II

Treatment of Experimental Data

The usual objective of performing an experiment in physical chemistry is to obtain one or more numerical results. Between the recording of measured values and the reporting of results there are processes of numerical calculations, some of which may involve averaging or smoothing the measured values but most of which involve the application of formulas derived from theory. Part A of this chapter is devoted to a discussion of general techniques for carrying out such calculations. The use of computer spreadsheet programs such as Lotus 1-2-3, Quattro-Pro, Excel, and others is particularly advantageous for data analysis. A discussion of the use of spreadsheets is given in Chapter III.

Our concern with the treatment of experimental data does not end when we have obtained a numerical result for the quantity of interest. We must also answer the question: ''How good is the numerical result?'' Without an answer to this question, the numerical result may be next to useless. The expression of how ''good'' the result may be is usually couched in terms of its *accuracy,* i.e., a statement of the degree of the uncertainty of the result. A related question, often to be asked before the experiment is begun, is ''How good does the result need to be?'' The answer to this question may influence important decisions as to the experimental design, equipment, and degree of effort required to achieve the desired accuracy.

Indeed, it may be useful to discuss the matter also in economic terms. The ''economic value'' of the numerical result of an experiment often depends on its degree of accuracy. To claim too high an accuracy through ignorance, carelessness, or self-deception is to cheat the ''consumer'' who makes decisions on the basis of this result. To claim too low an accuracy through overconservatism or intellectual laziness lessens the value of the result to the ''consumer'' and wastes resources that have been employed to achieve the accuracy that could rightfully have been claimed. These issues of errors and the accuracy of a result are dealt with in Part B of this chapter.

A. Calculations and Presentation of Data

This part of Chapter II is concerned with the mathematical processing of data to obtain the desired quantitative results. At an initial stage, numerical methods are intrinsic, since data consist of a set of observed ''points''—usually pairs x_i, y_i of a dependent quantity y_i

measured for a specified value of the independent variable x_i. Of course, y could be a function of two variables x and z, in which case one often holds z constant and measures y while varying x, then changes z to a new constant value and repeats the process. An example would be the measurement of the pressure of a gas as a function of volume at a series of constant temperatures. One can use numerical techniques to fit the data with a functional form: best line $y(x)$ through the points in an x–y plot or best surface $y(x, z)$ in a three-dimensional computer-graphics display. Then further processing of the data can be done with analytic methods. Or, in other cases, the data processing can be carried out completely with numerical methods.

SIGNIFICANT FIGURES

Many numerical data in the literature and in everyday use carry no explicit statement of uncertainty, such as an estimated standard deviation S or 95 percent confidence limit Δ. The definition and technique for evaluation of these error quantities are given in Part B of this chapter. Here we are concerned with more qualitative principles. Some idea of the uncertainty in the numerical value of a measurement (''reading'') should be conveyed by the number of digits employed to express that numerical value. The digits that are essential to conveying the numerical value to its full accuracy or precision are called *significant figures.* Thus the numbers 4521 and 6.784 each have four significant figures, provided all of the digits are meaningful. So also do 0.006784 and 4,521,000, since the leading and trailing zeros generally determine only the magnitude of the number and have nothing to do with the relative precision. In the latter case, however, there is some ambiguity because one or more of the trailing zeros may be significant. The ambiguity may be removed by expressing the number in exponential notation: 4.521×10^6 or 4.5210×10^6, etc., as the case may be.

This discussion applies to numerical values of variables whose possible range constitutes a continuum; in computer terminology these are called ''real'' numbers, as opposed to integers (or, of course, complex numbers). Integers should be identified as such, and all of the digits necessary to express them should be given.

The use of too few digits in reporting numerical values robs the user of valuable information. The usual fault, however, is the reckless use of too many digits, which conveys a false sense of the accuracy or precision of the numerical value.

In reporting numerical values, certain rules regarding significant figures should be followed.

1. Numerical values are considered to be uncertain in the last digit by ± 3 or more, and perhaps slightly uncertain in the next-to-last digit. Ordinarily the next-to-last digit should not be uncertain by more than ± 2.
2. In rounding off numbers, (*a*) increase the last retained digit by 1 if the leftmost digit to be dropped is more than 5, or is 5 followed by any nonzero digits:

$$3.457 \rightarrow 3.46 \qquad 52.6502 \rightarrow 52.7$$

(*b*) leave the last retained digit unchanged if the leftmost digit to be dropped is less than 5:

$$0.34648 \rightarrow 0.346$$

(*c*) if the leftmost digit to be dropped is 5 followed by no digits except zero, then increase the last retained digit by 1 if it is odd, and leave it unchanged if it is even:

$$73.135 \rightarrow 73.14 \qquad 48.725 \rightarrow 48.72$$

For a discussion of the retention of digits in numbers during computation, see the next section.

PRECISION OF CALCULATIONS

The rules given above for determining the number of significant figures to be retained in a numerical result should be kept in mind. However, these rules do not apply strictly to the retention of digits in intermediate values obtained during a computation. It is necessary to retain *at least* one additional digit and recommended to retain two or more additional digits during computations to avoid unnecessary round-off errors in the final results. There are generally no problems with precision when the calculation is carried out on a computer; for example, spreadsheet programs typically have 15 to 17 digits and display power limits of 10^{-99} to 10^{+99}, although a power range of ± 308 is used for internal steps in a calculation. One must, of course, remember to adjust properly the number of significant figures when reporting the final results.

Computations, whether by hand calculator or by computer, should be done with awareness of certain principles concerning the effects of arithmetic operations. These follow directly from the principles underlying the propagation of errors, discussed in Part B of this chapter.

When a number of numerical values are *added,* the precision of the result can be no greater than that of the least precise numerical value involved. Thus, generally, the number of decimal places in the result should be the same as the number of decimal places in the component with the fewest:

$$\begin{aligned} &32.7 \\ &+\ 3.62 \\ &+10.008 \\ &=46.328 \rightarrow 46.3 \end{aligned}$$

When the *difference* between two numbers yields a result that is relatively small, that difference not only has a precision limited to that of the less precise number but has also a relative precision much less than either number:

$$\begin{aligned} &673.425 \\ &-672.91 \\ &\quad 0.515 \rightarrow 0.52 \end{aligned}$$

The less precise of the two numbers has a relative precision of about 1 in 67,000 (0.0015 percent), while the difference has a relative precision of about 1 in 50 (2 percent). The effect can be particularly devastating in an intermediate calculation on a computer, where the user of the computer may be blissfully unaware of the loss of relative precision.

In *multiplication* and *division,* the relative precision of the result can be no greater than that of the least (relatively) precise number involved. Thus the number of significant figures in the result should be approximately that in the component having the fewest of them. In judging the appropriate number of significant figures in the result, bear in mind that the relative precision does not depend on the number of significant figures alone; 999, with three significant figures, is about as precise as 1001, with four. When in doubt, employ the larger number of significant figures. For example:

$$346 \times 121 \times 900.0 = 37{,}679{,}400 \rightarrow 3.77 \times 10^7$$

The uncertainty in 121, the least relatively precise figure, is about 0.8 percent. One in 377 is about 0.27 percent, which implies a greater relative precision, but if the result is rounded

to 3.8×10^7, with a relative precision of 2.6 percent, too much significance will be lost. The principles are no different when division is involved:

$$5.44 \times \frac{0.132}{8.092} = 0.0887395 \rightarrow 0.0887 \quad \text{or} \quad 0.089$$

The uncertainty in 0.132 is about 0.7 percent, that in 0.089 is about 1.1 percent, implying a small but tolerable loss of precision. If this is an intermediate calculation, keep the extra digit.

It should be kept in mind that in the case of a calculation with many numbers, round-off errors may accumulate to the extent that they reduce the precision significantly.

The above-mentioned considerations are not a substitute for a propagation-of-errors treatment to determine the uncertainty in the final result, especially when the calculations have been carried out on a computer and the result is given with an artificially large number of digits.

No matter what computational procedure has been used, there must be reliable assurance that mistakes in arithmetic and in the transcription of numerical data have not been made. Normally, an electronic calculator or computer can be trusted not to make mistakes in arithmetic, but mistakes in the manual transcription of input numbers and output numbers are all too easy to make. A printed record from a computer, which can be checked and double-checked, can provide satisfactory assurance against mistakes of this kind.

The best way to check a calculation is to repeat it, preferably in a different way, to the same precision. When an experiment has been performed by two persons as partners, the calculations should generally be done by each person independently, with crosschecking of intermediate and final numerical results.

ANALYTICAL METHODS

Much of the mathematical analysis required in physical chemistry can be handled by analytical methods. Throughout this book and in all physical chemistry textbooks, a variety of calculus techniques are used freely: differentiation and integration of functions of several variables; solution of ordinary and partial differential equations, including eigenvalue problems; some integral equations, mostly linear. There is occasional use of other tools such as vectors and vector analysis, coordinate transformations, matrices, determinants, and Fourier methods. Discussion of all these topics will be found in calculus textbooks and in other standard mathematical texts.

Accounts of the sort of applied mathematics useful in physics and chemistry can also be found in specialized monographs such as those listed below.

H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry,* 2d ed., Van Nostrand, New York (1965) or Krieger Publ. Co., Huntington, N.Y. (1976).

G. A. Korn and T. M. Korn, *Mathematical Handbook for Science and Engineering,* 2d ed., McGraw-Hill, New York (1968).

M. E. Starzak, *Mathematical Methods in Chemistry and Physics,* Plenum, New York (1989).

G. Strang, *Introduction to Applied Mathematics,* Wellesley-Cambridge Press, Wellesley, Mass. (1986).

F. W. Byron, Jr., and R. W. Fuller, *Mathematics of Classical and Quantum Physics,* vols. 1 and 2, Addison-Wesley, Reading, Mass. (1969).

These are generally more advanced presentations with a strongly applied flavor. Margenau and Murphy is an old but classic book that is still of value, as is Korn and Korn. The

next two books are more modern in their selection of techniques and topics. Byron and Fuller is on quite an advanced level.

It must be kept in mind that experimental situations do not yield directly an analytical function for the measured observable. Thus numerical methods are needed at least to establish the parameters of some theoretical form or to find some empirical approximation function. After a functional form $y = f(x)$ is found, one can often use analytical methods to evaluate quantities such as dy/dx, $\int y\,dx$ or to find the roots of $f(x) = 0$. However, some functions $f(x)$ are difficult to integrate analytically, and the roots of some equations are difficult to find. In such cases, numerical methods are very useful, especially since they can be implemented on computers with the use of very powerful algorithms.

NUMERICAL METHODS

Numerical calculations may be performed by any method that does not introduce round-off errors that are significant in comparison with the experimental errors.[1–5] A hand calculator will suffice for most of the experiments in this book. For lengthy or repetitive calculations, the use of a spreadsheet program is strongly recommended (see Chapter III). In some cases it may be advantageous to write special programs based on the mathematical procedures described in each experiment. As an aid to this, many useful algorithmic routines are available in Refs. 2–4. Reference 4 is particularly good because it contains 200 algorithms with a numerical-analysis text explaining the basic methods used. It is wise to carry out one typical calculation with a hand calculator in order to test any new program, and then if trouble is observed you can debug the program before using it.

