

PHYSICS FOR ENGINEERING STUDENTS

Presentation of experimental data

Significant digits, tables, graphs and histograms

Lecture 3 — Book-style student notes

Learning focus: Significant digits · Rounding and uncertainty · Scientific graphs · Histogram normalization

University of Szeged · Institute of Physics

Presentation of Experimental Data

Measurement produces numerical values, but numbers become scientific information only after they are organized, processed, and presented in a form that preserves their meaning. Good presentation makes the precision, trend, comparison, and limitations visible. Poor presentation can hide uncertainty, exaggerate patterns, or imply an accuracy that the experiment does not support.

Learning objectives

- Distinguish exact and approximate numerical values.
- Apply significant-digit and rounding rules consistently.
- Report a measured value together with its uncertainty and unit.
- Construct clear tables and use linear interpolation appropriately.
- Choose linear, semi-logarithmic, or logarithmic graph scales.
- Display uncertainty with error bars.
- Construct and normalize histograms and interpret bin-width effects.

Central principle: A scientific figure or table is part of the analysis. Its labels, units, precision, scale, and uncertainty must allow the result to be interpreted without guessing.

Further interpretation

Scientific communication begins before the final report. Data structures, filenames, units, and metadata should be planned so that raw observations remain distinguishable from corrected and summarized values. A good workflow preserves the original records, performs calculations in reproducible steps, and creates tables and figures from the processed dataset rather than by manual copying. This reduces transcription errors and makes later corrections possible. Presentation choices should be linked to the question: a table supports exact lookup, a graph reveals relationships, and a histogram reveals distribution shape. Using all three without a purpose creates redundancy; choosing the appropriate form turns the result into evidence.

1. Exact and approximate numbers

Exact numbers are known without measurement uncertainty. They arise from counting, definitions, and exact mathematical relations. The number of specimens in a set is exact; the conversion $1\text{ m} = 100\text{ cm}$ is defined exactly. Approximate numbers come from measurement or estimation and have limited precision.

A measured length written as 12.40 cm communicates more than its numerical magnitude. The final zero indicates that the hundredth of a centimetre was resolved or estimated. Writing 12.4 cm would claim less precision, while 12.4000 cm would usually claim unjustified precision.

Number	Type	Reason
24 students	Exact	Result of counting
2π rad per revolution	Exact	Geometric definition
9.81 m s^{-2}	Approximate	Rounded local gravitational acceleration
20.0°C	Approximate	Measured temperature

Calculated results inherit limitations from measured inputs. A calculator may display many digits, but those digits are not automatically significant. The result should be rounded after the calculation, retaining guard digits in intermediate steps to avoid accumulated rounding error.

Interpretation: Precision is communicated by the reported digits only when the notation, unit, and uncertainty are stated consistently.

Further interpretation

Exactness belongs to the origin of a number, not merely to its appearance. A conversion factor defined by the SI remains exact even if written with few displayed digits, whereas a measured value remains approximate even when software prints many decimals. Constants taken from handbooks may also carry uncertainty unless defined exactly. When calculations mix exact and measured quantities, only the measured inputs limit experimental precision. Dimensional analysis provides a separate check: a numerically plausible answer with the wrong unit is still wrong. Recording units with raw values also prevents silent scale errors when data are combined, for example confusing millimetres with metres or milliseconds with seconds.

2. Significant digits

Significant digits are the digits that carry information about the precision of a number. All nonzero digits are significant. Zeros between nonzero digits are significant. Leading zeros only locate the decimal point and are not significant. Trailing zeros after a decimal point are significant because they express measured precision.

Value	Significant digits	Comment
0.00450	3	Leading zeros are placeholders; final zero is significant
100.2	4	Interior zeros are significant
1200	Ambiguous	Use scientific notation to show intended precision
1.200×10^3	4	Coefficient explicitly shows precision

For multiplication and division, the result is normally limited by the factor with the fewest significant digits. For addition and subtraction, the limiting factor is decimal place rather than digit count.

$$2.45 \times 3.1 = 7.595 \rightarrow 7.6$$

$$12.35 + 0.8 = 13.15 \rightarrow 13.2$$

These rules are convenient reporting conventions, not a replacement for uncertainty propagation. When uncertainties are known, the uncertainty determines the meaningful decimal place of the final result.

