SAPIENZA
UNIVERSITÀ DI ROMA

UNIVERSITÀ
DEGLI STUDI DELLA
TUSCIA

Statistics for business and decision making (SBDM)

Prof.ssa Ilaria Benedetti

6. Inferential Statistics: Point estimates and interval estimates

Statistics descriptive →

A set of methods for representing and interpreting a set of data with the aim of describing and summarising its characteristics.

Statistics inferential →

Methods for estimating a characteristic (parameter) of the population and making decisions about the population based on observation of the sample

Population, Population Parameters and Sample

A **population is the** set of all units under study

-All potential voters in the next election
-All customers of a certain company
-All invoices issued in 2011

Population **parameters** are constants describing characteristic aspects of the character distribution in the population (e.g. mean, variance and proportion).

A **random sample** is a subset of the population chosen so that the probability of drawing each unit is known.

-Some voters randomly selected for an interview
-Selected customers for a satisfaction interview
-Some invoices randomly selected for verification

Why is a sample used?

A sample allows statistical results to be obtained with sufficiently high precision

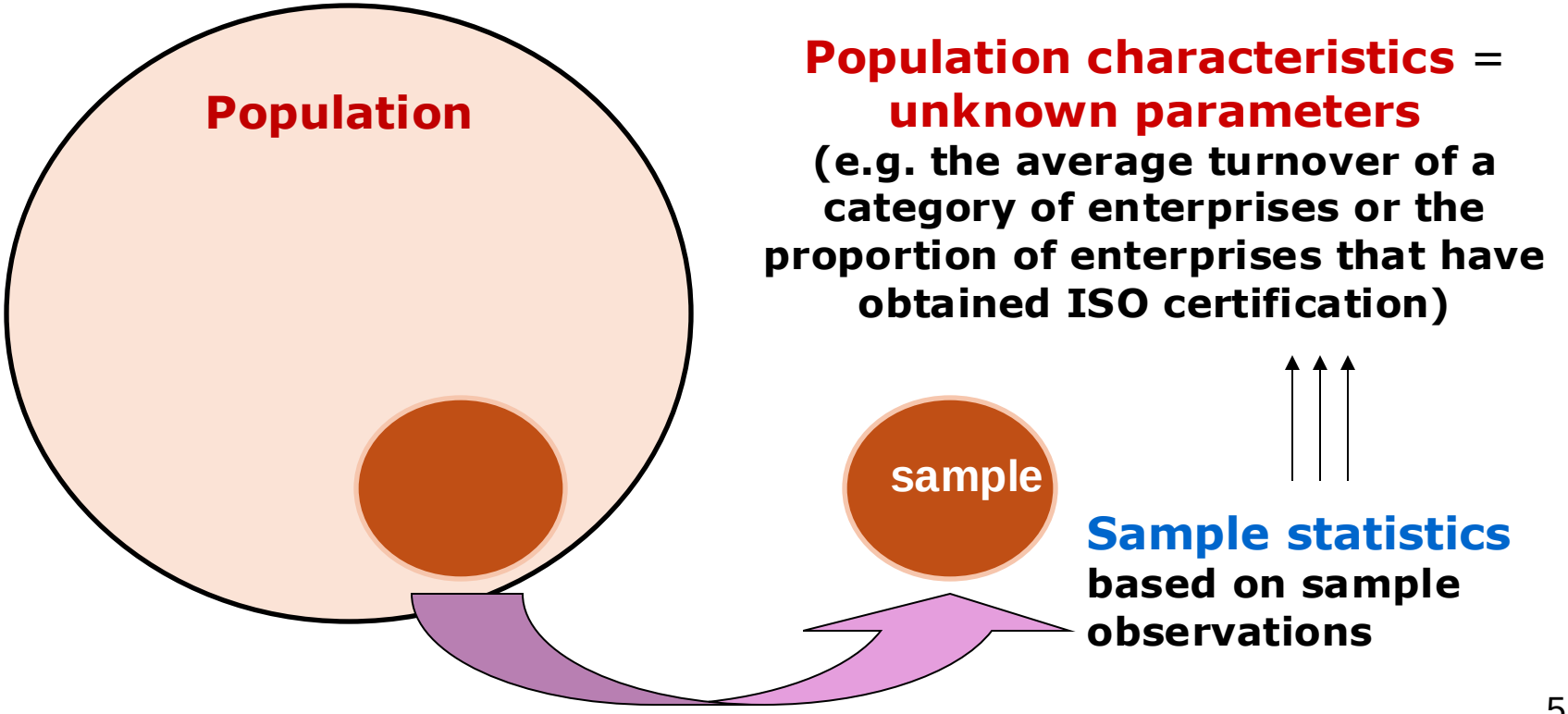
It has considerable advantages over a census

- Reduced costs
- Timeliness
- Checking the reliability of information
- Simpler organisation

Using sample data, inference can be made about population characteristics.

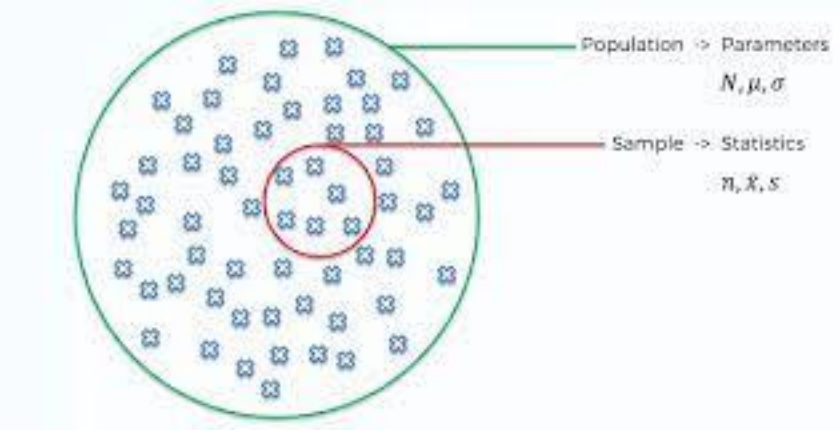
Statistical inference

Statistical inference is made by using information collected on a sample to learn about unknown parameters of the population.



Sample

- A sub-set of the units in a population, expected to represent the whole population
- By measuring data on a sample
  - Information on the entire population is gathered at a lower cost compared to censuses
  - Some margin of error is necessarily accepted



Definition: Parameters and Estimators

Every quantity in the population (parameter) has its analogue in the sample (statistic).

- POPULATION → *parameters* (μ, σ^2,π)
- SAMPLE → *statistics* ($\bar{x}$, S^2, p)

It is therefore natural to try to estimate a parameter of interest (e.g. μ) with the corresponding statistic calculated on the sample (e.g. $\bar{x}$).

When a statistic is used for inferential purposes to estimate a parameter, it is called an ***estimator*** (e.g. $\bar{x}$ is an estimator of μ).

Sample distribution

In practice, a single sample of fixed size n is randomly drawn from a population.

For inferential purposes, we hypothetically consider all samples of size n that can be extracted from the population (sample space or sample universe).

On each sample we calculate the sample statistic

The set of results constitutes the sample distribution (probability distribution of a statistic).

Sampling with repetition from a finished pop.

Population of N=4 companies

Company	Investments	Legal form
A	140	SpA
B	150	SpA
C	120	Srl
D	190	SpA

Parameters of interest:
- mean μ
- variance σ2

$$\mu = 150 \text{ (popolation mean)}$$
$$\sigma^2 = 650$$
$$\sigma = 25{,}50$$

All samples sorted with n=2 repetition are extracted…

Sample space (N=4;n=2)

Sample	Obs. In the sample	Average investments
1 AA	**140,140**	**140**
2 AB	**140,150**	**145**
3 AC	**140,120**	**130**
4 AD	**140,190**	**165**
5 BA	**150,140**	**145**
6 BB	**150,150**	**150**
7 BC	**150,120**	**135**
8 BD	**150,190**	**170**

Sample	Obs. In the sample	Average investments
9 CA	**120,140**	**130**
10 CB	**120,150**	**135**
11 CC	**120,120**	**120**
12 CDS	**120,190**	**155**
13 DA	**190,140**	**165**
14 DB	**190,150**	**170**
15 DC	**190,120**	**155**
16 DD	**190,190**	**190**

I may be particularly 'lucky' if I draw sample 6, with a mean of 150 (which coincides with the value of the population mean).
What is the probability of this occurrence? **1 in 16**

I may be particularly 'unlucky' if I draw sample 16, with a mean of 190 (which is very different from the value of the population mean). What is the probability of this occurring? **1 in 16**

Sample distribution of the mean

These are the distinct values of the mean that we would observe if we extracted all possible samples of 2 items.

Average investment values	Freq rel or probab.
120	1/16=0,062
130	2/16=0,125
135	2/16=0,125
140	1/16=0,062
145	2/16=0,125
150	1/16=0,062
155	2/16=0,125
165	2/16=0,125
170	2/16=0,125
190	1/16=0,062
Total	16/16=1,000

Indicates the relative freq. of samples on which the average calculation results in the corresponding value in the first column.

Inference and sampling error

- Prior knowledge on the *probability* of all potential samples allows ***statistical inference***

- *Statistical inference* is the generalization of sample statistics (*parameters*) to the target population, subject to a margin of uncertainty, or ***sampling error***.

