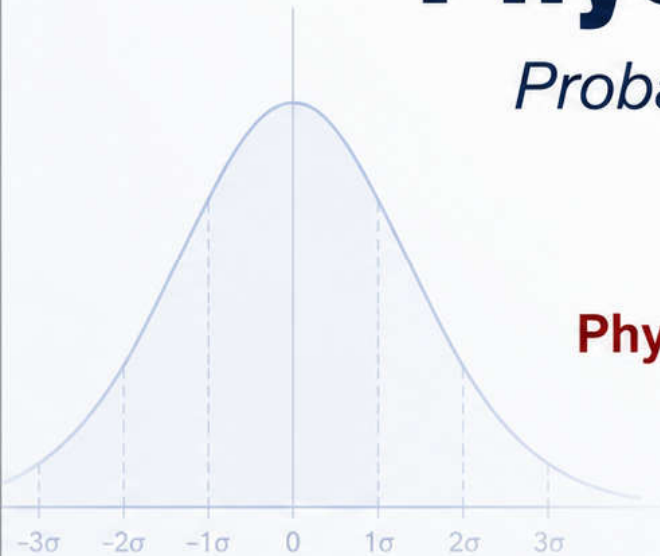


# Confidence Levels in Physical Measurements

*Probability Intervals, Confidence Intervals  
and Student's  $t$  Distribution*

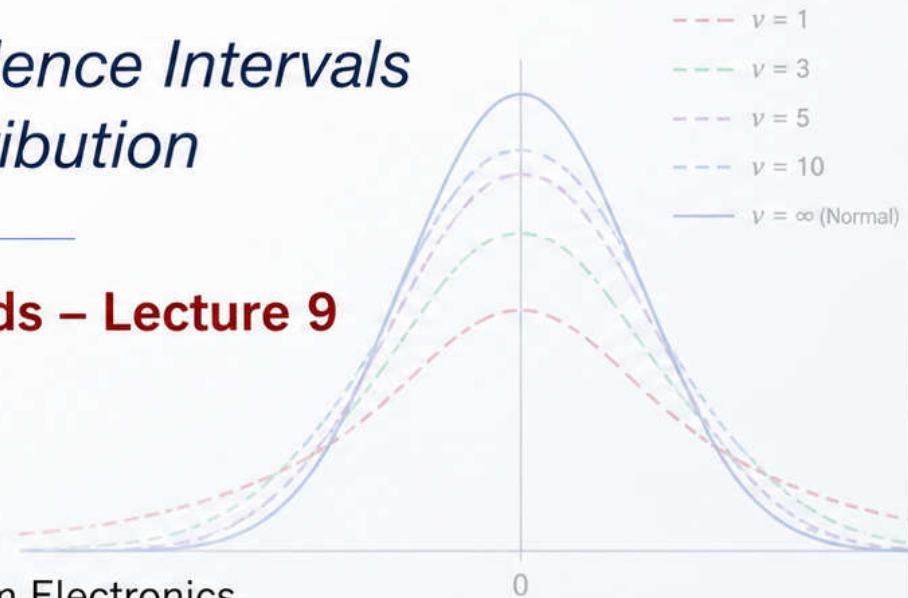


## Physical Measurement Methods – Lecture 9

**Dr. Tibor Ajtai**

University of Szeged

Department of Optics and Quantum Electronics



**UNIVERSITY OF SZEGED**  
UNIVERSITAS SCIENTIARUM SZEGEDIENSIS

# Recap: Reporting a Measurement

$$X = X_0 \pm \delta X$$

But what does  $\pm \delta X$  actually mean?



Does the true value necessarily lie inside this interval?



What probability is associated with the interval?

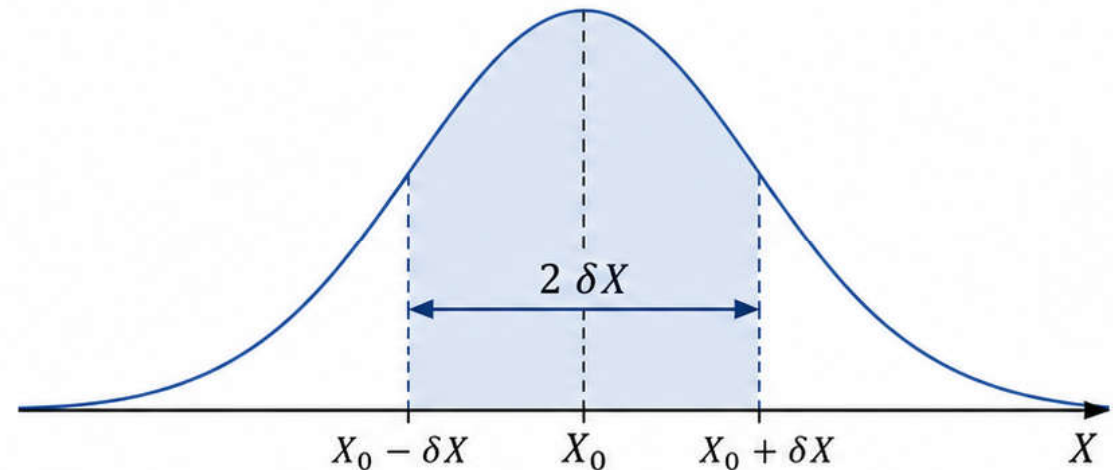


Does the answer depend on the number of measurements?



What changes when  $\sigma$  is unknown?

Probability distribution of the measured quantity  $X$



**Key question:**

*How confident are we that our interval contains the true value?*



### 3. Parent Population

We assume that the measured quantity  $X$  in the population is normally distributed.

#### Population model

Random variable:  $X \sim N(m, \sigma)$

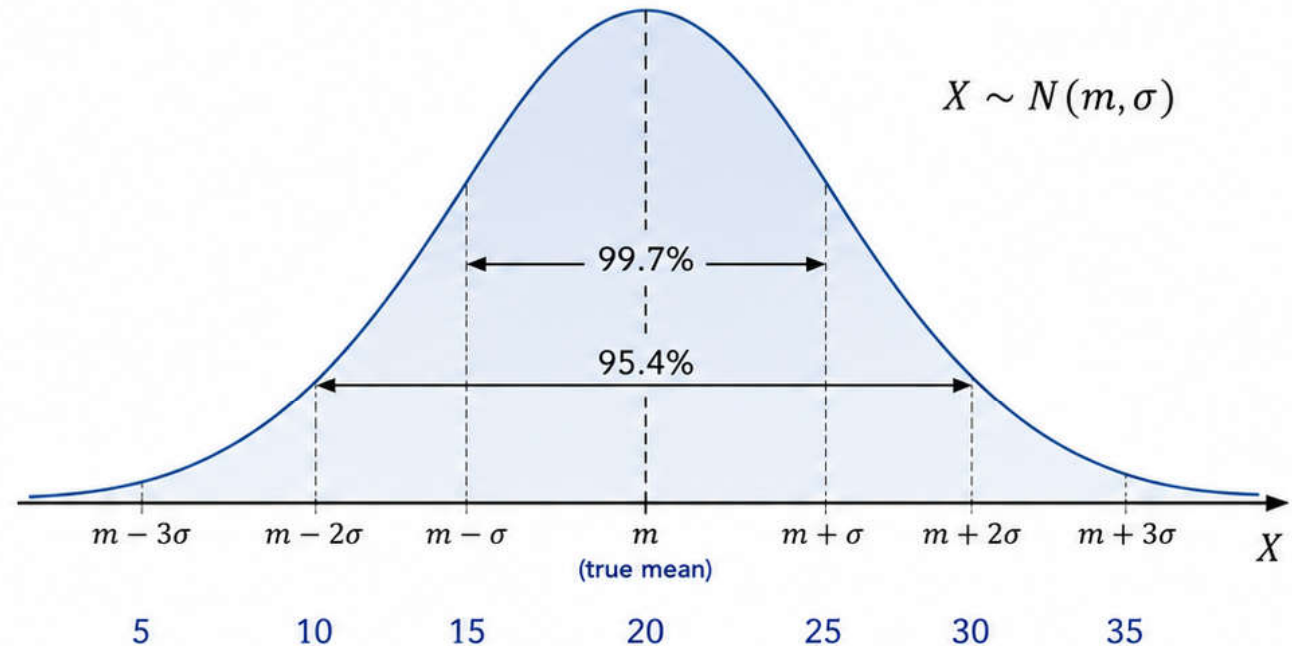
- $m$  : true (unknown) mean
- $\sigma$  : true (unknown) standard deviation

#### Example used in this lecture

$$m = 20, \quad \sigma = 5$$

The true mean is 20 and the true standard deviation is 5.

