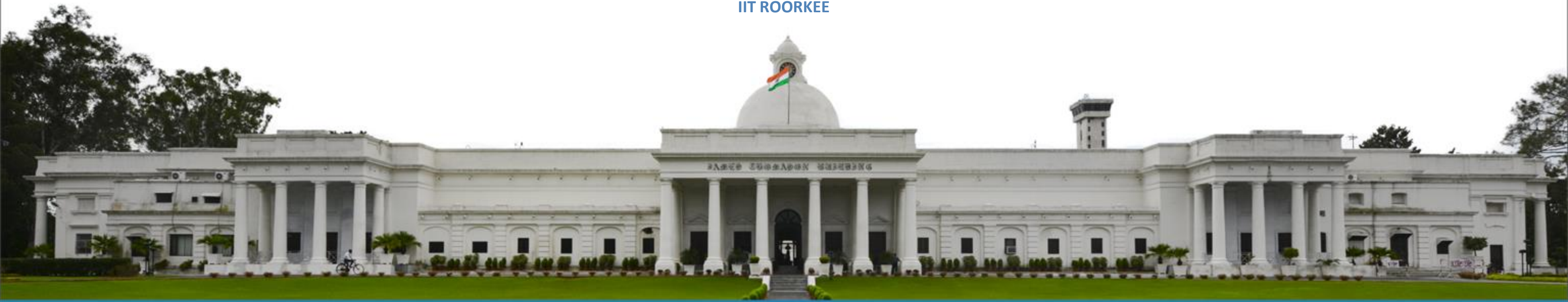
Confidence Interval Estimation: Single Population

Dr. A. Ramesh
Department of Management Studies
IIT ROORKEE



Goals

After completing this lecture, 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