- The sampling error can be estimated and used to assess the **precision** of sample estimates

- The ***sampling error*** is only a portion of the survey error (which also includes non-sampling error), but has the advantage that it can be estimated and controlled using the information on the *sampling method*

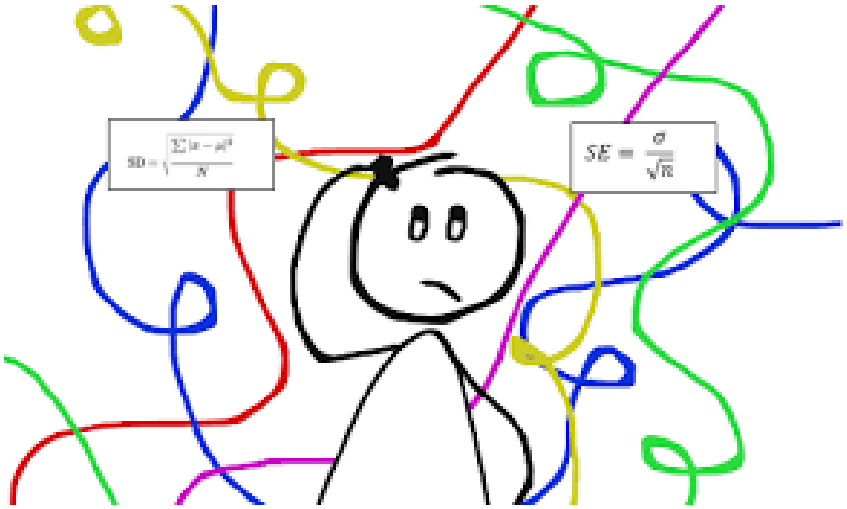
Sampling concepts

The sampling error depends on:

- The sampling fraction (ratio between sample size and population size)
  - Larger sampling fractions increase precision
  - However, the gain in precision decreases as the sampling fraction increases
- *The data variability in the population*
  - The less variable the population data, the more precise the sample estimates
  - However, population variability is rarely known and generally estimated
- *The precision of the sample estimator*
  - **Precision** of an estimator (measured through the **standard error of the estimator**) is the variability of an estimate across multiple measurements (across different samples)

Standard deviation vs. standard error

- The *standard deviation* measures the *variability* of a given variable (e.g. X) within the population or sample
- The ***standard error*** is a *precision measure* which refers to the variability of the *sample estimator* (e.g. the sample mean) across multiple estimates
- ***The standard error depends on the standard deviation but they are not the same concept***



Accuracy and precision

Accuracy: the degree to which the sample estimate is close to the true population value – maximum accuracy is obtained when the estimate equals the true population value *(st. deviation)*

Precision: the variability of the sample estimate in repeated measurements (across different samples) – maximum precision is obtained when the estimate is the same across all samples *(st. error)*

The standard error of an estimator is a measure of precision

The indirect problem and inference

- If one extracted **all potential samples** from a population (*the sampling space*) then the sampling distribution would be **exactly known**
- However, this would be a quite stupid exercise – given that the true population parameter would be already known

- Thus, statisticans are interested in the ***indirect problem***
  - **Only one sample is extracted**
  - **Only the sample statistics are known**
  - **The sampling distribution is not known exactly,** but it can be ascertained from the probabilistic sampling method

  - Given the sampling distribution and the sample statistics, one obtains estimates of the true population parameters through **statistical inference**

Example – sample mean

Extract two elements from a population of four

POPULATION: A=4; B=1; C=3; D=4

Pop. Average=3

SAMPLING SPACE:

AB – Sample mean= 2.5

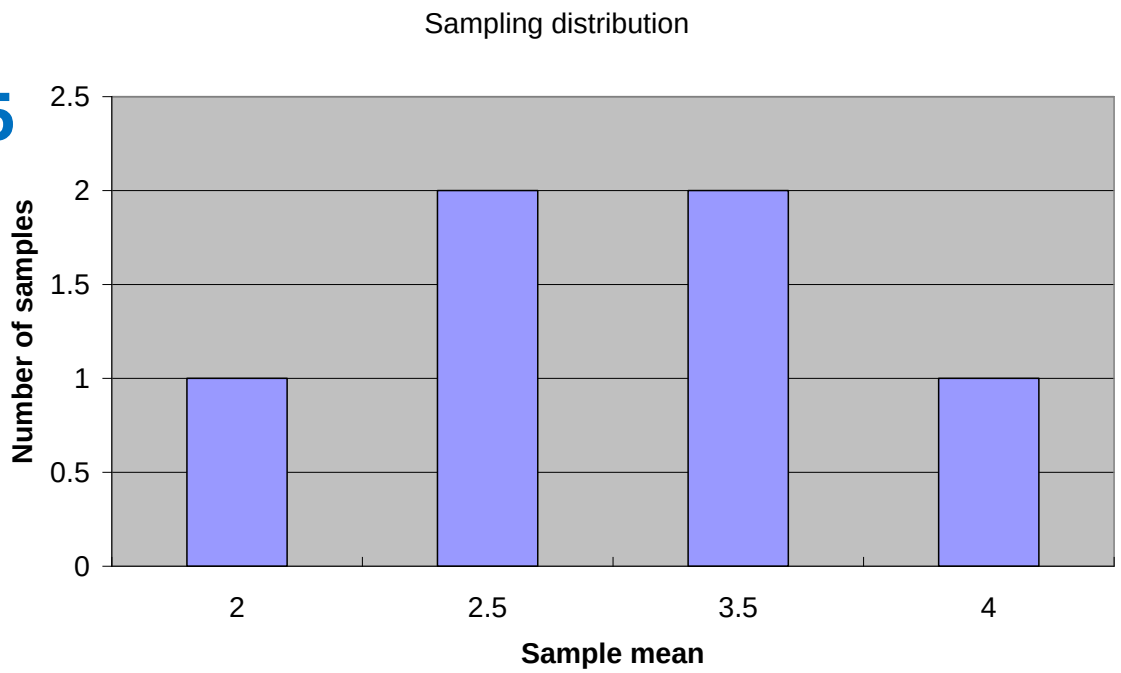
AC – 3.5

AD – 4

BC – 2

BD – 2.5

CD – 3.5



Example – the sampling distribution

- The ***average of all sample means*** in the sampling space is **3** (equal to the true population mean)
- None of the extracted samples exactly reflects the population (none has a mean of 3)
- The **mean absolute error** which we commit by observing only two out of four population units is 0.667 – this is a direct measure of ***sampling error***

Example – the indirect problem

- Suppose we have observed only one sample which is extracted randomly

- The probability extraction method and the sample observations allow us to
  - ***Obtain an estimate of the population mean***
  - ***Obtain a precision estimate (the sampling error)***

- By combining the sample estimate with the sampling error, one can draw ***inference*** on the true population value, for example by defining a bracket which is likely to include the true value

Example results

- Suppose we extract the sample AB with a simple random sampling
  - *The sample mean is 2.5*
  - *The mean error **within the sample** is 0.5*
- With a very rough (and inexact) assumption (that the mean error within the sample reflects the sampling error), we might claim that the true population value lies between [**2.5-0.5**] and [**2.5+0.5**], that is between two and three

- *This is a rough example, but with large samples and probability theory, knowledge based on a single sample can lead to accurate conclusions on the whole population, accounting for sampling error*

In practise…

- One sample is extracted
- Sample means and sample standard deviation are obtained
- An estimate of precision is obtained through an estimate of the *standard error of the mean*, which is a function of the sample standard deviation and the sample size
- Using the sample mean and the measure of precision one draws conclusion on the population mean

Population parameters
(in a population of N elements)

- **Mean** $$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- **Variance** $$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$$

- **Standard deviation** $$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2}$$

Sample statistics

- ***Sample mean***

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- ***Sample variance***

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

unbiasedness

- ***Sample standard deviation***

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Sample statistics (unknown pop. variance)

Mean case

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Proportion case (p)

Sample standard deviation of X

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

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

Standard error of the mean/proportion

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}}$$

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

PRECISION of sample

estimates

$\bar{x}, S^2$ and p are estimators

An estimator of the parameter $\boldsymbol{\theta}$ of the population of interest is Defined as a STATISTICS T=$t(X_1, X_2, \ldots, X_n)$ used to estimate The true parameter of the population $\boldsymbol{\theta}$

Properties of estimators

1- Unbiasedness

Considering a random sample $(X_1, X_2, \ldots, X_n)$ from a population Whose parameter θ is to be estimated.
Let T=$t(X_1, X_2, \ldots, X_n)$ one of the possible estimators of θ
Consider the estimation error (difference between the estimator T and true parameter θ (unknown): T- $\boldsymbol{\theta}$

The expected value of this random variable is:

E(T- θ)

It may be positive, negative or zero…

- In the first case T underestimate on average param θ
- In the second case T overestimate on average param θ