Probability density function of the parent population



This is the (unknown) distribution of the quantity in the entire population. We will never observe all values; we can observe only a finite sample.



## 4. One Experiment Is Only One Sample

We perform one experiment and take a sample of size  $N = 5$ .

Example measurements:

$x_1, x_2, x_3, x_4, x_5$

Sample mean:

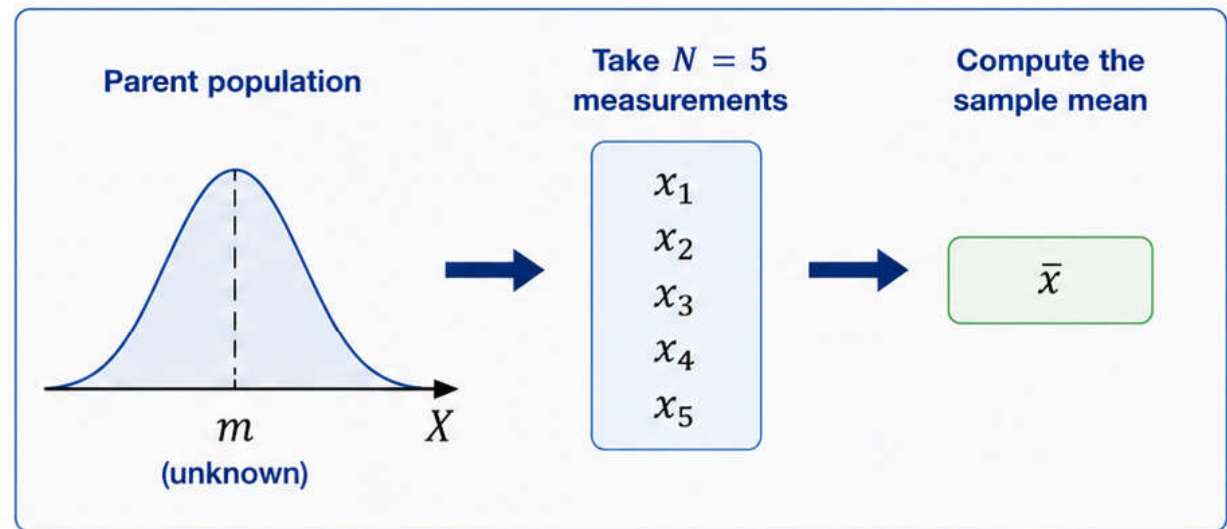
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$



**Important:**

A **different** experiment would give a **different** set of measurements and a **different** sample mean.

One experiment  $\rightarrow$  one sample  $\rightarrow$  one sample mean



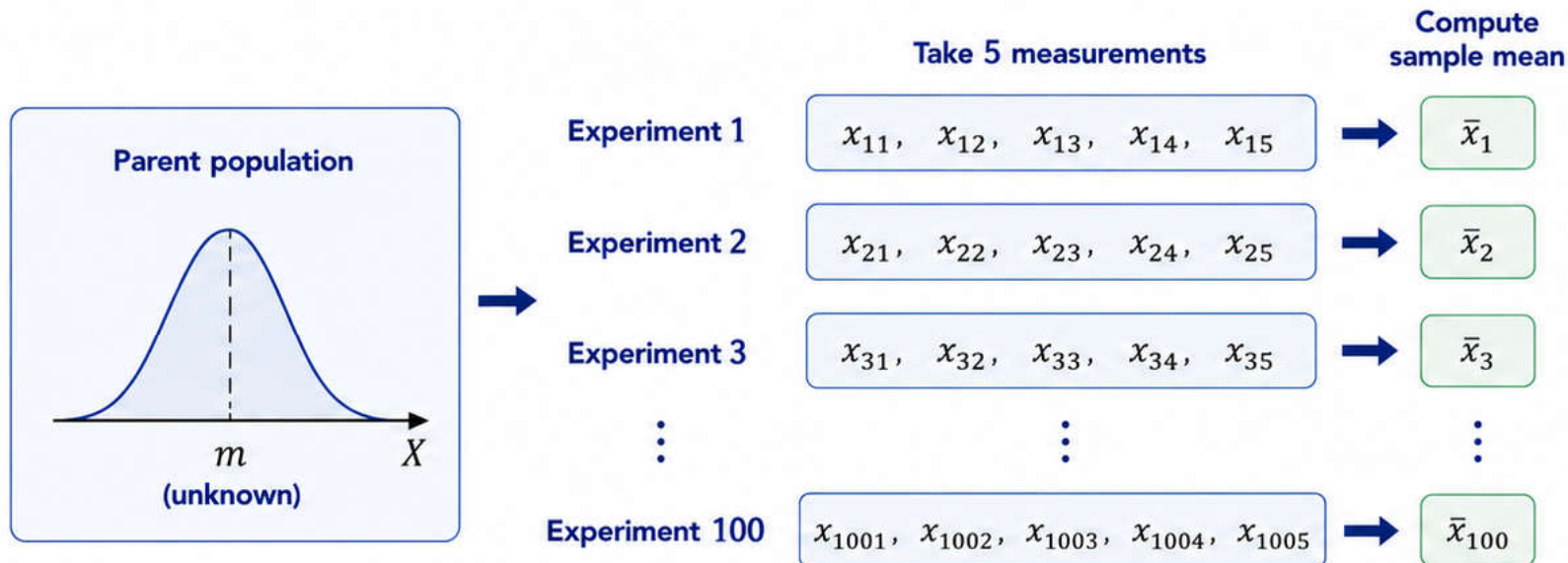
The sample mean  $\bar{x}$  is a **random quantity**: it would change from one experiment to another.



# 5. Repeat the Experiment

We repeat the same experiment many times.

Each time we take 5 measurements and compute the sample mean.



After many repetitions  
(e.g., 100 experiments):

We obtain 100 sample means:

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_{100}$$



These sample means form a distribution –  
the **sampling distribution**  
of the sample mean  $\bar{x}$ .

