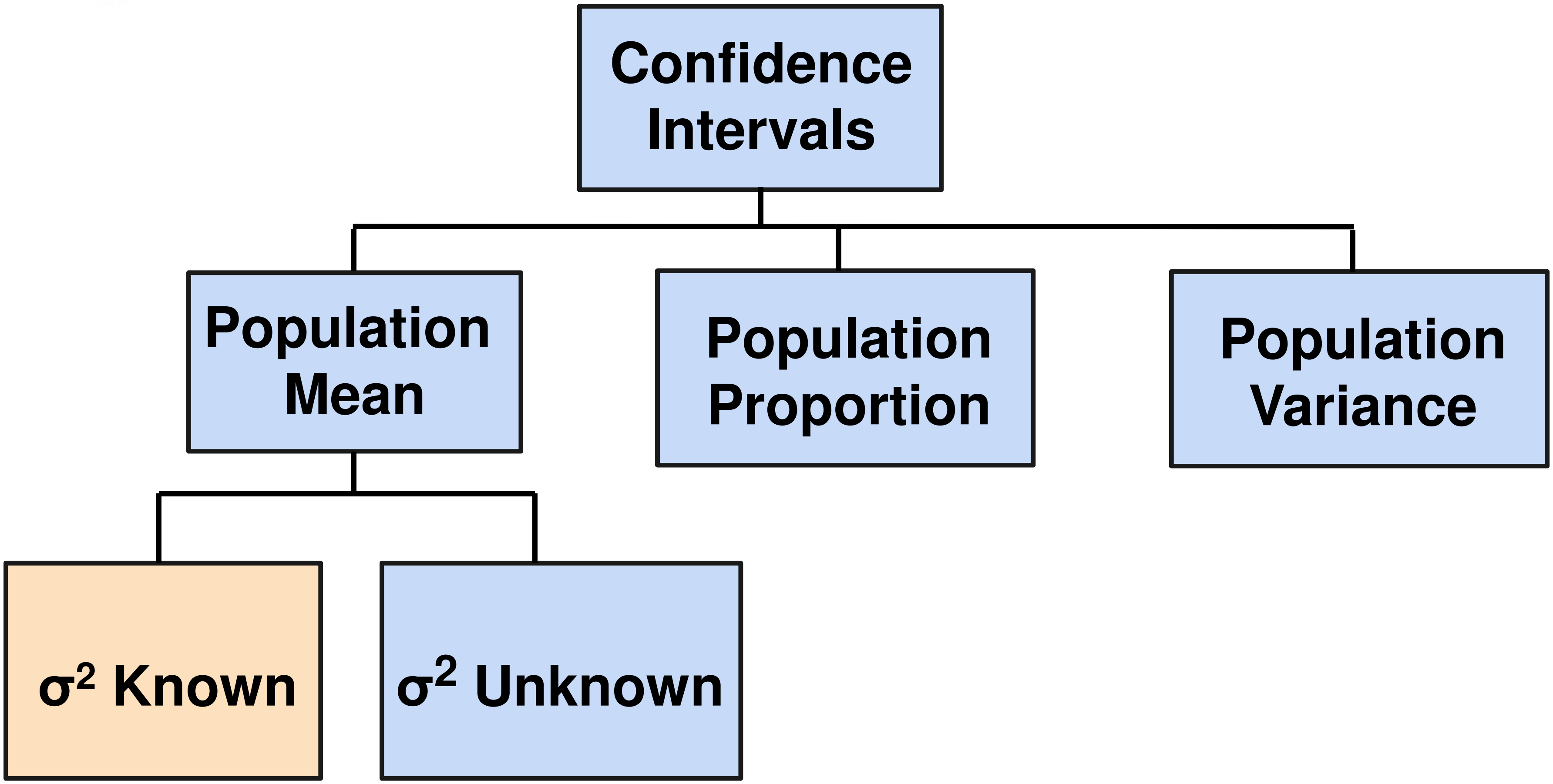
- The general formula for all confidence intervals is:

Point Estimate ± (Reliability Factor)(Standard Error)

- Example

$$P\left(\bar{X} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 0.95$$

Confidence Intervals



Confidence Interval for μ (σ^2 Known)