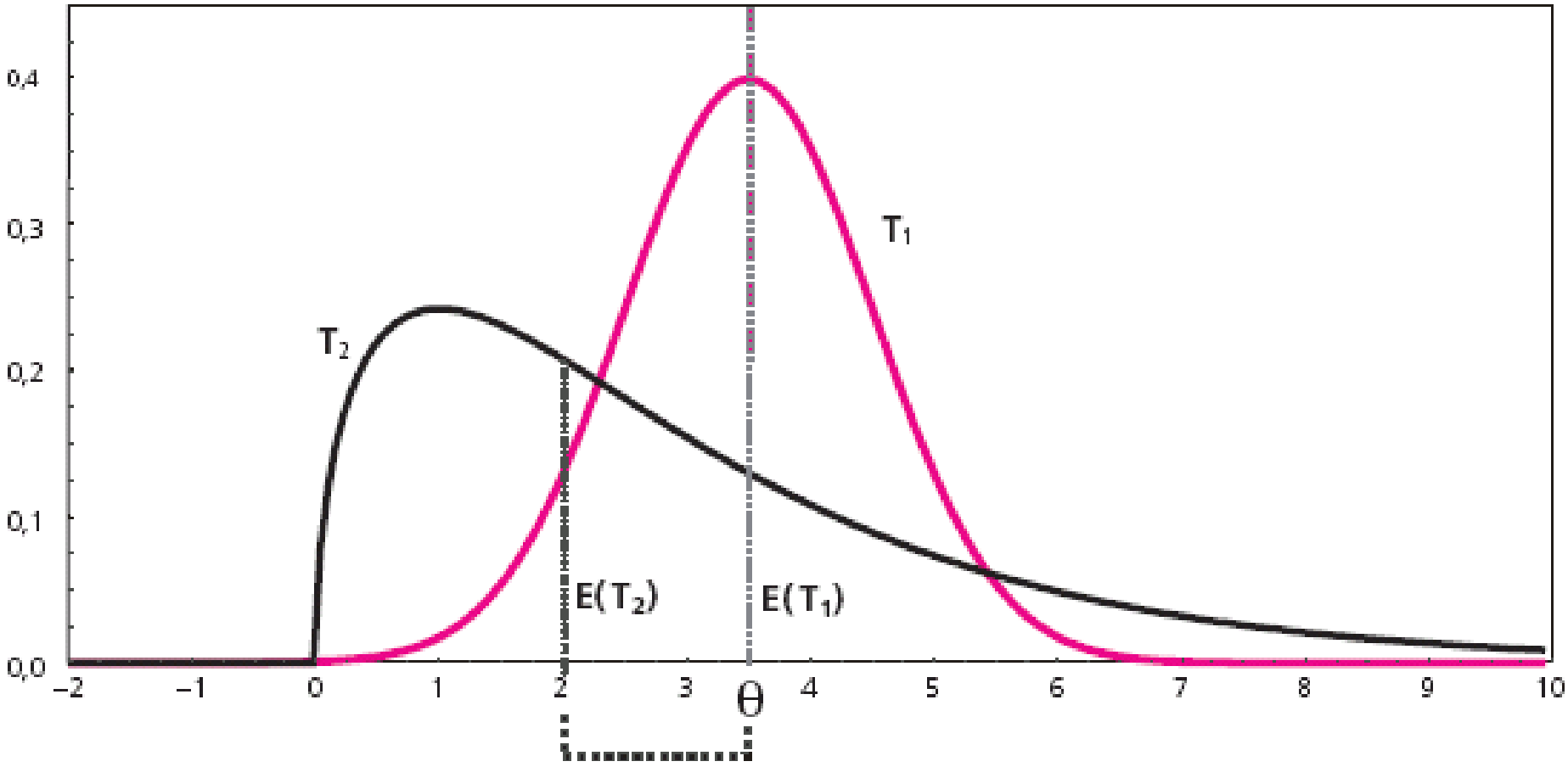
Consequently, it would be desiderable that:

E(T- θ)=0

or equivalently **E(T)=0**

Let $X_1, X_2, \ldots, X_n$ a random sample drawn from a population whose parameter θ is to be estimated.
An estimator T=$t(X_1, X_2, \ldots, X_n)$ of the parameter θ is unbiased if **E(T)= θ**
For any possible value of θ.
The difference is called **BIAS: B(T)=E(T)- θ**

1- Unbiasedness (cont..)



T2 Bias

2- EFFICIENCY

(In term of closeness of estimates to the true value θ).

The general used measure is: $E|T-\theta|^2$

Let $X_1, X_2, \ldots, X_n$ a random sample drawn from a population whose parameter θ is to be estimated. Then the mean square error of the estimator $T=t(X_1, X_2, \ldots, X_n)$ of the parameter θ is defined as:

$MSE=E(T-\theta)^2$

The MSE of an estimator T can be decomposed as:

$$MSE(T)=VAR(T)+[B(T)]^2$$

If an estimator is UNBIASED MSE and variance coincide

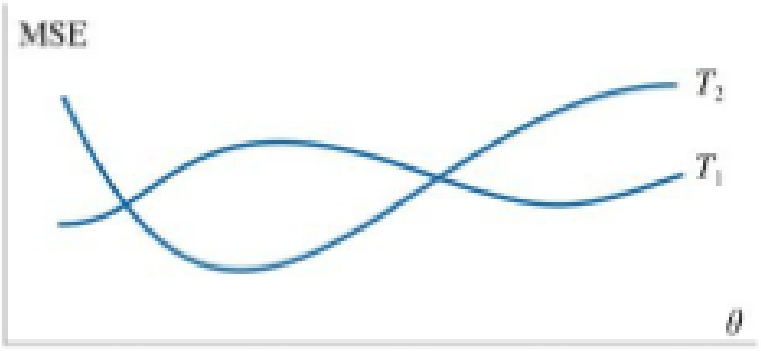
$$MSE(T)=VAR(T) \text{ and } E(T)=\theta$$

2- EFFICIENCY (cont…)

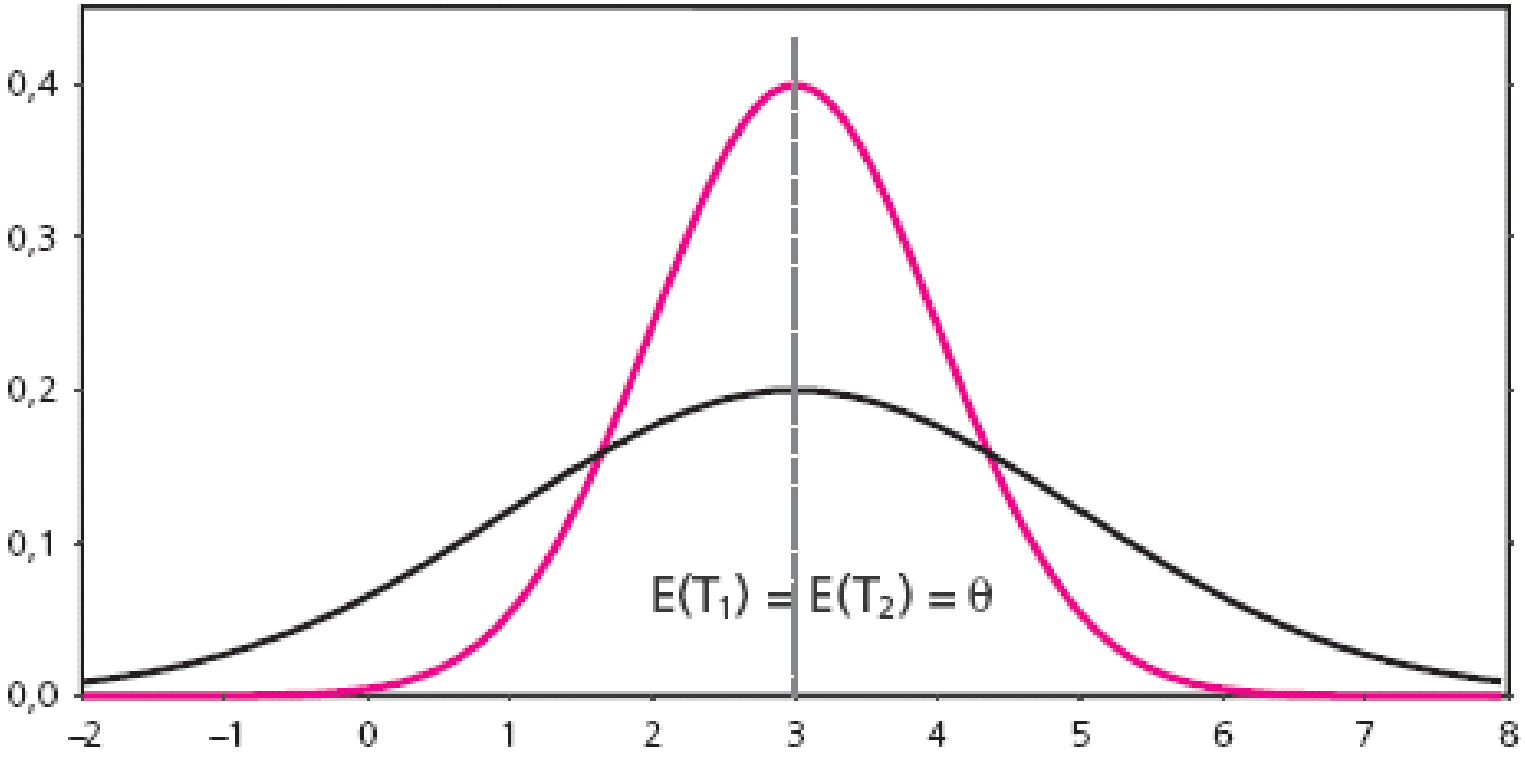
Let T1=$t1(X_1, X_2, \ldots, X_n)$ and T2=$t2(X_1, X_2, \ldots, X_n)$ be 2 alternative estimators of the parameter θ
T1 is said to be more efficient than T2 if
MSE(T1)≤ MSE(T2) *for all* θ

It is important to note that MSE may depend on θ
T1 is considered to be more efficient than T2 only if its MSE:

- Is not greater than of T2 for every value of θ
- Is lower than of T2 for at least one value of θ



2- EFFICIENCY (cont…)



MSE(T1)≤ MSE(T2)

Example 1

Given a sample of 5 measurements of the length in cm of gift boxes:

10	**12**	**13**	**16**	**15**

Finding correct and efficient estimators of the population mean and variance.