- 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

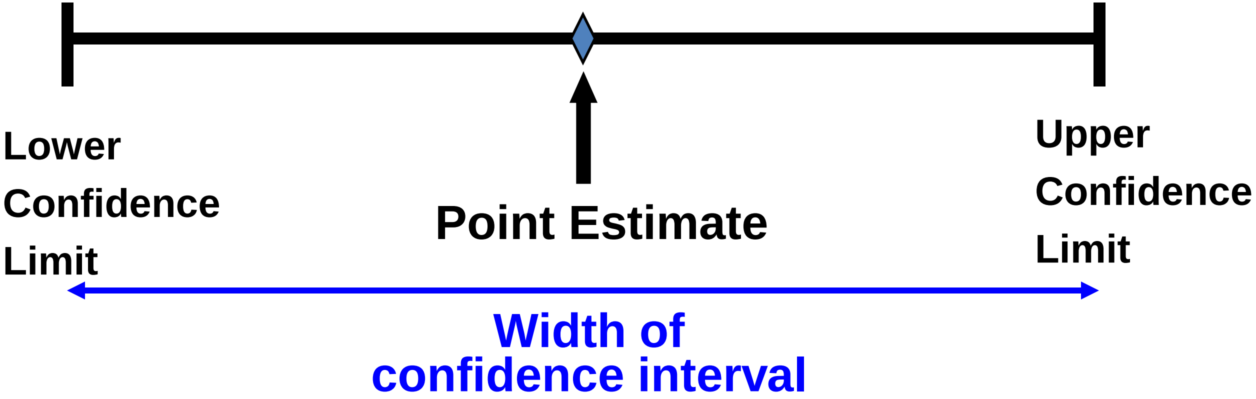
Definitions

- 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,

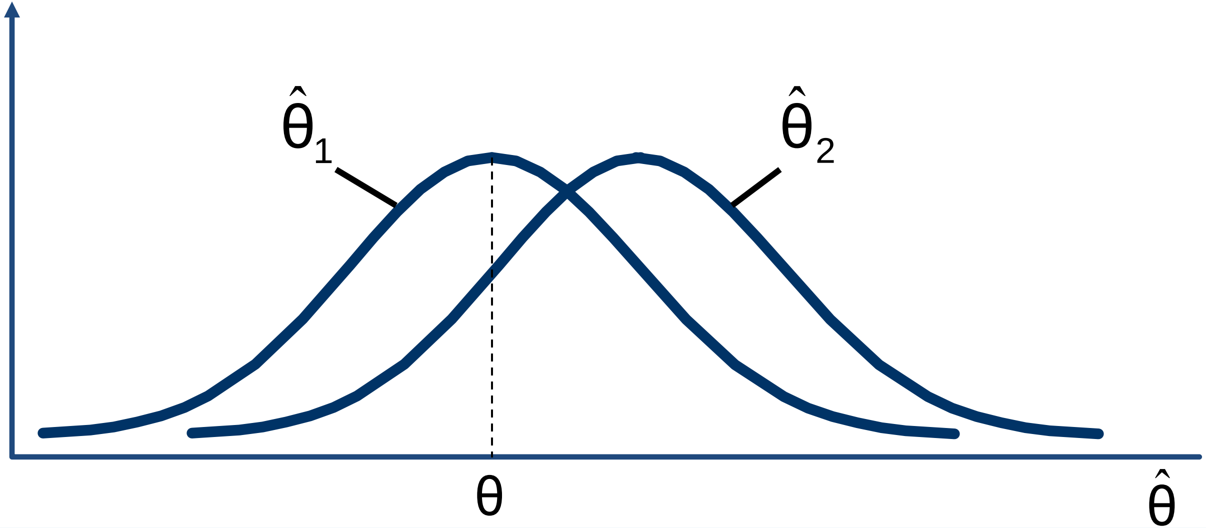
$$\boxed{E(\hat{\theta}) = \theta}$$

- 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

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

- 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 $\boxed{Var(\hat{\theta}_1) < Var(\hat{\theta}_2)}$
  - The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is the ratio of their variances:

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

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

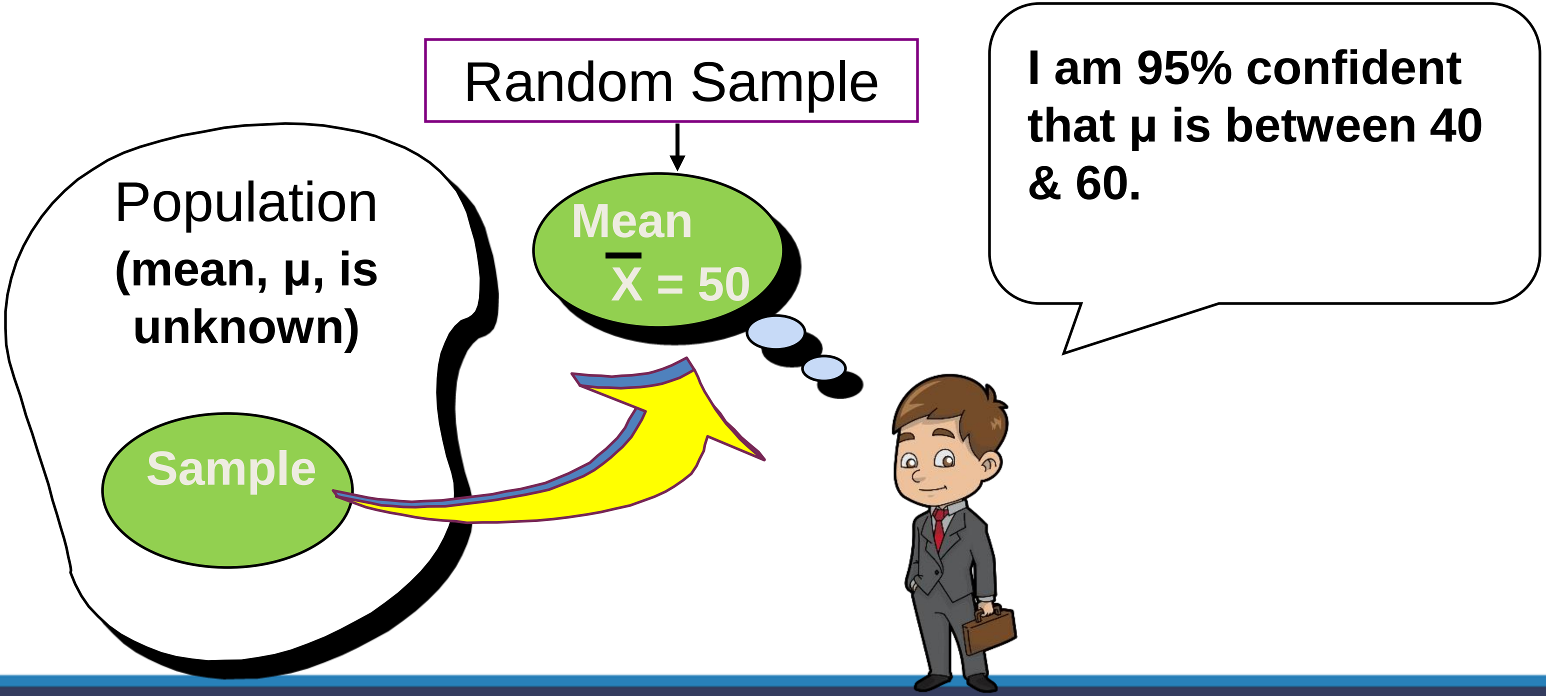
Confidence Interval Estimate

- 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

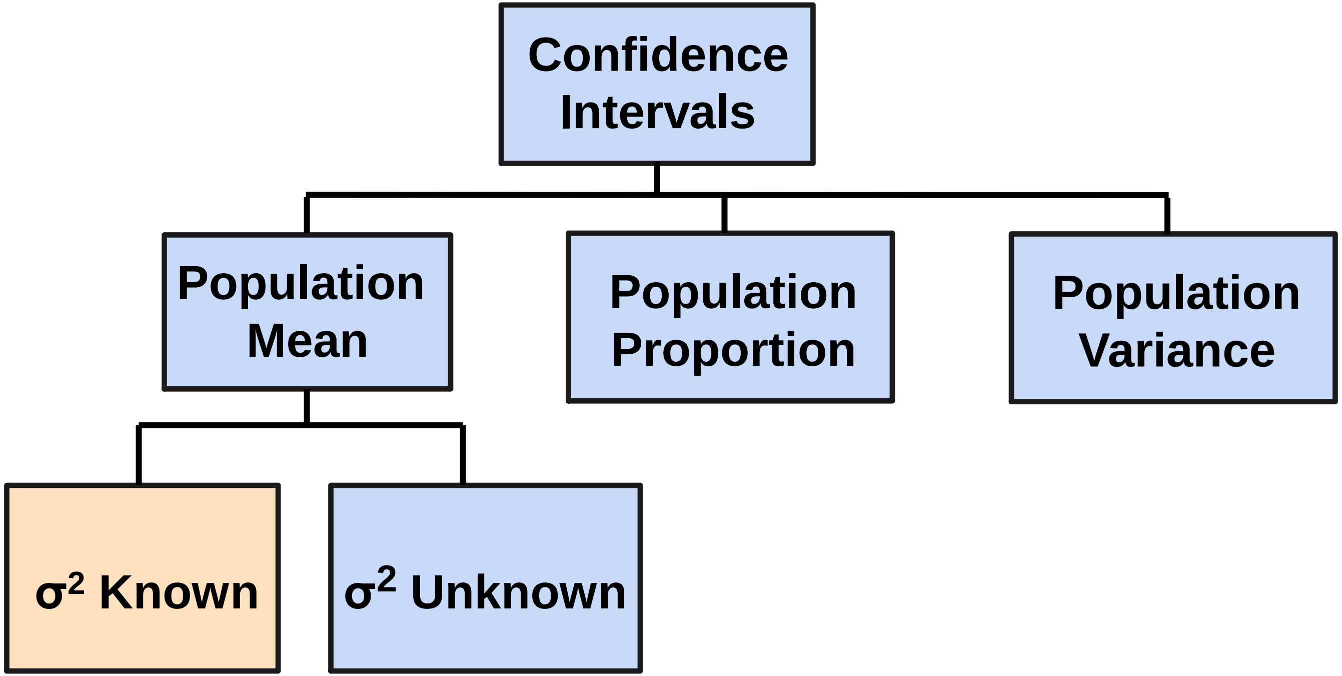
General Formula

- The general formula for all confidence intervals is:

$$\boxed{\textbf{Point Estimate} \pm \textbf{(Reliability Factor)(Standard Error)}}$$

- The value of the reliability factor 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:

$$\boxed{\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

(where $z_{\alpha/2}$ is the normal distribution value for a probability of $\alpha/2$ in each tail)

