Condidence Intervals
An Introduction

Loukia Meligkotsidou
Associate Professor of Statistics
Department of Mathematics
National and Kapodistrian
University of Athens

Confidence Intervals

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

Confidence Interval and Confidence Level

- If $P(LL < \mu < UL) = 1 - \alpha$ then the interval from the LL to UL is called a $100(1 - \alpha)\%$ confidence interval of μ.
- The quantity $(1 - \alpha)$ is called the confidence level of the interval (probability between 0 - 1)
  - In repeated samples of the population, the true value of the parameter μ would be contained in $100(1 - \alpha)\%$ of intervals calculated this way.
  - The confidence interval calculated in this manner is written as $LL < \mu < UL$ with $100(1 - \alpha)\%$ confidence

Confidence Level, $(1-\alpha)$

- Suppose the confidence level = 95%
- Also written $(1 - \alpha) = 0.95$ or $\alpha=0.05$
- 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

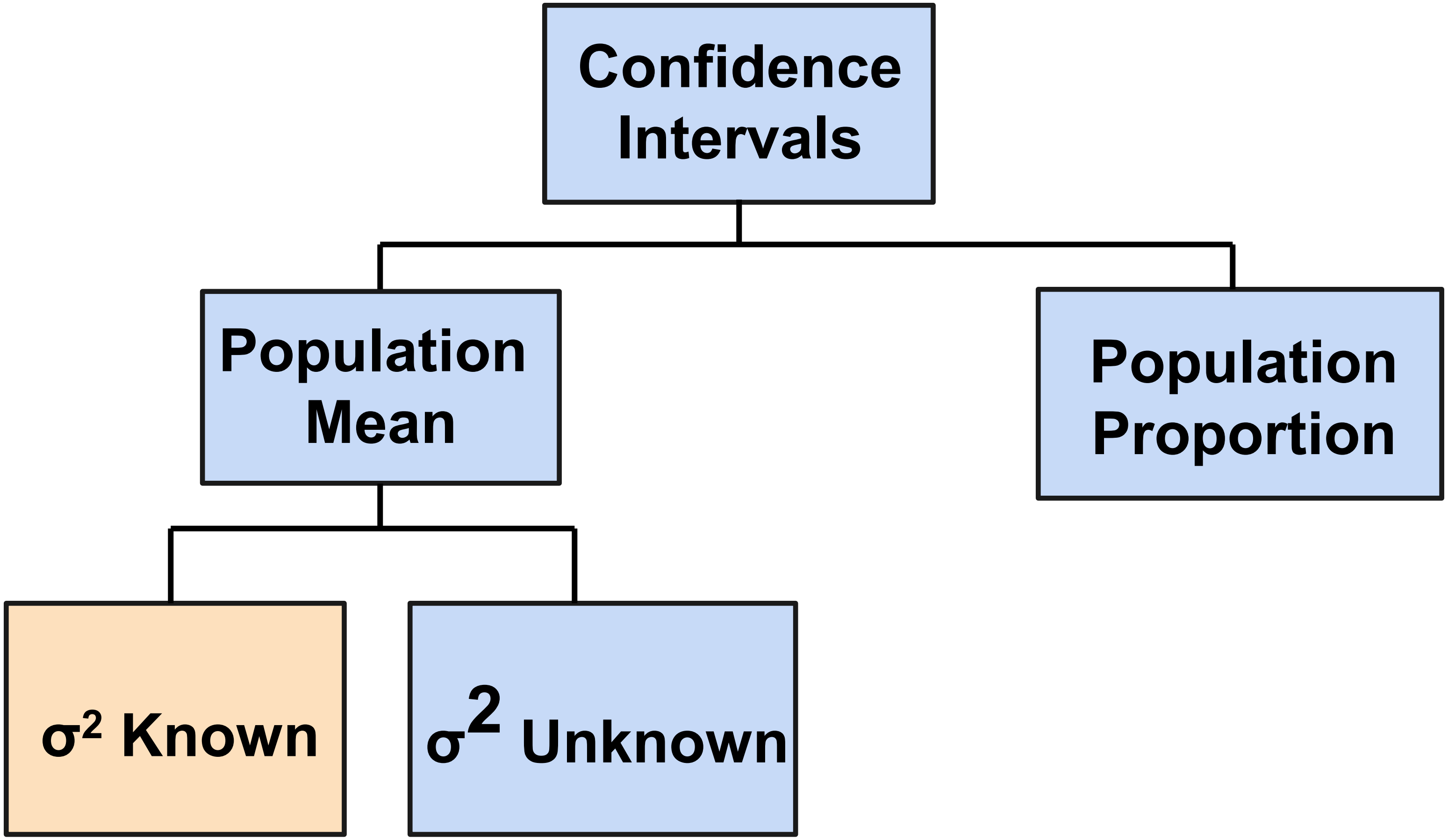
General Formula

- The general formula for all confidence intervals is:

Point Estimate ± k * Standard Error

- K often called reliability factor and depends on the desired level of confidence

Confidence Intervals



Confidence Interval for μ (σ^2 Known)

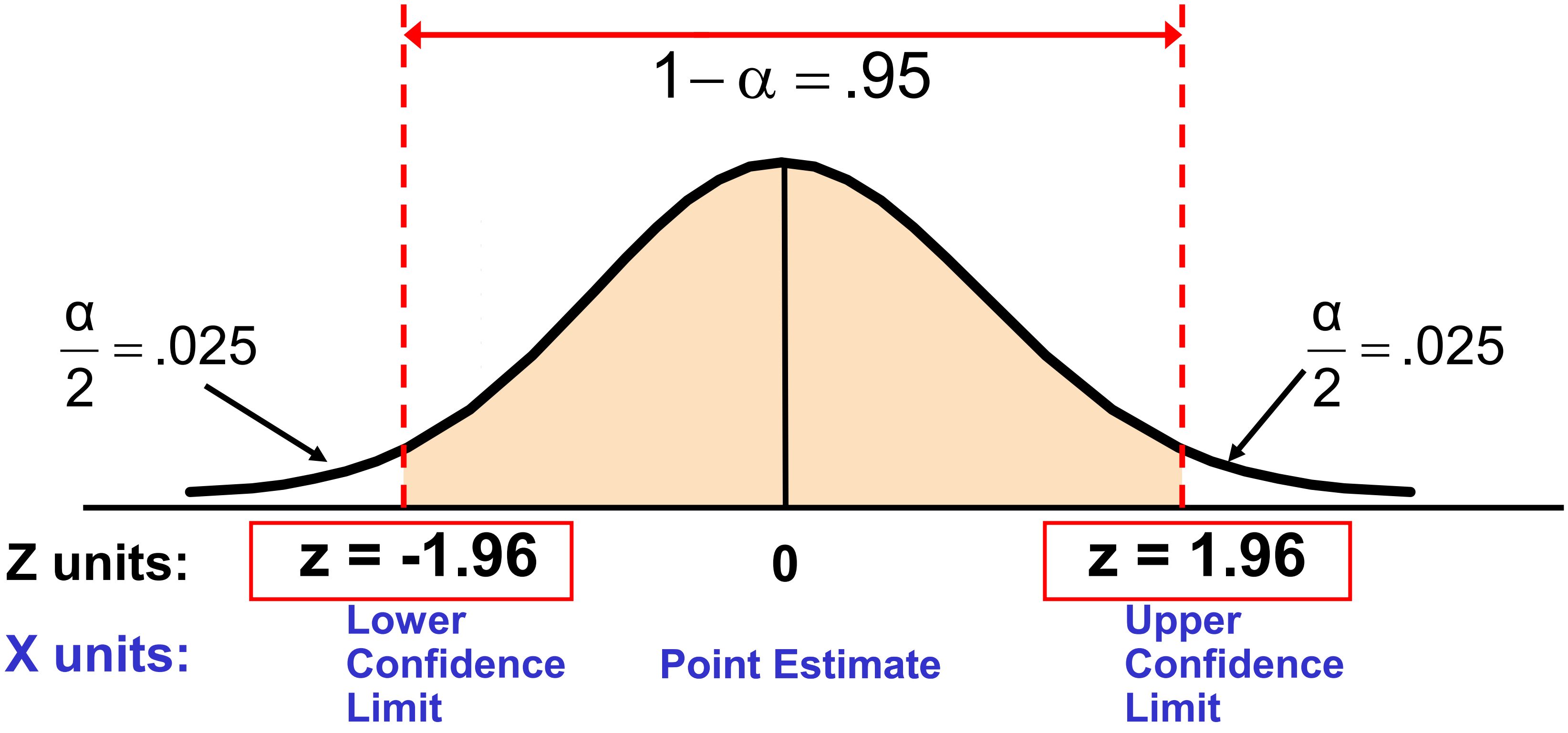
- Assumptions
  - Population variance σ^2 is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