Numerical methods can be applied to discrete (finite) data sets in order to carry out such procedures as differentiation, integration, solution of algebraic and differential equations, and data smoothing. Analytical methods, which deal with continuous functions, are exact or at least capable of being carried out to any arbitrary precision. Numerical methods applied to experimental data are necessarily approximate, being limited by the finite number of data points employed and their precision.

LEAST SQUARES

The *method of least squares* is a very powerful and widely used technique, and a detailed discussion of how to use it properly is given in Chapter XXII. Presented here is a brief commentary on this method.

Given a set of experimental data y_i measured at a series of values x_i for the independent variable x, it is often desired to construct a function $y(x)$ to represent them. If there is reason to believe that this function should conform to some analytic form depending on a few parameters, such as $y = (a/x) + b$, these parameters (a and b in the given case) can be adjusted to give the "best fit" to the data by using the method of least squares. The analytic form may be an equation arising from a theoretical model, which in principle may be exact. Alternatively, one may use some general "curve-fitting" function such as a polynomial that has no theoretical relationship to the experiment but will provide a useful empirical representation.

In general, the number of experimental points m is larger (often much larger) than the number of adjustable parameters. The resulting function is not an exact fit to each point but represents a best overall fit in the sense that the sum of the squares of the deviations $[y_i(\text{obs}) - y_{\text{calc}}(x_i)]$ is a minimum, hence the name. The fitting function $y(x)$ obtained by least squares can be differentiated, integrated, set equal to zero, and solved to find roots, etc., by analytical methods. Thus there is no need for the application of other numerical methods unless the function is very awkward to handle analytically in closed form.

A particularly easy type of least-squares analysis called *multiple linear regression* is possible for fitting data with a low-order polynomial, and this technique can be used for many of the experiments in this book. The use of spreadsheet programs, as discussed in Chapter III, is strongly recommended in such cases. In the case of more complicated nonlinear fitting procedures, other techniques are described in Chapter XXII.

In some situations, no theoretically motivated analytic form $y = f(x)$ is apparent, and low-order series approximations do not provide a good fit. You should then be cautious about using high-order empirical series fitting forms, since they may yield very poor extrapolations, interpolations, and derivatives. In these cases, direct numerical (or graphical) methods are available.

OTHER METHODS

Many other numerical methods are available, and some of these are described in Appendix F: smoothing, series approximations, differentiation, integration, and root finding. It should be noted that spreadsheet programs can be utilized easily for all of these numerical operations. For example, Quattro Pro contains standard operations for linear regression, matrix multiplication and inversion, and also has a ''Solve for'' procedure which will solve expressions such as $y = 10x^3 \sin x$ for the value of x corresponding to a given input value of y. These and other applications of spreadsheets are discussed in Chapter III.

Numerical methods beyond the scope of this book, such as the solution of linear equations, minimum/maximum problems, and Fourier transform methods, are also of value in some physical chemistry experiments. Algorithms for handling them with a computer or in some cases complete computer programs for the given task are described elsewhere.[1–4]

GRAPHS AND GRAPHICAL METHODS

By a *graph* we mean a representation of numerical values or functions by the positions of points and lines on a two-dimensional surface. A graph is inherently more limited in precision than a table of numerical values or an analytic equation, but it can contribute a ''feel'' for the behavior of data and functions that numerical tables and equations cannot. A graph reveals much more clearly such features as linearity or nonlinearity, maxima and minima, points of inflection, etc. Graphical methods of smoothing data and of differentiation and integration are sometimes easier than numerical methods. Graphs and graphical methods suffer, of course, from the limitations of a two-dimensional surface. Thus, normally, a plotted point has only two degrees of freedom, which we assume here to be represented by one independent variable x (increasing from left to right) and one dependent variable y (increasing from bottom to top). The word ''normally'' is required in the above sentence because three-dimensional plots, representing for example points with coordinates x, y, and z, can be represented in oblique projection on a two-dimensional surface by computer techniques.

Wherever possible, the data from an experiment should be plotted at an early stage, even if numerical methods will be used subsequently for greater accuracy. This is particularly true when functions are to be fitted by least-squares or other methods, since a graph may make evident special problems or requirements that might otherwise be missed.

MANUAL PREPARATION OF GRAPHS

For constructing a graph by hand, the following procedure is recommended.

1. *Design Your Graph First.* Plan it so as to make best use of the area available and present the data in a manner that will be clearest to the viewer. In most cases the entire

useful area of a normal-sized sheet of graph paper should be dedicated to the graph. In some cases, the paper should be turned sideways (i.e., with the binding margin on top) so that the independent variable, x, runs longways on the page, to provide an aspect ratio more favorable for displaying the desired features. Determine the maximum and minimum values of x and y (remember, it is *not* necessary that they both be zero at the lower left corner!) so as to fit the points, lines, and curves into the available area. Be sure to make adequate allowance of space for borders all around, scale numbers and scale labels, at least at the bottom and at the left, and a legend telling what the graph is about.

2. *Choose Your Graph Paper.* The graph paper should be of good quality, accurately ruled with thin, lightweight lines. In most cases these will constitute a grid of lines equally spaced in two perpendicular directions; special papers are available for plotting logarithmic functions in one direction ("semilog paper") or in both directions ("log–log paper"), or for plotting in polar coordinates, etc. Most good graph papers are ruled in colors that are soft enough that they do not distract the viewer from the plotted lines and points, which are usually in black. The most commonly used graph papers are ruled with a 1-mm spacing, the lines at 5 and 10 mm being heavier. The smallest divisions on the paper should preferably represent multiples of 1 or 2 or perhaps 5 in the plotted variable.
3. *Define the Plotting Area.* Draw, with heavy pencil lines, a rectangle within which all points and lines will be plotted, leaving sufficient margins all around, extra width for the binding margin, and also adequate allowance for the scale numbers and legends mentioned above. Indicate the major scale divisions with short, heavy pencil marks and label them with appropriate numerical values. Under the scale numbers for the independent variable enter the scale label, stating the quantity that is varying, and its units, e.g.,

$$T\ (\mathrm{K})$$

for absolute temperature in kelvins. To avoid scale numbers that are too large or too small for convenient use, multiply the quantity by a power of 10, e.g.,

$$\rho \times 10^4\ (\mathrm{g\ cm^{-3}})$$

for the density of a gas. This is equivalent to multiplying all the scale numbers by 10^{-4} but is much more convenient. Similarly, enter the scale label for the dependent variable along the side scale on the left; if necessary, turn the paper sideways to write the label so that it will run vertically.
4. *Plot the Points; Draw the Lines and Curves.* Using a sharp, hard pencil, plot the experimental points as small dots, as accurately as possible. Try to plot with an accuracy of one-fifth of the smallest grid division. Draw small circles (or squares, triangles, etc.) of uniform size (2 or 3 mm) around the points in ink to give them greater visibility.

   If it is desired to draw a straight line to provide a linear fit to the plotted points, draw a faint pencil line with a good straightedge such as a transparent ruler or draftsman's triangle, and do not hesitate to erase the line and try again. When you are satisfied that you have made the best possible visual fit, go over the line again with a softer pencil or a pen to make it sufficiently heavy; this time, do not allow the line to cross the circles or other symbols drawn around the experimental points. Smooth nonlinear curves may be drawn either freehand or with ships curves, splines, or other devices, with as much trial and error and erasing as required.

   If a "theoretical curve" is to be drawn based on an analytic function representing predicted behavior, plot the function at as many points, closely spaced, as reasonably possible; do not draw symbols around them. Carefully draw a smooth curve through the points, lightly at first, and then "heavy it up" as desired.

5. *Add a Legend.* Somewhere on the graph (if possible, at the bottom) enter a legend with an identifying figure number. The legend should state what the graph is about, identify symbols and line types used, and provide all needed information that will not be provided in the document to which the graph will be attached.
6. *Check the Overall Appearance.* After all points, lines, curves, scales, and legends are in place, check the overall clarity and legibility of the graph. If points or lines are too faint, make them heavier; if the graph is messy or confusing, attempt to improve it by erasing and redrawing the features concerned. If the end result falls seriously short in the expected qualities, *start over.* Your second attempt will probably take much less time to prepare, and it will look much better and will better suit its intended purpose. For examples of well-prepared graphs representing experimental data, consult current scientific journals such as those listed in Appendix E.

COMPUTER-GENERATED PLOTS

In most cases, you should take advantage of the ability of computer spreadsheet programs to generate graphs. These may be generated on the computer's monitor screen for inspection and editing and then produced on external graphic devices such as X–Y plotters, or they may be printed out on ink-jet or laser printers. Much of the above discussion of manual preparation of graphs applies also to computer graphs. Since computer graphs can be produced quickly, it may be worth while to make several tries to obtain the best results. The first plot may be a "default plot" to see the overall appearance of the plotted data; then user-chosen scales, symbols, aspect ratio, and labels may be introduced as needed. If the computer output uses generic labels such as x and y, be sure to relabel with appropriate variables.

GRAPHICAL DIFFERENTIATION AND INTEGRATION

The first derivative of a plotted function $y(x)$ with respect to x is given by the slope of a line drawn tangent to the curve at the point x, y concerned. Draw the tangent line long enough to maximize the accuracy of the slope determination. If two points x_1, y_1 and x_2, y_2 are accurately established at the two ends of the line, the slope and thus the derivative is given by

$$y'(x) = \frac{y_2 - y_1}{x_2 - x_1} \tag{1}$$

A good way to draw the tangent line is to hold a small rectangular mirror with its reflecting surface perpendicular to the paper, adjust it so that the reflection of the curve is tangent to and symmetrical with the curve itself, and then use the mirror as a straightedge to draw the tangent line. Another method is to use a compass to strike off arcs intersecting the curve on both sides and then draw a chord through the intersection points. The tangent can be drawn parallel to the chord if the curvature is uniform. If it is not, construct a second chord at a different distance in the same way and extend both lines to an intersection. Draw the (approximate) tangent line through this intersection point (see Fig. 1).

A definite integral, such as $\int_a^b y\,dx$, can be determined by measuring the area under the curve between the desired limits (see Fig. 1). The area can be measured by use of an instrument known as a *planimeter,* by cutting out the area concerned with a pair of scissors and weighing it (having also weighed a rectangle of known area), by approximating it as well as possible by a bar graph and adding the areas of the bars, or by "counting squares." In the last method a count is made of the smallest squares of the graph-paper grid that lie wholly or *more than half* inside the area concerned. Whatever the method used, the area to be measured should be minimized as much as possible; there is no need to run the planimeter around the rectangular area below $a'b'$ in Fig. 1, for instance.