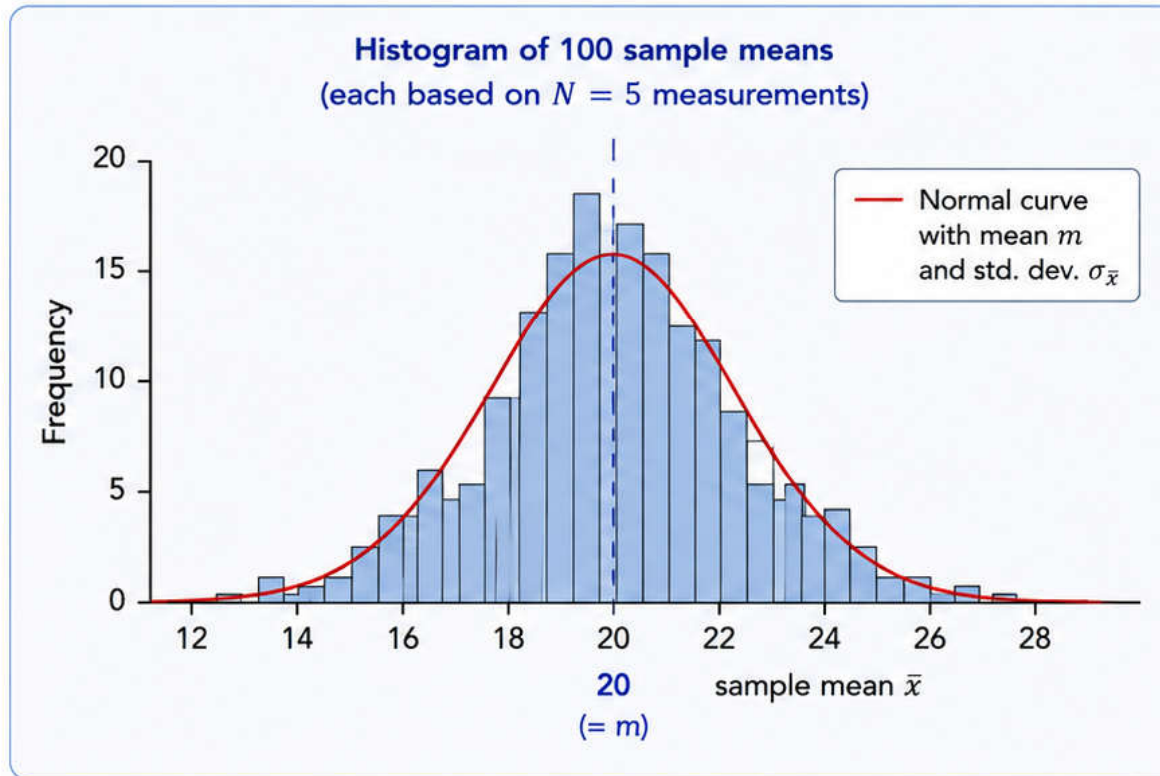


Although each sample contains only 5 measurements,  
the sample mean will **vary** from experiment to experiment.



# 6. Distribution of Sample Means

If we repeat the experiment many times and compute the sample mean each time, the sample means form a **distribution**.



## Theoretical result

If the parent population is normal  $X \sim N(m, \sigma)$ , then the distribution of the sample mean  $\bar{X}$  is also normal:

$$\bar{X} \sim N\left(m, \frac{\sigma}{\sqrt{N}}\right)$$

- $E(\bar{X}) = m$  (the **center** is the true mean  $m$ )
- Std. dev.  $(\bar{X}) = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$  (the **width**)

## What does this mean?



- On **average**, the sample mean **hits** the true mean  $m$ .
- But in any single experiment, it will typically **differ** from  $m$  by an amount of order  $\sigma/\sqrt{N}$ .

## Key takeaway



$m = 20, \sigma = 5, N = 5$ :  $SD(\bar{X}) = 5/\sqrt{5} \approx 2.24$ .  
Thus  $\bar{X} \sim N(20, 2.24)$ . Larger  $N$  narrows the distribution.



# 7. The Central Question

From one experiment we have measured  $N$  values and computed the sample mean  $\bar{x}$ .

We have also estimated its uncertainty (standard error)

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}.$$



?

But:

Where is the unknown true value  $m$ ?



This leads to two concepts that look similar but are **fundamentally different**:

## Probability interval

A statement about **random quantity**  $\bar{X}$  when the true value  $m$  is fixed but unknown.

versus

## Confidence interval

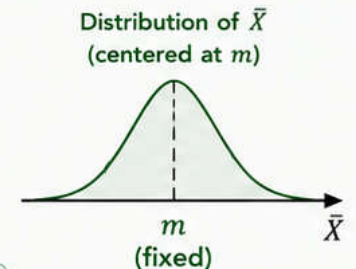
A statement about the **unknown true value**  $m$  based on the **observed data**.

### 1) Probability interval (fixed $m$ , random $\bar{X}$ )

- Assume the true mean  $m$  is fixed (but unknown).
- We repeat the experiment many times.
- Each time we get a different  $\bar{x}$ .
- The sample mean  $\bar{X}$  is a random variable with distribution  $N\left(m, \frac{\sigma}{\sqrt{N}}\right)$ .

**Example:** For  $N = 5$ ,  $\bar{X} \sim N\left(m, \frac{\sigma}{\sqrt{5}}\right)$ .

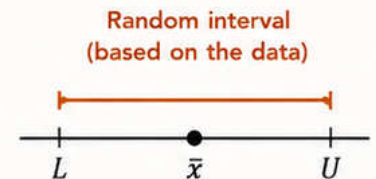
The probability that  $\bar{X}$  falls in a given interval can be computed.



### 2) Confidence interval (random interval, fixed data)

- We perform one experiment and obtain the data.
- The sample mean  $\bar{x}$  is now fixed (observed).
- We construct an interval  $(L, U)$  using this data.
- Different experiments would give different intervals.
- The true mean  $m$  is either inside or outside the interval (but we do not know which).

**Example:** A 95 % confidence interval means that if we repeated the whole procedure many times, 95 % of such intervals would contain  $m$ .



They answer **different questions** and are interpreted in completely **different ways**. Understanding this difference is essential for correct statistical reasoning.

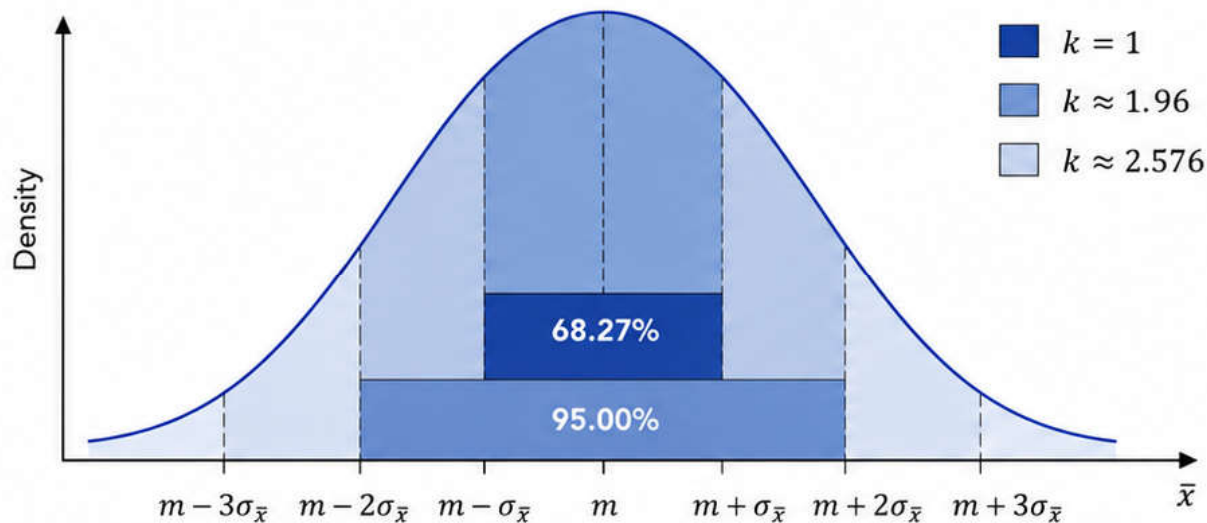


## 8. Coverage Factor $k$



The **coverage factor**  $k$  sets the width of the probability interval around the **known mean**  $m$ .  
Larger  $k \rightarrow$  wider interval  $\rightarrow$  higher probability.

Probability interval  $m - k\sigma_{\bar{x}} < \bar{x} < m + k\sigma_{\bar{x}}$



Typical values of  $k$

| $k$   | Interval                                                          | Probability $P$ |
|-------|-------------------------------------------------------------------|-----------------|
| 1     | $m - 1\sigma_{\bar{x}} < \bar{x} < m + 1\sigma_{\bar{x}}$         | 68.27%          |
| 1.96  | $m - 1.96\sigma_{\bar{x}} < \bar{x} < m + 1.96\sigma_{\bar{x}}$   | 95.00%          |
| 2     | $m - 2\sigma_{\bar{x}} < \bar{x} < m + 2\sigma_{\bar{x}}$         | 95.45%          |
| 2.576 | $m - 2.576\sigma_{\bar{x}} < \bar{x} < m + 2.576\sigma_{\bar{x}}$ | 99.00%          |



**Important:**

Often the approximation  $k = 2 \rightarrow 95\%$  is used in practice.