Determine the point estimate of the population mean and variance

(a) The correct and efficient estimator of the parameter μ is the sample mean

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

The estimated mean based on the observed sample is equal to 13.2

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10+12+13+16+15}{5} = 13{,}2$$

(b) The corrected and efficient estimator of the parameter σ2 is the corrected sample variance

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

The estimated variance based on the observed sample is equal to 5.7

$$s^2 = \frac{(10-13,2)^2 + (12-13,2)^2 + (13-13,2)^2 + (16-13,2)^2 + (15-13,2)^2}{5 - 1} = 5,7$$

Precision of estimators and sample size

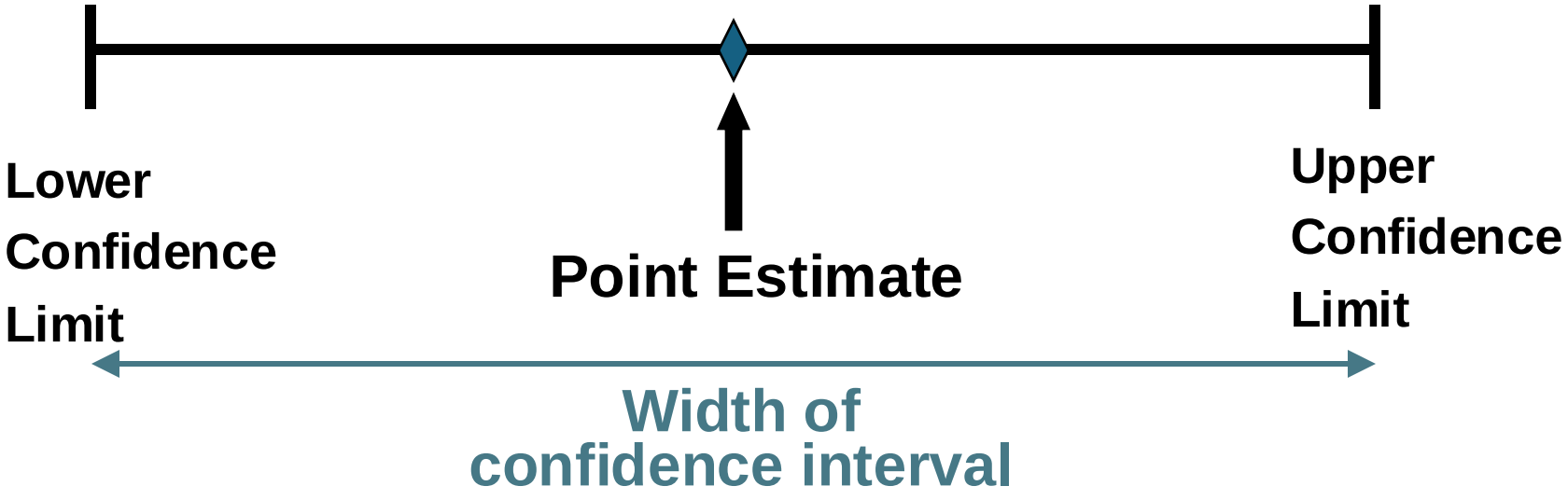
- *The standard error increases* with higher population variances and *decreases* with larger sample sizes

- However, the relative gain in precision *decreases* as sample size *increases*

- Very large sample sizes are not convenient, because the gain in precision is *very small* and the increase in costs is *very large*

Point and Interval Estimates

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



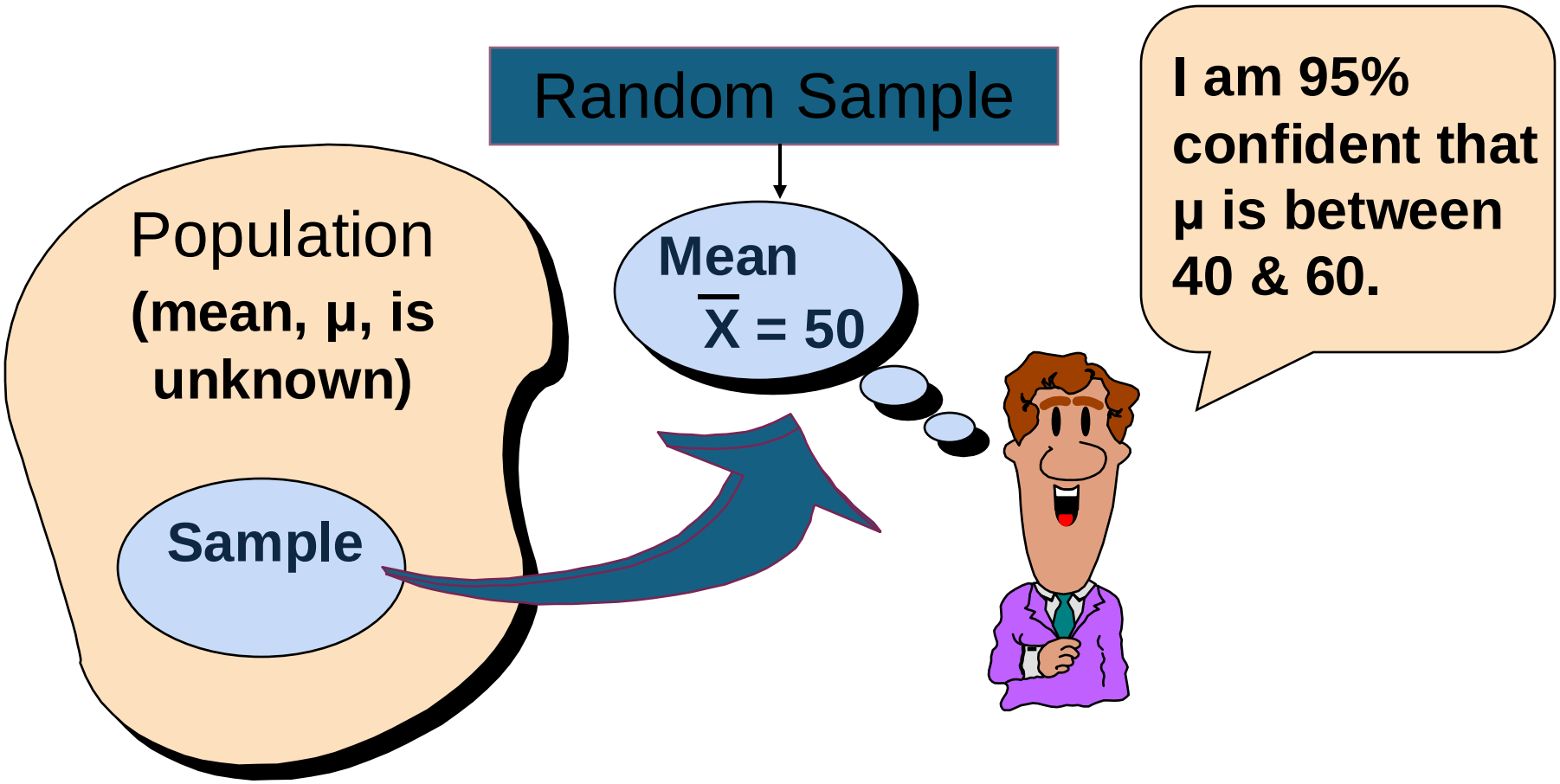
Confidence Intervals

- 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 observations from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - e.g. 95% confident, 99% confident
    - Can never be 100% confident

Estimation Process



- The general formula for all confidence intervals is:

Point Estimate ± (Critical Value)(Standard Error)

Where:

- Point Estimate is the sample statistic estimating the population parameter of interest
- Critical Value is a table value based on the sampling distribution of the point estimate and the desired confidence level
- Standard Error is the standard deviation of the point estimate

Confidence Level

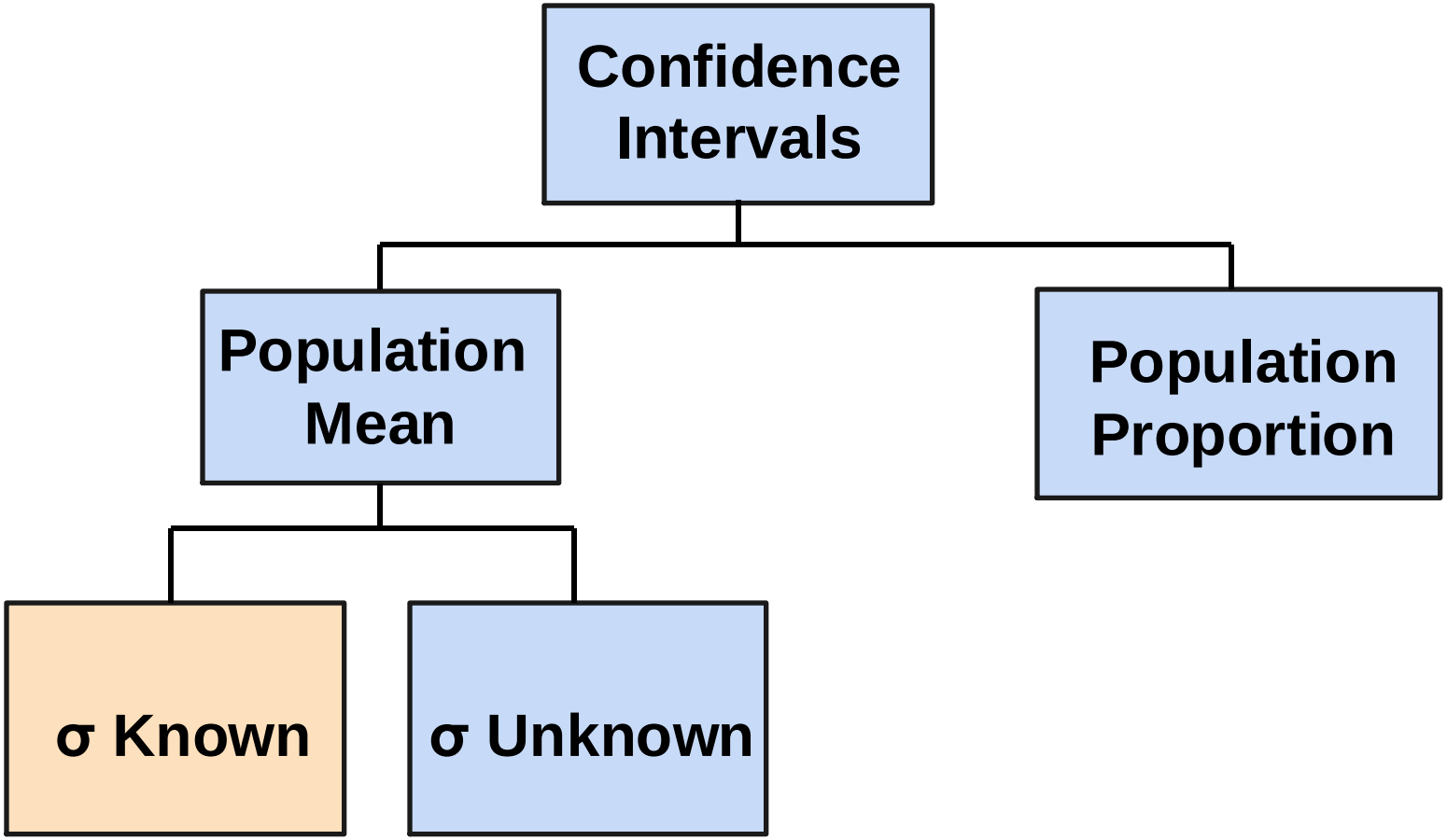
- Confidence Level
  - Confidence the interval will contain the unknown population parameter
  - A percentage (less than 100%)