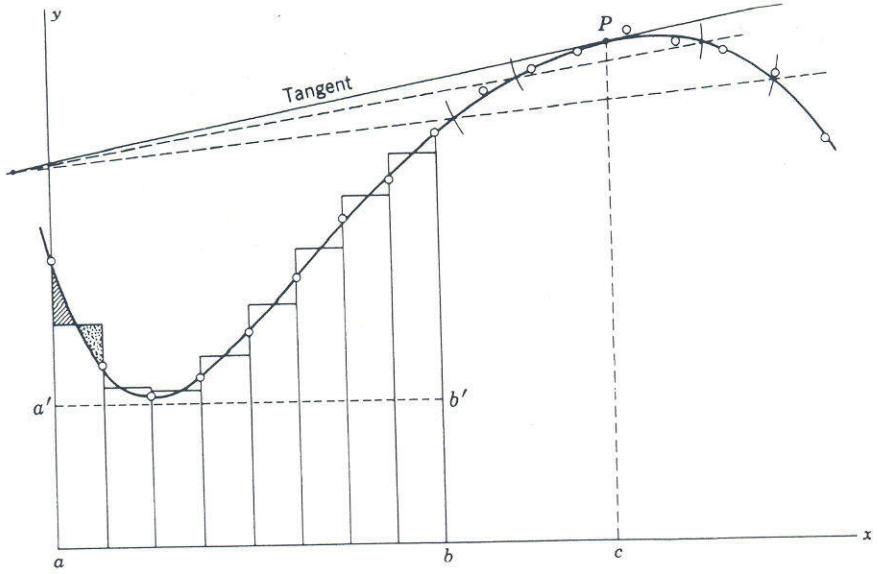


FIGURE 1
An illustration of two graphical procedures: (*a*) the construction of a tangent by the method of chords; (*b*) the evaluation of a definite integral by approximating the curve by a bar graph. The top of each bar is drawn so that the two small areas thereby defined (e.g., the shaded and stippled areas on the first bar) appear equal.

GRAPHICAL UNCERTAINTIES

It frequently happens that an intermediate or final result of the calculations in a given experiment is obtained from the slope or intercept of a straight line on a graph—say, a plot of y against x. In such a case it is desirable to evaluate the uncertainty in the slope or in the position of the intercept. A rough procedure for doing this is based on drawing a rectangle with width $2\Delta(x_i)$ and height $2\Delta(y_i)$ around each experimental point (x_i, y_i), with the point at its center. The assignments of the limits of error $\Delta(x_i)$ and $\Delta(y_i)$ are made as described in Part B. The significance of the rectangle is that any point contained in it represents a possible position of the "true" point (x_i, y_i) and all points outside are ruled out as possible positions. Having already drawn the best straight line through the experimental points, and having derived from this line the slope or intercept, draw now two other (dashed) lines representing maximum and minimum values of the slope or intercept, consistent with the requirement that both lines pass through every rectangle, as shown in Fig. 2. Where there are a dozen or more experimental points, it may be justifiable to neglect partially or completely one or more obviously "bad" points in drawing the original straight line and the limiting lines, provided that good judgment is exercised. The

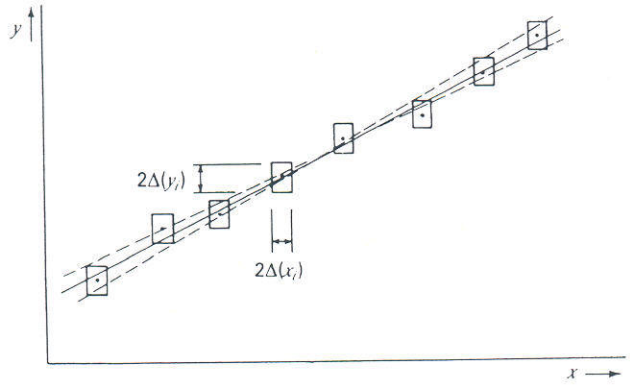


FIGURE 2
A graphical method of determining the limit of error in a slope.

difference between the two slopes or intercepts of the limiting lines can be taken as an estimate of twice the limit of error in the slope or intercept of the best straight line.

Where the number of points is sufficiently large, the limits of error of the position of plotted points can be inferred from their scatter. Thus, an upper bound and a lower bound can be drawn, and the lines of limiting slope drawn so as to lie within these bounds. Since the theory of least squares can be applied not only to yield the equation for the best straight line but also to estimate the uncertainties in the parameters entering into the equation (see Chapter XXII), such graphical methods are justifiable only for rough estimates. In either case, the possibility of systematic error should be kept in mind.

EXERCISES

1. Each of the following quantities is uncertain (95 percent confidence) by 0.20 percent. Express the quantities with the proper number of significant figures and a specification of the (absolute) uncertainty.
   (*a*) 2.05675 cm
   (*b*) 388.982 K
   (*c*) 1.21335×10^3 Torr

2. The viscosity η of pure ethanol in centipoise was determined at eight different temperatures at 1 atm pressure with the following results:

t, °C	0	10	20	30	40	50	60	70
η, cP	1.78	1.45	1.17	0.98	0.83	0.71	0.60	0.49

   Prepare a figure to display these data with a smooth curve drawn through the points, and then determine the temperature coefficient of viscosity $(\partial\eta/\partial t)_p$ at 25.0°C. It is recommended that this be done in two ways to allow a comparison: First, draw the smooth curve by eye and determine the temperature coefficient graphically; second, represent the data with an appropriate least-squares fitting function and differentiate analytically.

3. Bridgman obtained the following volume ratios V/V_0 for methanol, where $V_0 = 39.6\ \text{cm}^3$ is the molar volume at 0°C and a pressure of 1 kg/cm^2.

p (kg cm^{-2})	1	500	1000	2000	3000	4000	5000
$V(20°C)/V_0$	1.0238	0.9823	0.9530	0.9087	0.8792	0.8551	0.8354
$V(50°C)/V_0$	1.0610	1.0096	0.9763	0.9271	0.8947	0.8687	0.8476

   Use these data to evaluate the entropy change ΔS (in units of J K^{-1} mol^{-1}) when 1 mol of methanol at 35°C and 1 kg cm^{-2} is compressed isothermally to 5000 kg cm^{-2}. Recall that $\Delta S = -\int(\partial V/\partial T)_p\, dp \approx -\int(\Delta V/\Delta T)_p\, dp$. Thus you should determine values of $\Delta V/\Delta T$ and carry out the appropriate numerical integration.

B. Uncertainties in Data and Results

This part of the chapter is concerned with the evaluation of uncertainties in data and in calculated results. The concepts of random errors/precision and systematic errors/accuracy are discussed. Statistical theory for assessing random errors in finite data sets is summarized. Perhaps the most important topics are significance testing and propagation of errors. The first concerns how you can test the consistency of a given data set with either

an accepted value or a second set of observations. The latter shows how the error in an overall calculated result can be obtained from known or estimated errors in the input data. Examples are given throughout the text, and a convenient summary is given at the end.

ERRORS

RANDOM ERRORS AND PRECISION

We will assume that the numerical result to which our discussion applies is obtained with an instrument that measures a physical quantity for which the a priori range of possible values constitutes a continuum. This may be a continuous-reading instrument with a scale divided with numbered marks; customarily the scale reading is estimated to some fraction of the smallest interval ("scale division") between marks, typically one-fifth or one-tenth. The reading of the scale by a person with normal visual acuity is subject to a possible error (typically one- or two-tenths of a scale division). Even if the observer is capable of reading the scale with great precision, the measurement at any particular moment is subject to unpredictable, random fluctuations of the environmental conditions (temperature variations, voltage fluctuations in the electrical supply, electrical noise in the equipment, mechanical vibrations, etc.) which affect the instrumental response. Thus, repetitions of the measurement generally yield different readings, distributed at random over a small range. The same may apply to a digital instrument when the smallest increment in the numerical output, corresponding to ± 1 in the final digit, is smaller than the differences expected between different readings of the same quantity. If the smallest increment of numerical output is *not* smaller than expected random fluctuations, the instrument itself is making a contribution to measurement error because of round-off, and is perhaps not suitable for its intended purpose. Whatever the measuring instrument, we may say that the measurements are subject to *random error*. A statement of the *precision* of a measurement concerns its reproducibility when the measurement is made repeatedly with the same instrumentation and is therefore an expression of the uncertainty due to random error.

Faced with the inevitability of random error, the observer customarily makes several measurements of a given physical quantity and then averages them to obtain a value that is expected to be more reliable than any single measurement. The average or mean of N measurements of an experimental variable x is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2}$$

where x_i is the result of the ith measurement.

The observer also customarily reports in some manner the scatter of the measurements that go into this expression in order to give an indication of the precision. A crude way is to state the *range* of measured values:

$$R = x_{\text{largest}} - x_{\text{smallest}} \tag{3}$$

For several reasons this does not give a very clear measure of the precision of the individual measurements but can be used as the basis of quick estimates of more sophisticated measures. Very commonly the average deviation is cited:

$$\text{ave. dev.} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}| \tag{4}$$

Traditionally this has been cherished for its arithmetic simplicity, a virtue of declining value in this age of computers. Its usefulness for quantitative comparisons is limited because under given experimental conditions its value depends sensitively on sample size (the value for a pair of measurements being, on the average, about 30 percent smaller than the value calculated with a large group of measurements).

A measure of precision that is unbiased by sample size is the *variance,* denoted by S^2 and defined by

$$S^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{5a}$$

This quantity has the property of additivity so that estimates from several sources of random error may be combined, and it is useful in probability calculations. An alternative form of S^2, algebraically equivalent and often more convenient for use with a calculator or computer, is

$$S^2 = \frac{1}{N-1} \left(\Sigma x_i^2 - N\bar{x}^2 \right) = \frac{N}{N-1} (\overline{x^2} - \bar{x}^2) \tag{5b}$$

Particular attention should be paid to the divisor $(N - 1)$, which is known as the number of "degrees of freedom" and is equal to the number of *independent* data on which the calculation of S^2 is based. The N values of $x_i - \bar{x}$ are at this stage of the game not all independent; one degree of freedom has been used up in calculating the mean $\bar{x}$. That is, one of the N deviations, say the Nth one $(x_N - \bar{x})$, is not an independent variable since it can be calculated from the other deviations:

$$(x_N - \bar{x}) = -\sum_{i=1}^{N-1} (x_i - \bar{x}) \tag{6}$$

The difference between a factor of $N - 1$ or N in the denominators of Eqs. (5*a*) and (5*b*) is trivial if one has 100 measurements but becomes very important when N is small.

The square root of the variance,

$$S = \frac{1}{\sqrt{N-1}} \left[\sum_{i=1}^{N} (x_i - \bar{x})^2 \right]^{1/2} \tag{7}$$

is often called the *estimated standard deviation.* The significance of this name will become clear later. This parameter is widely used to indicate the precision of individual measurements. More important is the precision of the mean of the measurements. It will be shown later that the *estimated standard deviation of the mean* of N values is

$$S_m = \frac{S}{\sqrt{N}} = \frac{1}{\sqrt{N(N-1)}} \left[\sum_{i=1}^{N} (x_i - \bar{x})^2 \right]^{1/2} \tag{8}$$

It is apparent that the precision of the mean can be increased by increasing the number of individual measurements. The gain in going from $N = 2$ to $N = 10$ is quite dramatic; that in going from $N = 10$ to $N = 20$ is less so; and going from $N = 90$ to $N = 100$ results in only a small increase in precision. Thus, in principle the random error can be reduced to as small an amount as desired by repeated measurements. However, in practice there comes rapidly a "point of diminishing returns" in continuing to repeat measurements. It is rare to make more than about 10 measurements of the same physical quantity, and often the number of measurements is considerably smaller.