**More precisely:**

$k = 2$  gives  $P = 95.45\%$ ,  
while  $k = 1.96$  gives exactly 95.00%.



**Takeaway:** The coverage factor  $k$  controls the probability that a random sample mean  $\bar{x}$  falls within  $k$  standard errors of the **known mean**  $m$ .



# 9. Point Estimate



## Our best estimate

From a sample of size  $N$  we compute the sample mean  $\bar{x}$ .  
This is our best (unbiased) estimate of the population mean  $m$ .

$$\tilde{m} = \bar{x}$$

(point estimate)



## Why point estimate?

Among all unbiased estimators of  $m$ , the sample mean has the smallest variance:

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{N}$$

Therefore  $\bar{x}$  is the most precise single-number estimate of  $m$ .



## Limitation of a single number

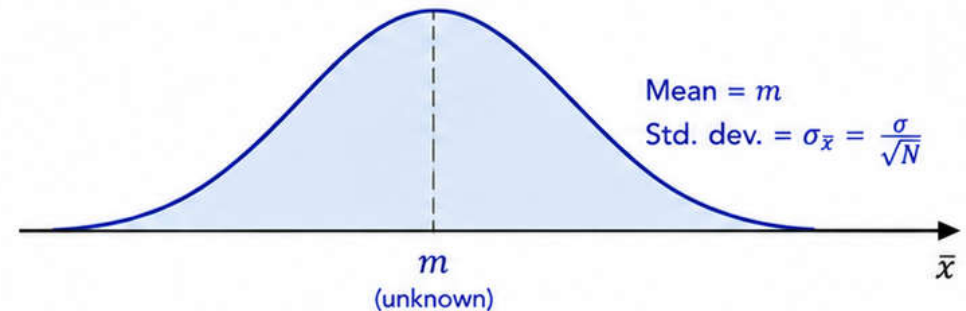
A single number does not tell us how reliable the estimate is.

Different samples give different  $\bar{x}$  values.



**Therefore we need an interval estimate** that quantifies the statistical uncertainty.

## Sampling distribution of the sample mean $\bar{x}$



**Key takeaway:** The sample mean  $\bar{x}$  is our best **point estimate** of the unknown population mean  $m$ , but we also need an **interval estimate** to express its reliability.



# 10. Confidence Interval ( $\sigma$ Known)



## Real experiment

In a real experiment:  
 $m$  = unknown (fixed),  
but after measurement:  
 $\bar{x}$  = known (observed).



## What we want

An interval that likely  
contains the true mean  $m$ .



## Assumption

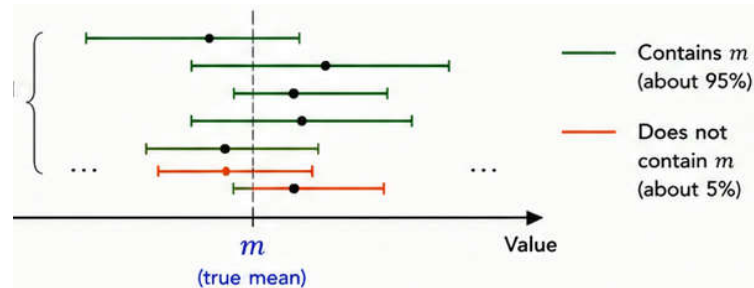
The population is normal  
and  $\sigma$  is known.

## Confidence interval for the population mean $m$ (when $\sigma$ is known)

$$\bar{x} - k \sigma_{\bar{x}} < m < \bar{x} + k \sigma_{\bar{x}}$$

$$\text{where } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

## Fixed $m$ ; intervals vary between samples



Over repeated samples, a fraction  $1 - \alpha$  of the  
constructed intervals covers  $m$ .



## Interpretation

This interval is centered on the  
measured sample mean  $\bar{x}$ .  
It extends  $k$  standard errors  
to the left and right.

$$\text{Width} = 2k \sigma_{\bar{x}} = 2k \frac{\sigma}{\sqrt{N}}$$

## Where do common levels come from?

$k$  is chosen to achieve a desired  
confidence level  $1 - \alpha$ :

- $k = 1 \Rightarrow 68.27\%$
- $k \approx 1.96 \Rightarrow 95.00\%$
- $k \approx 2.576 \Rightarrow 99.00\%$



**Key takeaway:** When  $\sigma$  is known, a  $100(1 - \alpha)\%$  confidence interval for the true mean  $m$  is  
centered on the observed sample mean  $\bar{x}$ .

$$\bar{x} - k \sigma_{\bar{x}} < m < \bar{x} + k \sigma_{\bar{x}}$$



# 11. What Does 95% Confidence Mean?



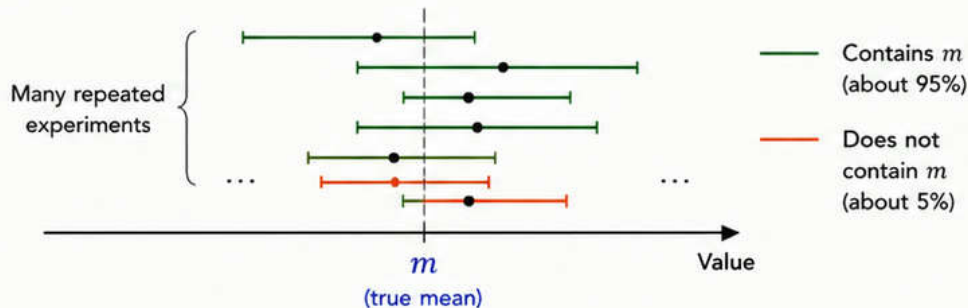
A common misunderstanding: It is **incorrect** to say:

*"There is a 95% probability that the true value is inside this particular interval."*

## The correct frequentist interpretation

If the same experiment were repeated many times under the same conditions, and a 95% confidence interval were constructed from each sample, then:

Approximately 95% of those intervals would contain the true population mean  $m$ , and about 5% would not.



## What it does NOT mean

For the interval you actually computed from your data, either the true mean is inside it or it is not—there is no probability attached to  $m$ . The true mean is fixed (unknown), not random.



After you have measured and computed the interval, the true mean is either inside or outside.

Probability refers to the **procedure**, not to this particular  $m$ .

## Intuition



It is like saying: if we roll a fair die many times,

- about  $1/6$  of the rolls will show a 6.

This does not mean that the next roll has a  $1/6$  probability of 6 after we see the number.



## Summary:

A 95% confidence interval is a statement about the **long-run** performance of the method. In repeated sampling, about 95% of the constructed intervals will capture the true mean. It is **not** a probability that the true mean lies in this specific interval.



# 12. But There Is Another Problem



In practice,  $\sigma$  is usually not known.

We must estimate it from the same data and use that estimate in the confidence interval.



## What we have

- We measure  $x_1, x_2, \dots, x_N$ .
- The true standard deviation  $\sigma$  is unknown.
- We estimate it with the sample standard deviation  $s$ .



## Sample standard deviation

Computed from the same measurements:

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

The factor  $N - 1$  (Bessel correction) gives an unbiased estimate of the variance.



## Key point

Because  $s$  is computed from the data, it is random. It varies from sample to sample.

This changes the distribution of the standardized variable.



**Therefore** we no longer use the known  $\sigma$ .

We substitute  $s$  and  $s_{\bar{x}} = \frac{s}{\sqrt{N}}$ . This leads to a different standardized variable.