Confidence Level, (1-α)

(continued)

- Suppose confidence level = 95%
- Also written (1 - α) = 0.95, (so α = 0.05)
- A relative frequency interpretation:
  - 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



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:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

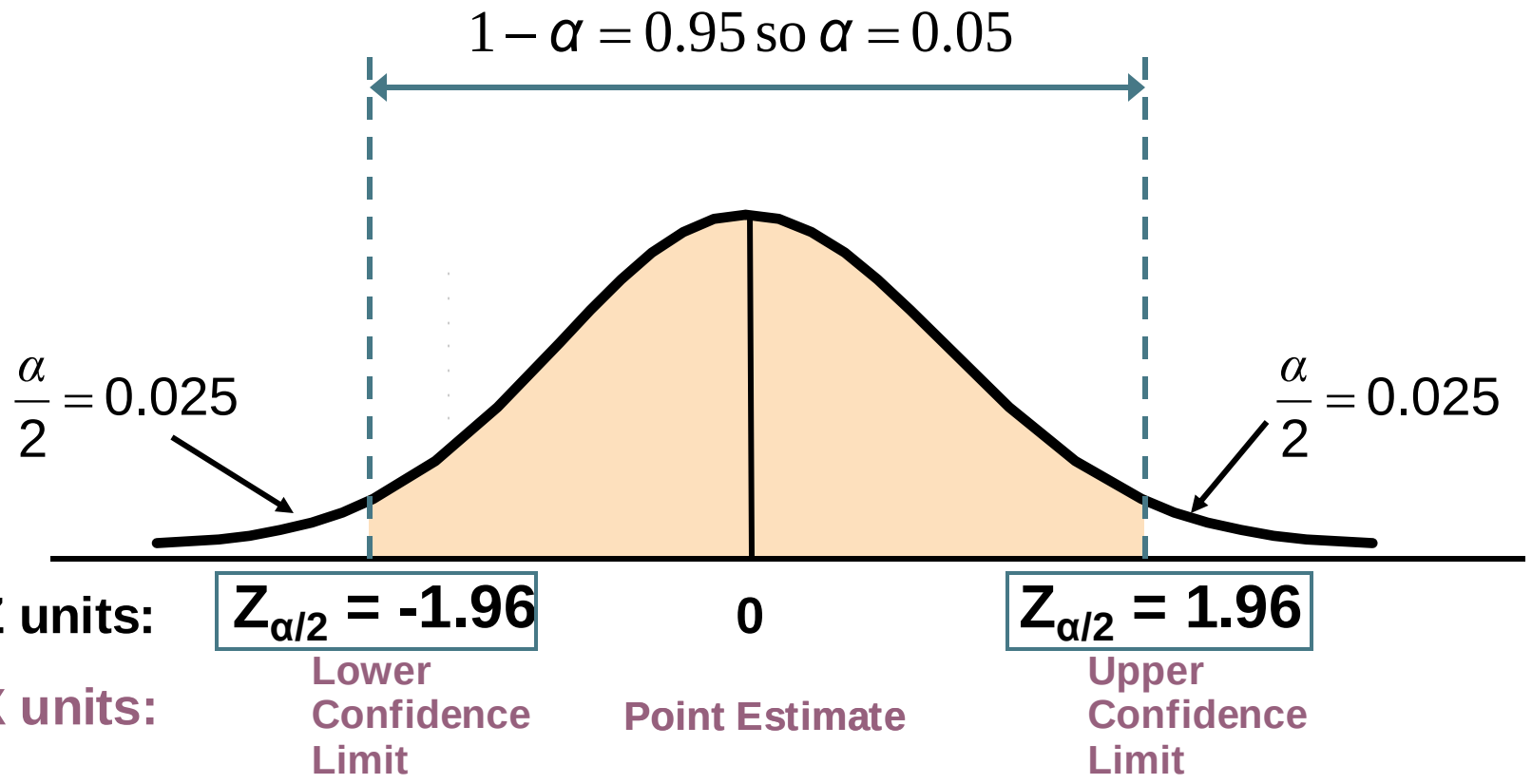
where:

- $\bar{X}$ is the point estimate of the mean
- $Z_{\alpha/2}$ is the normal distribution critical value for a probability of $\alpha/2$ in each tail
- $\frac{\sigma}{\sqrt{n}}$ is the standard error

Finding the Critical Value, $Z_{\alpha/2}$

- Consider a 95% confidence interval:

$$Z_{\alpha/2} = \pm 1.96$$

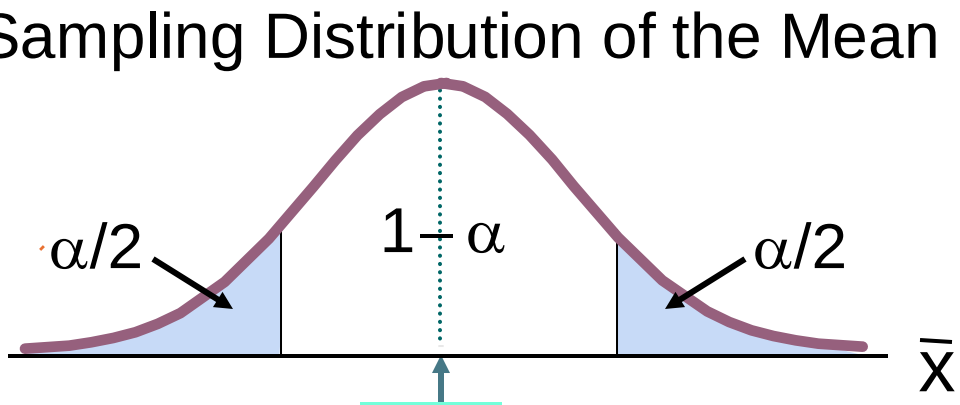


Common Levels of Confidence

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

Confidence Level	*Confidence Coefficient,* $1-\alpha$	$Z_{\alpha/2}$ *value*
80%	0.80	1.28
90%	0.90	1.645
95%	0.95	1.96
98%	0.98	2.33
99%	0.99	2.58
99.8%	0.998	3.08
99.9%	0.999	3.27

Intervals and Level of Confidence

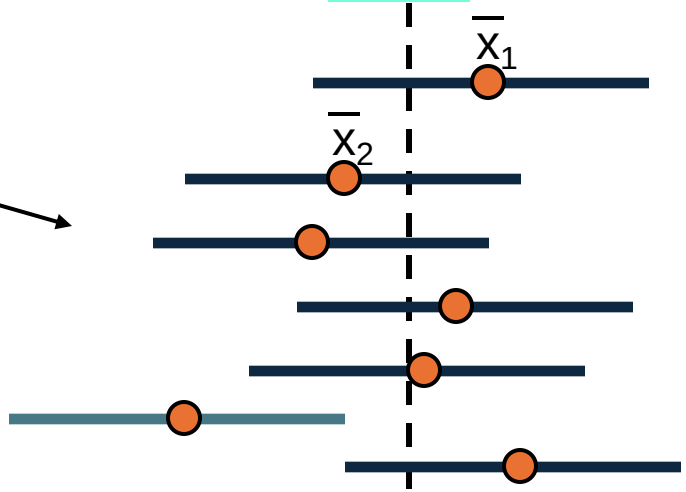


Intervals extend from

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

to

$$\overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$



Confidence Intervals

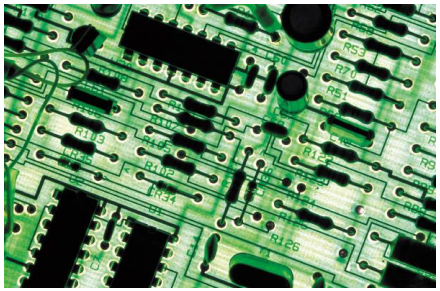
(1-α)x100% of intervals constructed contain μ;

(α)x100% 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 known and equal to 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 0.35 ohms.