(where $z_{1-\alpha/2}$ is $1-\alpha/2$ percentile of the standard normal distribution)

Percentiles of the standard normal

- Consider a 95% confidence interval:



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

Construction of confidence interval for the mean

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$P(Z_{a/2} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{1-a/2}) = 1-a \Rightarrow P(-Z_{1-a/2} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{1-a/2}) = 1-a \Rightarrow$$

$$P(-Z_{1-a/2}\frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < Z_{1-a/2}\frac{\sigma}{\sqrt{n}}) = 1-a \Rightarrow$$

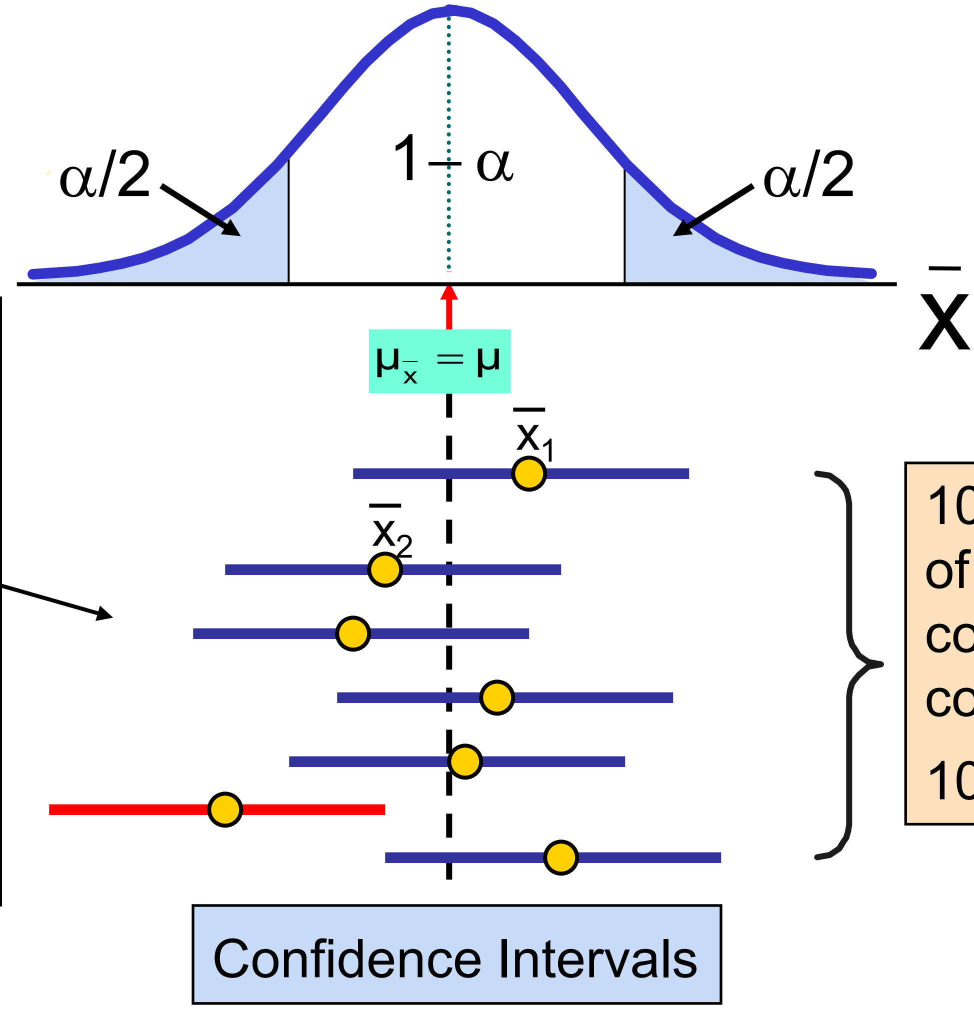
$$P(-\bar{x}-Z_{1-a/2}\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x}+Z_{1-a/2}\frac{\sigma}{\sqrt{n}}) = 1-a \Rightarrow$$