## Estimate of the standard error of the mean

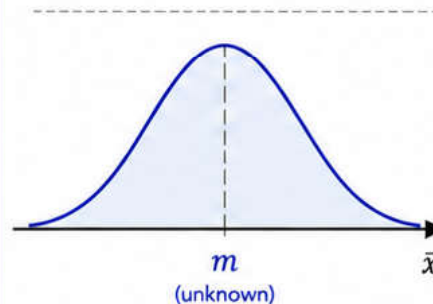
We replace  $\sigma$  in the standard error  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$

by its estimate  $s$ :

$$s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

## Interpretation

$s_{\bar{x}}$  is our estimate of the standard deviation of the sampling distribution of  $\bar{x}$ .



If  $\sigma$  were known:

- $\bar{x}$  is normally distributed with mean  $m$  and std. dev.  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ .

- The standardized variable

$$z = \frac{\bar{x} - m}{\sigma/\sqrt{N}}$$

follows the standard normal distribution  $N(0, 1)$ .



## Takeaway

Usually we do not know  $\sigma$ .

We estimate it with  $s$ , and use  $s_{\bar{x}} = \frac{s}{\sqrt{N}}$  in place of  $\sigma_{\bar{x}}$ .



# 13. From z to t

Define the  $t$  statistic

$$t = \frac{\bar{x} - m}{s_{\bar{x}}}$$

where

$$s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

$(s =$  sample standard deviation)

Compare

The  $z$  variable

$$z = \frac{\bar{x} - m}{\sigma_{\bar{x}}}$$

where

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

$(\sigma =$  population standard deviation)

|                                                                                                                        |                                                                                                                                                                               |                                                                                                                                                           |
|------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
|  <div>What is in the numerator?</div> | Both use the difference between the observed sample mean $\bar{x}$ and the true population mean $m$ .                                                                         |                                                                                                                                                           |
|  <div>The denominator</div>           | <div>Uses <math>\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}</math></div> <div>based on the <b>true</b> population standard deviation <math>\sigma</math> (<b>known</b>).</div> | <div>Uses <math>s_{\bar{x}} = \frac{s}{\sqrt{N}}</math></div> <div>based on the <b>sample</b> standard deviation <math>s</math> (<b>estimated</b>).</div> |
|  <div>Essential difference</div>     | <div><b><math>z</math></b>: population variance <b>known</b></div>                                                                                                            | <div><b><math>t</math></b>: population variance <b>estimated</b></div>                                                                                    |

★

Key message:

When the population standard deviation  $\sigma$  is known, we use  **$z$** .

When  $\sigma$  is **unknown** and must be **estimated** from the sample, we use  **$t$** .



# 14. Degrees of freedom



Degrees of freedom (d.f.) of Student's  $t$  distribution:

$$\nu = N - 1$$

where  $N$  = sample size

## Why $N - 1$ and not $N$ ?

Because once we compute the **sample mean**  $\bar{x}$ , the deviations from the mean are not all independent.

### Key identity

$$\sum_{i=1}^N (x_i - \bar{x}) = 0$$



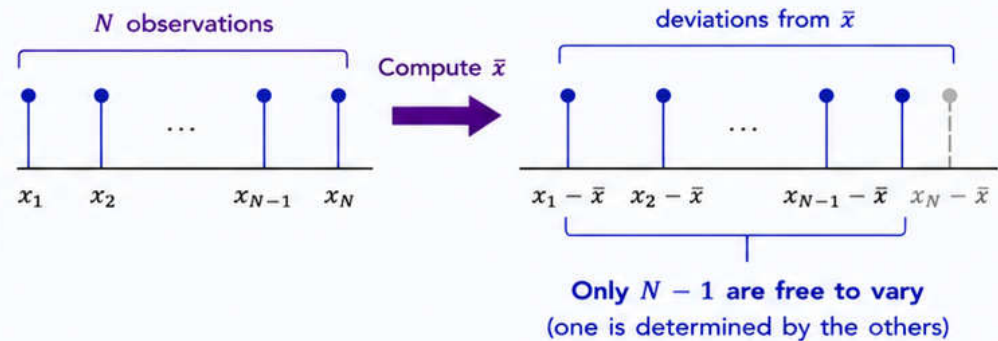
If any  $N - 1$  deviations are known, the last one is **fixed** by this equation.



Therefore, only  $N - 1$  deviations are independent. Hence the **degrees of freedom** are  $\nu = N - 1$ .

## Intuition

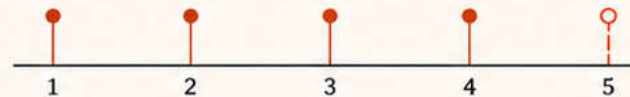
When we estimate the mean  $\bar{x}$  from the data, we “use up” one piece of information. One degree of freedom is lost.



### Example: $N = 5$

We compute  $\bar{x}$  from the five measurements. The deviations satisfy  $\sum_{i=1}^5 (x_i - \bar{x}) = 0$ .

Knowing any **four** deviations determines the fifth.



$$\nu = N - 1 = 4$$

degrees of freedom



**Summary:** Because the deviations must sum to zero, only  $N - 1$  of them can vary independently. This is the reason Student's  $t$  distribution has  $\nu = N - 1$  degrees of freedom.



# 15. What happens as $N$ increases?



As the sample size  $N$  increases, our **estimate of  $\sigma$  improves**, and Student's  $t$  distribution approaches the standard normal distribution.

## The chain of effects

$N \uparrow$

Sample size increases.



$\nu = N - 1 \uparrow$

Degrees of freedom increase.



Estimate of  $\sigma$  improves  
(less uncertainty in  $s$ ).

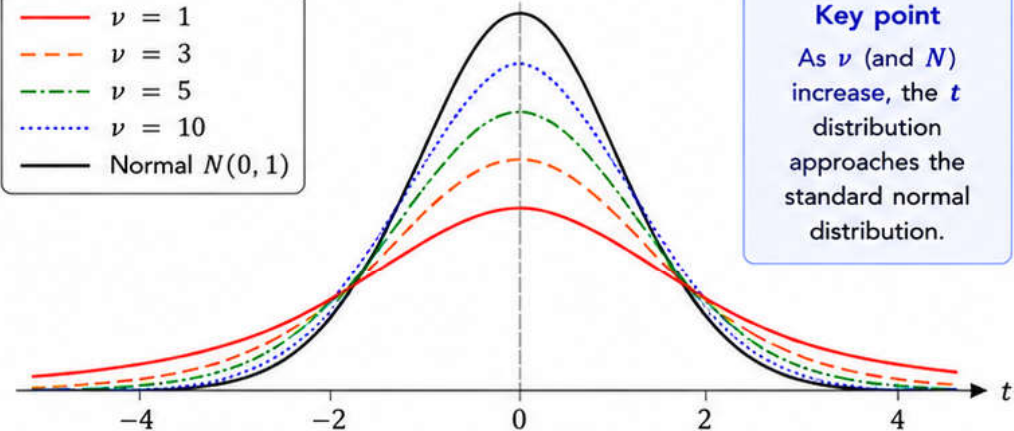
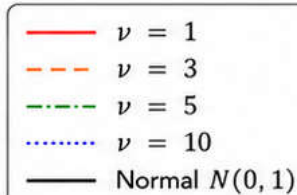


Student's  $t$  distribution narrows  
(less dispersion, lighter tails).



$t \rightarrow z$   
and  
 $t_\nu \rightarrow N(0, 1)$

## Visualizing the convergence



### Key point

As  $\nu$  (and  $N$ ) increase, the  $t$  distribution approaches the standard normal distribution.



In the limit of very large  $N$  (so  $\nu \rightarrow \infty$ ), Student's  $t$  distribution becomes indistinguishable from  $N(0, 1)$ .



**Take-home message:** More data  $\Rightarrow$  more degrees of freedom  $\Rightarrow$  better estimate of  $\sigma$   
 $\Rightarrow$  narrower  $t$  distribution  $\Rightarrow$   $t$  behaves like  $z$  (standard normal).



