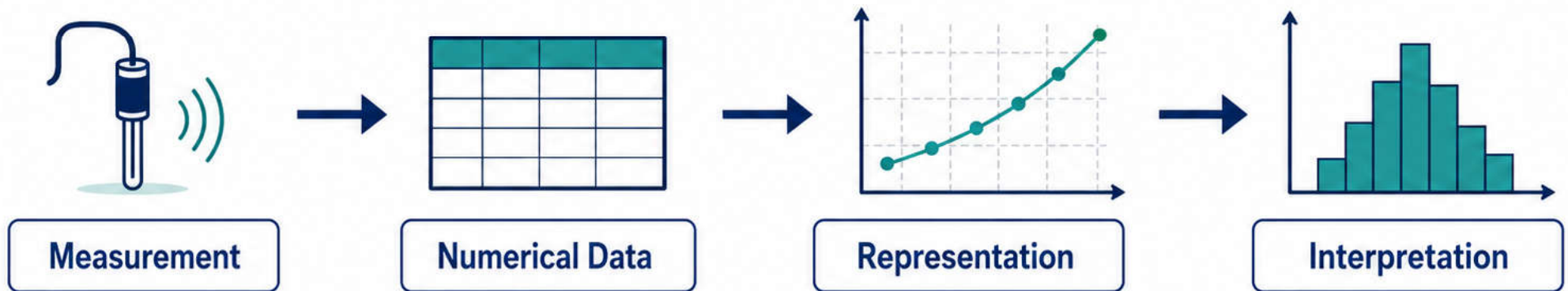




Presentation of Experimental Data

Significant Digits • Rounding • Tables • Graphs • Histograms



From measured values to meaningful scientific information

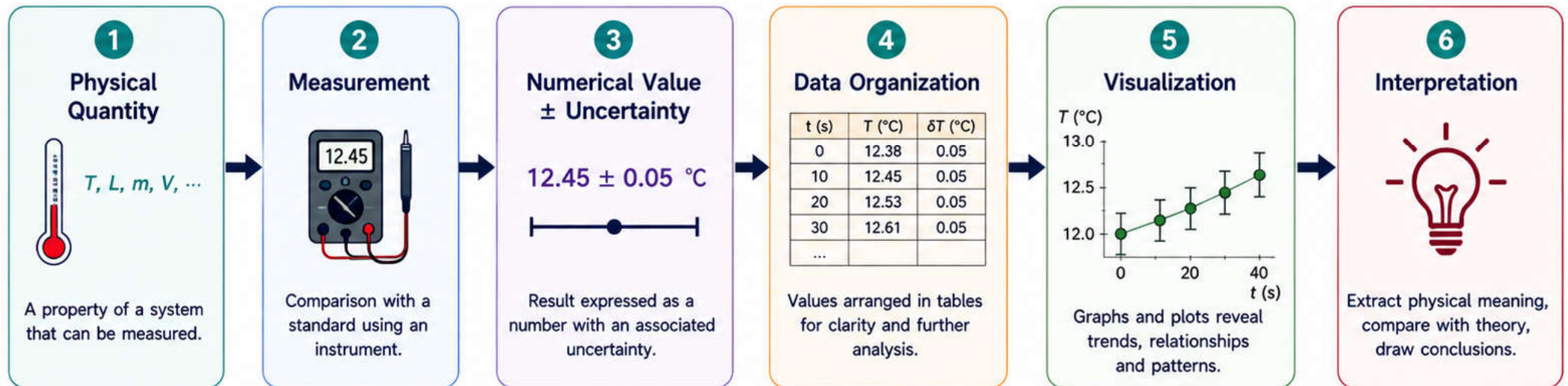




From Measurement to Scientific Information

A measurement alone is not information.

We must process, organize and present data to extract meaning.



Correct presentation of data is essential for **reliable** scientific conclusions and **effective communication**.

Good data presentation = Accuracy + Clarity + Transparency





Outline of Today's Lecture

We will learn how to properly present experimental data to communicate results accurately and effectively.

1

Significant Digits

2.0120

4 significant digits

- Definition and rules
- Zeros and scientific notation
- Precision and information
- Unit conversion

2

Rounding

12.34 → 12.3
12.36 → 12.4
12.355 → 12.4
12.45 → 12.4

- Rounding rules
- Rounding half to even
- Rounding after calculations
- Reporting with uncertainty

3

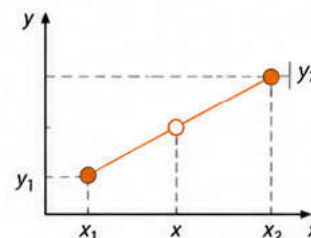
Tables

t (s)	T (°C)	ΔT (°C)
0	12.38	0.05
10	12.45	0.05
20	12.53	0.05
30	12.61	0.05
...	...	...

- Why and when to use tables
- Table structure and content
- Units and uncertainties
- Good practice examples

4

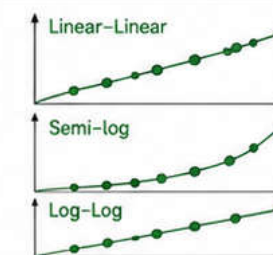
Linear Interpolation



- The interpolation problem
- Linear interpolation formula
- Worked example
- Limits and assumptions

5

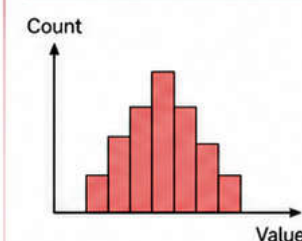
Graphs and Scales



- Graph elements and rules
- Linear and nonlinear scales
- Semi-log and log-log plots
- Choosing the right scale

6

Histograms



- What is a histogram
- Constructing histograms
- Normalization options
- Interpreting distributions



Goal: By the end of this lecture you will be able to present experimental data clearly and correctly using tables, graphs and histograms, and apply appropriate numerical practices.



Good data presentation strengthens your results!



Accurate measurement + Proper analysis + Clear presentation = Reliable science





Exact vs. Approximate Numerical Values

Numbers are either **exact** (known with certainty) or **approximate** (known with limited precision).



Exact Numerical Values

Values known without any uncertainty.
They are defined or counted, not measured.
They have an **infinite** number of significant digits.

Examples

- Counting items: 27 students
- Defined constants: $c = 299\,792\,458\text{ m/s}$ (speed of light in vacuum)
- Mathematical definitions: π , e
- Conversion factors: $1\text{ in} = 2.54\text{ cm}$ (by definition)
- Exact relationships: $V = IR$ (Ohm's law)



Approximate (Measured) Values

Values obtained by measurement or estimation.
They involve uncertainty and therefore have a **limited** number of significant digits.

Examples

- Length of a table: 1.23 m
- Mass of an object: 0.452 kg
- Time interval: 3.67 s
- Temperature: $22.4\text{ }^{\circ}\text{C}$
- Voltage reading: 5.12 V
- Any experimental result

Key Differences

Exact Values	Aspect	Approximate Values
Known with certainty	Nature	Known with limited precision
No uncertainty	Uncertainty	Have uncertainty (error)
Infinite significant digits	Significant Digits	Finite significant digits (reflect precision)
From definitions, counting, or exact relationships	Source	From measurements or estimations
Do not affect significant figures in calculations	Role in Calculations	Determine the precision of the result

Combined in Calculations

Both exact and approximate values can appear in the same calculation.

Examples

- Distance = $(25.0\text{ cm}) \times (2.54\text{ cm/in})$
 $\text{Measured} \times \text{exact} = \text{approximate}$
- Area = $(3.00\text{ m}) \times (4.00\text{ m})$
 $\text{Measured} \times \text{measured} = \text{approximate}$
- Density = $\frac{\text{Mass}}{\text{Volume}}$
 $\text{Measured} / \text{measured} = \text{approximate}$



Takeaway

Exact values are certain by definition or counting. Approximate values come from measurements and carry uncertainty.



Understanding the difference is essential for reporting data and using significant digits correctly.



Common Pitfall

Treating an exact value as approximate (or vice versa) can lead to incorrect rounding and misleading results.



Know the **difference**. Report it **right**.





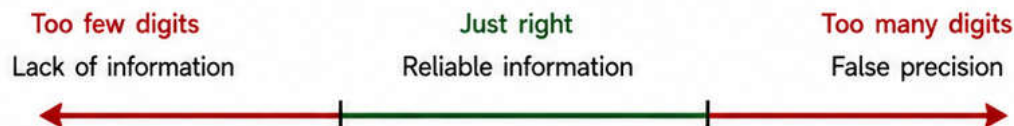
Significant Digits: Why They Matter

Significant digits indicate the **precision** of a measurement and help avoid giving a **false impression** of accuracy.



