Probability & Statistical Modelling

AQ077-3-2-PSMOD and Version VD1

ASIA PACIFIC UNIVERSITY OF TECHNOLOGY & INNOVATION

Estimation and Confidence Interval

Topic & Structure of The Lesson

- This topic is divided into 2 parts:
  - Part A: Sampling Distribution
  - Part B: Estimation & Confidence interval

Learning Outcomes

- **Part A: Sampling distribution:**
  - At the end of this section, you should be able to:
    - Recognise what a sampling distribution is and how it differs from other types of probability distribution
    - Make statistical statements that rely upon knowledge of the central limit theorem.

Key Terms You Must Be Able To Use

If you have mastered this topic, you should be able to use the following terms correctly in your assignments and exams:
(Prepare your own list)

- **Mean**
- **Proportion**
- **Sample size**
- **Central limit theorem**
- **Confidence interval for population mean**
- **Confidence interval for population proportion**

Sampling Distribution

- **Introduction**
  - Is the probability distribution of all possible sample means of n items drawn from a population.
  - Such a distribution exists not only for the mean but for any point estimate.
  - Properties
    - Very close to being normally distributed
    - The mean of the sampling is the same as the population mean.
    - It has a standard deviation which is called the standard error.

Standard Error

- The standard deviation of the sampling distribution.
- It measures the extent to which we expect the means from the different samples to vary because of the chance error in the sampling process.

Required to know how to:

- Determine the standard error of the
  - mean
    - with finite population size
      - And known population standard deviation
      - And unknown population standard deviation
    - With infinite population size
      - And known population standard deviation
      - And unknown population standard deviation
  - Proportion
    - With finite population size
    - With infinite population size

Standard error of the	Population finite	Population infinite
Mean - Known population standard deviation	$\frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$	$\frac{\sigma}{\sqrt{n}}$
- Unknown population standard deviation	$\frac{s}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$	$\frac{s}{\sqrt{n}}$
Proportion	$\sqrt{\frac{pq}{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$	$\sqrt{\frac{pq}{n}}$

Central limit theorem

- It states that as the sample size increases, the sampling distribution of the mean approaches the normal distribution in form, regardless of the form of the population distribution.
- For practical purposes, the sampling distribution of the mean can be assumed to be approximately normal, regardless of the population distribution whenever the sample size is at least 30.

- **Statistical inference**
  - Statistical inference can be defined as the process by which conclusions are drawn about some measure or attributes of a population based upon analysis of sample data.
  - Statistical inference can be divided into two types
    - Estimation
    - Hypothesis testing

Learning Outcomes

- **Part B: Estimation**
  - At the end of this section, you should be able to:
    - Make a confidence interval estimate of a population mean.
    - Make a confidence interval estimate of a population proportion,
    - Determine the appropriate sample size for interval estimation of means and proportions.

Key Terms you must be able to use

If you have mastered this section, you should be able to use the following terms correctly in your assignments and exams:
(Prepare your own list)

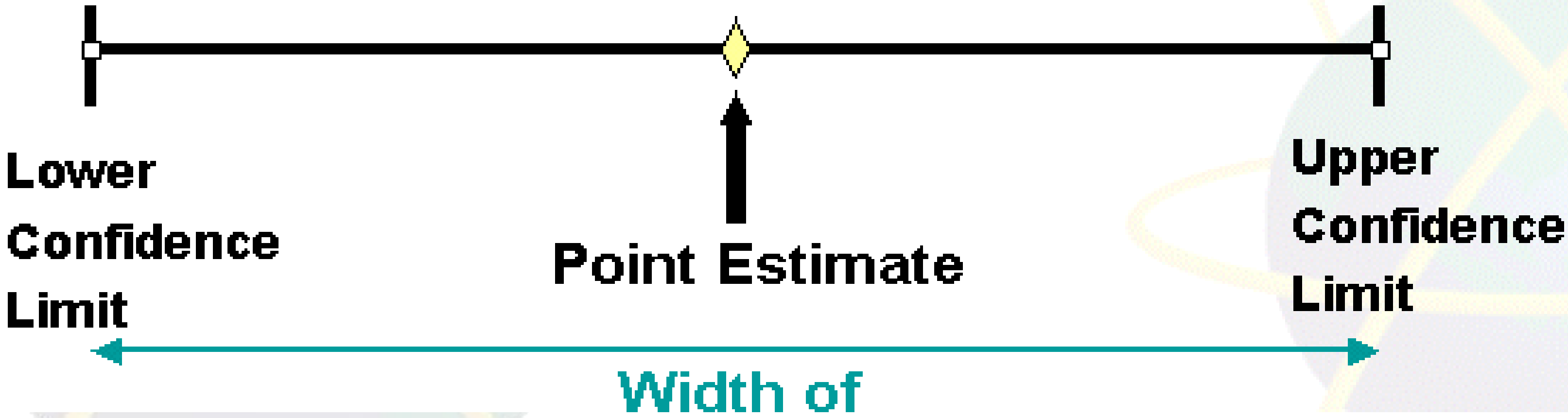
- **Confidence interval**
- **Level of significance**
- **Sample size**
- **Standard error.**