Good practice: Keep extra digits during calculation and round once, at the end. Repeated rounding can move the final result unnecessarily.

Further interpretation

Significant-figure rules are simplified versions of uncertainty reasoning. They work reasonably when input uncertainties are comparable to half a unit in their last reported digit, but can mislead when uncertainties differ strongly. For instance, two values with the same number of significant digits may have very different relative uncertainties. In formal analysis, propagate standard uncertainties first and then choose the final decimal place. Significant digits remain valuable as a communication safeguard: they prevent the report from implying detail unsupported by measurement. They should not be used to round raw data prematurely, because later averaging and model fitting benefit from retaining the original instrument indications.

3. Rounding and reporting

To round a value, first identify the last digit to be retained. If the first discarded digit is 0–4, the retained digit is unchanged. If it is 5–9, the retained digit is increased by one. All following digits are removed. This standard rule is adequate for ordinary experimental reporting.

The measured value and its uncertainty must be written to the same decimal place. The uncertainty is commonly reported with one or two significant digits; two are useful when the first digit is 1 or 2, or when loss of information would otherwise be excessive.

$$x = (12.347 \pm 0.126) \text{ mm} \rightarrow x = (12.35 \pm 0.13) \text{ mm}$$

Units should not be repeated inside and outside the parentheses. Scientific notation is especially useful for very large or small quantities and makes significant digits explicit.

$$(6.022 \pm 0.004) \times 10^{23} \text{ mol}^{-1}$$

Rounding negative numbers

Rounding concerns magnitude and place value. For example, -3.746 rounded to two decimal places is -3.75 . The minus sign is not a significant digit.

Avoid false precision: A long calculator output should never be copied directly into a report. The final digits must be supported by the experimental uncertainty.

Further interpretation

Rounding introduces a small additional error. If values are repeatedly rounded during a multi-step calculation, these effects can accumulate or create bias. Guard digits protect the calculation, while final rounding aligns the estimate and uncertainty. When thousands of values are rounded for storage, unbiased round-to-nearest rules are preferable to systematic truncation. Half-even rounding is sometimes used in computing to reduce aggregate bias, although ordinary half-up rounding is common in laboratory reports. Whichever convention is chosen, consistency matters. A final value should also be checked for unit prefixes: 0.00125 m , 1.25 mm , and $1250 \mu\text{m}$ represent the same magnitude but emphasize different practical scales.

4. Uncertainty in reported data

Uncertainty quantifies the dispersion of values that could reasonably be attributed to the measurand. It is not an admission of failure and it is not necessarily the same as error. Error is the difference between a measured value and a reference or true value; uncertainty describes limited knowledge of that value.

Absolute uncertainty has the same unit as the measured quantity. Relative uncertainty is dimensionless, and percentage uncertainty expresses it as a percentage.

$$u_r(x) = u(x)/|x|, \quad u_{\%} = 100u_r(x)\%$$

Suppose a diameter is reported as $d=(25.40 \pm 0.08) \text{ mm}$. The relative standard uncertainty is $0.08/25.40=0.00315$, or about 0.315% . A table should not mix values rounded to different positions when they arise from the same measurement process.

Symmetric and asymmetric intervals

The notation $x \pm u$ implies a symmetric interval. When the model or distribution is asymmetric, separate lower and upper limits are clearer, for example $4.2^{+0.8}_{-0.5} \mu\text{m}$. The reporting form must match the actual evaluation.

Honest communication: Uncertainty tells the reader how much confidence to place in distinctions between values. Differences smaller than their uncertainty may not be physically meaningful.

Further interpretation

Uncertainty notation should identify what the quoted number represents. A standard uncertainty corresponds to a one-standard-deviation scale, while an expanded uncertainty is $U = k u_c$ and is intended to define a wider coverage interval. A result written simply as $x \pm \text{number}$ is incomplete if the reader cannot tell which convention applies. Relative uncertainty is useful for comparison across magnitudes, but absolute uncertainty is necessary to identify the meaningful decimal place. When a transformed quantity is reported, such as logarithmic level or reciprocal time, symmetric uncertainty in the original variable may become asymmetric after transformation. The reported interval should follow the actual model rather than forcing symmetry for convenience.

5. Tables as scientific tools

A good table allows the reader to locate, compare, and reuse numerical values. It requires a descriptive title, clearly named quantities, units in the headings, consistent precision, and an ordering that supports interpretation. The table body should contain numbers rather than repeated units.