SYSTEMATIC ERRORS AND ACCURACY

There is an important reason why excessive efforts should not be made to reduce the magnitude of random errors by making a very large number of measurements of the same quantity. This reason is that measurements are also subject to other kinds of error, chiefly *systematic errors,* the magnitude of which can all too easily exceed that of the random errors. A systematic error is one that cannot be reduced or eliminated by any number of repetitions of a measurement because it is inherent in the method, the instrumentation, or occasionally the interpretation of data. While the precision of a result, as already stated, is an expression of uncertainty due to random error, the *accuracy* of a result is an expression of *overall uncertainty including that due to systematic error.* Accuracy is very much more difficult to assess than precision.

Examples of systematic error are a calibration error in the instrument, failure to establish properly a ''zero'' reading of the instrument scale, improper graduation or alignment of an instrumental scale, uncompensated instrumental drift, leakage of material (e.g., of gas in a pressure or vacuum system) or of electricity (in an electrical circuit), and incomplete fulfillment of necessary experimental conditions (e.g., incomplete reaction in a calorimeter, incomplete dehydration of a sample prior to weighing). An example of another kind is faulty theoretical treatment of the results of the measurements to obtain the desired result, perhaps through a faulty approximation in the phenomenological theory involved.

A particular kind of systematic error might be called ''personal error'' because it results from subjective judgments or personal idiosyncrasies on the part of the observer. Examples are consistent parallax errors in reading scales and setting cross-hairs, excessive reaction times in actuating a stopwatch or electronic timer, strong number prejudices that influence the reading of scales (tendency to favor 0s and 5s, or superstition in favor of 7 and against 13), and arbitrariness in rejecting or not rejecting readings that deviate somewhat from the mean of the others. Such errors can be particularly insidious if the observer has a personal stake in the outcome of the experiment (whether to confirm a theory and win a Nobel Prize, or to get the result thought to be expected by the instructor and win an ''A'' on the experiment). Persons of the highest character may not be altogether immune from this tendency, since biases are often subconscious.

Systematic errors are difficult to eliminate because they do not make their presence known as random errors do. They must be painstakingly sought out and eliminated through careful analysis of the entire experiment and of all assumptions on which the experimental method is founded. Some will be relatively easy to find, such as calibration and alignment errors; others, such as errors in the theory or the intervention of totally unexpected phenomena, can defy all attempts by ordinary means to identify them. All too often, important systematic errors are never found at all, and without doubt many are imbedded in the permanent scientific literature. Fortunately, some become apparent when new results, obtained by other methods, depart from the old ones by much more than the stated limits of error. As an example, R. A. Millikan's final value of Avogadro's number N_0, which he calculated as the ratio of the accurately known Faraday constant to his value of the electronic charge from the oil-drop experiment, was $(6.064 \pm 0.006) \times 10^{23}$ mol^{-1}. The currently recognized value is $(6.022137 \pm 0.000007) \times 10^{23}$, based on X-ray diffraction measurements of unit cell volume in a crystal together with the crystal density and formula weight. Millikan was a very careful experimenter; the systematic error in his value resided mainly in data reported by others that he used in his calculations.

As a rule, the best assurance of low systematic error is agreement among two or more entirely different methods of measurement. Further discussion of systematic errors can be found elsewhere.[6,7]

If all significant sources of systematic error cannot be eliminated, the experimentalist should at least seek to identify them and place limits on their possible magnitudes. However, to the maximum extent possible, you should aim to reduce the limits of systematic error to magnitudes that are small in comparison with the random errors of the experiment, so that statistical estimates made with equations in this chapter will be meaningful.

OTHER ERRORS

Another kind of error is important enough to mention. It is called by some "erratic error," which is a fancy name for "mistakes." Mistakes in arithmetic computation are an example, but these are inexcusable since all computations should be recorded in the laboratory notebook and carefully checked. One-time reading errors of instrument scales in which digits are read improperly (the numeral 5 being mistaken for a 6, for example), or in which pairs of digits are transposed or a decimal point is misplaced in recording, are examples of mistakes that can be very damaging because they usually cannot be traced without repeating the experiment and because they may be very much larger than any conceivable random error. One of the most frequent examples is weighing errors, since a sample of material being prepared for an experiment is weighed only once. Other mistakes are failure to execute some essential step in the procedure (such as turning on a heater or a circulating pump) or faulty execution of the procedure (such as use of a wrong optical filter or even a wrong chemical substance for investigation). These mistakes may be untraceable, since clues may not exist in the laboratory notebook if the experimenter has been less than meticulous. Serious mistakes may be revealed only when the consequent numerical result appears to be impossible or ridiculous or is in wide disagreement with a theoretical value or prior experimental result.

Any measurement that will of its nature be done only once should provoke in you a strong sense of insecurity that will motivate you to be very deliberate in recording the data in the notebook and checking them thoroughly before going on to something else. Wherever possible, measurements should be done at least in duplicate. Duplicate measurements are NOT, as many suppose, for the purpose of enabling statistical treatment of random error! For that purpose, at least four measurements should be made, preferably more. Mistakes in procedure can only be avoided by exercise of due care in planning and executing the experiment. If the procedure is excessively complicated, prepare a checklist ahead of time. Make sure that illumination is adequate, labels are clear, scales are clean. Record in the notebook the identification numbers of optical filters, tared vessels, and all relevant data from labels of sample containers. Keep the experimental setup orderly and as uncomplicated as possible.

REJECTION OF DISCORDANT DATA

It occasionally happens in making multiple measurements that one value differs from the rest considerably more than they differ from one another. Should the discordant value be rejected before an average is taken? This question provides one of the most severe tests of your scientific objectivity, for you may (indeed, you should) reject any measurement, whether concordant or discordant, if there is a valid reason to believe that it is defective.

The first step should always be to check for evidence of a determinate error (used the wrong pipette, added the wrong reagent, inverted the digits in writing down the number, etc.) or poorly controlled conditions (temperature control erratic during that measurement, forgot to check instrument zero setting, etc.). If no objective mistake or operating difficulty can be identified, the suspect value must be included unless valid statistical arguments can be presented to show that such a large deviation on the part of a member of the

TABLE 1 Critical Q values for rejection of a discordant value at 90 percent confidence level[8]

N	3	4	5	6	7	8	9	10
Q_c	0.94	0.76	0.64	0.56	0.51	0.47	0.44	0.41

hypothetical ideal population of similar measurements is highly improbable. There are a number of ways of going about this. Most of them—such as the traditional rule about discarding values which differ from the mean of the others by more than four times the average deviation of the others—are based on faulty application of large-number statistics to small-sample problems and are not to be trusted. There is, however, a very simple approach, known as the Q test, which is both statistically sound and straightforward to use.

Q Test. When one in a series of 3 to 10 measurements appears to deviate from the mean by more than seems reasonable, calculate the quantity Q, defined by

$$Q \equiv \frac{|(\text{suspect value}) - (\text{value closest to it})|}{(\text{highest value}) - (\text{lowest value})} \tag{9}$$

Compare this value of Q with the critical value Q_c in Table 1 corresponding to the number of observations in the series. If Q is equal to or larger than Q_c, the suspect measurement should be rejected. If Q is less than Q_c, this measurement must be retained.

EXAMPLE. Five determinations of the baseline reading of a spectrophotometer at a standardizing wavelength were 0.32, 0.38, 0.21, 0.35, and 0.34 absorbance units. May the 0.21 be discarded?

$$Q = \frac{|0.21 - 0.32|}{0.38 - 0.21} = 0.65$$

From Table 1, the critical value Q_c for $N = 5$ is 0.64. The value of Q is higher, hence the 0.21 value may be discarded with 90 percent confidence.

It should be noted that only *one* value in a series may be discarded in this manner. If there is more than one divergent value in a small series, it means that the data really do show a lot of scatter. When one value in a small series is considerably off but not sufficiently to justify rejection by the Q test, there is good reason to consider reporting the median value (i.e., the middle value) instead of the mean. As an illustration of this, consider the situation if the five spectrophotometer readings had been 0.32, 0.38, 0.23, 0.35, and 0.34. The value of Q for the suspect member of this set is 0.60, which does not quite exceed the critical value Q_c; and the 0.23 value cannot be rejected. However, reporting the median value 0.34 avoids the large effect that 0.23 would have if included in calculating the mean, while still allowing it a "vote." The mean of all five values is 0.32, whereas the mean without the suspect value is 0.35.

When N exceeds 10 it is suggested that an observation be rejected if its deviation from the mean of the others exceeds $2.6S$, where S is the estimated standard deviation of the others from their mean; this corresponds to about a 1 percent probability that the suspect observation is valid.

Whether the criteria for rejection are those recommended above or alternative criteria, *it is important that the chosen criteria be used consistently.* The criteria must be decided upon ahead of time; selection of criteria after the measurements have been made runs the grave risk of introducing bias. If a series of measurements presents a problem that

cannot be treated objectively with previously selected criteria, then if at all possible the entire series should be rejected and a new series of measurements made after careful review of the experimental procedure to minimize the possibility of large errors, mistakes, or bias.

STATISTICAL TREATMENT OF RANDOM ERRORS

Before presenting the results of statistical analysis of random errors, let us recall the definition of a normalized probability P. As an example, consider flipping a two-sided coin (heads or tails) 10 times and recording the number h of heads. Clearly, the variable h observed in that measurement must have one of the 11 values between $h_1 = 0$ and $h_{11} = 10$. Now carry out many measurements (each a set of 10 flips) and keep track of the number of times n_i that a given value h_i is observed. The probability P_i of observing the value h_i is defined by $P_i = n_i/\Sigma n_i = n_i/N$, where N is the total number of observations. It is clear that $\Sigma P_i = 1$, which corresponds to the *certainty* that all h_i values must lie somewhere between 0 and 10. In fact, we expect the most likely average of many measurements to be 5; that is, P_i should be sharply peaked at 5. But P is not a delta function; there is some spread about 5, and any single measurement may yield an h value of 3 or 6 or even 10!

The next three subsections describe the background and principles of random error treatment, and they introduce two important quantities: standard deviation σ and 95 percent confidence limits. The four subsections following these—Uncertainty in Mean Value, Small Samples, Estimation of Limits of Error, and Presentation of Numerical Results—are essential for the kind of random error analysis most frequently required in the experiments given in this book. The Student t distribution is particularly important and useful.