# 16. The key lesson



## The essential message:

**Small samples require wider intervals.**

Not because the measured scatter must be larger, but because **our estimate of the scatter is less reliable.**

### 1) The root cause

We estimate the population standard deviation  $\sigma$  by the sample standard deviation  $s$ .

#### Key point:

$s$  fluctuates from sample to sample.  
Therefore our estimate of the mean  $\bar{x}$  is less certain.

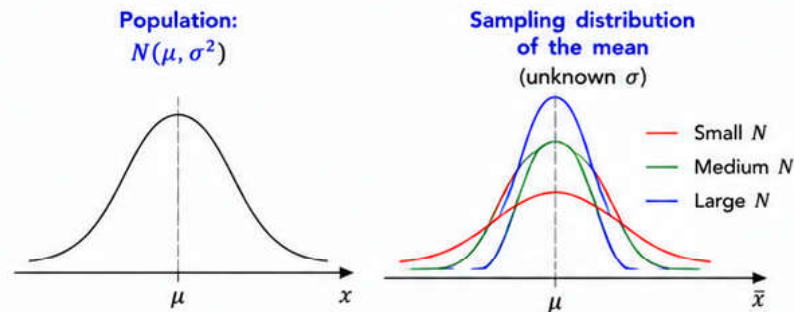


**Less certainty  $\Rightarrow$  Wider interval needed**  
(for the same confidence level)

#### Intuition

When we are less sure about the scatter, we must allow for more extreme possibilities.

### 2) Visual summary



- **Small  $N$  (small  $\nu$ ):**  
heavier tails  $\Rightarrow$  more probability in the tails  $\Rightarrow$  **wider interval**
- **Large  $N$  (large  $\nu$ ):**  
lighter tails  $\Rightarrow$  less probability in the tails  $\Rightarrow$  **narrower interval**

### 3) Practical consequence

To achieve a given confidence level  $P$ , we must use a larger  $k$  when  $N$  is small.

#### Example for $P \approx 68\%$ ( $k = 1$ case)

| Degrees of freedom $\nu$          | Actual coverage $P( t_\nu  \leq 1)$ |
|-----------------------------------|-------------------------------------|
| $\nu = 1$                         | 50.00%                              |
| $\nu = 5$                         | 63.68%                              |
| $\nu = 10$                        | 65.91%                              |
| $\nu \rightarrow \infty$ (normal) | 68.27%                              |



**Same  $k$  does NOT mean same confidence**  
**when  $\nu$  is finite.**



### Take-home message

- Student's  $t$  distribution accounts for the extra uncertainty in estimating  $\sigma$  from the data.
- Smaller samples  $\Rightarrow$  heavier tails  $\Rightarrow$  more extreme values  $\Rightarrow$  wider intervals.
- As  $N$  increases,  $s$  becomes a better estimate of  $\sigma$ , the  $t$  distribution approaches the normal distribution, and intervals become narrower.

In the limit:  
 $t_\nu \rightarrow N(0,1)$   
as  $\nu \rightarrow \infty$



# 17. Confidence level

For Student's  $t$  distribution with  $\nu$  degrees of freedom,  
the probability that  $|t|$  is less than  $k$  is

$$P = P(|t| < k)$$

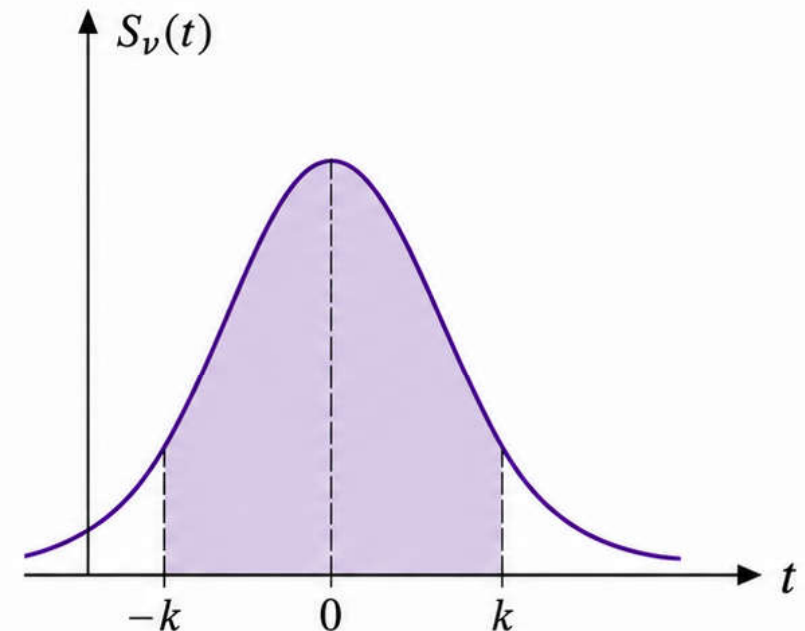
or

$$P = \int_{-k}^k S_{\nu}(t) dt$$

Therefore:

$$P = f(k, \nu)$$

The coverage probability depends on both  
the interval factor  $k$  and the degrees of freedom  $\nu$ .



# 18. 95% confidence interval

For a chosen confidence level:

$$P = 95\%$$

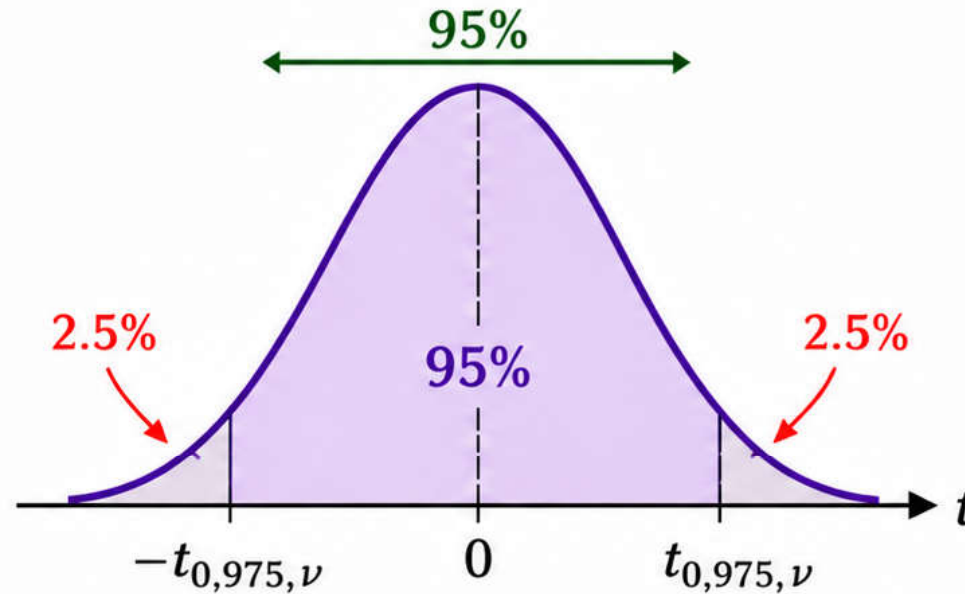
find:

$$t_{0,975,\nu}$$

Then:

$$\bar{x} \pm t_{0,975,\nu} \frac{s}{\sqrt{N}}$$

This should become the  
**central practical equation**  
of the lecture.