Key Idea:

We should report all the digits that are meaningful, but no more.



Example Measurement



Length measured:

12.34 mm

- What is the true value?
- What is the uncertainty?
- How many digits are meaningful?

Same Value, Different Reporting

12 mm



Too few digits
Information is lost.



12.34 mm



Just right
Balanced and reliable.



12.340000 mm



Too many digits
False impression
of higher precision.



Takeaway



Significant digits reflect the quality of a measurement.

They help ensure that:

- Results are realistic
- Comparisons are meaningful
- Uncertainty is properly communicated



Report all meaningful digits. Avoid false precision. **Be accurate and transparent.**





What Are Significant Digits?

Significant digits are the digits in a number that **carry meaning** contributing to its **precision**.



Definition

Significant digits (or significant figures) are all the certain digits in a number plus the first uncertain digit. They reflect the **precision** of a measurement.



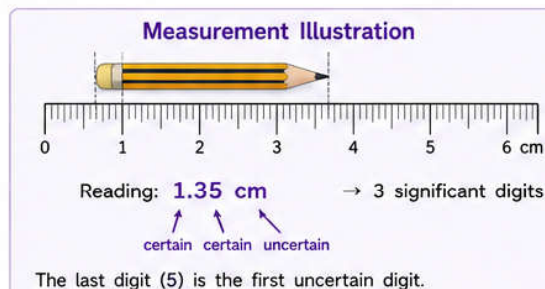
Why It Matters

- They indicate the precision of a measurement.
- They determine how results should be reported.
- They control rounding and calculations.

More significant digits
≠ more accuracy,
but they show
more **precision**.

How to Identify Significant Digits

- 1 All nonzero digits are significant.
(1, 2, 3, 4, 5, 6, 7, 8, 9)
- 2 Zeros between nonzero digits are significant.
(e.g., 1002 → 4 sig. figs.)
- 3 Leading zeros (to the left of the first nonzero digit) are NOT significant.
(e.g., 0.034 → 2 sig. figs.)



- 4 Trailing zeros in a decimal number are significant.
(e.g., 2.50 → 3 sig. figs.)
- 5 Trailing zeros in a whole number WITHOUT a decimal point are NOT significant.
(e.g., 2500 → 2 sig. figs.)

Examples

Number	Significant Digits
7	1
42.8	3
0.0045	2
1002	4
2.50	3
2500	2
2500.	4
0.0700	3

Counting Significant Digits: More Examples

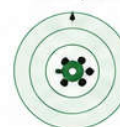
0.000507	600.0	0.30	12000.
0 0 0 0 5 0 7	6 0 0 0 . 0	0 . 3 0	1 2 0 0 0 .
↑ ↑ ↑	↑ ↑ ↑ ↑	↑ ↑	↑ ↑ ↑ ↑
2 significant digits	4 significant digits	2 significant digits	5 significant digits

Significant Digits Indicate Precision

Fewer significant digits
→ less precise



More significant digits
→ more precise



Key Takeaway

Significant digits tell us how precisely a number is known.

Report data with the correct number of significant digits—no more, no less!



Important Note

Significant digits do not guarantee accuracy—they only reflect **precision**. A measurement can be precise (many significant digits) but still **inaccurate** if there is a systematic error.



Count carefully. Report precisely. Communicate clearly.





Rounding Rules

*Rounding is used to report a value with the appropriate number of significant digits.
Look at the digit to be dropped and apply these simple rules.*

- 1** If the first dropped digit is 0, 1, 2, 3, or 4:
Keep the last retained digit the same.

2.461 → 2.46 (to 3 sig. figs.)
12.374 → 12.37 (to 4 sig. figs.)
0.00456 → 0.0046 (to 2 sig. figs.)
7.8184 → 7.82 (to 3 sig. figs.)



Dropped digit < 5 → last retained digit stays the same.

- 2** If the first dropped digit is 5, 6, 7, 8, or 9:
Increase the last retained digit by 1.

2.467 → 2.47 (to 3 sig. figs.)
12.375 → 12.38 (to 4 sig. figs.)
0.00456 → 0.0046 (to 2 sig. figs.)
7.819 → 7.82 (to 3 sig. figs.)



Dropped digit ≥ 5 → round up
(add 1 to the last retained digit).

- 3** All following digits are dropped
after rounding.

3.14159 → 3.14 (to 3 sig. figs.)
25.6789 → 25.68 (to 4 sig. figs.)
1.999 → 2.00 (to 3 sig. figs.)
9.995 → 10.00 (to 4 sig. figs.)



After rounding, keep only the chosen
number of significant digits.



Special Case: All 9's

If rounding causes a carry through (e.g., 9.99 → 10.0),
the number of digits may increase.

Examples: 9.99 → 10.0 (to 3 sig. figs.)
99.95 → 100.0 (to 4 sig. figs.)



Why Rounding Matters

- Communicates precision honestly.
- Prevents false accuracy.
- Ensures consistency in calculations and reporting.
- Makes comparisons meaningful.

Round your results, **not** your **intermediate calculations**.





Uncertainty: Types and How to Report

All measurements have **uncertainty**. Understanding and reporting it is essential for **honest** and **meaningful** results.



What is Uncertainty?

Uncertainty (or error) quantifies the range within which the true value is expected to lie.

A result is incomplete without its uncertainty.



Why Report Uncertainty?

- Shows the reliability of your measurement
- Allows meaningful comparisons
- Helps identify significant differences
- Supports scientific transparency and reproducibility



Types of Uncertainty

- **Random Uncertainty:** Due to unpredictable fluctuations; can be reduced by repeated measurements.
- **Systematic Uncertainty:** Due to biases in instruments or methods; affects accuracy; not reduced by repetition.

Visualizing Uncertainty

Precise and Accurate

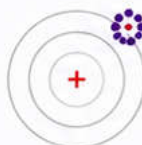
Small spread; close to true value.



+ True value
• Measurements

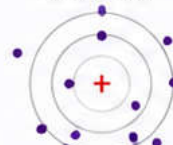
Precise but Not Accurate

Small spread; far from true value.



Not Precise

Large spread; may or may not be accurate.



Good measurements aim for both **accuracy** (closeness to true value) and **precision** (small spread).

How to Report Uncertainty

Report measurements as:

value $\pm$ uncertainty

Use the correct number of significant digits.

Examples:

- Length: $L = 12.34 \pm 0.02$ cm
- Time: $t = (1.562 \pm 0.002)$ s
- Temperature: $T = (22.1 \pm 0.5)$ °C

Rules

- 1 Report uncertainty with 1–2 significant digits.
- 2 Match the decimal place to the value.
- 3 Use the same units.
- 4 State the confidence level if relevant (e.g., 95%).



Example: Multiple Measurements

Five measurements of a length (cm):

12.31 12.33 12.29 12.32 12.30

Mean: $\bar{x} = 12.31$ cm

Standard deviation: $s = 0.015$ cm



Report:

$L = 12.31 \pm 0.02$ cm
(uncertainty rounded to 1–2 sig. figs.)



Key Takeaway

Uncertainty is not a weakness—it is a strength. It shows you understand the limits of your measurement.



Measure **carefully**. Estimate **honestly**. Report **clearly**.





Tables: Organizing Data Clearly

Good tables make your data easy to read, understand and use.



What is a Table?

A table is an organized arrangement of data in rows and columns with clear headings and units.



Why Use Tables?

- Present many values compactly
- Reveal patterns and trends
- Facilitate comparisons
- Serve as a basis for graphs and further analysis

Example: Experimental Data Table

Clear, descriptive title Units in column headings Uncertainty (if applicable)

Period of a Simple Pendulum vs. Length

Trial	Length, L (m)	Time for 10 oscillations, t_{10} (s)	Period, T (s) ($= t_{10}/10$)
1	0.200	9.04	0.904 ± 0.002
2	0.300	11.02	1.102 ± 0.002
3	0.400	12.76	1.276 ± 0.002
4	0.500	14.30	1.430 ± 0.002
5	0.600	15.62	1.562 ± 0.002

Row headings (if needed) Consistent number of significant digits Aligned decimal points Uncertainty reported with the value



Best Practices

- **Clear title:** Summarizes what the table shows.
- **Column headings:** Include quantity and unit.
- **Units:** Use SI units and write them in the headings.
- **Significant digits:** Be consistent within each column.
- **Uncertainties:** Report alongside the value or in a separate column.
- **Alignment:** Align decimal points for easy comparison.
- **Simplicity:** Include only necessary data.
- **Footnotes:** Use symbols (*) and notes for additional information if needed.