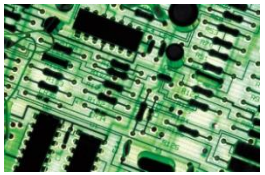
- Solution:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96\,(0.35/\sqrt{11})$$

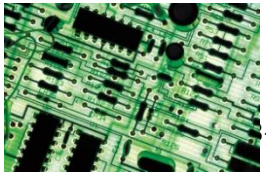
$$= 2.20 \pm 0.2068$$

$$1.9932 \le \mu \le 2.4068$$

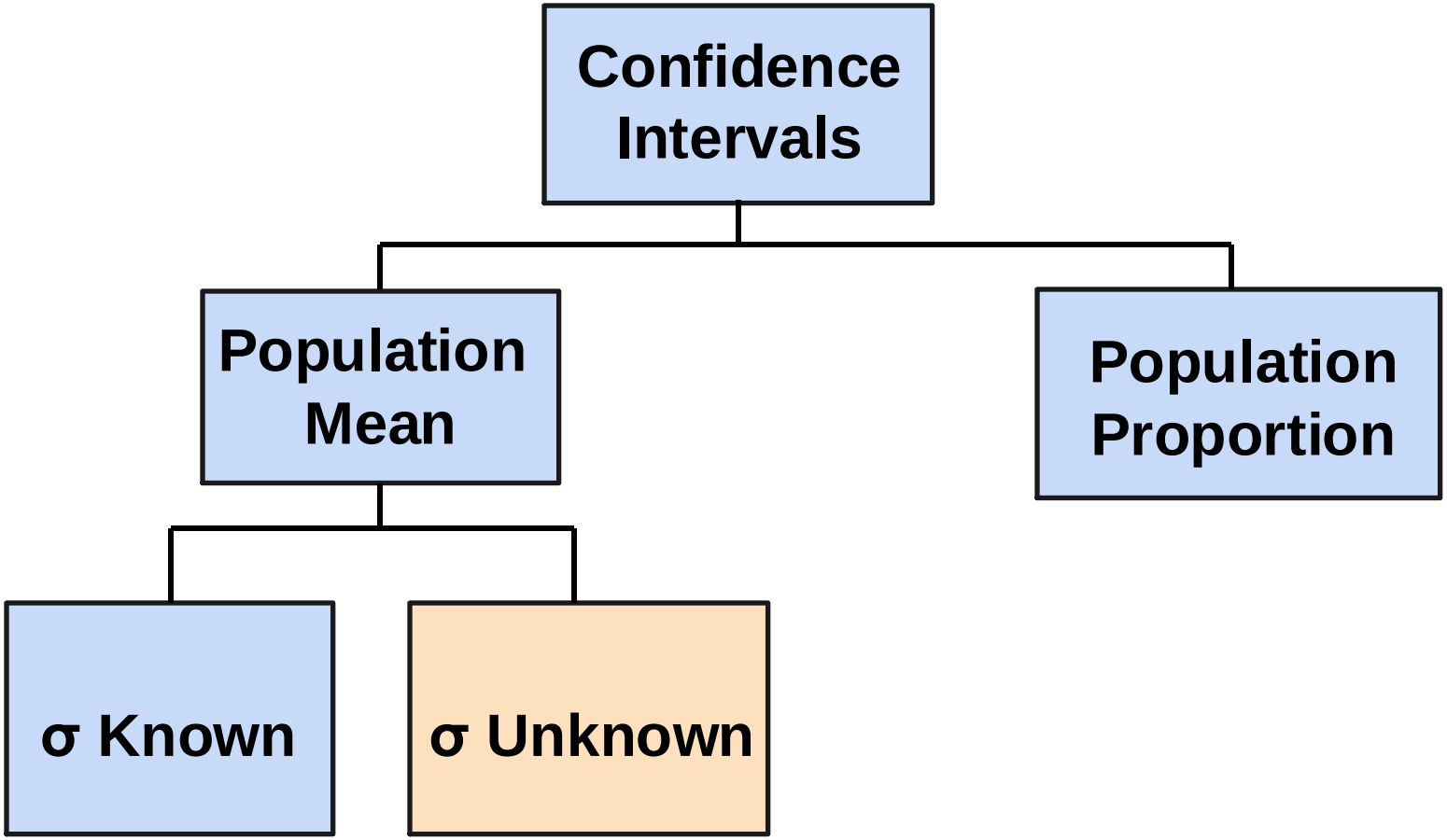


Interpretation

- 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



Do You Ever Truly Know σ?

- Probably not!
- In virtually all real world business situations, σ is not known.
- If there is a situation where σ is known then μ is also known (since to calculate σ you need to know μ.)
- If you truly know μ there would be no need to gather a sample to estimate it.

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

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:

$$\overline{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

(where $t_{\alpha/2}$ is the critical value of the t distribution with n -1 degrees of freedom and an area of α/2 in each tail)

Student's t Distribution

- The t is a family of distributions
- The $t_{\alpha/2}$ value depends on degrees of freedom (d.f.)
  - Number of observations that are free to vary after sample mean has been calculated

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

Degrees of Freedom (df)

Idea: Number of observations that are free to vary
after sample mean has been calculated

Example: Suppose the mean of 3 numbers is 8.0

Let $X_1 = 7$
Let $X_2 = 8$
What is X_3?

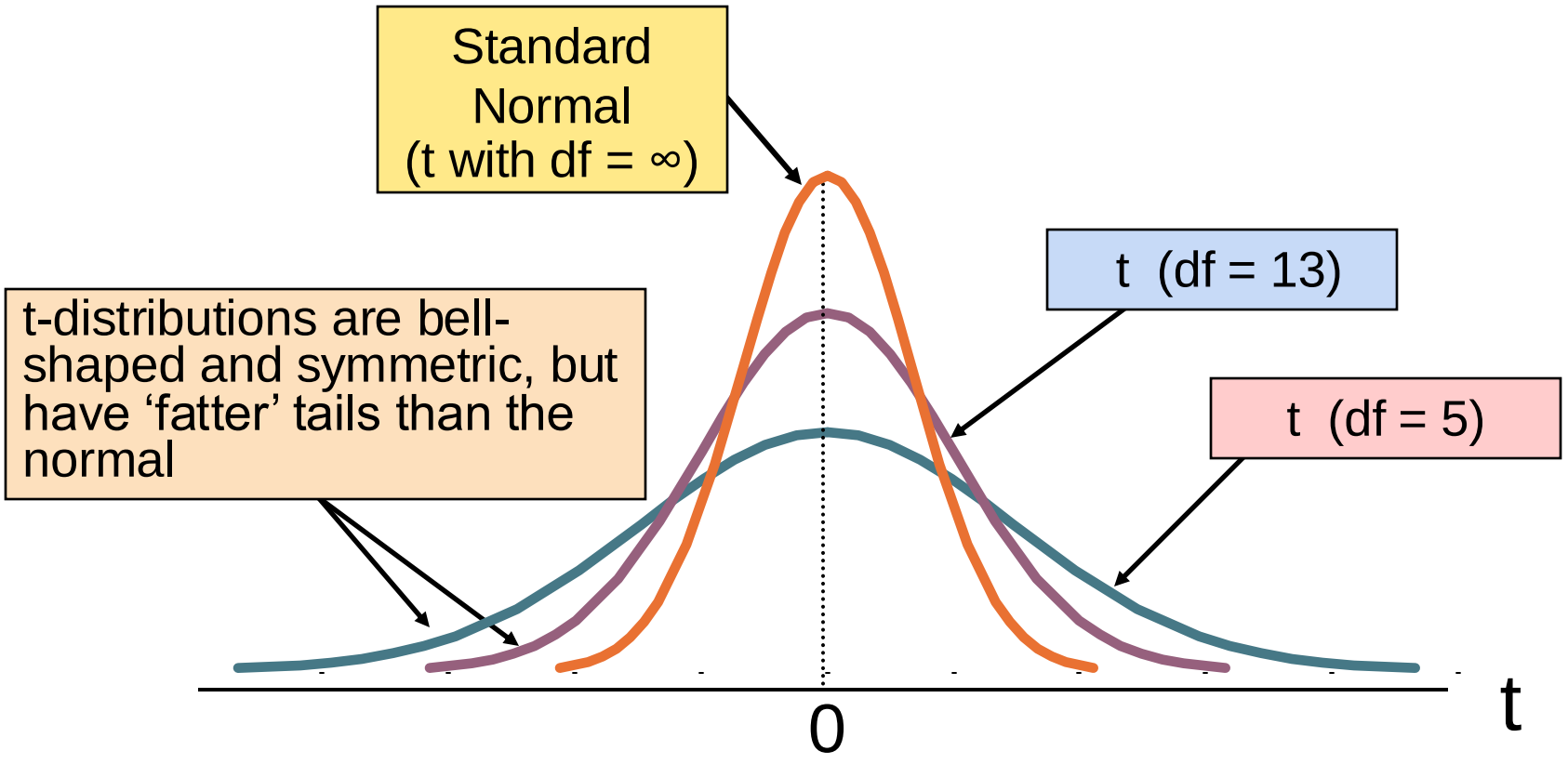
If the mean of these three values is 8.0,
then X_3 must be 9
(i.e., X_3 is not free to vary)

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 → Z as n increases

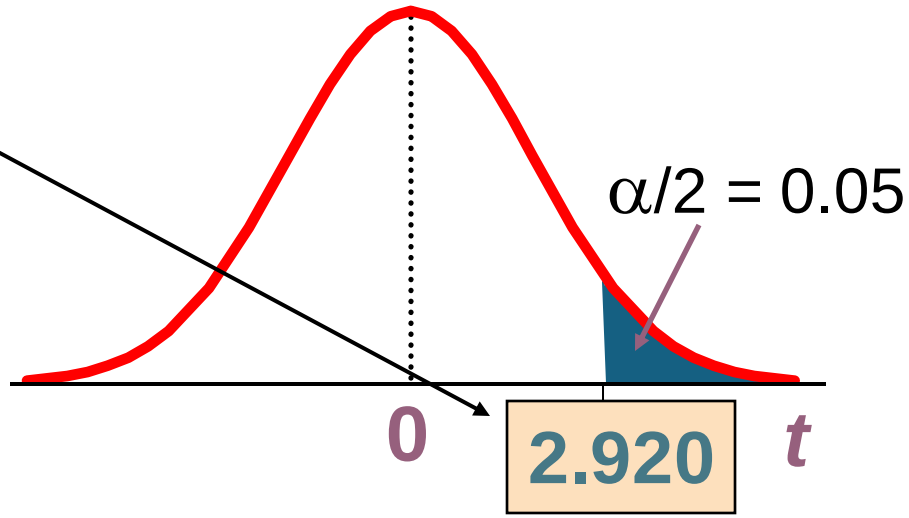


Student's t Table

	Upper Tail Area		
df	.25	.10	**.05**
1	1.000	3.078	6.314
2	0.817	1.886	**2.920**
3	0.765	1.638	2.353