**Normal /  $t$  distribution:**

Central area: **95%**

Leaves: **5%**

Split between  
two tails: **2.5% + 2.5%**

$$\text{Therefore: } 1 - \frac{\alpha}{2} = 0.975$$



# 19. Critical $t$ values

Two-sided 95% confidence level ( $P = 95\%$ )

| Sample size $N$         | Degrees of freedom<br>$\nu = N - 1$ | Critical value<br>$t_{0,975,\nu}$ |
|-------------------------|-------------------------------------|-----------------------------------|
| 2                       | 1                                   | 12.706                            |
| 3                       | 2                                   | 4.303                             |
| 5                       | 4                                   | 2.776                             |
| 6                       | 5                                   | 2.571                             |
| 10                      | 9                                   | 2.262                             |
| 20                      | 19                                  | 2.093                             |
| 30                      | 29                                  | 2.045                             |
| $\infty$ (normal limit) | $\infty$                            | 1.960                             |

As  $N$  (and  $\nu$ ) increase,  $t_{0,975,\nu}$  decreases and approaches the normal value 1.960.



The penalty for small  $N$  can be enormous.

Examples (95% two-sided):

- At  $N = 3$  ( $\nu = 2$ ),  $t_{0,975,2} = 4.303$  is more than  $2\times$  larger than the normal value 1.960.
- At  $N = 5$  ( $\nu = 4$ ),  $t_{0,975,4} = 2.776$  is about 42% larger than 1.960.
- Even at  $N = 10$  ( $\nu = 9$ ),  $t_{0,975,9} = 2.262$  is still 15% larger than 1.960.



Small sample sizes require a much larger coverage factor  $k$ , i.e. wider intervals.



# 20. Standard vs expanded uncertainty

A **measurement result** is an estimate of the true value together with a quantitative statement of the uncertainty associated with it.

## 1) Standard uncertainty

The **standard uncertainty** of the mean (standard error of the mean) is

$$u(\bar{x}) = s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

where  $s$  is the sample standard deviation and  $N$  is the sample size.

## 2) Expanded uncertainty

The expanded uncertainty is obtained by multiplying the standard uncertainty by a coverage factor  $k$ :

$$U = k u(\bar{x}) = k \frac{s}{\sqrt{N}}$$

The coverage factor  $k$  is chosen to achieve a stated confidence level  $P$  (e.g. 95%).

## 3) Measurement result

The measurement result is reported as

$$X = \bar{x} \pm U$$

This means that we are  $P \times 100\%$  confident that the true value lies within the interval  $(\bar{x} - U, \bar{x} + U)$ .



**Always report the following with your result  $X = \bar{x} \pm U$ !**

- ① Coverage factor  $k$  (or equivalently the confidence level  $P$ ).
- ② Degrees of freedom  $\nu = N - 1$  (or the sample size  $N$ ).
- ③ Any other relevant information (e.g. method, units, conditions).

All three are **essential** for the correct interpretation of the result.



**Key message (Fornasini)**

To interpret a result written as  $X_0 \pm \delta X$ , the sample size  $N$  (or the degrees of freedom  $\nu$ ) **must always be provided.**



## 21. Example: repeated length measurement

### Suppose:

- We perform repeated measurements of a length.
- The results of  $N = 6$  measurements are summarized by:

$$\bar{x} = 25.42 \text{ mm}$$

$$s = 0.18 \text{ mm}$$

➡ Find the **95%** confidence interval for the true mean length.

### What we want:

$$\bar{x} \pm t_{0.975, \nu} \frac{s}{\sqrt{N}}$$

$$\text{with } \nu = N - 1$$

### Given:

$$N = 6$$

$$\bar{x} = 25.42 \text{ mm}$$

$$s = 0.18 \text{ mm}$$



**Goal:** Compute the **95%** confidence interval for the true mean length based on these repeated measurements.



## 22. Step 2: Student factor



To obtain the confidence interval, we need the **Student factor** (*t* factor).

### 1) Degrees of freedom

$$\nu = N - 1 = 6 - 1 = 5$$

**i** For  $N = 6$  measurements, the degrees of freedom is  $\nu = 5$ .

### 2) Student factor for 95% confidence

From the Student *t* distribution table:

$$t_{0,975,5} = 2.571$$

(Two-sided 95% confidence corresponds to the quantile  $t_{0,975,\nu}$ .)

### 3) Half-width of the confidence interval

$$U = \underbrace{t_{0,975,5}}_{\substack{\text{Student factor} \\ \text{(for } \nu = 5)}} \underbrace{s_{\bar{x}}}_{\substack{\text{standard error from} \\ \text{Step 1}}} = \underbrace{2.571}_{\substack{\text{Student factor} \\ \text{(for } \nu = 5)}} \times 0.0735 \text{ mm} = 0.189 \text{ mm}$$

$$U = 0.189 \text{ mm}$$



## 23. Final result

The 95% confidence interval for the true mean length is

$$\bar{x} = (25.42 \pm 0.19) \text{ mm}$$

at 95% confidence, with:

$$\nu = 5$$

- Number of measurements:  $N = 6$
- Standard deviation of the sample:  $s = 0.18 \text{ mm}$
- Standard error of the mean:  $s_{\bar{x}} = 0.0735 \text{ mm}$
- Student factor (two-sided 95%):  $t_{0.975,5} = 2.571$
- Half-width of the interval:  $U = 0.189 \text{ mm}$

**Comparison: using normal factor incorrectly ( $k = 1.96$ )**

If we ignore the finite sample size and use  $k = 1.96$  (instead of the Student factor), we would get:

$$U = 1.96 \times 0.0735 \text{ mm} = 0.144 \text{ mm}$$

➡ This would give the (incorrectly) narrower interval:

$$\bar{x} = (25.42 \pm 0.14) \text{ mm}$$



**Ignoring finite sample size would underestimate the interval.**

Using the Student factor correctly accounts for the additional uncertainty when  $N$  is small.



## 24. Are two results compatible?

Suppose we have two results:

$$X_1 = x_1 \pm u_1$$

and

$$X_2 = x_2 \pm u_2$$



**Question:**

Is their difference compatible with statistical fluctuations?

**What does “compatible” mean?**

Two results are compatible if any discrepancy between them can reasonably be attributed to random statistical fluctuations.



We test this by comparing the difference to the combined uncertainty of the two results.



This is essential for deciding whether measurements agree, or whether a discrepancy is **statistically surprising**.



This application is explicitly emphasized by **Fornasini**.



## 25. Example from Fornasini

### Given:

- Measured result:

$$\bar{x} = 2.5, \quad u(\bar{x}) = 0.5$$

- Reference (accepted value):

$$m = 1.8$$

### Relative discrepancy ( $t$ ):

$$t = \frac{|\bar{x} - m|}{u(\bar{x})} = \frac{|2.5 - 1.8|}{0.5} = 1.4$$

### Interpretation:

The quantity  $t$  tells us how many “standard uncertainties” separate the two values.

If the results are compatible, we expect  $t$  to be of order 1 (not much larger than 1).

### Question:



Is this discrepancy statistically surprising?



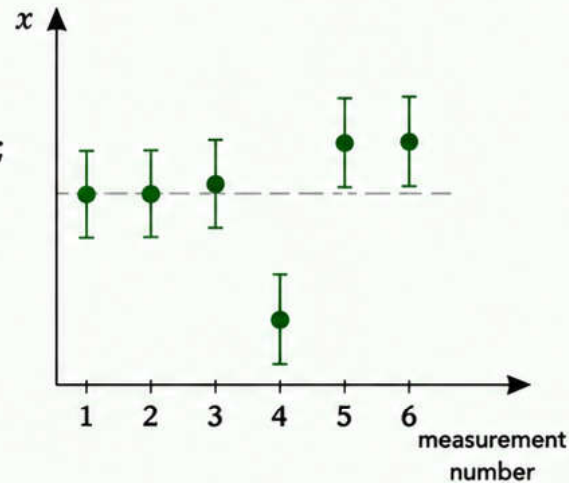
This question leads naturally to the topic of **outliers** and **statistical tests**.



## 26. Why can a point be unusual?

### Three possibilities:

- 1 unusually large statistical fluctuation;
- 2 unidentified systematic error;
- 3 unexpected physical phenomenon.



### Therefore:



**Never remove**  
a data point simply  
because it does not  
fit your expectation.



An unusual value does not automatically mean that it is **wrong** or should be **discarded**.  
We must analyze it statistically.