Avoid: Overcrowding, unclear headings, mixed units, inconsistent digits.

Typical Table Structure

Title (what, where, when) →

Column headings (quantity and unit) →

Data (organized) with consistent format →

Table 1. Measurement of Resistance vs. Length

Length, L (m)	Voltage, V (V)	Current, I (mA)	Resistance, R (Ω)
...	...	...	...
...	...	...	...
...	...	...	...

Notes and Footnotes

- Use notes below the table for explanations.
- Use symbols (*, †, ‡, ...) for footnotes.
- Example:

* Uncertainty is expanded uncertainty, $k = 2$ ($\approx 95\%$ confidence).

Example of Note:

All measurements were made at room temperature (22 ± 1 °C).

Well-organized tables lead to *clear insights* and *reliable conclusions*.





Good Table vs. Bad Table

Well-designed tables make data easy to read, understand, and use. Poorly designed tables hide information and lead to mistakes.



Good Table

Clear, complete, consistent, and easy to understand.

Example: Resistance of a Wire vs. Length

Length, L (m)	Diameter, d (mm)	Resistance, R (Ω)	Uncertainty in R (Ω)
0.50	0.50	1.26	± 0.02
1.00	0.50	2.53	± 0.02
1.50	0.50	3.79	± 0.03
2.00	0.50	5.06	± 0.03
2.50	0.50	6.32	± 0.03

- ✓ Clear and descriptive column headings with units.
- ✓ Units are shown in parentheses.
- ✓ Consistent number of significant figures.
- ✓ Data is aligned in columns (easy to read).
- ✓ Uncertainty is included with the data.
- ✓ Table has a concise, descriptive title.
- ✓ No unnecessary decimal places.



Principles of Good Tables

- 1 Give the table a clear, concise title.
- 2 Use column headings that describe the data.
- 3 Include units in the headings (in parentheses).
- 4 Report uncertainties when appropriate.
- 5 Use a consistent number of significant figures.
- 6 Align numbers by decimal points.
- 7 Use appropriate number of rows and columns.
- 8 Keep the table simple and uncluttered.



Why It Matters

Good tables communicate information clearly and reduce the chance of misinterpretation.



Bad Table

Unclear, inconsistent, and hard to understand.

Example: resistance vs length

L (m)	dia (mm)	R (ohms)	error
.5	0.5	1.2589	.02
1	.50	2.53	0.02
1.5	0.500	3.7893	
2	.5	5.1	.03
2.50	0.50	6.317	0.03000
3	0.5	7.58	0.03
3,5	.500	8,84	.03

- ✗ Unclear or incomplete column headings.
- ✗ Units are missing or not shown consistently.
- ✗ Inconsistent significant figures and decimal places.
- ✗ Numbers are not aligned by decimal points.
- ✗ Uncertainty is missing in some rows.
- ✗ Too many decimal places in some entries.
- ✗ Inconsistent notation (0.5 vs .5, commas used).
- ✗ Extra or unnecessary rows make it cluttered.



Key Takeaway

A well-designed table is not just about how the data looks—it's about making the data **easy to understand and use**.



Think: If someone else read your table for the first time, would they understand it without asking questions?



Good tables tell the story. Bad tables hide the truth.





Linear Interpolation: Estimating Between Points

Interpolation is used to estimate a value that lies between measured data points.



What is Interpolation?

Interpolation estimates an intermediate value between two known data points assuming a smooth (often linear) trend.

When to Use Linear Interpolation

- Data points are close together
- The trend between points is approximately linear
- No sharp changes or discontinuities
- You need an estimate within the range of your data

How Linear Interpolation Works

- 1 Identify the two data points that surround the desired value: (x_1, y_1) and (x_2, y_2)

- 2 Calculate the slope between the points:

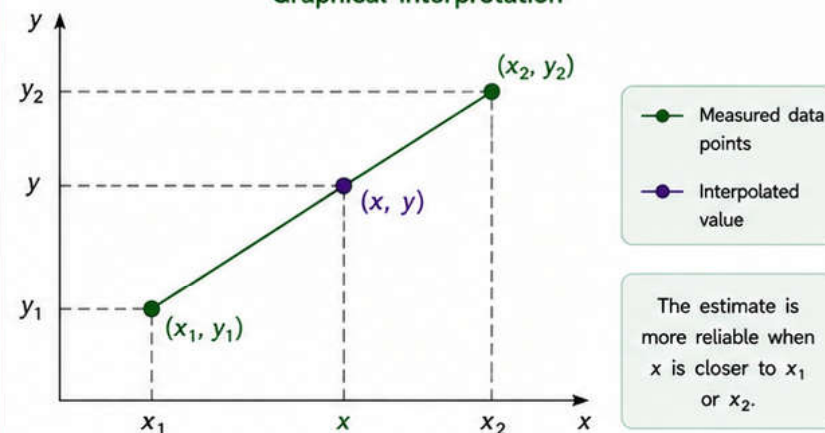
$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

- 3 Use the point-slope form to estimate y at x :

$$\begin{aligned} y &= y_1 + m(x - x_1) \\ &= y_1 + \left(\frac{y_2 - y_1}{x_2 - x_1} \right) (x - x_1) \end{aligned}$$

- 4 Report the interpolated value with appropriate significant digits.

Graphical Interpretation



Example

A thermistor has the following calibration data:

T (°C)	R (kΩ)
20	12.50
30	9.20
40	6.80

Estimate R at $T = 35$ °C.

Solution:

- 1 Points: (30 °C, 9.20 kΩ) and (40 °C, 6.80 kΩ)
- 2 $m = (6.80 - 9.20) / (40 - 30) = -0.24$ kΩ/°C
- 3 $R(35) = 9.20 + (-0.24)(35 - 30) = 8.00$ kΩ
- 4 Report: $R(35$ °C) = 8.00 kΩ (3 significant digits)



Limitations

- Outside the data range → use extrapolation (risky).
- Nonlinear behavior → linear interpolation may be inaccurate.
- Large gaps between points → higher uncertainty.

Always consider the uncertainty of the interpolated value. It is typically larger than that of measured points.



Interpolate *within the data range*, with care and awareness of *uncertainty*.





When Does Linear Interpolation Fail?

Linear interpolation assumes a straight-line relationship between data points.

It can give large errors or even misleading results when this assumption is not valid.



Linear Interpolation Works Well When...

- The underlying relationship is approximately linear over the interval.
- Data points are close together.
- The function changes smoothly (no sharp bends, jumps, or thresholds).

Always check if the data behave linearly before trusting the result.



Linear Interpolation Fails When...

- The relationship is curved (nonlinear).
- The interval between data points is large.
- There are abrupt changes, discontinuities, or thresholds.
- Near a maximum, minimum, or inflection point.
- The data contain large scatter or measurement error.

In these cases, linear interpolation can produce large errors and false confidence.



What to Do Instead

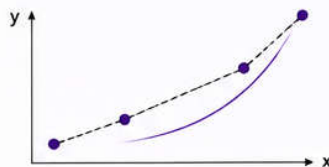
- Use more data points (smaller intervals).
- Use a more appropriate model (polynomial, logarithmic, exponential, etc.).
- Use graphical methods to understand the trend.
- Report the uncertainty and potential error.

Understand the behavior of the data—don't rely on straight lines blindly.

Examples of Failure

1

Strong Curvature

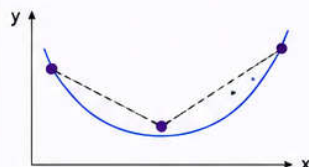


The true relationship is curved. Linear interpolation underestimates (or overestimates) in the middle.

→ **Large error**

2

Large Interval

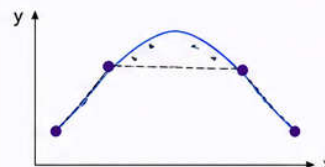


The interval between data points is too large to capture the trend. The straight line misses the shape.

→ **Large error**

3

Near a Maximum / Minimum

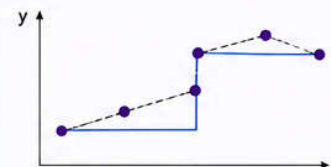


Near peaks or valleys, the slope changes sign. Linear interpolation cannot capture this.