Temperature, $T / ^\circ\text{C}$	Resistance, R / Ω	Standard uncertainty, $u(R) / \Omega$
20.0	107.79	0.05
30.0	111.67	0.05
40.0	115.54	0.06
50.0	119.40	0.06

Column headings can use the form quantity / unit or quantity (unit), but the chosen convention should be consistent. Decimal points should align visually where possible. Empty cells must be explained: a dash, “not measured,” and zero do not mean the same thing.

The title should state what was measured and under what relevant condition, without repeating every heading. Notes below the table may define symbols, explain exclusions, or identify uncertainty coverage.

Table test: A reader should be able to understand each column, its unit, and its precision without searching the main text for missing definitions.

Further interpretation

Tables should be designed for the reader's task. Sorting by the independent variable supports interpolation and trend recognition, whereas sorting by category may support comparison. Excessive grid lines compete with the data; restrained rules and alignment usually improve readability. Uncertainty can be shown in a separate column when it varies, or stated in a note when it is common to all entries. Repeated text should be moved to headings, but explanatory notes should remain close enough to prevent ambiguity. Machine-readable supplementary data may contain more digits than the printed table, provided metadata clearly distinguish stored numerical precision from justified measurement precision.

6. Linear interpolation

Interpolation estimates a value between two known data points. Linear interpolation assumes that the relation is adequately represented by a straight line over the interval from (x_1, y_1) to (x_2, y_2) .

$$y(x) = y_1 + (x - x_1)(y_2 - y_1)/(x_2 - x_1)$$

Example: a calibration table gives $R=111.67 \, \Omega$ at $30.0 \, ^\circ\text{C}$ and $R=115.54 \, \Omega$ at $40.0 \, ^\circ\text{C}$. At $35.0 \, ^\circ\text{C}$, linear interpolation gives $R=111.67+0.5(115.54-111.67)=113.605 \, \Omega$, reported as $113.61 \, \Omega$ if the calibration supports that precision.

Interpolation is generally safer than extrapolation because it stays within a region constrained by data. Nevertheless, it can fail when the underlying curve is strongly nonlinear, the interval is too wide, a discontinuity occurs, or the selected points belong to different physical regimes.

Practical checks

- Plot the data and inspect curvature.
- Use nearby points that bracket the target.
- Compare the interpolation error with measurement uncertainty.
- Do not infer extra precision from the formula.

Model condition: Linear interpolation is a local model, not a universal property of the experiment. State or justify why a straight segment is adequate.

Further interpretation

Interpolation uncertainty has two components: uncertainty in the tabulated points and model error from assuming a straight segment. The first can be propagated through the interpolation weights. The second is small only when curvature across the interval is negligible relative to required accuracy. Using closer points reduces curvature error but may make noise more influential. Polynomial splines can produce smoother estimates, yet they may overshoot or create artificial structure if applied without physical judgement. Extrapolation is much more dangerous because no bracketing data constrain it. When extrapolation is unavoidable, the distance beyond the measured range and the physical basis of the model should be stated explicitly.

7. Anatomy of a scientific graph

A scientific graph should reveal a relationship while retaining enough information to evaluate it. The independent or controlled variable normally appears on the horizontal axis, and the dependent or measured response on the vertical axis. Each axis requires the quantity name or symbol and unit.

- Use a descriptive title or, in reports, an informative figure caption.
- Choose axis limits that include the data without excessive empty space.
- Use readable tick intervals and avoid unjustified digits.
- Plot measured points distinctly; do not hide them beneath a fitted curve.
- Identify multiple series with a legend or direct labels.
- Explain symbols, error bars, and fitted models in the caption.

Connecting points is appropriate when the line represents an ordered continuous path or guides the eye. A smooth curve should not imply a fitted physical model unless one was actually used. Likewise, a trend line must not be extrapolated beyond the measured range without justification.

Slope and intercept

For a linear model $y=a+bx$, the slope $b=\Delta y/\Delta x$ has units equal to the y-unit divided by the x-unit, while the intercept a has the unit of y .

$$b = \Delta y/\Delta x, \quad a = y(0)$$

Visual integrity: Graphical choices should clarify the data, not manufacture a stronger relationship than the measurements support.