ERROR FREQUENCY DISTRIBUTION

Let it be supposed that a very large number of measurements x_i ($i = 1, 2, \ldots, N$) are made of a physical quantity x and that these are subject to random errors ϵ_i. For simplicity we shall assume that the true value x_0 of this quantity is known and therefore that the errors are known. We are here concerned with the frequency $n(\epsilon)$ of occurrence of errors of size ϵ. This can be shown by means of a bar graph, like that of Fig. 3, in which the

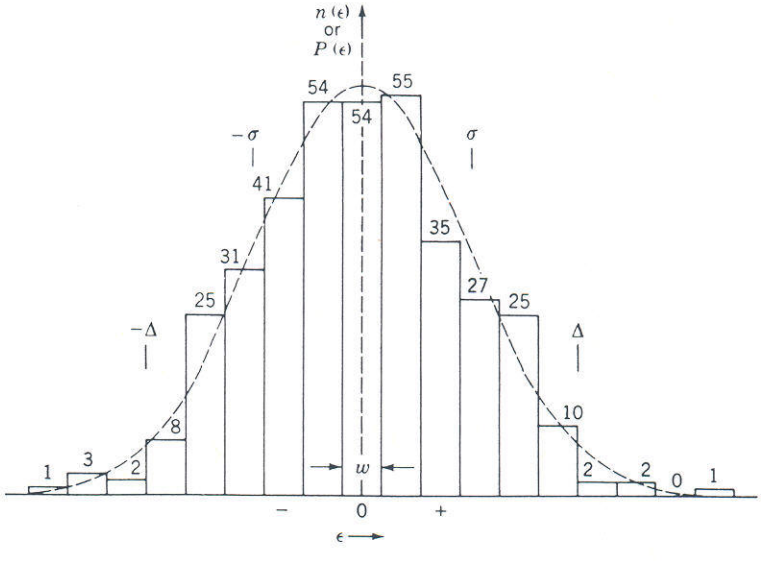


FIGURE 3
A typical distribution of errors. The bar graph represents the actual error frequency distribution $n(\epsilon)$ for 376 measurements; the estimated normal error probability function $P(\epsilon)$ is given by the dashed curve. Estimated values of the standard deviation σ and the 95 percent confidence limit Δ are indicated in relation to the normal error curve.

error scale is divided into ranges of equal width and the height of each bar represents the number of measurements yielding errors that fall within the respective range. The width w is chosen as a compromise between the desirability of having the numbers in each bar as large as possible and the desirability of having the number of bars as large as possible.

It is seen that even with as many as 376 measurements the graph shows irregularities, owing to the fact that the number of measurements represented by each bar is subject to statistical fluctuations that are not small in comparison with the number itself. From this graph we can make rough predictions concerning the probability that a measurement will yield an error of a given size. If we greatly increase the number of measurements represented (while perhaps decreasing the width w of the range more slowly than the rate of increase in the total number of measurements), the statistical fluctuations will become smaller in relation to the heights of the bars and our probability predictions are improved; we may even draw a smooth curve through the tops of the bars and assume it to represent an *error probability function* $P(\epsilon)$. The vertical scale of the $P(\epsilon)$ distribution should be adjusted by multiplication with an appropriate factor so that the function $P(\epsilon)$ is normalized, i.e., so that

$$\int_{-\infty}^{\infty} P(\epsilon)\, d\epsilon = 1 \tag{10}$$

Its significance is that the probability that a single measurement will be in error by an amount lying in the range between ϵ and $\epsilon + d\epsilon$ is equal to $P(\epsilon)\, d\epsilon$.

A probability function derived in this way is approximate; the true probability function cannot be inferred from any finite number of measurements. However, it can often be assumed that the probability function is represented by a Gaussian distribution called the *normal error probability function,*

$$P(\epsilon) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\epsilon^2/2\sigma^2} \tag{11}$$

where the *standard deviation* σ is a parameter that characterizes the width of the distribution.[1,7,9] It is the root-mean-square error expected with this probability function:

$$\sigma \equiv (\overline{\epsilon^2})^{1/2} = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \epsilon^2 e^{-\epsilon^2/2\sigma^2}\, d\epsilon\right)^{1/2} \tag{12}$$

If the true value x_0 and thus the errors ϵ_i themselves are known, σ can be estimated from

$$\sigma = \left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2\right)^{1/2} \tag{13}$$

The dashed curve in Fig. 3 represents a normal error probability function, with a value of σ calculated with Eq. (13) from the 376 errors ϵ_i.

The usual assumptions leading to the normal error probability function are those required for the validity of the *central limit theorem.* This remarkable theorem states that if (1) an error in measurement is compounded of a large number of unpredictably variable contributions (which need *not* conform to a normal error probability distribution function or any other particular distribution function), if (2) these contributions are mutually independent, and if (3) all of them are small compared with the expected error itself, then the probability distribution function for the error itself will be approximately normal and will approach a normal distribution in the limit as the number of contributions increases indefinitely. The proof of this theorem, which we will not give here,[10,11] begins with an expression of the *joint probability* function for all of the individual contributions to the error, see Eq. (24), and continues with manipulation of this expression by Fourier transformation

and other techniques. The assumptions leading to this theorem are sufficient but not always altogether necessary; the normal error probability function may arise at least in part from circumstances different from those described. The factors which in fact determine the distribution are seldom known in detail. Thus, it is common practice to assume that the normal error probability function is applicable even in the absence of valid a priori reasons. Indeed the distributions obtained frequently conform to the normal error probability function to within ordinary statistical error, as in the case of the distribution shown in Fig. 3. However, even with 376 measurements the normal error probability function is not proved to be valid in all respects; with a much larger number of measurements it might become apparent that the true probability function is slightly skewed or flat-topped or double-peaked (bimodal), etc.

INFINITELY LARGE SAMPLE

So far the discussion has dealt with the errors themselves, as if we knew their magnitudes. In actual circumstances we cannot know the errors ϵ_i by which the measurements x_i deviate from the true value x_0, but only the deviations $(x_i - \bar{x})$ from the mean $\bar{x}$ of a given set of measurements. If the random errors follow a Gaussian distribution and the systematic errors are negligible, the best estimate of the true value x_0 of an experimentally measured quantity is the arithmetic mean $\bar{x}$. If you as an experimenter were able to make a very large (theoretically infinite) number of measurements, you could determine the *true mean* μ exactly, and the spread of the data points about this mean would indicate the precision of the observation. Indeed, the probability function for the deviations would be

$$P(x - \mu) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \tag{14}$$

where μ is the mean and σ the standard deviation of the distribution of the hypothetical *infinite population* of all possible observations, given by

$$\sigma = \lim_{N\to\infty} \left[\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2\right]^{1/2} \tag{15}$$

In the absence of systematic errors, μ should be equal to the true value x_0.

The normal probability function as expressed by Eq. (14) is useful in theoretical treatments of random errors. For example, the normal probability distribution function is used to establish the probability P that an error is less than a certain magnitude δ, or, conversely, to establish the limiting width of the range, $-\delta$ to δ, within which the integrated probability P has a certain value:

$$P = \int_{-\delta}^{\delta} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{\epsilon^2}{2\sigma^2}\right] d\epsilon \tag{16}$$

If δ is specified as equal to σ, it is found that $P = 0.6826$. That is, 68.26 percent of all errors are less than the standard deviation in magnitude; see Fig. 4*a*. If P is specified as 0.95 (95 percent), then

$$\delta_{0.95} = 1.96\sigma \simeq 2\sigma \tag{17}$$

This may be described as a "95 percent confidence limit" for an individual measurement and is illustrated in Fig. 4*b*. The relationship between a specified uncertainty and the corresponding confidence limit is shown in Table 2. There is a 68.26% probability that the x value obtained in a single measurement will lie between $\mu - \sigma$ and $\mu + \sigma$. In contrast, the level of confidence that a single x value will lie between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 95 percent.

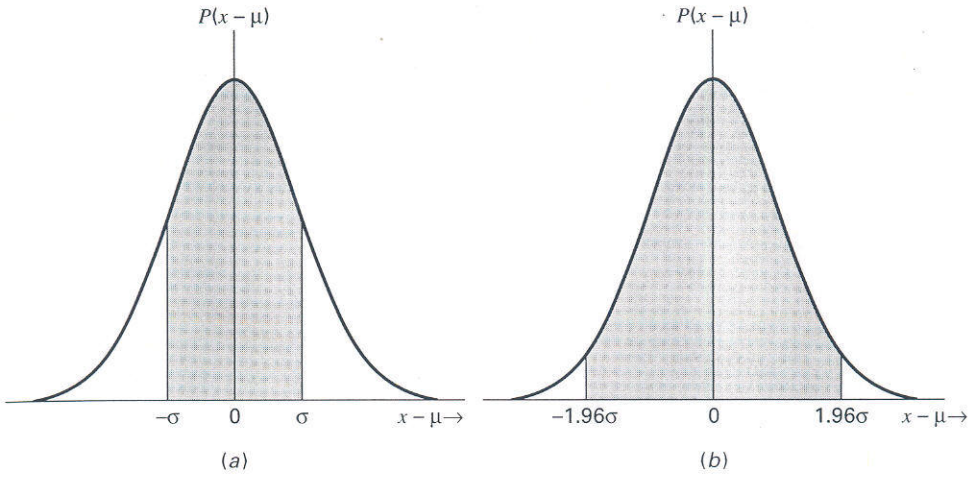


FIGURE 4
The integrated probability P for a normal error distribution, i.e., the statistical probability that the error lies between the specified limits. The value of P is given by the shaded area: (*a*) standard deviation error limits, $\pm\sigma$; (*b*) 95 percent confidence limits, $\pm 1.96\sigma$.

TABLE 2 Correspondence between uncertainty value and level of confidence

Uncertainty	$\pm\sigma$	$\pm 1.64\sigma$	$\pm 1.96\sigma$	$\pm 2.58\sigma$	$\pm 3.29\sigma$
Confidence level	68.26	90	95	99	99.9

Confidence limits will be discussed more fully in the section on uncertainties in mean values. It must be emphasized that Eq. (17) can be used only when σ is known. For σ to be known to a satisfactory approximation from the scatter of N measurements alone, N should be at least 20, preferably more.

LARGE FINITE SAMPLE

If a large enough number of data points are available (at least 20 and preferably 100), classical probability calculations provide a very satisfactory description of the precision of the measurements in question. In practice, one rarely has anything like 20 measurements of the same quantity; most experiments are based on averages of 6 or less. An important exception to this is computer-controlled automated experiments, where a digital reading (say, a voltage drop or the resistance of a thermistor as given by a high-resolution multimeter) can be made rapidly—say, within 100 ms. In this case, it is feasible and attractive to make 20 or more sequential readings and average them.

Classical probability statistics are inadequate for the treatment of small numbers of observations, and techniques developed only within recent years are necessary to avoid large errors in the estimates of error. For a finite (and usually small) number of observations of a quantity, one obtains data that show a certain amount of spread. The true mean μ and the true spread of the hypothetical infinite population of measurements are what one wishes to have. Usually the best that we can actually have with a finite number N of measurements are *estimates* of μ and σ. These are, respectively, the mean $\bar{x}$, and the spread of the measurements for which we use the quantity S, the estimated standard deviation, defined in Eq. (7). Clearly,

$$\text{as } N \to \infty, \qquad \bar{x} \to \mu \qquad \text{and} \qquad S \to \sigma \tag{18}$$

and the *experimental* probability function may be written

$$P_{\text{exp}}(x - \bar{x}) = \frac{1}{\sqrt{2\pi}\, S} \exp\left[-\frac{(x - \bar{x})^2}{2S^2}\right] \tag{19}$$

If N is larger than 20, S provides a satisfactory estimate of σ. Note that the estimated standard deviation S is a finite sample standard deviation, and the so-called standard deviation σ is more precisely the standard deviation of an infinite sample set.