$$P(\bar{x}+Z_{1-a/2}\frac{\sigma}{\sqrt{n}} > \mu > \bar{x}-Z_{1-a/2}\frac{\sigma}{\sqrt{n}}) = 1-a \Rightarrow$$

$$P(\bar{x}-Z_{1-a/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+Z_{1-a/2}\frac{\sigma}{\sqrt{n}}) = 1-a$$

Intervals and Level of Confidence

Sampling Distribution of the Mean



Intervals extend from

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

to

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

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

100(α)% do not.

Example

- Consider the grades of 9 students:

  3, 8, 4, 11, 8, 6, 9, 10, 5

  We know from past exams that the population standard deviation is 1.5 (assume Normality)

- Construct a 95% confidence interval for the true mean grade of the population.

Example

- Solution:

$$\bar{x} = \sum_{i=1}^{n} x_i / n = (3+8+4+11+8+6+9+10+5)/9 = 64/9 = 7.11$$

$$Z_{1-a/2} = Z_{1-0.05/2} = Z_{0.975} = 1.96$$

$$\bar{x} \pm Z_{1\text{-}a/2} \frac{\sigma}{\sqrt{n}} = 7.11 \pm 1.96\,(1.5/\sqrt{9}) = 7.11 \pm 1.96 \cdot 0.5 =$$