Estimation

- Introduction
  - Deals with the estimation of population characteristics from sample statistics
  - The distribution of sample means follows a normal curve.

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	$\bar{p}$

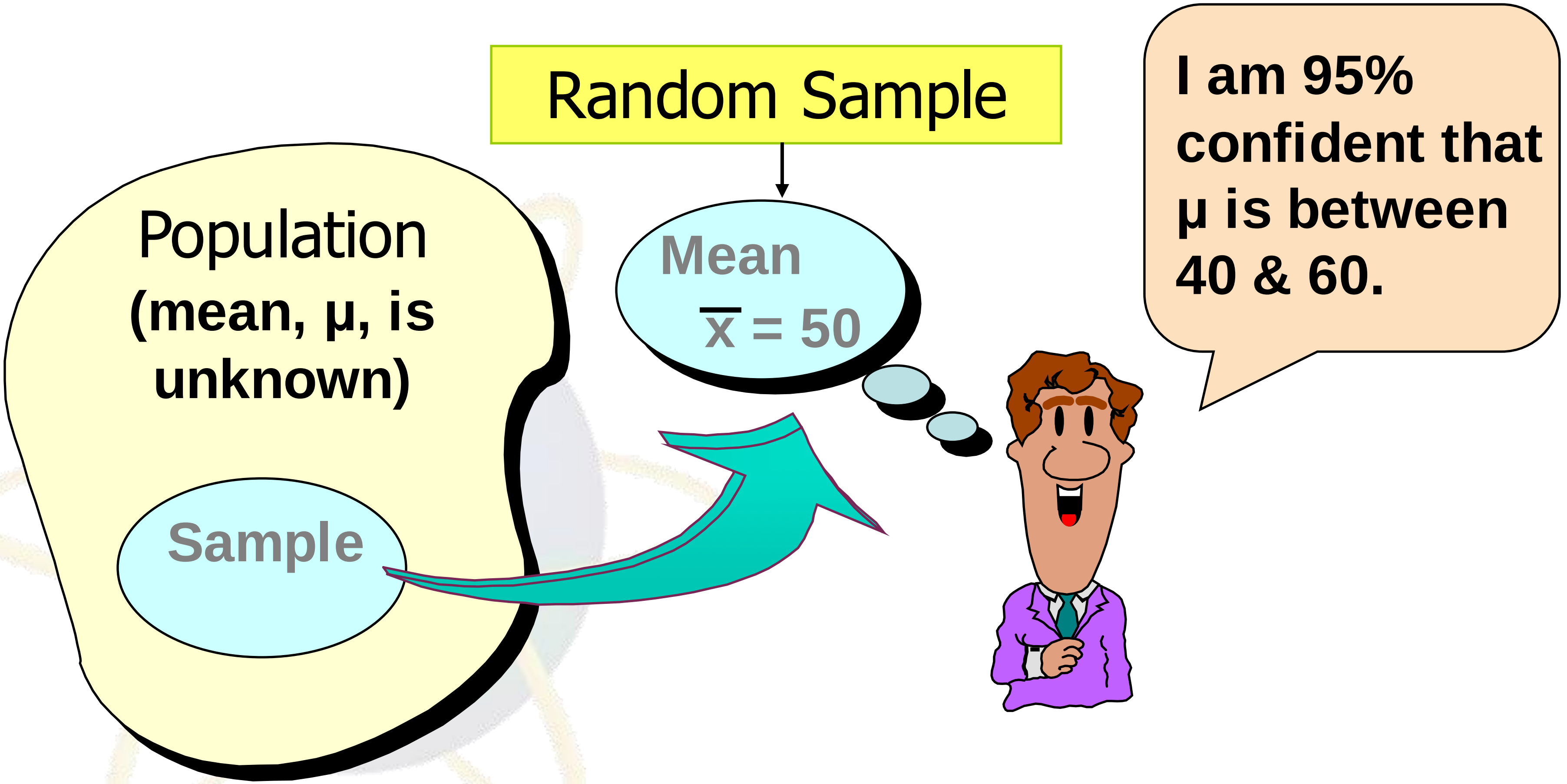
- How much uncertainty is associated with a point estimate of a population parameter?