The body of the table contains t values, not probabilities

Let: n = 3
df = *n* - 1 = 2
α = 0.10
$\alpha/2$ = 0.05



Selected t distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z (∞ d.f.)
0.80	1.372	1.325	1.310	1.28
0.90	1.812	1.725	1.697	1.645
0.95	2.228	2.086	2.042	1.96
0.99	3.169	2.845	2.750	2.58

Note: $t \rightarrow Z$ as n increases

Example of t distribution confidence interval

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

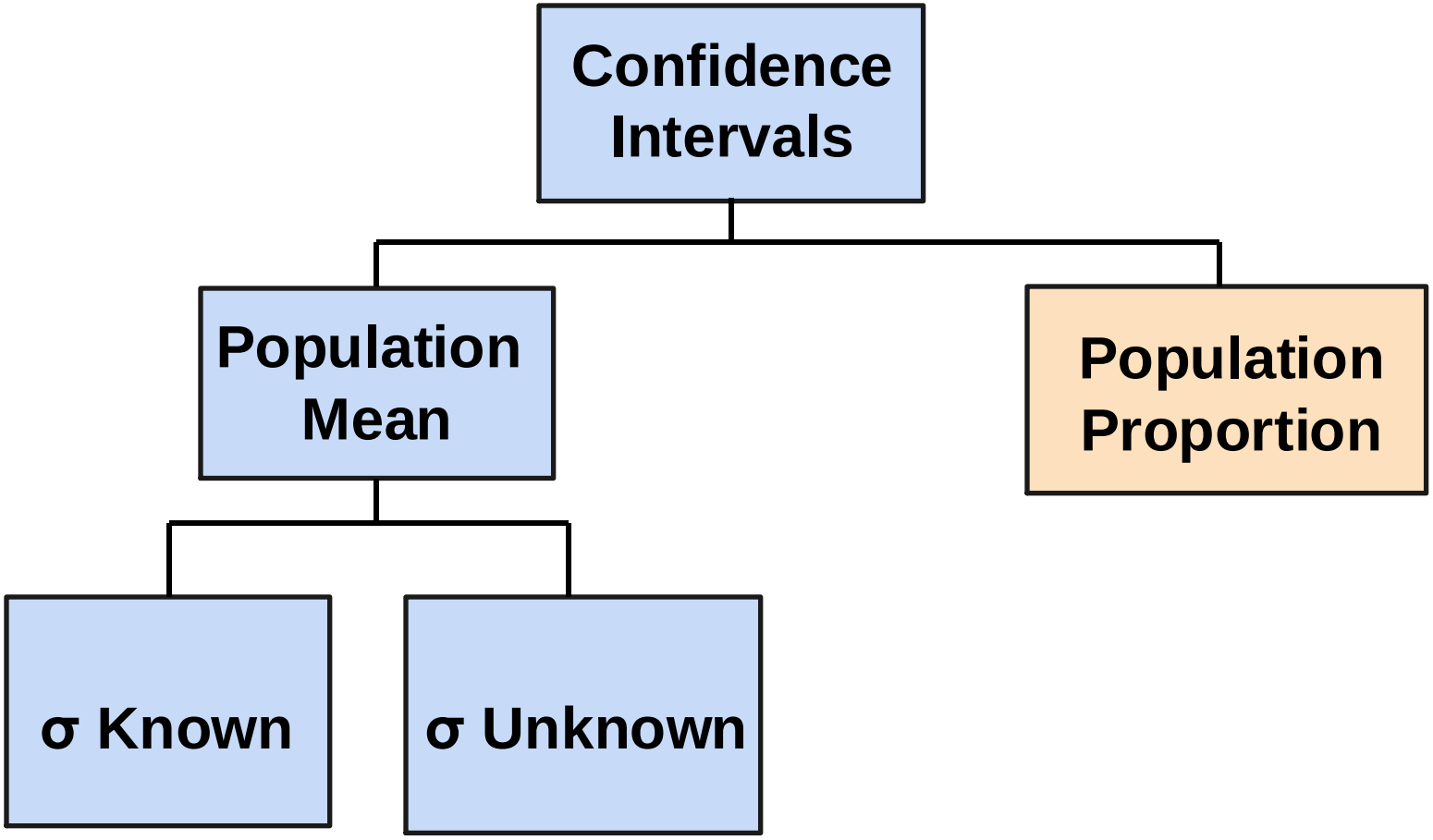
- d.f. = n – 1 = 24, so $t_{\alpha/2} = t_{0.025} = 2.0639$

The confidence interval is

$$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

$$46.698 \le \mu \le 53.302$$

Confidence Intervals



Confidence Intervals for the Population Proportion, π

- An interval estimate for the population proportion (π) can be calculated by adding an allowance for uncertainty to the sample proportion (p)

Confidence Intervals for the Population Proportion, π

(continued)

- Recall that the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

$$\sigma_p = \sqrt{\pi(1-\pi)}$$

- We will estimate this with sample data:

$$\sqrt{p(1-p)}$$

Confidence Interval 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
  - $Z_{\alpha/2}$ is the standard normal value for the level of confidence desired
  - p is the sample proportion
  - n is the sample size
- Note: must have np > 5 and n(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.

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

$$= 25/100 \pm 1.96\sqrt{0.25(0.75)/100}$$

$$= 0.25 \pm 1.96\,(0.0433)$$

$$0.1651 \le \pi \le 0.3349$$

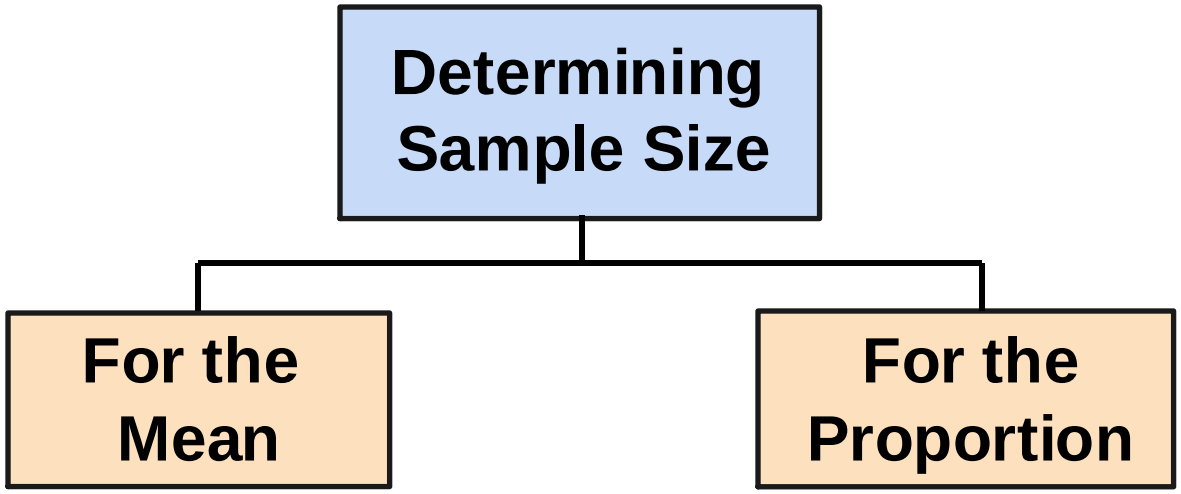
Interpretation

- 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.