UNCERTAINTY IN MEAN VALUE

The error ϵ_m in the mean of N observations [Eq. (2)] is the mean of the individual errors ϵ_i:

$$\epsilon_m = \frac{1}{N}\sum_{i=1}^{N} \epsilon_i \tag{20}$$

The mean square error is obtained by squaring both sides of this expression and averaging over an infinite population of sets of N measurements. The averaging is done by multiplying the square of each side of Eq. (20) by the probability distribution function $P(\epsilon)$ and integrating from $-\infty$ to ∞.

$$S_m^2 = \overline{\epsilon_m^2} = \frac{1}{N^2}\left[\sum_{i=1}^{N} \overline{\epsilon_i^2} + \sum_{i=1}^{N}\sum_{j=1}^{N}{}' \overline{\epsilon_i \epsilon_j}\right] \tag{21}$$

where the prime indicates a sum over j values that are not equal to i. The mean of $\epsilon_i\epsilon_j$ vanishes because $\epsilon_i\epsilon_j$ is as often negative as positive. Therefore

$$S_m^2 = \frac{1}{N^2} N S^2$$

or

$$S_m = \frac{S}{\sqrt{N}} \tag{22}$$

where S_m is the estimated standard deviation of the mean, in agreement with Eq. (8). This is a quantitative statement of the intuitively obvious conviction that the mean of a group of N independent measurements of equal weight has a higher precision than any single one of the measurements.

Previously we were concerned with estimating μ and σ, which are parameters of an infinite sample set. Now we are dealing with a large but finite set—say, 20 or more measurements of x. Consider several such finite sets, each with its average $\bar{x}$. The distribution of these *averages* follows a normal distribution characterized by a sample standard deviation S_m. That is, there is a 68.26 percent probability that the true value of x lies between $\bar{x} - S_m$ and $\bar{x} + S_m$ for a sample of $N \geq 20$ measurements.

As before, we can also determine a 95 percent confidence limit *in the mean,* to be denoted as Δ:

$$\Delta \equiv \delta_{m,0.95} = \frac{1.96S}{\sqrt{N}} \simeq \frac{2S}{\sqrt{N}} \tag{23}$$

When N is less than about 20, small sample statistics are needed and the Student t distribution described below should be used instead of the normal distribution.

The general results given above can be proved through the use of the normal probability distribution, Eq. (11).[8–10] The *joint probability* that of N observations the first lies in the range x_1 to $x_1 + dx_1$, the second lies in the range x_2 to $x_2 + dx_2$, and so on, is the product of the individual probabilities:

$$\begin{aligned} dP_{\text{joint}} &= P_{\text{joint}}(x_1, x_2, \ldots, x_N)\, dx_1\, dx_2 \cdots dx_N \\ &= P_1(x_1)\, dx_1\, P_2(x_2)\, dx_2 \cdots P_N(x_N)\, dx_N \end{aligned} \tag{24}$$

On substituting the normal probability distributions into this equation, we obtain

$$dP_{\text{joint}} = (2\pi)^{-N/2}\sigma^{-N}\exp\left[-\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}\right]dx_1\,dx_2\cdots dx_N \tag{25}$$

We now change from the set of variables $x_1, x_2, \ldots, x_N$ to a new set of N variables. Of these the only one we need to identify here is $\bar{x}$ [defined by Eq. (2)]; we integrate over the rest of them and obtain (with procedures not described here)[8]

$$dP_m = P_m(\bar{x})\,d\bar{x} \tag{26}$$

where

$$P_m(\bar{x}) = \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{N})}\exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right] \tag{27}$$

is the probability distribution function for the distribution of mean values $\bar{x}$ around the true mean of the distribution μ. Note that this equation resembles Eq. (11) except that $\sigma/\sqrt{N}$ replaces σ. Therefore, in the case where σ is known, we may write

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \tag{28}$$

In a similar way, using the *experimental* error distribution function of Eq. (19) instead of the true one of Eq. (11), we can derive Eq. (22).

SMALL SAMPLES

In the analysis presented above, we have discussed infinite ($N \to \infty$) and large finite ($N \geq 20$) samples of multiple measurements of some observable quantity x. Now we need to consider small samples ($1 < N < 20$), since in many cases N is as small as 3 or 4. As before, the mean $\bar{x}$ given by Eq. (2) is still our best estimate of the true value. The sample standard deviation S and the standard deviation of the mean S_m can still be calculated with Eqs. (7) and (22), but their usefulness is not clear since the probability distribution is not known. The problem is that the standard deviation σ is not known a priori, and we need to find some kind of distribution function that will apply for a small sample with unknown σ.

Student *t* Distribution. The way to proceed for small sample sets is to consider the joint probability given in Eqs. (24) and (25) and integrate it after a different change of variables than that used above.[7] In the present case, the only variable that we do not integrate over is

$$\tau \equiv \frac{\bar{x}-\mu}{(S/\sqrt{N})} = \frac{\bar{x}-\mu}{S_m} \tag{29}$$

After integration over all new variables except τ we obtain

$$dP = P(\tau)\,d\tau \tag{30}$$

where

$$P(\tau) = \left[\frac{\Gamma(N/2)}{\sqrt{(N-1)\pi}\,\Gamma([N-1]/2)}\right]\left(1+\frac{\tau^2}{N-1}\right)^{-N/2} \tag{31}$$

The factor in large square brackets, containing the gamma functions, is a normalization constant, with a value which assures that

$$\int_{-\infty}^{\infty} P(\tau)\,d\tau = 1 \tag{32}$$

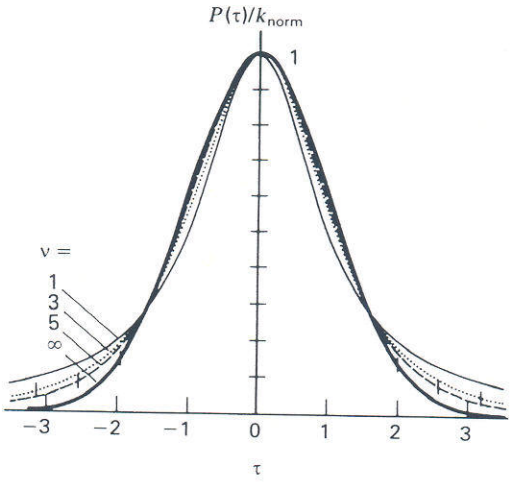


FIGURE 5
Student t distribution functions $P(\tau)$ for $\nu = 1, 3, 5, \infty$, where ν = number of degrees of freedom. The quantities actually plotted are

$$P(\tau)/k_{norm} = \left[1 + \frac{\tau^2}{N-1}\right]^{-N/2} = [1 + \tau^2/\nu]^{-(\nu+1)/2}$$

which are not normalized; they need to be multipled by the normalizing constant k_{norm}, which is the bracketed quantity in Eq. (31) with $N - 1 = \nu$. For $\nu = 1$ this is 0.3183, for $\nu = 3$ it is 0.3676, for $\nu = 5$ it is 0.3796, and for $\nu = \infty$ it is $1/\sqrt{2\pi} = 0.3989$. The curve for $\nu = \infty$ is the normal error probability curve. The short vertical bars mark the 95 percent confidence limits. These are not indicated for $\nu = 1$, as they are at $\tau = \pm 12.7$, far outside the plotting range.

The distribution represented in Eqs. (30) and (31) is known as the *Student t distribution;* "Student" is a pseudonym for W. S. Gosset, who developed this distribution for application to biometrics. Note that σ does not appear in the expression. The shape of the Student t distribution, unlike that of the normal distribution, is variable, depending on N. When N is large, it rapidly approaches the normal distribution. The distribution functions for several values of the number of degrees of freedom ν ($\nu = N - 1$ for determining the mean $\bar{x}$) are shown in Fig. 5.

We may use this distribution in much the same way we use the normal distribution. Suppose we wish to find the (symmetrical) range of values of τ over which the integral of the Student probability function is a fraction P (say 0.95, or 95 percent). Let us define this range to be from $-t$ to t:

$$\int_{-t}^{t} P(\tau)\, d\tau = P \tag{33}$$

"Critical" values of t for a given number of degrees of freedom ν and a given P are given in Table 3 and in published tables.[12] Let us define a limit of error δ as the value of $\bar{x} - \mu$ that corresponds to the limit of integration t. Then, from Eq. (29),

$$\delta = tS_m = t\frac{S}{\sqrt{N}} \qquad \Delta = t_{0.95}S_m = t_{0.95}\frac{S}{\sqrt{N}} \tag{34}$$

where again we use Δ to mean a 95 percent confidence limit in the mean. We emphasize the use of ν rather than $N - 1$ in Fig. 5 and Table 3 because the figure and the table are then relevant to cases in which several parameters are determined from the N measurements, rather than only one (in the present case, the mean $\bar{x}$). Such a case is least-squares curve fitting (Chapter XXII). If the number of parameters to be determined by least squares is n and the number of data points is N, then $\nu = N - n$.

Caution is required when using tables from some sources,[12] since for those tables the integration of Eq. (33) is from $-\infty$ to t rather than from $-t$ to t. If we designate the corresponding integrated probability as P', then

$$P' = (P + 1)/2 \qquad \text{or} \qquad P = 2P' - 1 \tag{35}$$

TABLE 3 Critical values of t (taken from Ref. 12)[a]

ν	P	0.50	0.80	0.90	0.95	0.98	0.99	0.999
	P'	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1		1.00	3.08	6.31	12.7	31.8	63.7	637.0
2		0.816	1.89	2.92	4.30	6.96	9.92	31.6
3		0.765	1.64	2.35	3.18	4.54	5.84	12.9
4		0.741	1.53	2.13	2.78	3.75	4.60	8.61
5		0.727	1.48	2.02	2.57	3.36	4.03	6.87
6		0.718	1.44	1.94	2.45	3.14	3.71	5.96
7		0.711	1.41	1.89	2.36	3.00	3.50	5.41
8		0.706	1.40	1.86	2.31	2.90	3.36	5.04
9		0.703	1.38	1.83	2.26	2.82	3.25	4.78
10		0.700	1.37	1.81	2.23	2.76	3.17	4.59
15		0.691	1.34	1.75	2.13	2.60	2.95	4.07
20		0.687	1.33	1.72	2.09	2.53	2.85	3.85
30		0.683	1.31	1.70	2.04	2.46	2.75	3.65
∞		0.674	1.28	1.64	1.96	2.33	2.58	3.29

[a] P is the probability that the mean μ of the population does not *differ from* the sample mean $\bar{x}$ by a factor of more than t, the tabular entry for a given number of degrees of freedom ν. P' is the probability that μ does not *exceed* $\bar{x}$ (or, alternatively, the probability that $\bar{x}$ does not *exceed* μ) by a factor of more than t for a given number of degrees of freedom ν.