- 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
      - Never 100% sure

Estimation Process



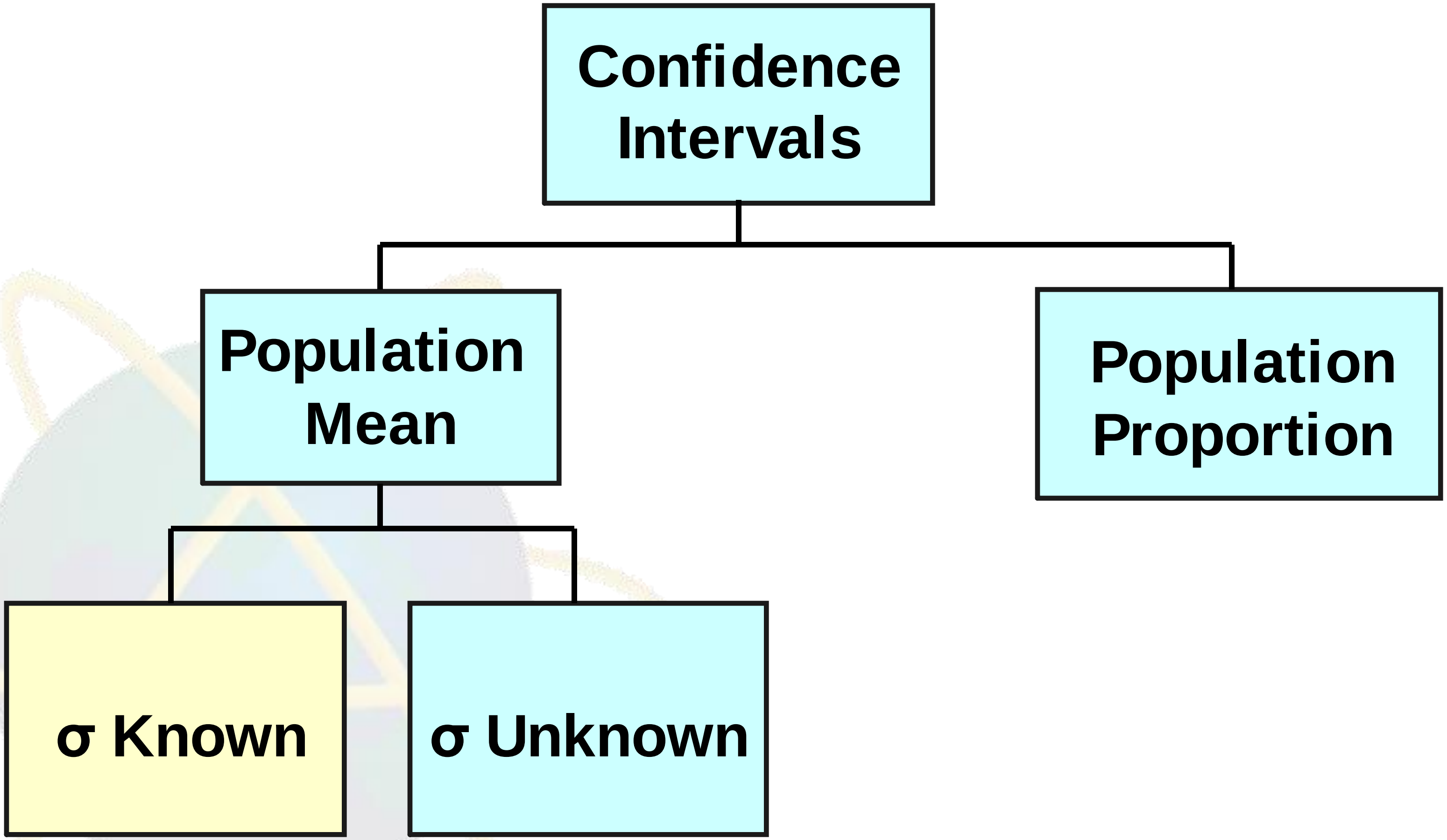
- **The general formula for all confidence intervals is:**