→ **Wrong trend**

4

Discontinuity / Threshold

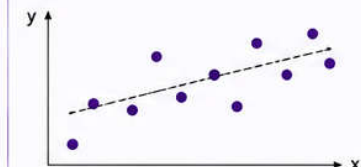


A sudden jump or threshold exists. Linear interpolation predicts values that never occur.

→ **Misleading result**

5

Noisy Data (Large Scatter)



High scatter hides the true trend. Linear interpolation may follow noise instead of the underlying behavior.

→ **Unreliable result**



Key Takeaway

Linear interpolation is a useful tool, but it is not always valid.

Always evaluate the **behavior of your data** and the **size of the interval** before trusting an interpolated value.



Before You Interpolate, Ask Yourself

- Does the data look linear in this region?
- Are the data points close enough together?
- Could there be curvature, thresholds, or jumps?
- How large could the interpolation error be?



Best Practice

- Plot the data first.
- Use the simplest method that is appropriate.
- When in doubt, estimate the possible error and report it.



Interpolate **carefully**. Understand your data. Avoid **false confidence**.





Scientific Notation: Expressing Very Large and Very Small Numbers

A compact way to write numbers in the form $a \times 10^n$.



What is Scientific Notation?

Scientific notation expresses a number as the product of a number between **1 and 10** (the **significant**) and a power of 10.

$$N = a \times 10^n$$

$1 \leq a < 10$, n is an integer

Examples

$$6.022 \times 10^{23}$$

- Avogadro constant
- 602,200,000,000,000,000,000,000

$$1.5 \times 10^{-12}$$

- Pico-scale quantity
- 0.0000000000015

$$3.0 \times 10^{-6}$$

- Micro-scale quantity
- 0.000003

$$4.7 \times 10^5$$

- Large number
- 470,000



How to Convert to Scientific Notation

- 1 Move the decimal point so that the significant is between 1 and 10.
- 2 Count the number of places the decimal was moved.
- 3 Use a positive exponent if the decimal moved to the left (large number).
- 4 Use a negative exponent if the decimal moved to the right (small number).

Examples:

- $65000 = 6.5 \times 10^4$ (moved 4 places left)
- $0.0042 = 4.2 \times 10^{-3}$ (moved 3 places right)
- $7.6 = 7.6 \times 10^0$ (already in range)



Benefits

- ✓ Makes very large and very small numbers easy to read and compare.
- ✓ Helps avoid long strings of zeros.
- ✓ Improves clarity in scientific and technical work.



Quick Example

Write 0.00056 in scientific notation:

- Move the decimal 4 places right → 5.6×10^{-4}



Common Prefixes

Tera (T) = 10^{12}
Giga (G) = 10^9
Mega (M) = 10^6
Kilo (k) = 10^3

Milli (m) = 10^{-3}
Micro (μ) = 10^{-6}
Nano (n) = 10^{-9}
Pico (p) = 10^{-12}



Key Takeaway

Use **scientific notation** ($a \times 10^n$) for very large or very small numbers to keep values compact and clear.



Scientific notation makes **big** and **tiny** numbers easy to read.



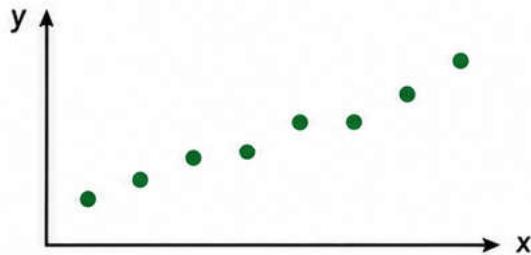


Why Do We Plot Experimental Data?

Plotting experimental data turns numbers into understanding.



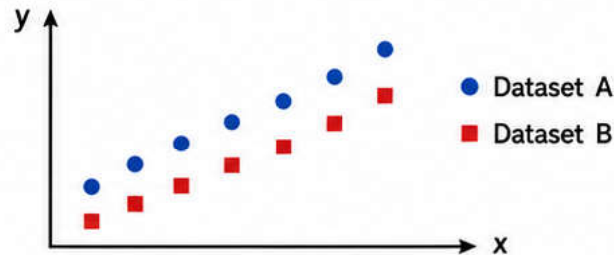
VISUALIZE



Graphs let us **see** the overall behavior of the data at a glance.



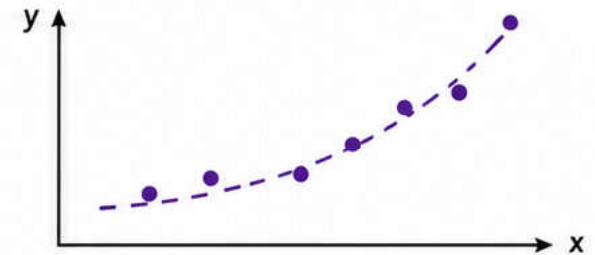
COMPARE



Graphs make it easy to **compare** different datasets or conditions.



DISCOVER RELATIONSHIPS



Graphs help **reveal** relationships, trends, and structures that may not be obvious in tables of numbers.



Data



Pattern



Physical interpretation

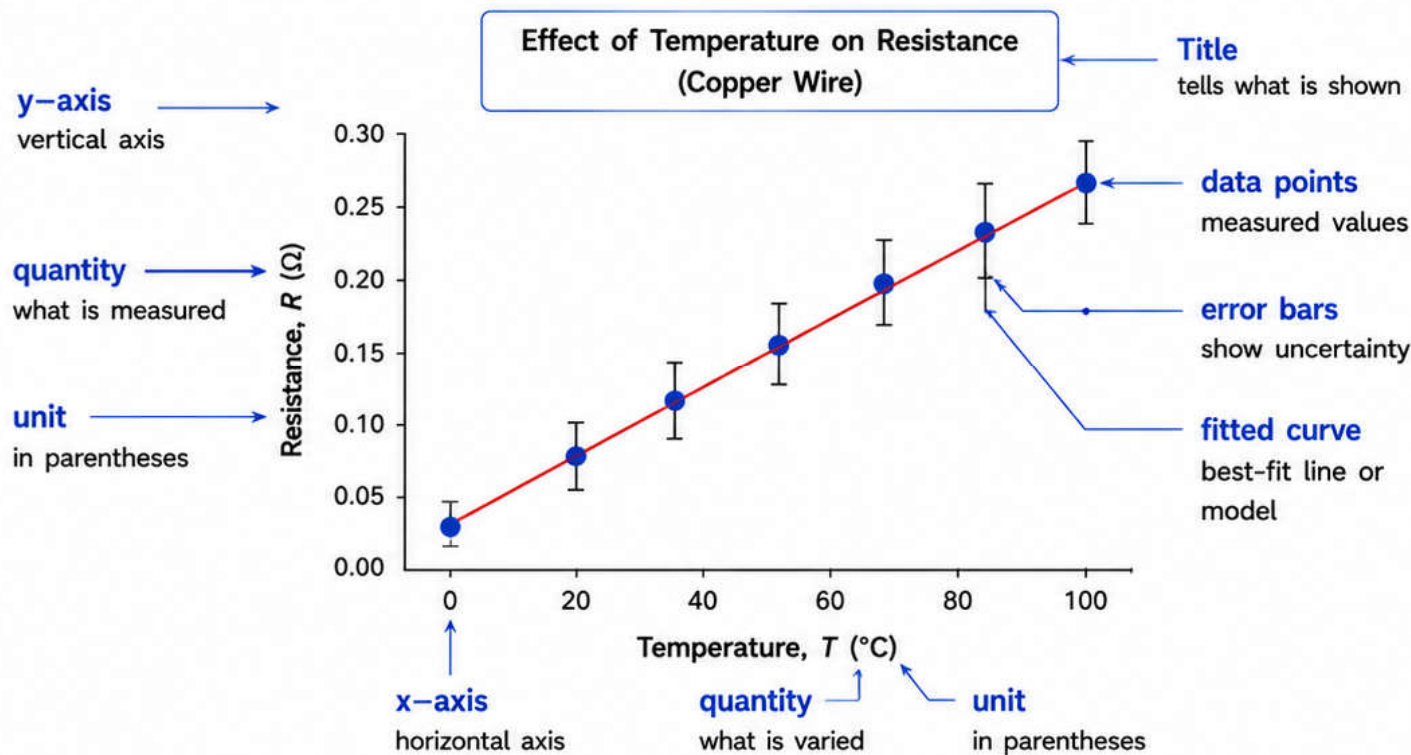
Good graphs don't just show data. *They tell a story.*





Anatomy of a Scientific Graph

A correct scientific graph contains all the essential elements.



x-axis

Horizontal axis.



y-axis

Vertical axis.



quantity

What is measured or plotted.



unit

Written in parentheses.



data points

Measured values.



error bars

Represent uncertainty.



fitted curve

Best-fit line or theoretical model.