Determining Sample Size

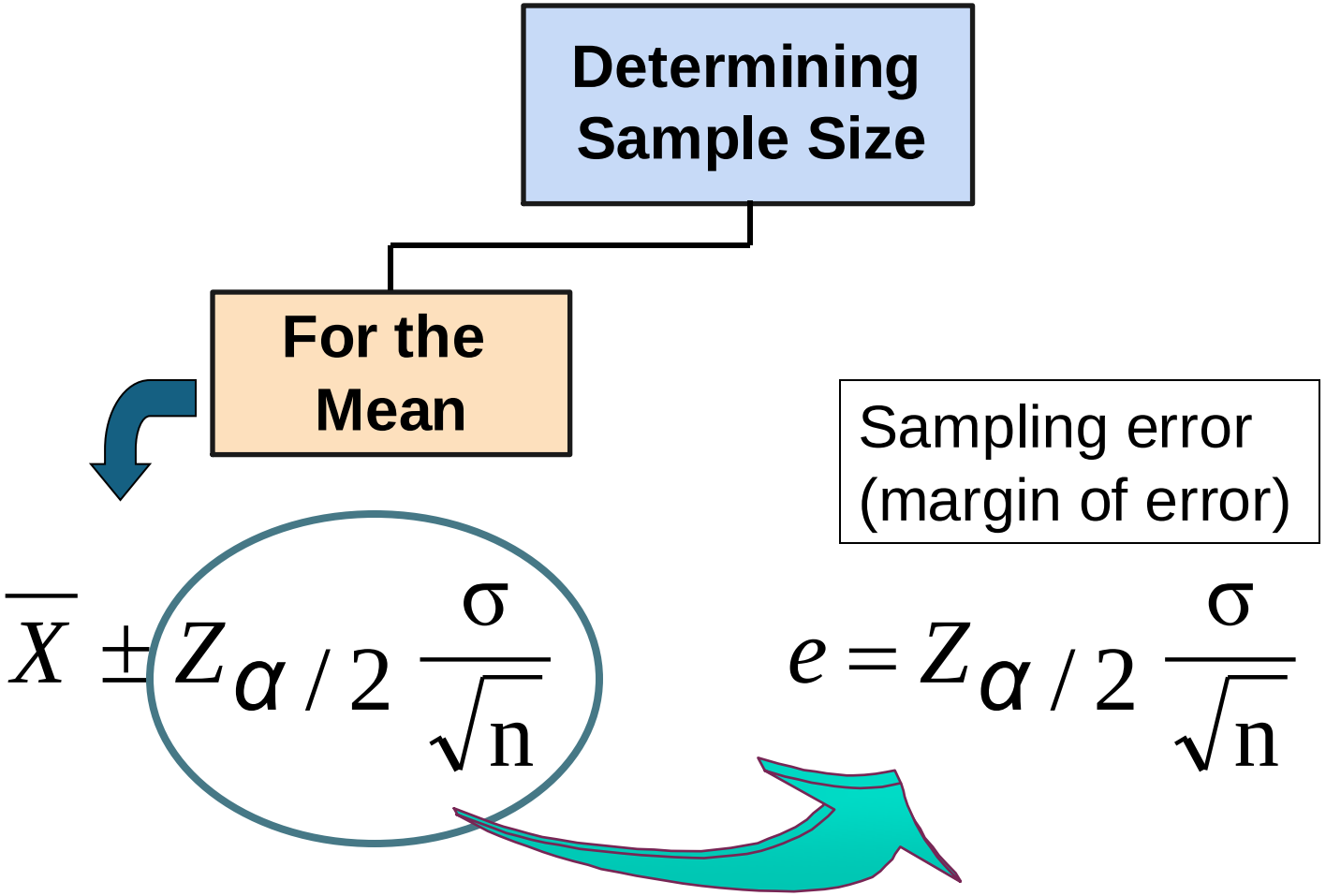


Sampling Error

- The required sample size can be found to reach a desired margin of error (e) with a specified level of confidence (1 - α)

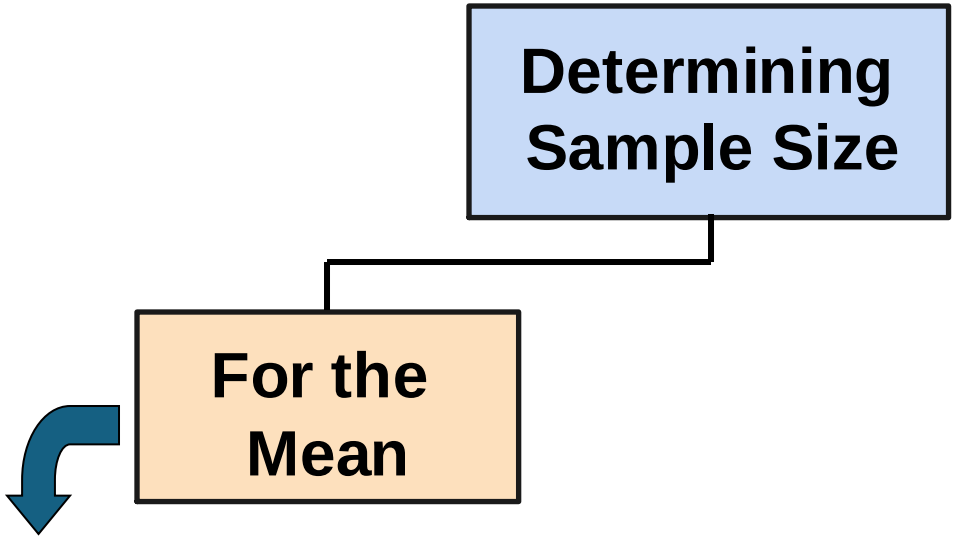
- The margin of error is also called sampling error
  - the amount of imprecision in the estimate of the population parameter
  - the amount added and subtracted to the point estimate to form the confidence interval

Determining Sample Size



Determining Sample Size

(continued)



$$e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \rightarrow \text{Now solve for n to get} \rightarrow n = \frac{Z_{\alpha/2}{}^2\,\sigma^2}{e^2}$$

Determining Sample Size

(continued)

- To determine the required sample size for the mean, you must know:

  - The desired level of confidence (1 - α), which determines the critical value, $Z_{\alpha/2}$
  - The acceptable sampling error, e
  - The standard deviation, σ

Required Sample Size Example

If $\sigma = 45$, what sample size is needed to estimate the mean within ± 5 with 90% confidence?

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.645)^2(45)^2}{5^2} = 219.19$$

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

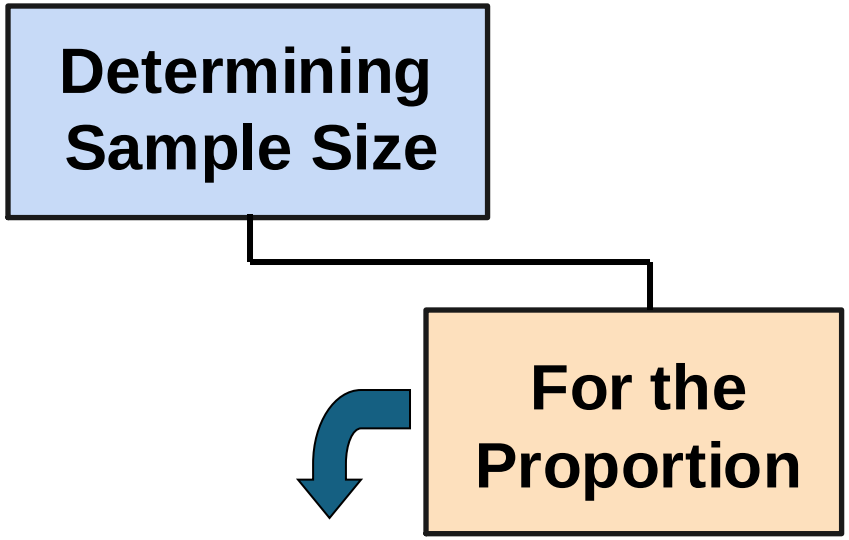
(Always round up)

If σ is unknown

- If unknown, σ can be estimated when using the required sample size formula
  - Use a value for σ that is expected to be at least as large as the true σ
  - Select a pilot sample and estimate σ with the sample standard deviation, S

Determining Sample Size

(continued)



$$e = Z\sqrt{\frac{\pi(1-\pi)}{n}}$$

Now solve for n to get

$$n = \frac{Z^2\,\pi(1-\pi)}{e^2}$$

Determining Sample Size

(continued)

- To determine the required sample size for the proportion, you must know:
  - The desired level of confidence $(1 - \alpha)$, which determines the critical value, $Z_{\alpha/2}$
  - The acceptable sampling error, e
  - The true proportion of events of interest, π
    - π can be estimated with a pilot sample if necessary (or conservatively use 0.5 as an estimate of π)

Exercises to do
Point and interval estimates

BDM course - Prof. Ilaria Benedetti

Ex 1

A sample of 25 firms is drawn from a population of firms on which turnover is observed.

Assuming that the population has normal distribution with unknown μ e and $\sigma^2 = 16$ that the average turnover in the sample is 173 (€ thousand),

determine the 95% interval estimate for the mean of the population.

Variance in the pop is known , interval for μ is equal to:

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

where $z_{\alpha/2}$

is the value of the standardised normal variable Z such that:

$$P\left(Z \le -z_{\alpha/2}\right) = P\left(Z \ge z_{\alpha/2}\right) = \alpha/2$$

By solving, we have: $173 - 1{,}96\frac{4}{5} \le \mu \le 173 + 1{,}96\frac{4}{5}$

Thus, the sample data suggest that the average turnover of the population of enterprises is between 171.4 and 174.6

Ex 2

A sample of twelve pieces was taken from the population of pasta packages whose actual net weights, in grams, are:

498	498	503	493	491	499
512	504	483	506	510	509

Assuming that the actual net weight in the population has a normal distribution, determine the point estimate for the population mean and the estimate for 95% and 99% interval.

First, point estimates of the mean and variance are determined, using its corrected estimate for the variance S^2

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 500{,}5 \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 73{,}727$$

As the variance is unknown, it is required the use of Student's T-distribution with 11 degrees of freedom. At the 95% level this results in:

$$\bar{x} - t_{0,025}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{0,025}\frac{s}{\sqrt{n}}$$

Since $t_{0,025}$=2,201, the result is:

$$500{,}5 \pm 2{,}201\frac{8{,}59}{\sqrt{12}}$$

so the average weight range between 495 and 505.9 grams.

Al 99% confidence level, we consider: $t_{0,005}=3,1058$

By substituting, we have:

$$500,5 \pm 3,1058 \frac{8,59}{\sqrt{12}}$$

The interval is: $(492,8\,;\ 508,2)$

Ex 3.

On a random sample of 2000 regular readers of one newspaper, a proportion of regular readers of a second newspaper of 0.12 was calculated; construct the interval estimate for this proportion at the 95% confidence level.

Bernoullian populations are characterised by the fact that the units of the population are classified into only two categories, say success and failure.

Considering a sample $(X_1, X_2, ..., X_n)$

extracted from a Bernoullian population with parameter π , then:

$$P = \overline{X} = \frac{1}{n}\sum x_i$$

i.e. the relative frequency of cases with the characteristic of interest is the estimator of π.

If the sample size is sufficiently large, the normal approximation can be invoked. Thus, given 1-α, we can write the relation

$$P\left(-z_{\alpha/2} \le \frac{\overline{X} - \pi}{\sqrt{\overline{X}(1-\overline{X})/n}} \le z_{\alpha/2}\right) \cong 1 - \alpha$$

The interval for π with a confidence level 1-α :

$$\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}}, \bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}}$$

Considering: $\alpha = 0,05$ then: $z_{\alpha/2} = \pm 1,96$

By substituting:

$$0,12 - 1,96\sqrt{\frac{0,12(1-0,12)}{2000}} \leq \pi \leq 0,12 + 1,96\sqrt{\frac{0,12(1-0,12)}{2000}}$$

At the end:

$$0,1058 \leq \pi \leq 0,1342$$