Further interpretation

The graph's aspect ratio influences perceived slope, so it should be chosen to support reading rather than dramatize change. Starting a linear axis above zero can be legitimate when small variations are important, but the truncation must be unmistakable; bar charts generally require a zero baseline because length encodes magnitude. Data markers, fitted lines, and theoretical predictions should be visually distinct. Colour must not be the only identifier, since printing and colour-vision differences can remove that distinction. A strong caption states the system, principal conditions, meaning of symbols, and central observation. It complements the graph instead of merely repeating its title.

8. Error bars carry physical information

Error bars display uncertainty or variability associated with plotted values. Horizontal bars represent uncertainty in x and vertical bars represent uncertainty in y . Their meaning must be stated: standard uncertainty, standard deviation, standard error, or a confidence interval are not interchangeable.

A point may be written conceptually as $(x \pm u_x, y \pm u_y)$. The bars define plausible intervals along each coordinate, but they do not automatically represent a rectangular confidence region, especially when x and y are correlated.

$$(x, y) \rightarrow (x \pm u_x, y \pm u_y)$$

Overlapping error bars do not provide a universal significance test, and non-overlap is not always required for a meaningful difference. Formal inference depends on the definition of the bars, the statistical model, and the dependence between observations.

Fitting with uncertainties

When vertical uncertainties vary substantially, weighted least squares gives more influence to more precise points, commonly using weights $w_i = 1/u_i^2$. Uncertainty in x may require an errors-in-variables or orthogonal-distance approach.

$$\chi^2 = \sum w_i [y_i - f(x_i)]^2, \quad w_i = 1/u_i^2$$

Caption requirement: Always state exactly what an error bar represents and, for confidence intervals, the confidence level and method used.

Further interpretation

Error bars can represent several different statistical objects. Standard deviation describes variability among observations, whereas standard error describes uncertainty of an estimated mean and decreases with sample size. A confidence interval includes both an estimator and a confidence procedure; its width depends on the chosen level and sample assumptions. For a single observation, instrument

uncertainty may be more relevant than sample statistics. Correlated errors can move several points together, so independent-looking vertical bars may not capture uncertainty in slope or differences. When a calibration offset affects every point, the relative trend can be precise even though the absolute vertical location is uncertain.

9. Choosing linear and logarithmic scales

A linear–linear graph is appropriate when equal numerical differences should occupy equal distances and the relationship is approximately additive or linear. A semi-logarithmic graph uses a logarithmic scale on one axis; a log–log graph uses logarithmic scales on both axes.

Model	Linearized plot	Slope
$y = a + bx$	y versus x	b
$y = ae^{(bx)}$	$\ln y$ versus x	b
$y = ax^b$	$\ln y$ versus $\ln x$	b

Taking logarithms transforms multiplicative relations into additive ones.

$$y = ae^{bx} \Rightarrow \ln y = \ln a + bx$$

$$y = ax^b \Rightarrow \ln y = \ln a + b \ln x$$

Logarithmic axes are useful when values span several orders of magnitude or when exponential and power-law models are physically motivated. Zero and negative values cannot be plotted on an ordinary logarithmic scale. A straight line on transformed axes suggests a model but does not prove it; residuals and uncertainty must still be examined.

Axis choice: Select the scale from the physical question and model, then label it clearly. Never use a log scale merely to make inconvenient data appear linear.

Further interpretation

Axis transformation changes both visual spacing and statistical behaviour. On a log scale, equal distances represent equal ratios rather than equal differences. Error bars that are symmetric in linear space become asymmetric after taking logarithms. Fitting a straight line to logged data is not always equivalent to fitting the original nonlinear model, because the transformation changes residual weighting and assumed noise distribution. If measurement errors are additive and constant in y, direct nonlinear fitting may be more appropriate; if relative errors are roughly constant, logarithmic treatment may be natural. Residual plots on the relevant scale help decide whether the chosen representation is defensible.

10. Histograms and bin construction

A histogram displays the distribution of numerical observations by dividing the data range into intervals called bins and showing the number of observations in each bin. Unlike a bar chart, the horizontal axis is quantitative and adjacent bins represent contiguous intervals.

Construction procedure

- Find the minimum, maximum, and range.
- Choose bin boundaries and a bin width Δx .
- Count observations n_j in each bin.
- Plot adjacent rectangles over the intervals.
- State the sample size N and boundary convention.