7.2

- 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

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

Reducing the Margin of Error

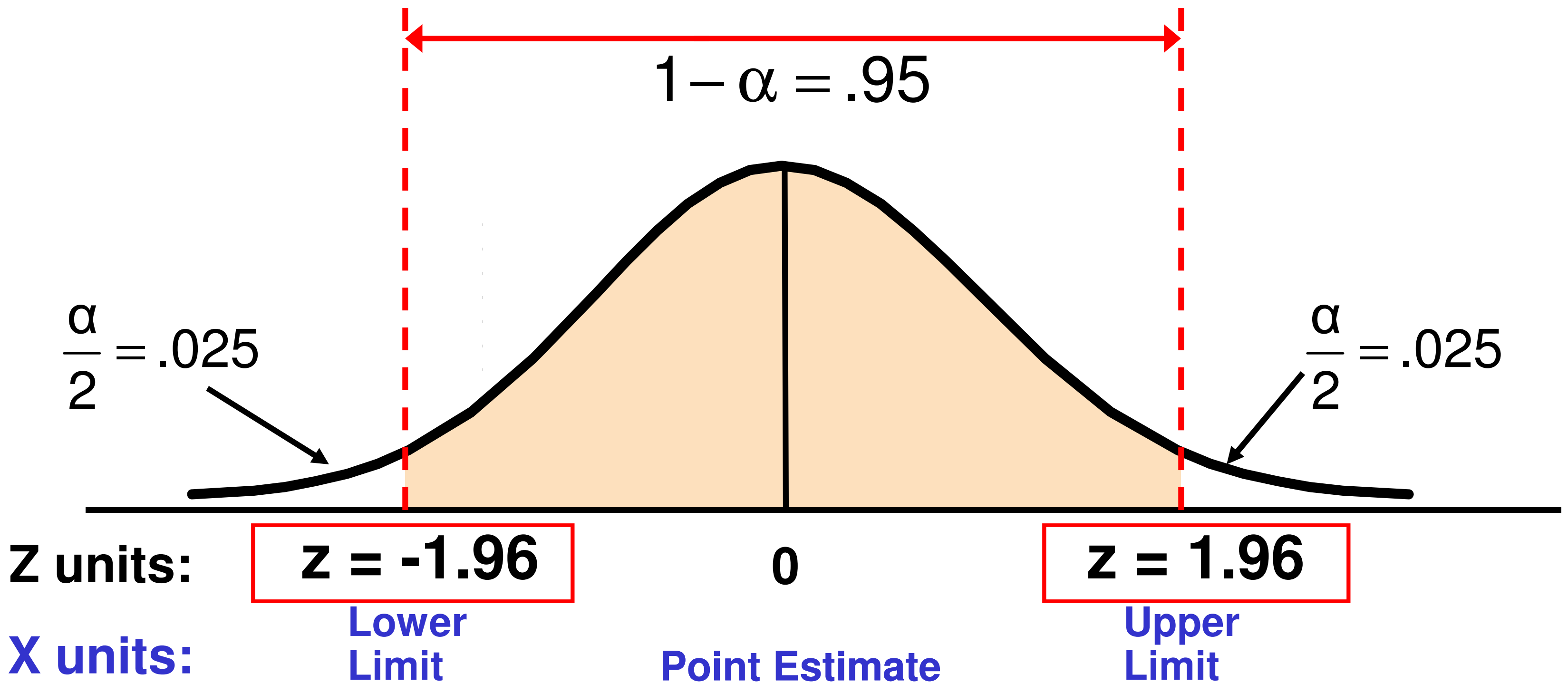
$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