$$\boxed{\textbf{Point Estimate} \pm \textbf{(Critical Value)(Standard Error)}}$$

Confidence Level

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

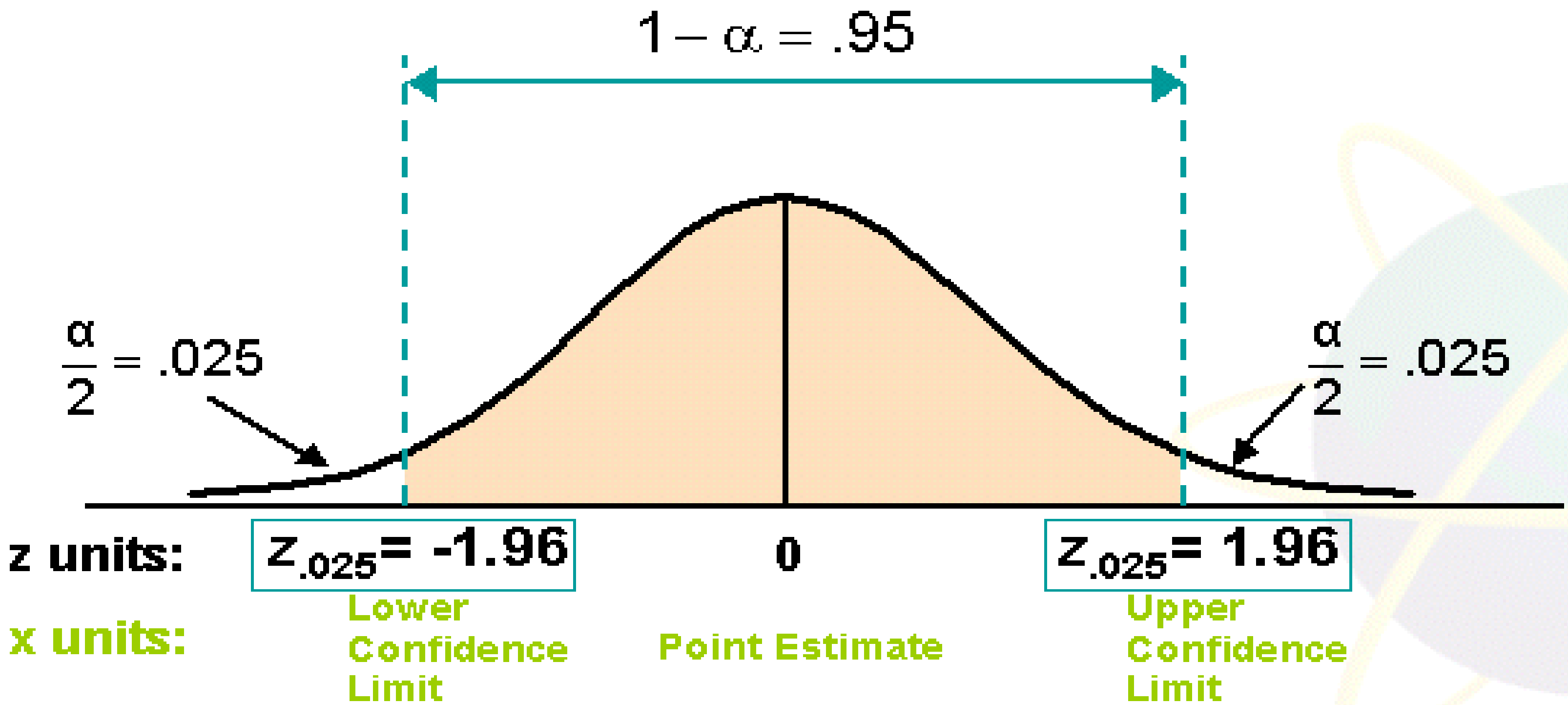
- Suppose confidence level = 95%
- Also written $(1 - \alpha) = .95$
- A relative frequency interpretation:
  - In the long run, 95% of all the confidence intervals that can be constructed will contain the unknown true parameter
- A specific interval either will contain or will not contain the true parameter
  - No probability involved in a specific interval