Key idea: Every element of a scientific graph has a purpose. Together, they make the data clear, complete, and interpretable.

A well-made graph is *easy to read* and *hard to misunderstand*.





General Rules for Scientific Graphs

Follow these rules to make your graphs clear, correct, and meaningful.



Independent variable → x



Put the independent (controlled or varied) variable on the x-axis. It does not depend on the measured outcome.



Dependent variable → y



Put the dependent (measured) variable on the y-axis. It depends on the independent variable.



Axes → quantity + unit



Label every axis with the quantity name and the unit in parentheses. Example: Time, t (s)



Measured values → symbols



Plot measured data as symbols (points), not connected lines. Choose clear, consistent symbols.



Uncertainty → error bars



Show uncertainty with error bars in both x and/or y directions, whenever it is known.



Fit/model → line



If a model or fit is appropriate, draw the best-fit line or curve. Do not connect points unless it is a model.



Remember: A good graph follows the rules so others can trust your data and understand your results.



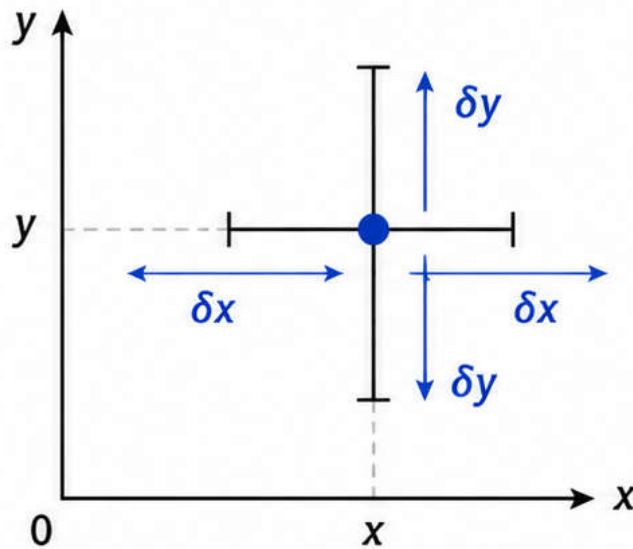
Good graphs *communicate* science.





Error Bars Carry Physical Information

Every measurement has uncertainty. Error bars show it.



A measured data point represents:

$$(x \pm \delta x, y \pm \delta y)$$

horizontal error bar $\rightarrow \delta x$

vertical error bar $\rightarrow \delta y$



What error bars tell us:



They indicate the precision of the measurement.



They determine how much weight the point has in analysis.



They affect the reliability of comparisons and fits.



A data point without uncertainty contains incomplete measurement information.

Report *uncertainties*. They are *part of the measurement*, not an afterthought.





Characterizing Measurement Uncertainties: Box Plot

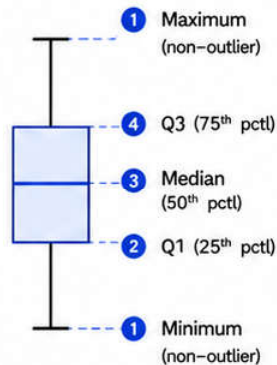


Key idea: Box plots provide a clear visual summary of the spread, central value, and potential outliers in repeated measurements.

What is a Box Plot?

A box plot (or box-and-whisker plot) displays the distribution of data using five key summary statistics:

- | | |
|-----------------------|--|
| 1 Minimum | Smallest (non-outlier) value |
| 2 Q1 (First quartile) | 25 th percentile |
| 3 Median | 50 th percentile (middle value) |
| 4 Q3 (Third quartile) | 75 th percentile |
| 5 Maximum | Largest (non-outlier) value |



Example: Repeated Length Measurements

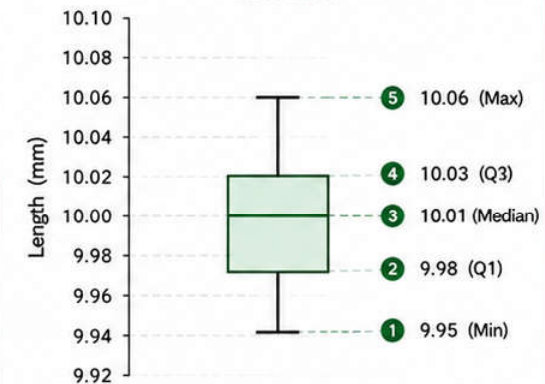
Data (mm):

9.96, 10.02, 10.01, 9.98,
10.05, 9.99, 10.00, 10.03,
9.97, 10.02, 10.04, 10.01,
9.95, 10.00, 10.06

Five-number summary (mm):

Min	9.95
Q1	9.98
Median	10.01
Q3	10.03
Max	10.06

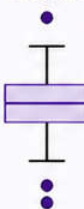
Box Plot



Interpreting Box Plots

- The box contains the middle 50% of the data (Q1 to Q3).
- The line inside the box is the median.
- The whiskers show the typical spread of the data.
- Points beyond the whiskers are potential outliers.

Outliers



Outliers are usually defined as values more than $1.5 \times \text{IQR}$ beyond Q1 or Q3, where $\text{IQR} = Q3 - Q1$.



Why use box plots?

- ✓ Summarize data compactly.
- ✓ Reveal spread and central tendency.
- ✓ Identify outliers and asymmetric distributions.
- ✓ Ideal for comparing multiple measurements.



Box plots make measurement uncertainties *visible, interpretable, and easy to compare.*

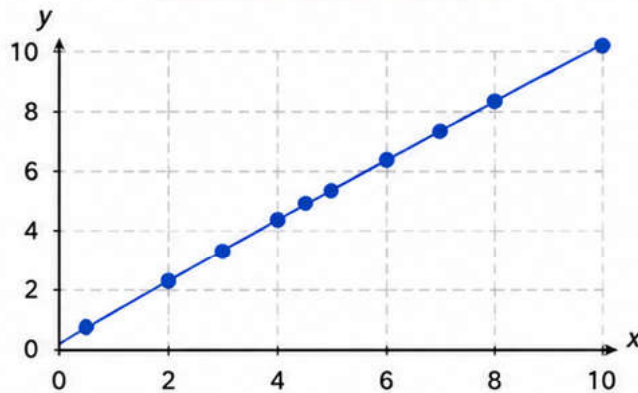




Choosing the Right Scale

The scale you choose should make the relationship easy to see and easy to understand.

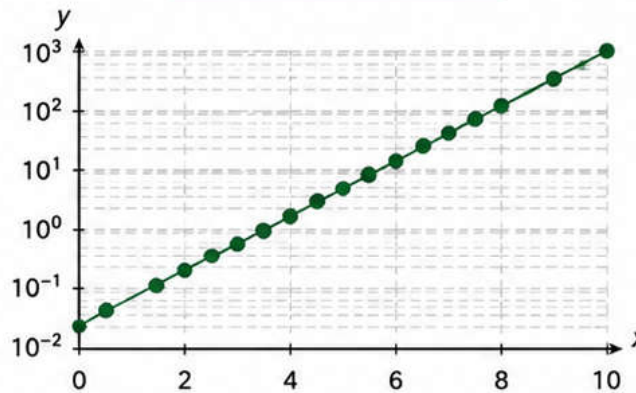
Linear-Linear



Linear relation

Use linear-linear scales when the data vary linearly.
Both axes are **linear**.

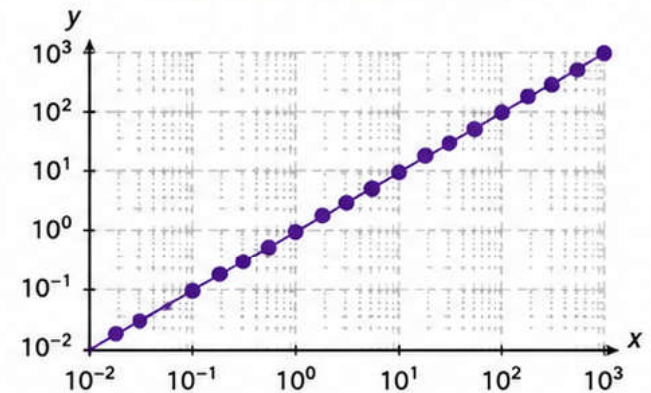
Semi-log



Exponential relation

Use semi-log (log y) scales when the data follow an exponential trend.
One axis is **logarithmic**, the other is **linear**.