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

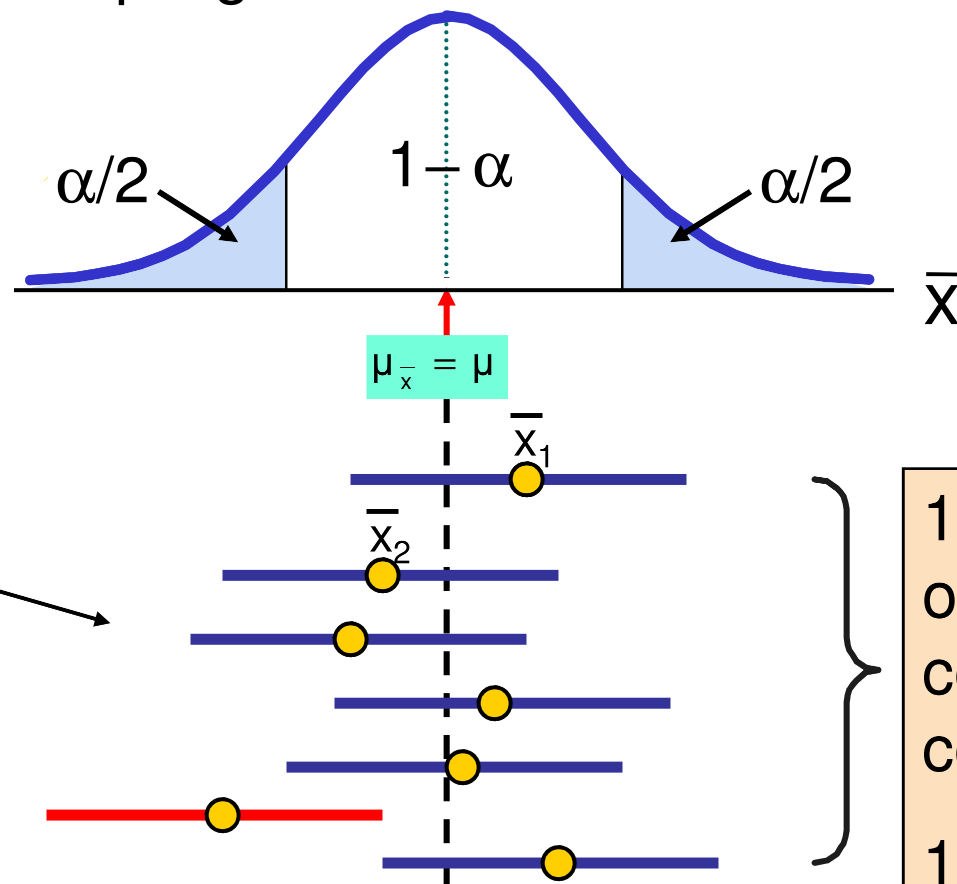
Common Levels of Confidence

- 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

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

to

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

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

$100(\alpha)\%$ do not.

Example

- A sample of 27 light bulb from a large normal population has a mean life length of 1478 hours. We know that the population standard deviation is 36 hours.

- Determine a 95% confidence interval for the true mean length of life in the population.

Example

(continued)

- Solution:

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

$$= 1478 \pm 1.96\,(36/\sqrt{27})$$

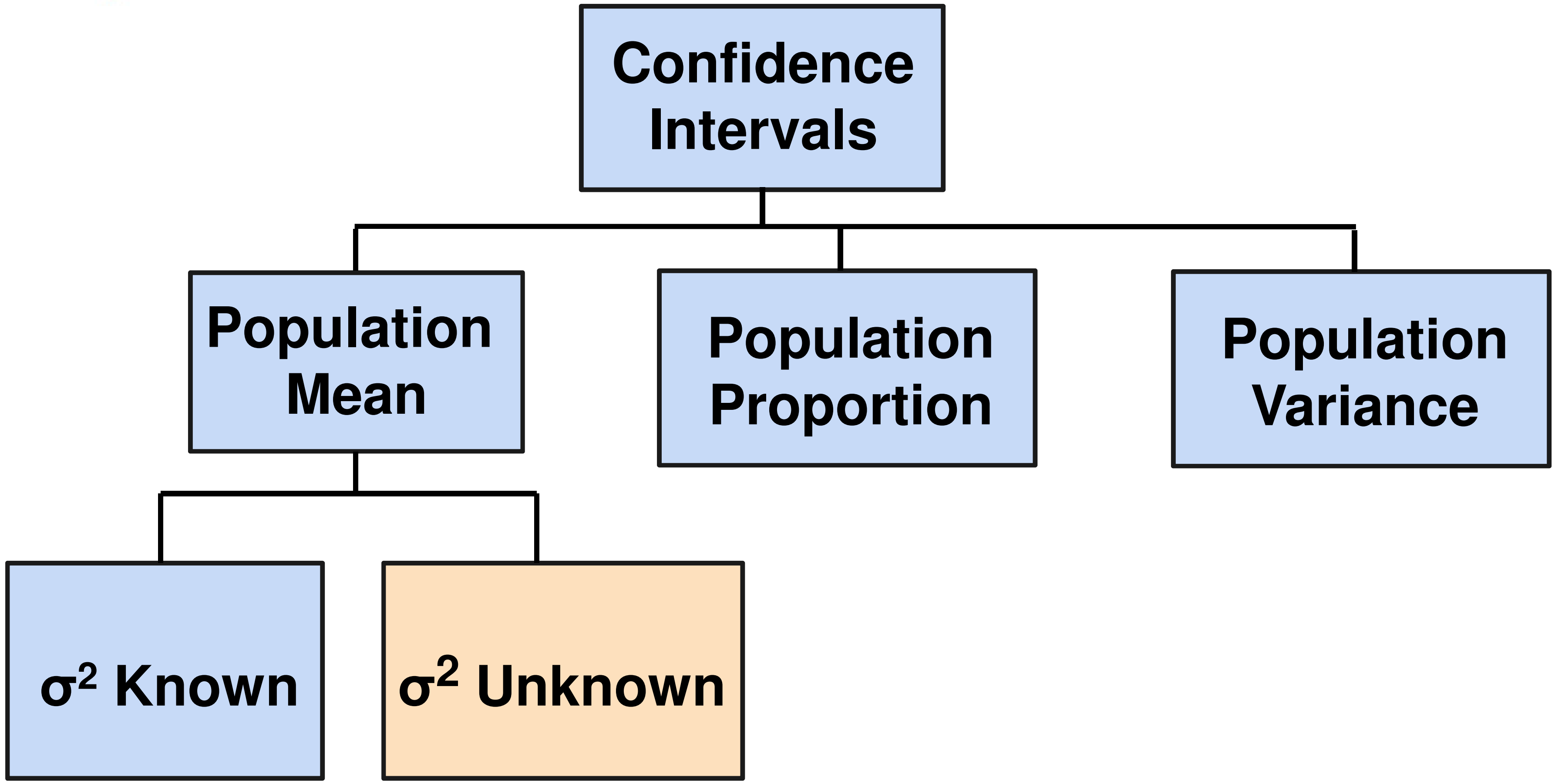
$$= 1478 \pm 13.58$$

$$1464.42 < \mu < 1491.58$$

Interpretation

- We are 95% confident that the true mean life time is between 1464.42 and 1491.58
- 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 sample mean $\bar{x}$ and standard deviation s
  - from a normally distributed population with mean μ

- Then, the random variable

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

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

Confidence Interval for μ (σ^2 Unknown)

- If 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

Student's t Distribution

Let $Z \sim N(0,1)$ and χ^2_v follows Chi-square distribution with degrees of freedom v. Then, a random variable

$$t_v = \frac{Z}{\sqrt{\chi^2_v / v}}$$

follows Student's t distribution with degrees of freedom v.

Student's t Distribution

$$t = \frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}$$

$$= \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s_n^2}{\sigma^2}/(n-1)}}$$

$$= \frac{Z}{\sqrt{\chi_v^2/v}} \quad with\ v = n - 1$$

Confidence Interval for μ (σ Unknown)

(continued)

- Assume **population is normally distributed**
- Confidence Interval:

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

where $t_{n-1,\alpha/2}$ is the critical value of the t distribution with (n-1) d.f. such that