Confidence Intervals
Population Mean
Population Proportion
σ Known
σ Unknown

- 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

$$\boxed{\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

- Finding the critical value $\boxed{z_{\alpha/2} = \pm 1.96}$
  - Consider a 95% confidence interval:



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

Confidence Level	*Confidence Coefficient,* $1-\alpha$	*z value,* $z_{\alpha/2}$
80%	.80	1.28
90%	.90	1.645
95%	.95	1.96
98%	.98	2.33
99%	.99	2.575
99.8%	.998	3.08
99.9%	.999	3.27

- **Margin of Error (e): the amount added and subtracted to the point estimate to form the confidence interval**

Example: Margin of error for estimating μ, σ known:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

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

Factors Affecting Margin of Error

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

- **Data variation, σ :** **e $\Downarrow$ as σ $\Downarrow$**
- **Sample size, n :** **e $\Downarrow$ as n $\Uparrow$**
- **Level of confidence, 1 - α :** e $\Downarrow$ if 1 - α $\Downarrow$

Quick Review Question

- **Example 5.1:**
  - **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.**

- **95% confidence interval:**

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

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

$$= 2.20 \pm .2068$$

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
- An incorrect interpretation is that there is 95% probability that this interval contains the true population mean.

(This interval either does or does not contain the true mean, there is no probability for a single interval)

Example 2

After a particularly wet nights, 10 worms surfaced on the lawn. Their lengths, measured in cm, were

9.5 9.5 11.2 10.6 9.9 11.1 10.9

9.8 10.1 10.2

Assuming that this sample came from a normal population with variance 4, calculate a 95% confidence interval for the mean length of all the worms in the garden.

- **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**

- 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

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

- **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

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

Quick Review Question

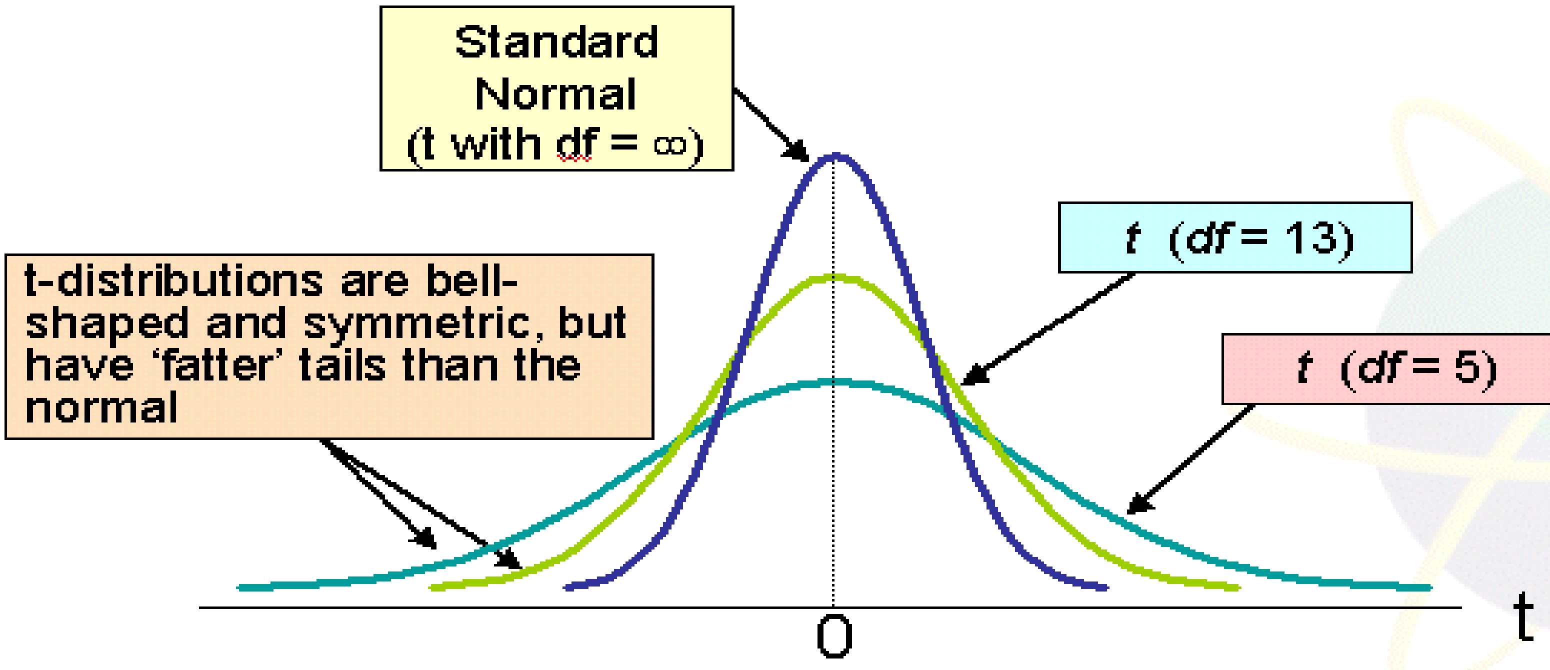
- **Example:** Suppose the mean of 3 numbers is 8.0
- Let $x_1 = 7$
  Let $x_2 = 8$
  What is x_3?