Log-Log



Power-law relation

Use log-log scales when the data follow a power-law (scale-free) trend.
Both axes are **logarithmic**.



Key idea: Choose the scale that makes your data look as straight as possible.
A straight line reveals the simplest underlying relationship.



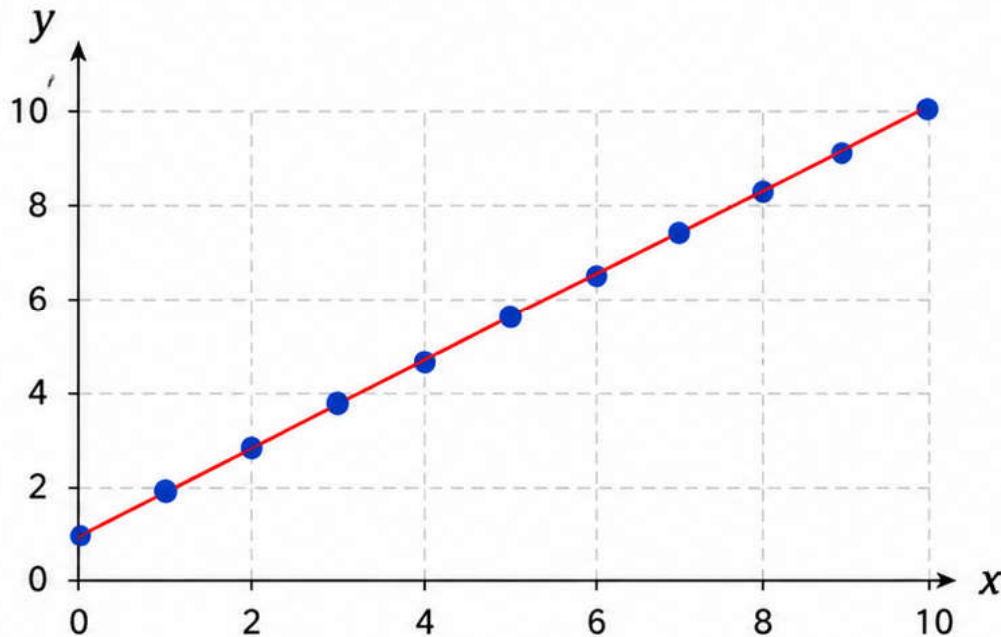
The right scale turns **complexity** into **clarity**.





Linear Scale

On a linear-linear scale, a linear relationship appears as a straight line.



$$y = a + bx$$

Slope:

$$b = \frac{\Delta y}{\Delta x}$$

The slope b tells how much y changes for a unit change in x .

Intercept:

$$a = y(0)$$

The intercept a is the value of y when $x = 0$.



Key idea: On a linear scale, the equation $y = a + bx$ describes a straight line. The slope and intercept summarize the relationship.

Linear scale reveals *linear relationships*.





Why Transform the Axes?

Make nonlinear relationships visually linear.

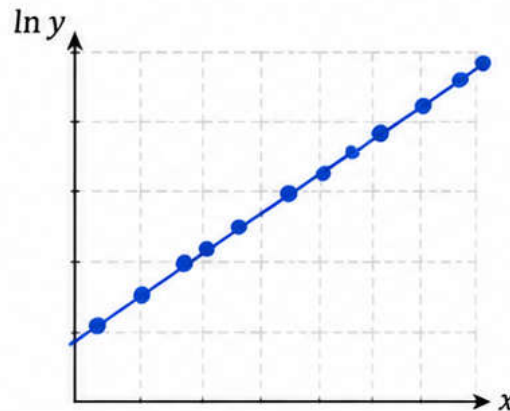
Exponential relation

$$y = ae^{bx}$$



Plot $\ln y$ vs. x

Then $\ln y = \ln a + bx$,
which is a straight line.



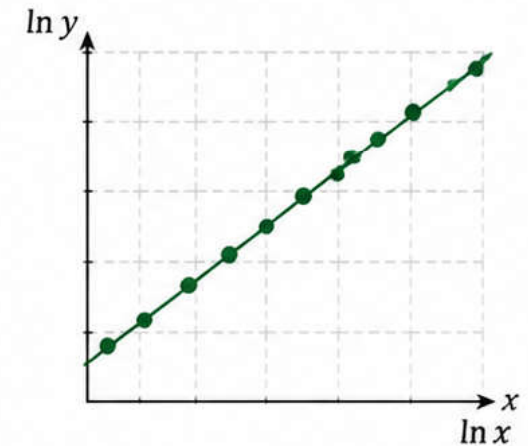
Power law

$$y = ax^b$$



Plot $\ln y$ vs. $\ln x$

Then $\ln y = \ln a + b \ln x$,
which is a straight line.



Key idea: By transforming the axes, we can reveal simple linear behavior and determine the model parameters from a straight line.



Choose the right scale to *see the physics*.





Semi-log and Log-log Scales



Key idea: Use logarithmic scales when data span several orders of magnitude or follow exponential or power-law behavior.

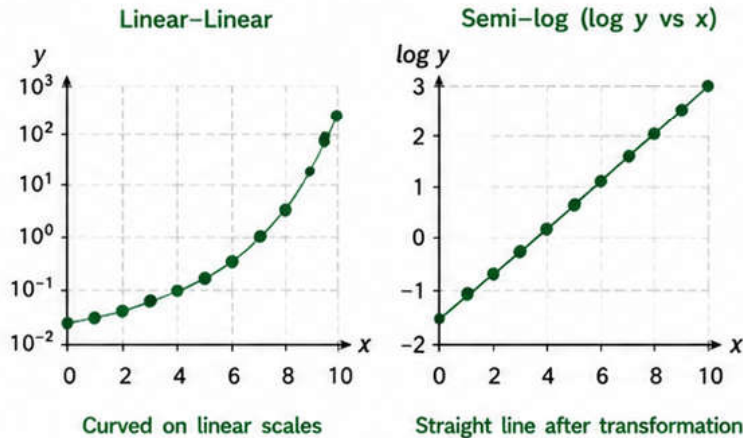
Semi-log Scale (log y vs x)

Use when y changes exponentially with x .

Example:

$$y = ae^{bx}$$

An exponential relationship is curved on linear scales but becomes a straight line on a semi-log plot.



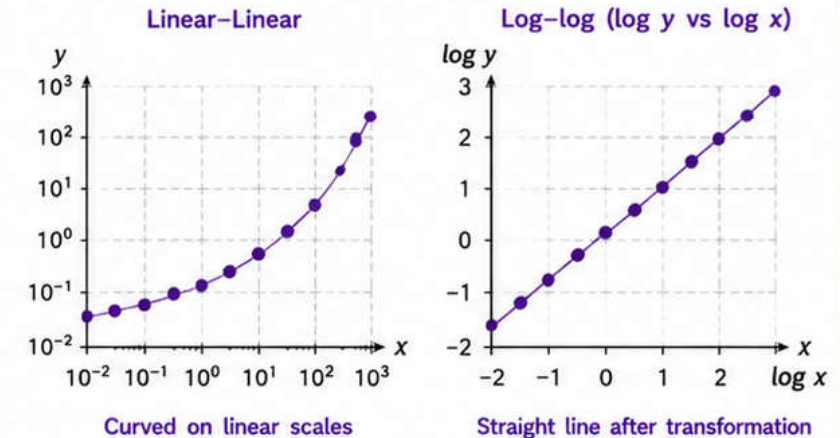
Log-log Scale (log y vs log x)

Use when y follows a power law with x .

Example:

$$y = ax^b$$

A power-law relationship is curved on linear scales but becomes a straight line on a log-log plot.



Why use log scales?



They handle data that vary over many orders of magnitude.



They reveal simple relationships hidden in complex data.



They allow easier comparison and modeling.

Logarithmic scales *compress large ranges* and *reveal simple underlying patterns*.





What Is a Histogram?



Key idea: A histogram displays the distribution of numerical data by grouping values into bins and showing how often they occur.