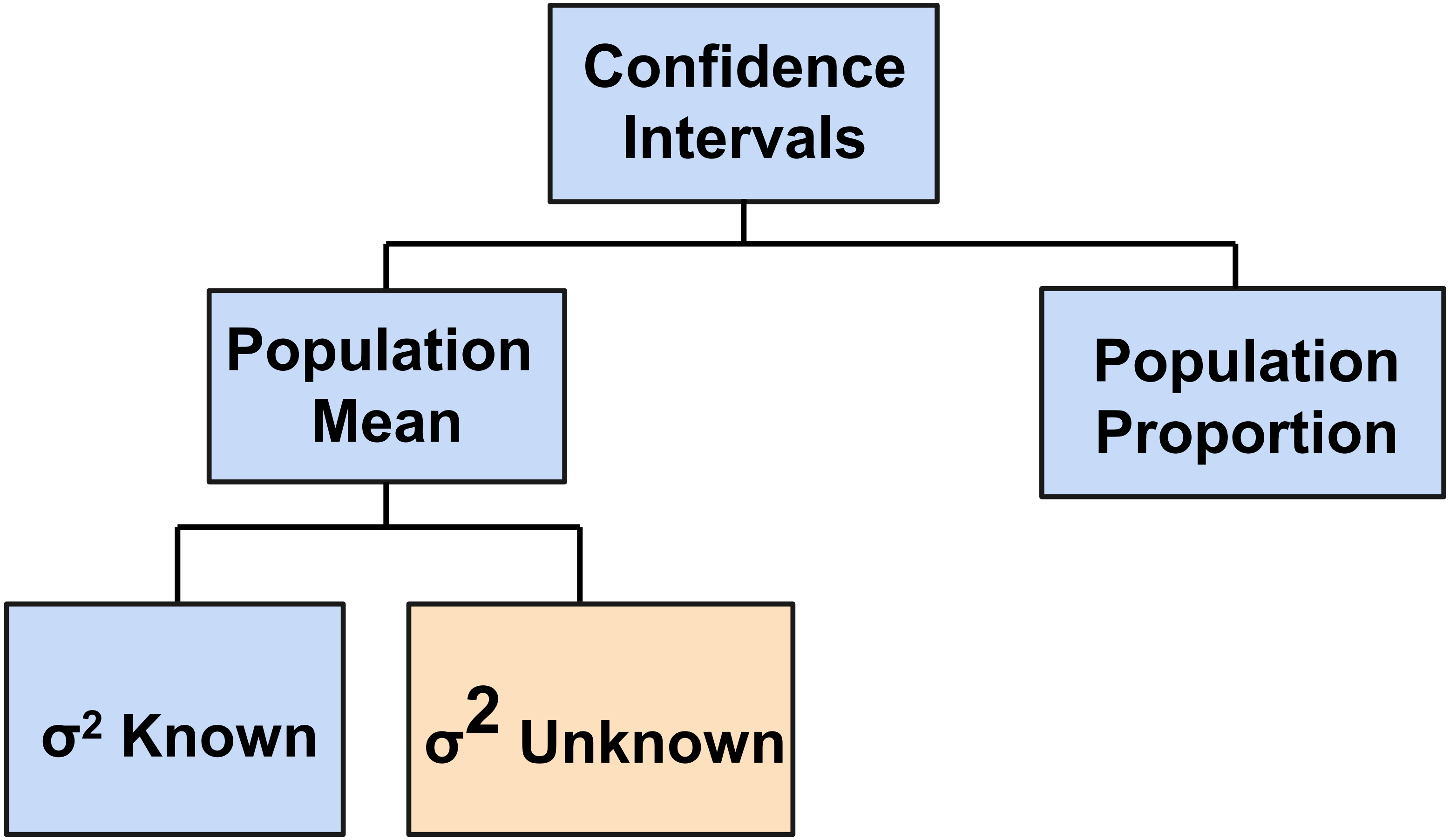
$$= 7.11 \pm 0.98$$

$$6.13 < \mu < 8.09$$

Interpretation

- We are 95% confident that the true mean grade is between 6.13 and 8.09
- 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)

- 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:

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

where $t_{n-1,1-\alpha/2}$ is the 1-a/2 percentile of the t distribution with n-1 degress of freedom