Values of P' are also given in Table 3, and are useful in significance tests under certain circumstances to be described later.

Short-Cut Methods Based on the Range. For small samples (N from 3 to about 10) the range, $R = x(\text{largest}) - x(\text{smallest})$, can be used to obtain rough uncertainty estimates. The range, when multiplied by the appropriate value of K_2 from Table 4, gives an approximate but useful estimate of the standard deviation. The factor J in Table 4 is equivalent to $tK_2/\sqrt{N}$, where t is the critical value from the $P = 95$ percent column in Table 3 for the appropriate degree of freedom. Hence, to get a quick estimate of the 95 percent confidence limits of the mean of a few observations, simply calculate $\pm JR$. In summary,

$$\sigma \simeq S \simeq K_2 R \qquad \text{and} \qquad \Delta(95\%\ \text{confidence}) \simeq JR \tag{36}$$

TABLE 4 Factors for small-sample estimates based on the range R

N	K_2	J (95%)
3	0.59	1.3
4	0.49	0.72
5	0.43	0.51
6	0.40	0.40
7	0.37	0.33
8	0.35	0.29
9	0.34	0.26
10	0.33	0.23

These techniques, which have developed out of the statistical methodology known as quality control, are becoming widely adopted because of their convenience. They are, however, decidedly inferior to the methods already described.

ESTIMATION OF LIMITS OF ERROR

It must be kept in mind that the above discussion of random errors is based on the notion that one has no prior information about the precision of the measurements, in which case the uncertainty must be estimated from the scatter of the data in a given sample. In practice the number of independent measurements of an experimental quantity is often very small (sometimes two or only one) owing to time limitations, and it is unrealistic or impossible to estimate the confidence limits from the scatter or the range. However, the experimenter frequently has a good idea of the precision of the instrument or method being used. In this case, and also in cases where systematic errors must be estimated as well, the 95 percent confidence limits must be assigned on the basis of the individual judgment and experience of the experimenter. The ability to assign realistic limits of error, large enough to be safe but not so large as to detract unnecessarily from the value of the measurement, is one of the marks of a good experimentalist. It would be very difficult to formulate a set of rules that would be applicable in all cases; the confidence limits assigned in any given case will depend on the characteristics and capabilities of the instrument, its age and condition, the reproducibility of its reading, the quality of its calibration, and the user's experience and familiarity with it. A reasonable 95 percent confidence limit in weighing a small solid object on an analytical balance with weights calibrated to 0.05 mg might be 0.3 mg; in determining a short time interval with a manually actuated timer it might be 0.4 s; in reading an instrumental scale it might be 0.2 of the smallest scale division, or more depending on the instrument; etc.

Confidence limits assigned in this way will be represented in later sections by Δ on the rough assumption that they are 95 percent, based on long-term experience. However, one person's 95 percent estimate might be another's 90 percent, or 99 percent; this is unavoidable, and you as the experimenter simply have to do the best you can. Estimation of meaningful standard deviations in this manner is more difficult; if these are needed, rough values can be obtained as the estimated 95 percent confidence limits divided by 2.

PRESENTATION OF NUMERICAL RESULTS

At some point the final numerical result or results of an experiment and their uncertainties must be presented in a report or a publication, and it is important that this be done in a clear manner. Uncertainties are often presented (usually with "$\pm$") without making clear what kind of uncertainty estimates are being employed. Some conventions have arisen, which must be used with caution since they are by no means universally accepted; see Table 5.

TABLE 5 Most common modes of uncertainty presentation

RESULT	UNCERTAINTY	PRESENTATION
In terms of the estimated standard deviation, S, with parentheses around the uncertain digits		
3.42 bar	$S = 0.04$ bar	3.42(4) bar
0.451 nm	$S = 0.012$ nm	0.451(12) nm
In terms of the 95 percent confidence limit, Δ, with $\pm$		
66.32 J K^{-1}	$\Delta = 0.21$ J K^{-1}	66.32 ± 0.21 J K^{-1}
5.02×10^{-4} V	$\Delta = 5 \times 10^{-6}$ V	$(5.02 \pm 0.05) \times 10^{-4}$ V

Whatever the uncertainty figure, it is important that the number of decimal places given or implied by the uncertainty figure be the same as that of the numerical value to which it applies. Obviously, the same units apply to the uncertainty figure as to the numerical value. However, it is usually not necessary to state the units twice; it is sufficient to state them following the uncertainty figure. To be sure to avoid confusion, even with the confidence level at the usual 95 percent, it is becoming increasingly common to state the level of confidence and the number of measurements N or the number of degrees of freedom ν, e.g.,

$$w = 142.8 \pm 0.3 \text{ mg} \qquad (95 \text{ percent}, N = 7)$$

or

$$w = 142.8 \pm 0.3 \text{ mg} \qquad (95 \text{ percent}, \nu = 6)$$

when the result presented is a mean of several independent measurements. Otherwise, when the limits are estimated on the basis of judgment and experience,

$$w = 142.8 \pm 0.3 \text{ mg} \qquad (95 \text{ percent})$$

is better than a presentation with no confidence limit explicitly stated.

As results are prepared for presentation the relationship between the 95 percent confidence limit and the number of significant figures retained in the final result should be recognized: Generally, the confidence limit should be expressed as a single digit equal to or greater than 3 or as a two-digit number not greater than 25. Correspondingly, if the uncertainty is expressed as an estimated standard deviation S, it should be expressed as a single digit equal to or greater than 2 or a two-digit number not greater than 15.

SIGNIFICANCE TESTING

A frequent question that arises in quantitative experimentation is whether the observed difference between an experimentally measured quantity and the accepted, theoretical, or standard value is representative of a genuine difference or merely the natural result of random experimental error. Similarly, the question may arise whether two experimental studies differ in their results because of real differences either in the systems studied or in the methods used, or whether the apparent discrepancy is no greater than might be found by chance simply because of natural variability in the data.

The basic philosophy of significance testing is simple. Two series of measurements are made and their results are observed to differ by an amount we shall call D. The maximum difference that could be expected to occur if both series represent random samples of the same thing (i.e., no real difference in the quantity being measured in the two series) at a given probability level is calculated from probability theory. If the observed difference D is smaller, it is concluded that there is no real evidence of difference. If it is larger, it is concluded that the difference is probably real (at whatever probability level was chosen). The conventionally accepted probability level for considering differences significant is 95 percent. It is based on the experience that if lower levels of significance (e.g., 90 or 80 percent) are chosen, there are too many embarrassing false alarms, whereas if the requirements for significance are made too stringent (e.g., at the 99 percent level), too many real differences are missed. Under circumstances in which it is very important not to claim a significant difference when none is really there, the 99 percent or even the 99.9 percent level may be set as the basis for taking action, but in ordinary scientific practice the 95 percent level has been found to be best.

We will here show how the Student t statistic can be used to answer the question as to whether the above-mentioned difference D is statistically significant at a specified probability level. We begin by defining the effective uncertainty in the difference D between the two values being compared, using statistical data obtained in calculating mean values. This uncertainty, expressed as an estimated standard deviation, S_D, is then converted to a confidence limit Δ_D with the help of Eq. (34) written in the form

$$\Delta_D = tS_D \tag{37}$$

where t is found in Table 3. If the difference D exceeds the confidence limit Δ_D, the difference between the two values is said to be significant at the specified confidence level; otherwise, it is said to be not significant.

An equivalent approach is to determine t from $t = D/S_D$ and find in Table 3 for the applicable value of ν the values of t that bracket it. The heading of the column with the *lower* t entry represents the probability level at which the difference is *significant;* the heading of the column with the *higher* t entry represents the probability level at which the difference is *not significant.*

We will examine two different cases.

Case 1. The mean $\bar{x}$ of N experimental observations is being compared with a theoretical, accepted, standard, or "true" value x_0 that may be regarded as exact. Is the difference D between $\bar{x}$ and x_0 statistically significant? The uncertainty in $\bar{x}$ is characterized by S_m as calculated with Eq. (8). Given the desired confidence level P (95 percent, or whatever) and the number of degrees of freedom ν, t is found from Table 3 and Δ_D is calculated with Eq. (37). If $D \equiv |\bar{x} - x_0| > \Delta_D$, the difference D is significant.

A related situation occurs when the question asked is whether $\bar{x}$ is significantly *greater than* x_0. (The same question could be asked with *less than.*) In this situation the value of t is found under P' rather than under P; if $D \equiv \bar{x} - x_0 > \Delta_D$, the difference D is significant.

EXAMPLE. The accepted value of the weight percent of bromine in a standard KBr sample is 67.14, which may be regarded as exact. Four analyses of this sample by a new method gave a mean value $\bar{x}$ of 67.04 and an S value of 0.04. Is the difference $D = 0.10$ significant at the 95 percent level? Here $S_D = S_m = 0.04/\sqrt{4} = 0.02$, and with $\nu = 3$ we find from Table 3 and $P = 0.95$ that $t = 3.18$. Thus:

$$\Delta_D = 3.18 \times 0.02 = 0.063$$

Since $D > \Delta_D$, we conclude that the difference D *is* significant at the 95 percent confidence level. (It is, however, *not* significant at the 99 percent level.) We therefore conclude with 95 percent confidence that the new method of analysis is subject to systematic error.

EXAMPLE. The accepted value of the density of pure silver metal at 25°C is $x_0 = 10.501$ g cm^{-3}. This value, based on X-ray diffraction measurements, may be regarded as exact for the purposes of this example because the uncertainty, ± 0.0002 g cm^{-3}, is quite small compared to the random errors inherent in routine density measurements. It may also be regarded as an upper limit for pure silver metal. Low values can be given by a routine pycnometric determination if, for example, tiny air bubbles are entrained in the sample (see Sample Report, Chapter I). The density of a sample of finely divided pure silver metal was determined six times with a pycnometer. The mean density and the estimated standard deviation of the mean were $\bar{x} = 10.446$ g cm^{-3} and $S_m = 0.024$ g cm^{-3}; thus $D = -0.055$ g cm^{-3}. Is the measured density significantly lower than the accepted value? Setting $S_D = S_m = 0.024$ g cm^{-3} we obtain, with $\nu = 5$, $P' = 0.95$,

$$\Delta_D = 2.01 \times 0.024 = 0.048$$

Since the magnitude of D is greater than Δ_D, the difference is significant at the 95 percent confidence level, and systematic errors in the method, perhaps due to the presence of tiny air bubbles, are to be suspected. Notice that if, instead of taking $P' = 0.95$, we were to take $P = 0.95$, we would obtain

$$\Delta_D = 2.57 \times 0.024 = 0.062$$

indicating erroneously that the difference is *not* significant at the 95 percent level.