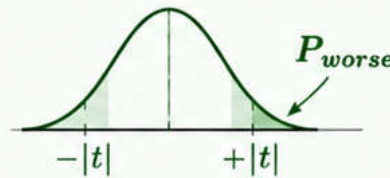


## 27. Chauvenet's criterion

- 1 For each suspected value calculate its **normalized deviation**  $t$ .

$$t = \frac{x_i - \bar{x}}{s}$$

- 2 Then determine the probability  $P_{worse}$  of observing an equal or more extreme deviation.



- 3 Expected number of such observations:

$$N_{worse} = NP_{worse}$$

- 4 Chauvenet criterion:

$$N_{worse} < 0.5$$

➡ **rejection** may be justified.

### Idea:

If such a deviation is so unlikely that we would expect to see less than half a case of it in  $N$  observations, then the value is very probably not a random fluctuation.

### Interpretation:

The smaller  $NP_{worse}$ , the stronger the evidence that the point is an outlier.



This is a statistical rule of thumb proposed by **Chauvenet** (1863).



## 28. Apply it to the example

Given measurements (Fornasini example):

2.7, 2.5, 2.8, **1.5**, 2.4, 2.9

- 1 Sample size:  $N = 6$
- 2 Suspected value:  $x_s = 1.5$
- 3 From the six values:  $|z| = |1.5 - 2.467| / 0.509 \approx 1.90$
- 4 Probability of a deviation this large (two-sided):  $P(|Z| \geq 1.90) \approx 0.057$
- 5 Expected number of such or more extreme observations:  $N P \approx 6 \times 0.057 = 0.344$

**Chauvenet criterion:**

$$N P_{\text{worse}} < 0.5$$

→ rejection may be justified.

**Here:  $N P \approx 0.344 < 0.5$**

### Conclusion

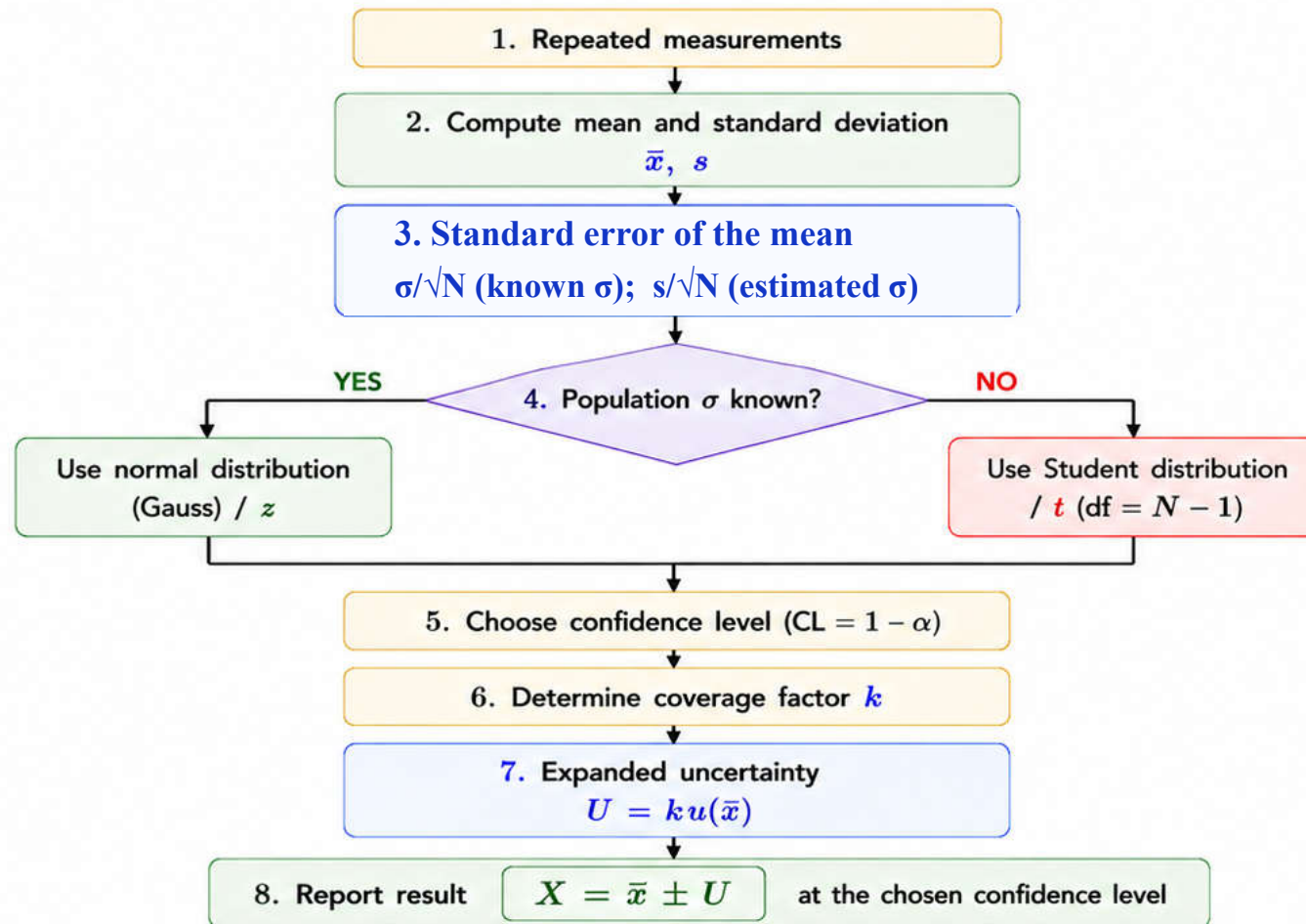
Screening flags 1.5. The mean and s came from these six values, so inspect the record and use a justified outlier method before exclusion.



A statistical flag prompts investigation; it is not an automatic instruction to delete a measurement.



## 29. The complete logic



### Symbols

|              |                                        |
|--------------|----------------------------------------|
| $\bar{x}$    | mean of the measurements               |
| $s$          | sample standard deviation              |
| $N$          | number of measurements                 |
| $u(\bar{x})$ | standard uncertainty of the mean       |
| $k$          | coverage factor                        |
| $U$          | expanded uncertainty                   |
| $X$          | measured quantity                      |
| CL           | confidence level                       |
| $\alpha$     | significance level ( $1 - \text{CL}$ ) |
| df           | degrees of freedom = $N - 1$           |

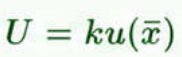


### Key idea:

The type of distribution determines  $k$ , which sets the size of  $U$ .



## 30. Key takeaways

- 1  **Repeat measurements.** Random errors decrease and reproducibility can be checked.
- 2  **Compute  $\bar{x}$  and  $s$ .** They summarize your data.
- 3  **Standard uncertainty of the mean:**  $u(\bar{x}) = \frac{s}{\sqrt{N}}$ .
- 4  **Choose the right distribution.** Normal ( $z$ ) if  $\sigma$  known; Student ( $t$ ) if  $\sigma$  unknown.
- 5  **Choose the confidence level** ( $CL = 1 - \alpha$ ). Typical choices: 68.3%, 95%, 99%.
- 6  **Find the coverage factor  $k$**  from the appropriate table ( $z$  or  $t$ ).
- 7  **Expanded uncertainty:**  $U = ku(\bar{x})$ .
- 8  **Report the result:**  $X = \bar{x} \pm U$  at the chosen CL.

### Good practice

- ✓ **Plan measurements carefully.**  
Good data start with good planning.
- ✓ **Record everything.**  
Raw data, conditions, units, and assumptions.
- ✓ **Analyze objectively.**  
Let statistics decide, not your expectation.
- ✓ **Do not discard data without justification.**  
Outliers may reveal mistakes or new physics.
- ✓ **Report clearly and completely.**  
Others must be able to reproduce your result.