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

Margin of Error

- The confidence interval,

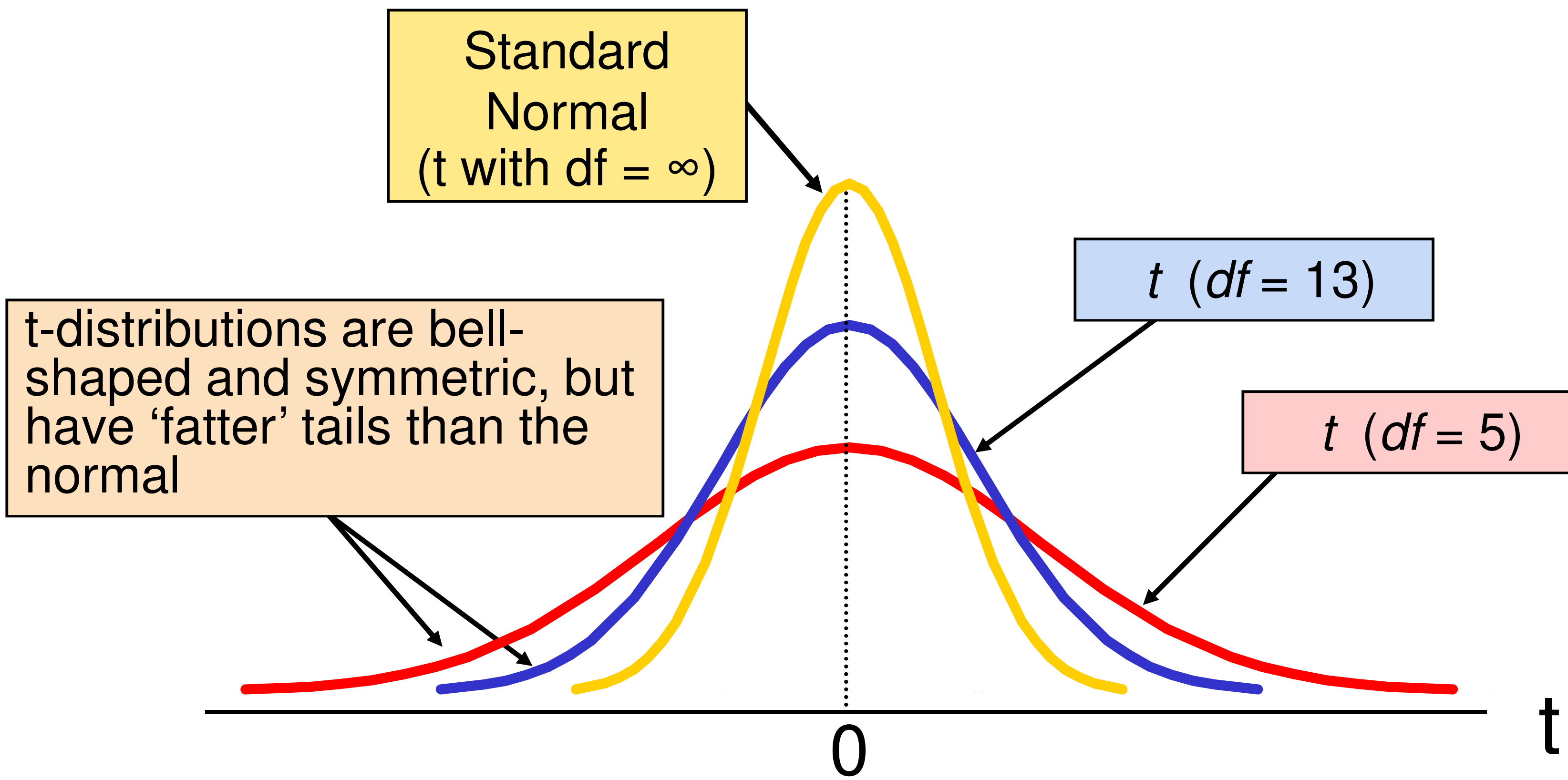
$$\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