The bin convention must avoid double counting. A common choice is left-closed and right-open intervals $[b_j, b_{j+1})$, except possibly the final bin. The displayed pattern depends on the bin origin as well as the width.

$$N = \sum_j n_j$$

Too few wide bins conceal structure; too many narrow bins emphasize sampling noise. Rules such as Sturges' rule or the Freedman–Diaconis rule can provide starting values, but scientific judgment and the purpose of the analysis remain important.

Interpretation: A histogram is an estimate of distribution shape, not the distribution itself. Features based on only one or two counts should be interpreted cautiously.

Further interpretation

Histogram bins define a measurement of the empirical distribution. Shifting the bin origin by half a width can change visible peaks, particularly for small samples. Therefore, important conclusions should not depend on one arbitrary bin choice. A dot plot, strip chart, or empirical cumulative distribution can supplement the histogram without binning. Outliers deserve investigation but should not be erased merely to produce a smoother shape. They may reflect transcription mistakes, transient instrument faults, or genuine rare states. The decision to exclude a value should follow a documented criterion and be reported with the result.

11. Histogram normalization

Counts are useful for describing one sample, but normalized forms are needed when sample sizes or bin widths differ. Relative frequency divides each count by the total number of observations.

$$p_j = n_j/N, \quad \sum_j p_j = 1$$

For unequal bin widths, height should represent density rather than count. The density height h_j is chosen so that rectangle area equals relative frequency.

$$h_j = n_j/(N\Delta x_j), \quad \sum_j h_j \Delta x_j = 1$$

This distinction is crucial: on a density histogram, probability is represented by area, not height alone. A narrow bin can be tall while containing fewer observations than a wider neighbouring bin.

Display	Bar height	Total
Count	n_j	N
Relative frequency	n_j/N	1
Density	$n_j/(N\Delta x_j)$	Area = 1

As the sample grows and bins become suitably narrow, the normalized histogram may approximate an underlying probability density. The limit is conceptual; choosing bins that shrink too quickly leaves noisy counts.

Comparison rule: Use relative frequency for equal-width bins and density when widths differ or when comparing with a probability-density function.

Further interpretation

Normalization enables comparison but changes the meaning of the vertical axis. Counts answer how many observations occurred; relative frequency answers what fraction occurred; density estimates probability per unit of x . For density, changing the unit rescales the vertical values while leaving bar areas unchanged. This is why the axis must include inverse units, such as mm^{-1} . When comparing groups, identical boundaries and widths make visual comparison easier, but a common binning scheme may still obscure features in a narrow distribution. Transparent overlays can become confusing; aligned panels or outline histograms often communicate differences more honestly.

12. From measurements to a scientific figure

A complete experiment does not end with collecting numbers. Raw observations must be checked, organized, rounded, displayed, and interpreted without losing traceability to the original data. The final presentation should make both the central result and its limitations visible.

Integrated workflow

- Preserve raw data and metadata before editing or rounding.
- Distinguish exact constants from measured inputs.
- Calculate with guard digits and round the final result consistently with uncertainty.
- Use a table for precise lookup and a graph for patterns or relationships.
- Choose axes and transformations from the physical model.
- Show uncertainty and define error bars.
- Use histograms to study repeated-measurement distributions.
- Write a caption that states the main message and essential conditions.

A clear result is reproducible: another reader can identify quantities, units, sample size, processing, and uncertainty. Cosmetic polish cannot compensate for missing definitions, misleading scales, or unsupported precision.

Key equations

$$y = y_1 + (x - x_1)(y_2 - y_1)/(x_2 - x_1)$$

$$h_j = n_j/(N\Delta x_j)$$

Take-home message: Present only as many digits as the measurement supports, and design every table and figure so that meaning, uncertainty, and physical context remain inseparable.

Further interpretation

The final figure should be checked as both a scientific object and a standalone communication object. Recalculate a few plotted points from the raw data, verify axis units and transformations, and confirm that uncertainty bars correspond to the intended statistic. Read the caption without the main text and ask whether symbols and conditions are identifiable. Then inspect the graph without the caption and ask whether scale choices could be misread. This two-way check catches many common errors. A reproducible script or spreadsheet should generate the final table and figure directly, with versioned data and explicit processing steps, so visual revisions never silently alter numerical content.