Raw Measurements

2.15 2.14 2.17 2.15 2.16 ...
2.18 2.13 2.16 2.14 2.15 ...
2.17 2.16 2.15 2.18 2.14 ...
⋮

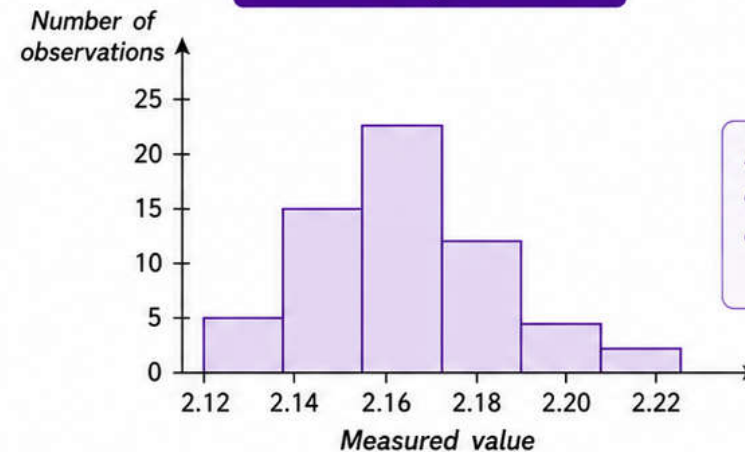


Group into Bins

2.12 – 2.14
2.14 – 2.16
2.16 – 2.18
2.18 – 2.20
⋮



Histogram



Axes:

- $x \rightarrow$ measured value
- $y \rightarrow$ number of observations



A histogram turns a list of measurements into a clear **visual summary** of how the data are distributed.

Histograms are the bridge from **raw measurements** to **statistical thinking**.





Constructing a Histogram



Key idea: A histogram is built by grouping data into bins (intervals) and counting how many observations fall in each bin.

1



Determine range

Find the minimum and maximum values.

$$\text{Range} = x_{\max} - x_{\min}$$

2



Choose bin width Δx

Select a convenient bin width that balances detail and clarity.

3



Assign values to bins

Decide the bin edges and place each value into the appropriate bin.

4



Count occurrences n_j

Count how many observations fall in each bin j .

5



Plot bars

Plot adjacent bars with height n_j (or other normalization).

$$N = \sum_{j=1}^J n_j \quad \text{Total number of observations}$$

Example

Raw data (20 measurements)

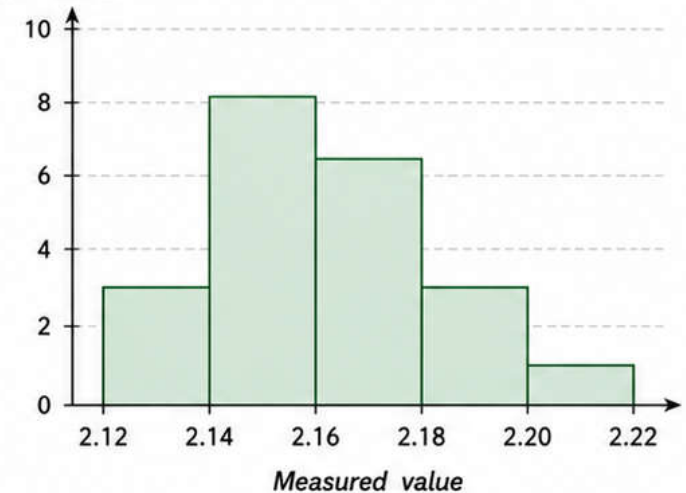
2.15	2.14	2.17	2.15	2.16
2.18	2.13	2.16	2.14	2.15
2.17	2.16	2.15	2.18	2.14
2.19	2.17	2.16	2.20	2.15

Binning and counts

Bin j	Interval	Count n_j
1	2.12 – 2.14	3
2	2.14 – 2.16	7
3	2.16 – 2.18	6
4	2.18 – 2.20	3
5	2.20 – 2.22	1
Total N		20

Number of observations

Histogram (counts)



Good practice



Choose bin width thoughtfully.



Show bin edges clearly.



Always report N and Δx .



A histogram turns *raw measurements* into a clear picture of the data's *distribution*.





Why Bin Width Matters

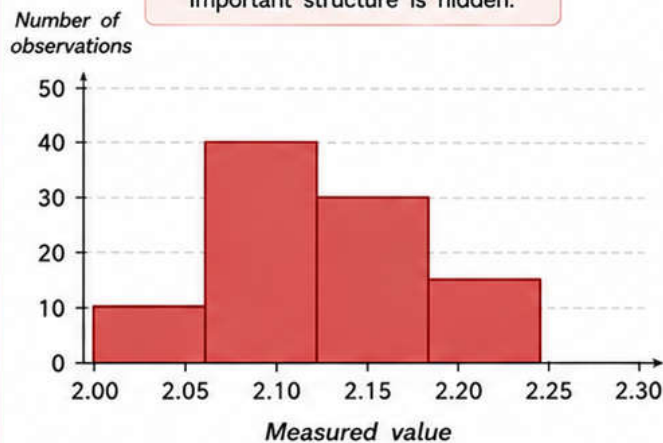


Key idea: The choice of bin width Δx affects how much detail or noise appears in the histogram. There is no single “best” bin width—use judgment and consider the purpose of the analysis.

The Same Data (100 Measurements)

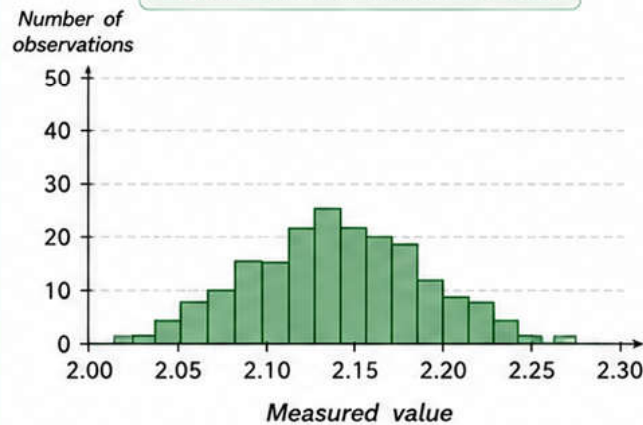
Too Few Bins

→ Information lost
Important structure is hidden.



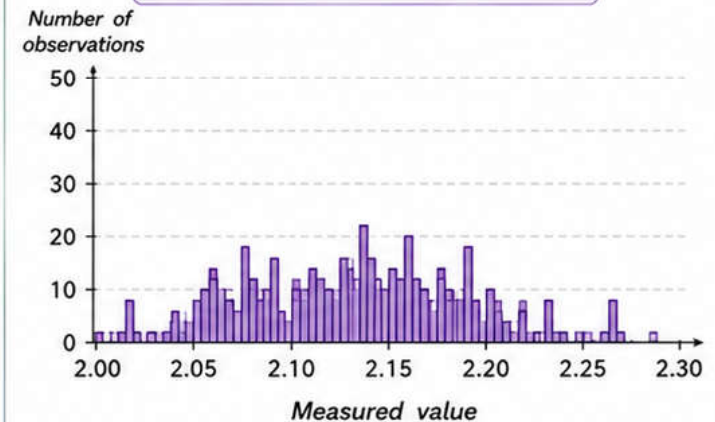
Reasonable Binning

→ Structure visible
Overall shape and features are clear.



Too Many Bins

→ Noisy appearance
Random fluctuations obscure the pattern.



Guidelines

- Start with a **moderate** number of bins (e.g., 10–20).
- Adjust bin width to reveal **real structure** without adding excessive noise.
- Consider **sample size**: larger N allows smaller bin widths.

Bin width controls the balance between **information** and **noise**. Choose wisely!





Histogram Normalization

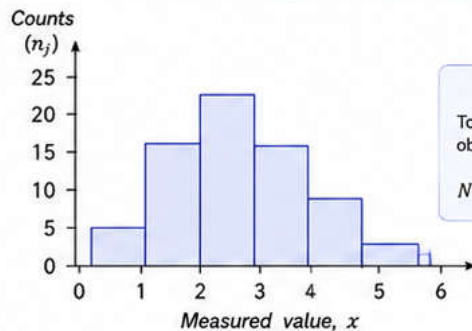


Key idea: Histograms can be displayed in different ways. Normalization allows us to compare distributions, estimate probabilities, and relate histograms to probability density.