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

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

Student's t Distribution

Note: $t \longrightarrow Z$ as n increases

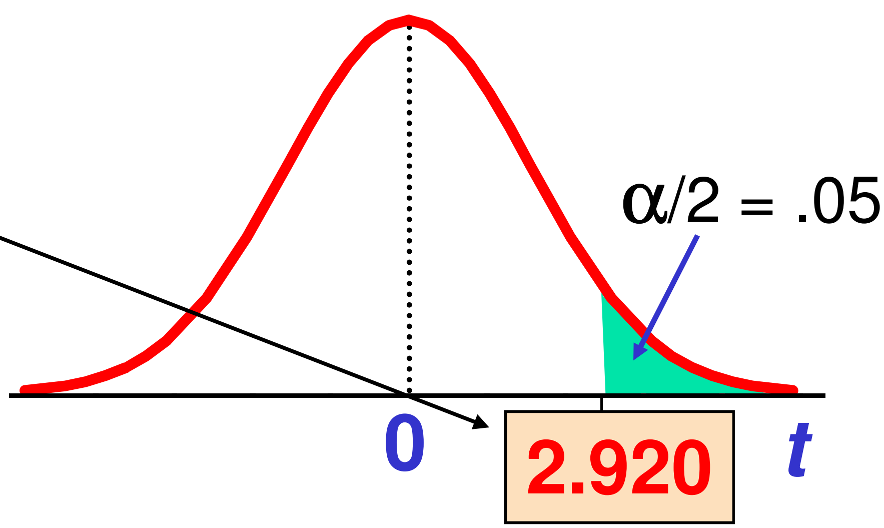


Student's t Table

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

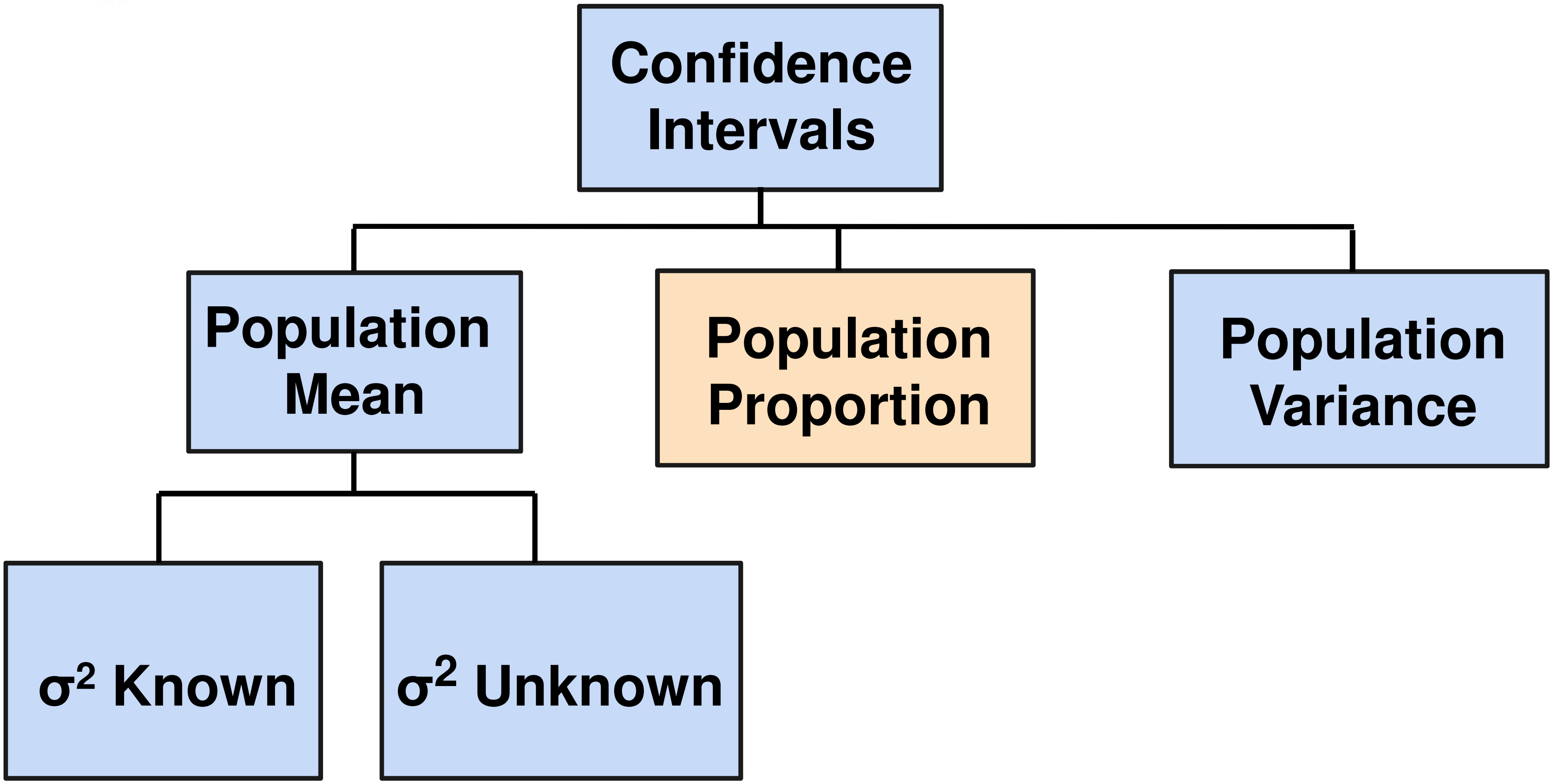
The confidence interval is

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

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

$$46.698 < \mu < 53.302$$

Confidence Intervals



Confidence Intervals for the Population Proportion, p

(continued)

- By the Central Limit Theorem,

$$\hat{p} - p \sim N(0, \sigma_p^2)$$

where