Margin of Error

- The confidence interval,

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

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

  where ME is called the margin of error

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

Reducing the Margin of Error

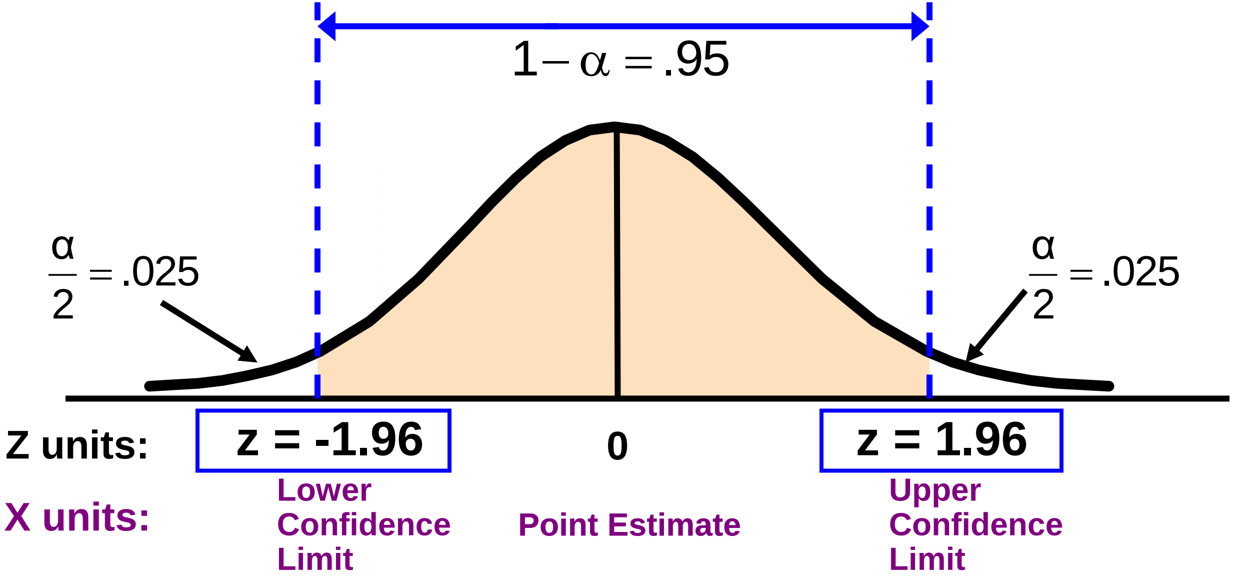
$$\boxed{ME = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

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

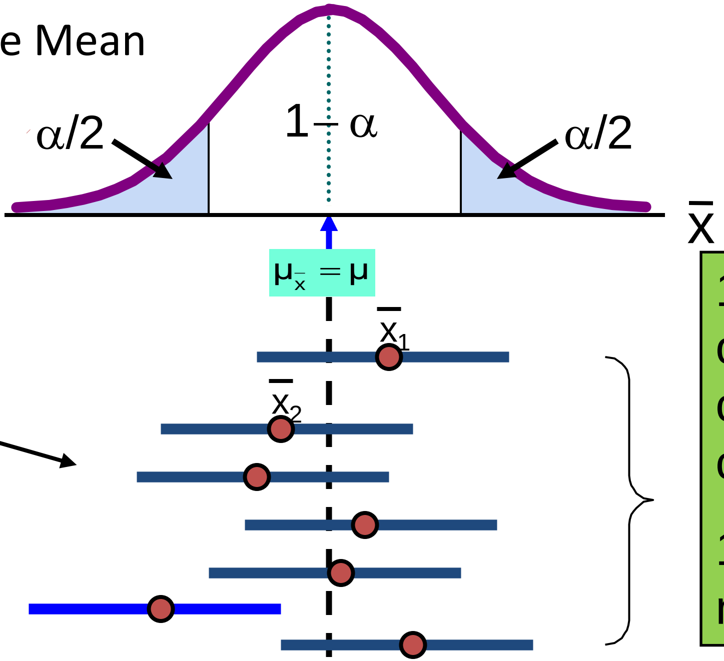
Common Levels of Confidence

- 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-α)% of intervals constructed contain μ;

100(α)% 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