1) Counts

The bar height is the number of observations in each bin.

n_j = number of observations in bin j



Example

Total number of observations:

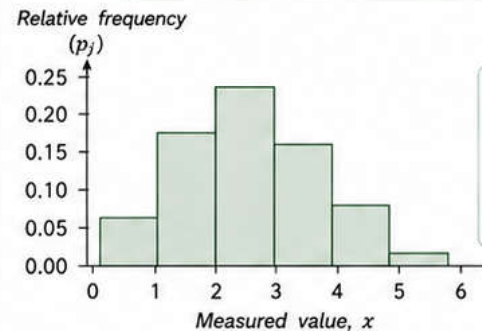
$$N = \sum_j n_j = 100$$

- Bar height depends on both the frequency and the bin width.
- Not suitable for comparing histograms with different N or Δx .

2) Relative Frequency

The bar height is the fraction of observations in each bin.

$$p_j = \frac{n_j}{N} = \text{relative frequency in bin } j$$



Properties

$$0 \leq p_j \leq 1$$

$$\sum_j p_j = 1$$

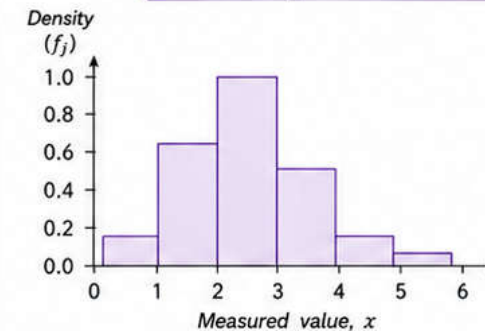
- Useful when all histograms use the same bin width.
- Shows the proportion of data in each bin.

Key result: Relative-frequency histogram: $\sum_j p_j = 1$

3) Density (Height-Normalized)

The bar height is the frequency per unit measured value.

$$f_j = \frac{n_j}{N \Delta x_j} = \text{density in bin } j$$



Properties

$$f_j \geq 0$$

$$\sum_j f_j \Delta x_j = 1$$

- Required when bin widths are different.
- Corresponds to a probability density function.

Key result: Density histogram: $\sum_j f_j \Delta x_j = 1$



Important:

Counts → how many

Relative frequency → what fraction

Density → how concentrated per unit value

Normalization makes histograms comparable and connects them to probability.





Why Normalize Histograms?



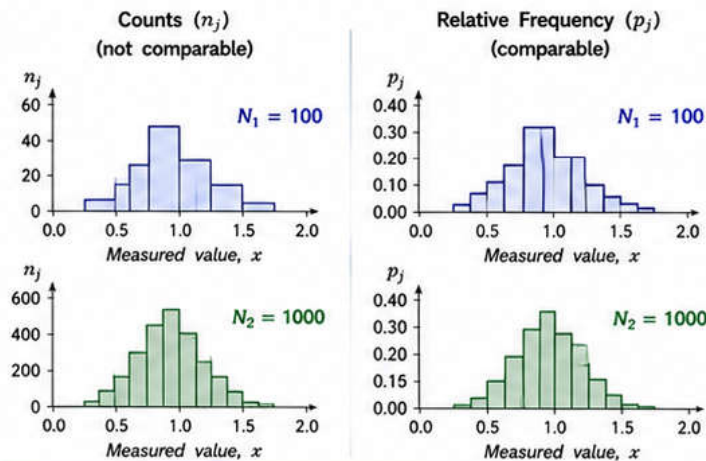
Key idea: Normalization allows meaningful comparison between histograms, even when sample sizes or bin widths differ.

1) Same Bin Width, Different Sample Sizes (N)

Relative frequency (or density) allows comparison of distributions, independent of the total number of measurements.

Same bin width: $\Delta x = 0.2$
Same underlying distribution
Different sample sizes

- $N_1 = 100$
- $N_2 = 1000$



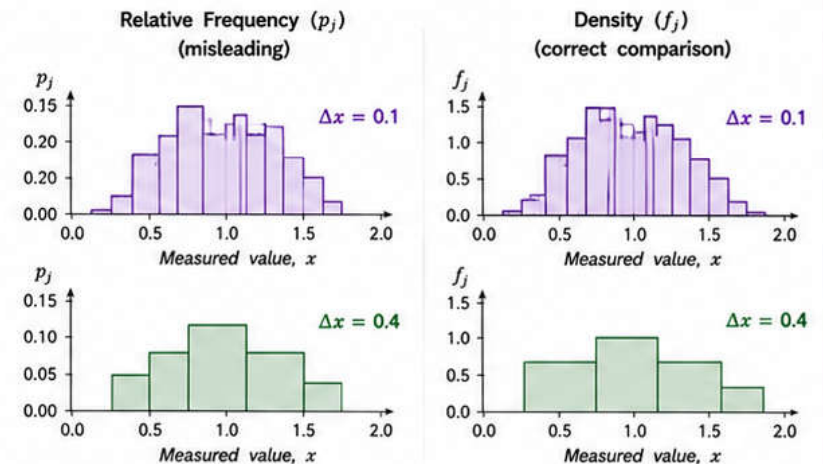
Conclusion: Counts depend on N . Relative frequency (or density) does not.
Use relative-frequency histograms to compare distributions with the same bin width.

2) Different Bin Widths (Δx)

Density (height-normalized) histogram is required when bin widths differ. Counts or relative frequency can be misleading.

Same data ($N = 300$)
Different bin widths

- Narrow bins: $\Delta x = 0.1$
- Wide bins: $\Delta x = 0.4$



Conclusion: With different bin widths, only the density histogram provides a valid comparison.
Use density (height-normalized) histograms when Δx varies.



Histogram **height** and histogram **area** do not necessarily represent the same quantity.

Height $\rightarrow$ frequency or density

Area $\rightarrow$ probability

Normalization is essential for **meaningful comparison** and for connecting histograms to **probability**.





Worked Example: From Measurement to Scientific Figure



Key idea: A complete experiment does not end with taking measurements. It ends with clear, correct, and interpretable communication of the results.

1. Raw measurements

Example: Mass of object (g) measured 15 times

12.3741
12.3750
12.3734
12.3748
12.3742
12.3739
12.3746
12.3735
12.3743
12.3740
12.3747
12.3736
12.3744
12.3742
12.3738

Instrument resolution: 0.001 g
Uncertainty type B: 0.002 g

2. Correct rounding

Round according to the certainty of the result

Rounded values

12.374
12.375
12.373
12.375
12.374
12.374
12.375
12.374
12.375
12.374
12.374
12.375
12.374
12.374
12.374

Report at most to 0.001 g

3. Table

Organize the data and summarize the result

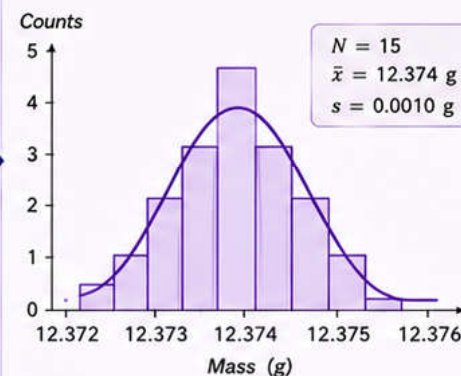
Quantity	Value	Uncertainty	Unit
Mean ($\bar{x}$)	12.374	0.001	g
Std. dev. (s)	0.0010		g
Std. uncertainty of mean ($s_{\bar{x}}$)	0.00026		g
Expanded uncertainty (U) ($k = 2$)	0.0005		g
Result	12.374 ± 0.001		g

$$\text{Mean } \bar{x} = \frac{\sum x_i}{N} = 12.3741 \text{ g}$$

$$U = k \cdot s_{\bar{x}}, \quad k = 2 \text{ (approx. 95\% confidence)}$$

4. Histogram / graph

Visualize the distribution (and compare if needed)



Histogram with bin width 0.0005 g
(density-normalized could be used if bin widths differ)

5. Interpretation

Draw conclusions based on the data and its uncertainty

- The mass is (12.374 ± 0.001) g ($k = 2$, ~95% confidence).
- The distribution is approximately normal.
- No outliers detected.
- The measurement is consistent and precise for the purpose of the experiment.



The quality of a scientific result depends on the entire chain:

MEASURE CAREFULLY → RECORD CORRECTLY → ORGANIZE CLEARLY → VISUALIZE APPROPRIATELY → INTERPRET WISELY

Good data presentation turns measurements into *meaningful scientific information*.