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

- The sample analogue estimator of σ_p is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval Endpoints

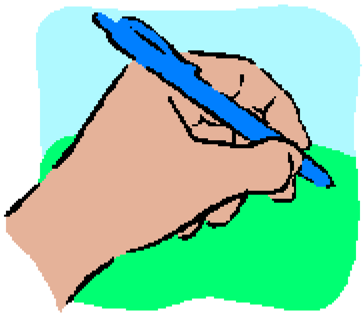
- Upper and lower confidence limits for the population proportion are calculated with the formula

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

- 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.

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

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

$$0.1651 < p < 0.3349$$

Interpretation

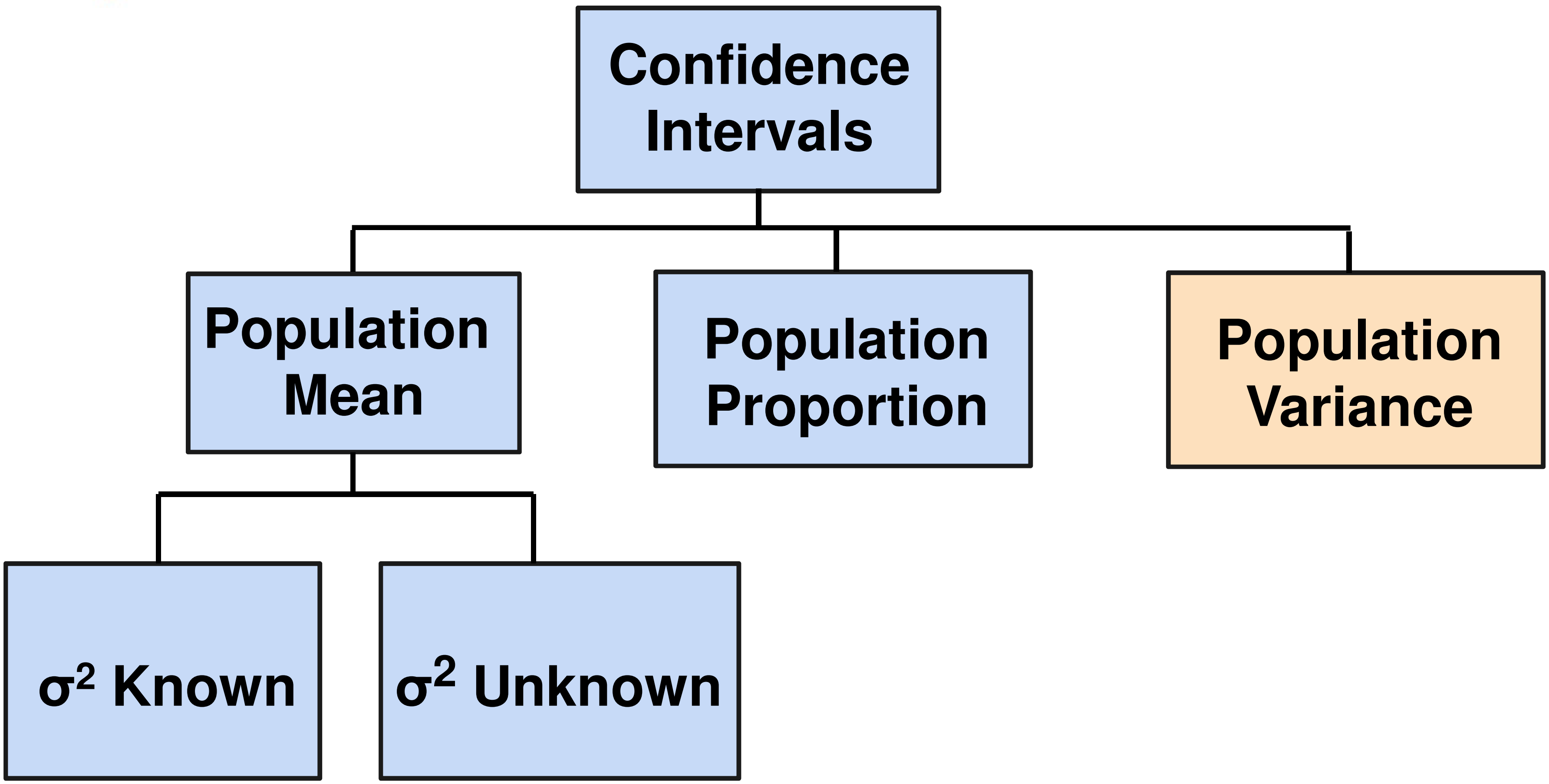
- We are 95% confident that the true percentage of left-handers in the population is between