  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



Example 3

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

Approximation for Large samples

- Since t approaches z as the sample size increases, an approximation is sometimes used when $n \geq 30$:

Technically correct	Approximation for large n
$\bar{x} \pm t_{\alpha/2} \dfrac{s}{\sqrt{n}}$	$\bar{x} \pm z_{\alpha/2} \dfrac{s}{\sqrt{n}}$

Example 4

A random sample of 120 measurements taken from a normal population gave the following data:

n = 120, $\sum x = 1008$, s = 1.44

Find

(a) a 97% confidence interval

(b) a 99% confidence interval

Example 5

The 95% confidence interval for the mean paper length is (10.994 inches, 11.002 inches). This interval is based on results from a random sample of 100 sheets. Find the 99% confidence interval for the mean paper length, assuming that the paper length is normally distributed.

Determining Sample Size

- The required sample size can be found to reach a desired margin of error (e) and level of confidence (1 - α)
- **Required sample size, σ known:**

$$n = \frac{z_{\alpha/2}^{2}\,\sigma^{2}}{e^{2}} = \left(\frac{z_{\alpha/2}\,\sigma}{e}\right)^{2}$$

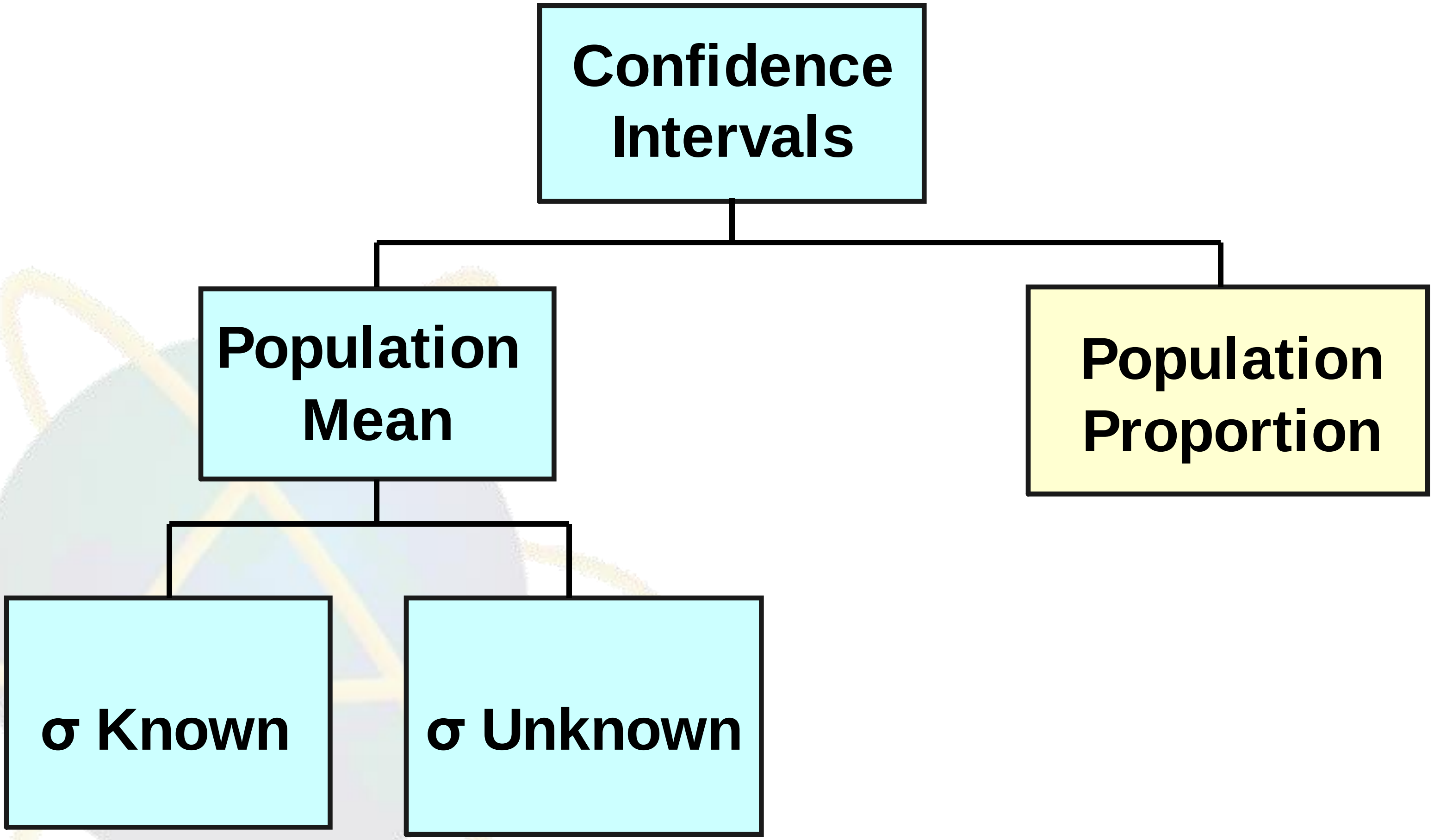
Example 6

- Example:
  - If σ = 45, what sample size is needed to be 90% confident of being correct within ± 5?

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{e} \right)^2 = \left(\frac{1.645(45)}{5} \right)^2 = 219.19$$

So the required sample size is **n = 220**

(Always round up)


Confidence Intervals
Population Mean
Population Proportion
σ Known
σ Unknown

- **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 ($\bar{p}$)**

- **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:

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

- **Confidence intervals endpoints**
  - Upper and lower confidence limits for the population proportion are calculated with the formula

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

where

- z is the standard normal value for the level of confidence desired
- $\bar{p}$ is the sample proportion
- n is the sample size

Example 7

- 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

$$p = 25/100 = .25$$

$$S_p = \sqrt{p(1-p)/n} = \sqrt{.25(.75)/n} = .0433$$

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

$$0.1651 \ldots\ldots 0.3349$$

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

- **Changing the sample size**
  - **Increases in the sample size reduce the width of the confidence interval.**
  - **Example:**
    - **If the sample size in the above example is doubled to 200, and if 50 are left-handed in the sample, then the interval is still centered at .25, but the width shrinks to**

.19 …… .31

Finding the required sample size for proportion problems

Define the margin of error:

$$e = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Solve for n:

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

p can be estimated with a pilot sample, if necessary (or conservatively use p = .50)

Example 8

- **What Sample size ?**
  - **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 = .12)

For 95% confidence, use $Z = 1.96$

$E = .03$

$\bar{p} = .12$, so use this to estimate p

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

Summary of Main Teaching Points

- Estimation on population mean and proportion
- Confidence interval on population mean and proportion

Question and Answer Session

Q & A

What we will cover next

- Hypothesis Testing