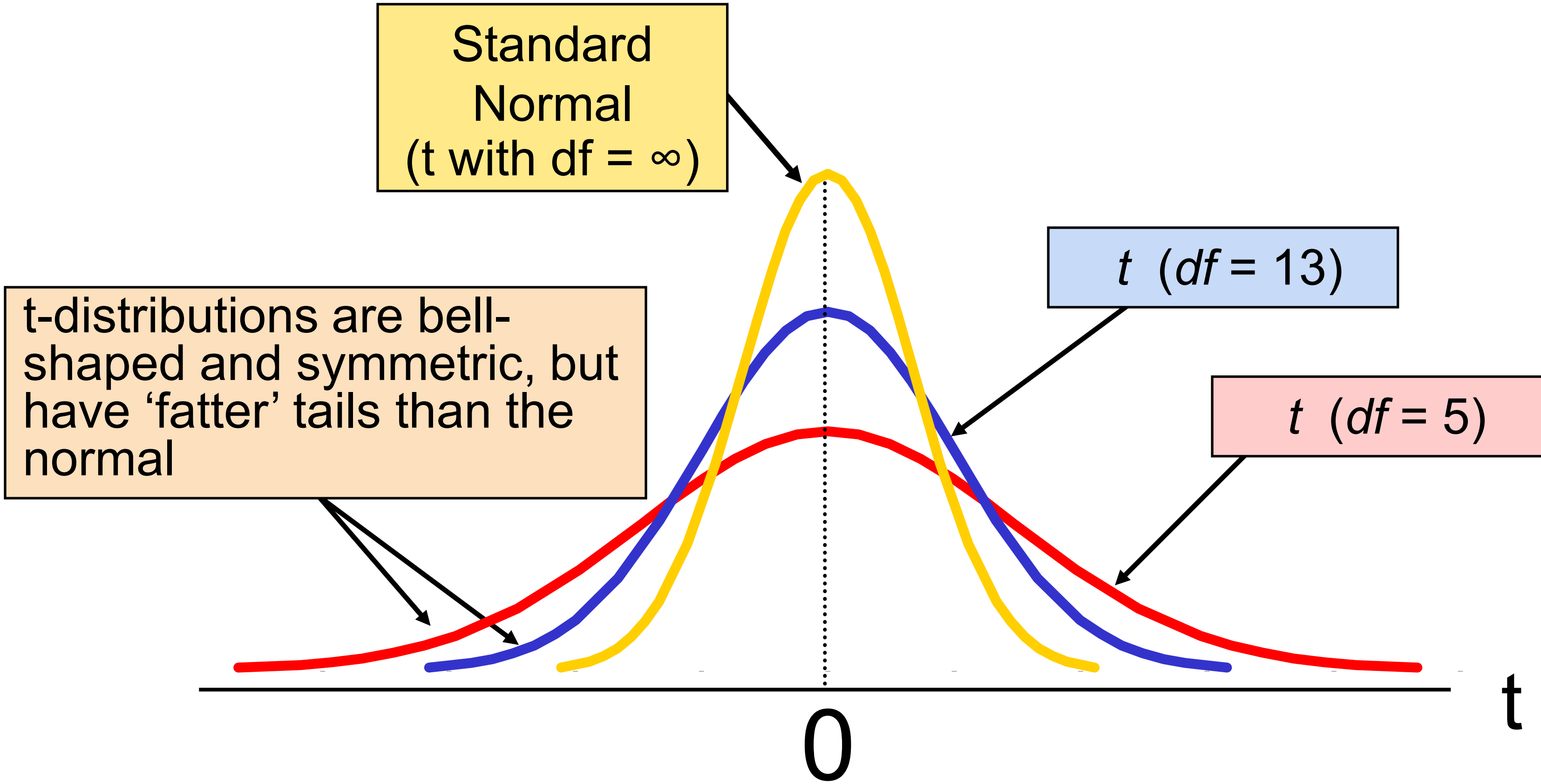
Student’s t Distribution

- The t is a family of distributions
- The t value depends on the degrees of freedom (d.f.)
  - The number of observations that are free to vary after the sample mean has been calculated

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

Student's t Distribution

Note: $t \longrightarrow Z$ as n increases



Example

- Consider the grades of 9 students:

  3, 8, 4, 11, 8, 6, 9, 10, 5

  (assume normality)

- Construct a 95% confidence interval for the true mean grade of the population.

Example

- Solution:

$$\overline{\mathrm{x}} = \sum_{i=1}^{n} x_i / n = (3+8+4+11+8+6+9+10+5)/9 = 64/9 = 7.11$$

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} =$$

$$= [(3-7.11)^2 + (8-7.11)^2 + (4-7.11)^2 + (11-7.11)^2 + (8-7.11)^2 +$$

$$+ (6-7.11)^2 + (9-7.11)^2 + (10-7.11)^2 + (5-7.11)^2]/8 =$$

$$= 60.89/8 = 7.61$$

$$\hat{\sigma} = \sqrt{7.61} = 2.76$$

Example

$$t_{n-1,1-\alpha/2} = t_{8,0.975} = 2.306$$

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

$$7.11 - (2.306)\frac{2.76}{\sqrt{9}} < \mu < 7.11 + (2.306)\frac{2.76}{\sqrt{9}}$$

$$7.11 - 2.12 < \mu < 7.11 + 2.12$$

$$4.99 < \mu < 9.23$$