  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.

7.5

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)

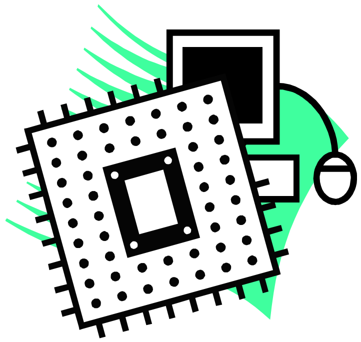
The (1 - α)% confidence interval for the population variance is

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

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**



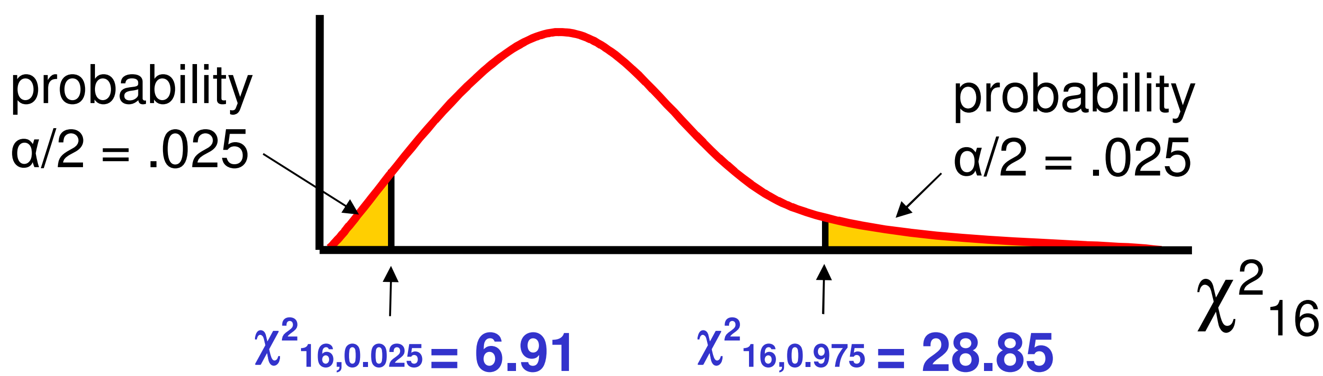
Assume the population is normal.
Determine the 95% confidence interval for σ_x^2

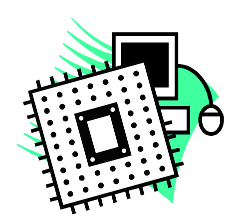
Finding the Chi-square Values

- 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:

$$\chi^2_{n-1,\,\alpha/2} = \chi^2_{16,\,0.025} = 6.91$$

$$\chi^2_{n-1,\,1-\alpha/2} = \chi^2_{16,\,0.975} = 28.85$$





Calculating the Confidence Limits

- The 95% confidence interval is

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

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

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

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