Case 2. The mean $\bar{x}_1$ of N_1 observations of a physical variable is being compared with the mean $\bar{x}_2$ of N_2 observations of the same physical variable made by a different method. Are the two means significantly different (suggesting that at least one may be subject to systematic error)? We need to have statistical estimates of uncertainty for both data sets; let the estimated standard deviations be S_1 and S_2. The square root of the *pooled* variance of the two data sets is[6]

$$S_{12} = \left[\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}\right]^{1/2} \tag{38}$$

and the estimated standard deviation in the *difference between the two means, D*, is

$$S_D = S_{12}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{1/2} = S_{12}\left(\frac{N_1 + N_2}{N_1 N_2}\right)^{1/2} \tag{39}$$

Again, this S_D value is multiplied by the t value appropriate to the desired confidence level, but here the number of degrees of freedom ν is $N_1 + N_2 - 2$. If $D \equiv |\bar{x}_1 - \bar{x}_2| > \Delta_D = tS_D$ then the difference D is significant.

If $N_1 = N_2 = N$, Eqs. (38) and (39) can be greatly simplified:

$$S_D = (S_{m1}^2 + S_{m2}^2)^{1/2} \tag{40}$$

and if in addition $S_{m1} = S_{m2} = S_m$, then

$$S_D = \sqrt{2}\, S_m \tag{41}$$

In either case the appropriate number of degrees of freedom ν is $2N - 2$.

Again, the situation may instead be one in which the question to be asked is whether one mean is significantly *greater than* (or, alternatively, *less than*) the other mean at the stated probability level. Here again P' should be used instead of P.

EXAMPLE. The heat of neutralization of an amine is measured four times in a large calorimeter and six times in a small one with the following results:

Large:	$N = 4$,	$\overline{\Delta H_1} = 30.5$ kJ mol^{-1},	$S_1 = 1.3$ kJ mol^{-1}
Small:	$N = 6$,	$\overline{\Delta H_2} = 28.9$ kJ mol^{-1},	$S_2 = 1.7$ kJ mol^{-1}

Is the difference $D = 1.6$ kJ mol^{-1} significant? The estimated standard deviation of the difference is, from Eqs. (38) and (39)

$$S_D = \left(\frac{3 \times 1.69 + 5 \times 2.89}{4 + 6 - 2} \times \frac{4 + 6}{4 \times 6}\right)^{1/2} = 1.01 \text{ kJ mol}^{-1}$$

For $\nu = 4 + 6 - 2 = 8$ and $P = 0.95$, t from Table 3 is 2.31. Therefore

$$\Delta_D = 2.31 \times 1.01 = 2.33 \text{ kJ mol}^{-1}$$

Since $D < \Delta_D$ the difference is not significant at the 95 percent level. If we set $t = D/S_D = 1.6/1.01 = 1.58$, we see that this is bracketed in Table 3 between 1.40 for $P = 0.80$ and 1.86 for $P = 0.90$. Therefore the difference is significant at the 80 percent confidence level but not at the 90 percent level. Thus there may be systematic errors in one (or both) calorimetric methods, but this is not well established by the available data.

EXAMPLE. A chemical company was trying to decide between Company A and Company B for a contract to provide a certain catalytic product at a set price. The chemical company was leaning strongly toward Company B for reasons not connected with the quality of the product (i.e., one of the directors had a brother-in-law working for Company B), but could not justify signing with that company if Company A could be shown to provide a product that was significantly better in its catalytic yield at the 95 percent confidence level. The crucial test agreed upon by all parties was to measure, in both products under identical conditions and under inspection by representatives of both suppliers, the value of the catalytic activity as measured by the conversion of CO and H_2 to methanol. In each case, 11 measurements of the yield variable were made. For Company A's product the mean was 53.23 (in the appropriate units, which we are not concerned with here) with $S_m = 0.75$; for Company B's product the mean was 50.87 with $S_m = 0.85$. Should Company A be selected to provide the product? Here $D = 2.36$, and $\nu = 20$. At $P = 0.95$, $t = 2.09$. Using Eqs. (40) and (37) we obtain

$$S_D = (0.75^2 + 0.85^2)^{1/2} = 1.25$$

$$\Delta_D = 2.09 \times 1.25 = 2.61 \qquad (95 \text{ percent}, \nu = 20)$$

Since $D < \Delta_D$, the directors of the chemical company were happily preparing to award the contract to Company B, until a bright young Ph.D. chemist recently added to their research division pointed out that P' should have been used instead of P. The question being asked was *not* "Is A's product significantly *different* from B's product?" but rather "Is A's product significantly *better* than B's product?," it having been presumed ahead of time that otherwise B's product would be the one chosen. With $t = 1.72$,

$$\Delta_D = 1.72 \times 1.25 = 2.15$$

Glumly, the directors signed the contract with Company A. The record does not reveal whether the bright young Ph.D. was promoted or is now working for Company A.

The above discussion and examples are based on the application of statistical theory. Something should be said about situations in which the limits of error for one or both of the values that are being compared have been estimated by the experimenter on the basis of judgment and experience rather than with equations derived from statistics. Here, the number of degrees of freedom is unknown or without meaning. In the situation analogous to Case 1, if the difference between the measured and "true" values exceeds the estimated Δ for the former, the difference is regarded as significant at the 95 percent confidence level. In the situation analogous to Case 2, the pooled uncertainty may be taken as the square root of the sum of the squares of the individual estimated confidence limits, weights being used if judged necessary:

$$\Delta_D = (\Delta_1^2 + \Delta_2^2)^{1/2} \tag{42}$$

or

$$\Delta_D = \left[\frac{w_1\Delta_1^2 + w_2\Delta_2^2}{w_1 + w_2}\right]^{1/2} \tag{43}$$

If the difference between the means exceeds Δ_D, then the difference is regarded as significant at the 95 percent confidence level. In these situations there is little point in distin-

guishing between ''different from'' and ''greater (less) than'' in view of the roughness of the 95 percent confidence estimate.

PROPAGATION OF ERRORS

An experiment has been performed; a variety of direct measurements have been made—weights, volumes, temperatures, measured electromotive forces, spectral absorbances, etc.—and uncertainties in all of them have been estimated, either from statistical data obtained by repeated measurements or from judgment and experience. From the values obtained by direct measurement, and with the aid of a phenomenological theory, a final numerical result is calculated. Let the desired numerical result be designated by F and the directly measured quantities be designated by $x, y, z, \ldots$. *The latter quantities are assumed to be mutually independent.* Let their uncertainties, usually in the form of 95 percent confidence limits, be designated $\Delta(x), \Delta(y), \Delta(z), \ldots$. The value of F is determined by substituting the experimentally determined values of $x, y, z, \ldots$ into a mathematical formula, which we write symbolically as

$$F = f(x, y, z, \ldots) \tag{44}$$

The issue to be discussed here is how can one estimate the uncertainty of the final result F, usually in the form of a 95 percent confidence limit.

Infinitesimal changes dx, dy, etc., in the experimentally determined values will produce in F the infinitesimal change dF, where

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial z}dz + \ldots \tag{45}$$

If the changes are finite rather than infinitesimal, but are small enough that the values of the partial derivatives are not appreciably affected by the changes, we have approximately

$$\Delta F = \frac{\partial F}{\partial x}\Delta x + \frac{\partial F}{\partial y}\Delta y + \frac{\partial F}{\partial z}\Delta z + \ldots \tag{46}$$

This is equivalent to a Taylor expansion in which only the first-power terms have been retained. Now suppose that Δx represents the experimental error $\epsilon(x)$ in the quantity x:

$$\Delta x = \epsilon(x) \equiv x \text{ (measured)} - x \text{ (true)} \tag{47}$$

Such errors will produce an error in F,

$$\Delta F = \epsilon(F) \equiv F \text{ (calculated from measured } x, y, z, \ldots) - F \text{ (true)} \tag{48}$$

the value of which is given by

$$\epsilon(F) = \frac{\partial F}{\partial x}\epsilon(x) + \frac{\partial F}{\partial y}\epsilon(y) + \frac{\partial F}{\partial z}\epsilon(z) + \ldots \tag{49}$$

PROPAGATION OF SYSTEMATIC ERRORS

Systematic or ''determinate'' errors are not governed by probability statistics and do not involve such concepts as the variance or standard deviation. If a systematic error in a quantity x is established, the sign and magnitude of $\epsilon(x)$ is known, and Eq. (49) can be used for the propagation of the systematic error in F. Two important examples of this are:

$$F = x + y - z \qquad \epsilon(F) = \epsilon(x) + \epsilon(y) - \epsilon(z) \tag{50}$$

$$F = \frac{xy}{z} \qquad \frac{\epsilon(F)}{F} = \frac{\epsilon(x)}{x} + \frac{\epsilon(y)}{y} - \frac{\epsilon(z)}{z} \tag{51}$$

Of course, if systematic errors are known, they should be corrected. Equation (49) provides one way of calculating the correction; normally it would be better simply to recalculate the result. Much more often, systematic errors are unknown; if suspected and of unknown sign, allowance for their possible magnitudes can be included in the estimated limits of "random" error.

PROPAGATION OF RANDOM ERRORS

In the case of random errors, we do not know the actual values of $\epsilon(x)$, $\epsilon(y)$, etc., and we cannot determine the actual value of $\epsilon(F)$. However, we have already assigned to each experimental variable an estimated standard deviation S or a confidence limit Δ. We now wish to deduce the corresponding uncertainty in the final result F, taking into consideration the high probability that the errors in the several variables will tend somewhat to cancel one another out. Let us square both sides of Eq. (49):

$$[\epsilon(F)]^2 = \left(\frac{\partial F}{\partial x}\right)^2 [\epsilon(x)]^2 + \left(\frac{\partial F}{\partial y}\right)^2 [\epsilon(y)]^2 + \cdots + 2\left(\frac{\partial F}{\partial x}\right)\left(\frac{\partial F}{\partial y}\right)\epsilon(x)\epsilon(y) + \ldots \quad (52)$$

Now let us average this expression over all values expected for $\epsilon(x)$, $\epsilon(y)$, . . . , in accordance with the normal frequency distribution. Since the $\epsilon(x)$ independently have average value zero, we expect the cross-terms to vanish. However, the squared terms are always positive and will not vanish. If we replace the average of each squared error by the variance, we obtain

$$S^2(F) = \left(\frac{\partial F}{\partial x}\right)^2 S^2(x) + \left(\frac{\partial F}{\partial y}\right)^2 S^2(y) + \ldots \quad (53)$$

It can be seen immediately that the variance in the mean of N measurements $S^2(\bar{x}) = S^2(x)/N$, which was discussed previously, can be obtained as a trivial example of Eq. (53) by taking

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

Equation (53), which is not limited to normal distributions, is an equation that can be used when all or substantially all of the experimental variables are means of sets of measurements from which statistical measures of variance are available. From $S(F)$ a confidence limit $\Delta(F)$ can be derived if the total number of degrees of freedom can be ascertained. In some cases this may be the sum of the number of degrees of freedom from all of the experimental variables; in general it will be larger than the number of degrees of freedom from any one variable. This complicated problem may be bypassed by assuming that the central limit theorem is at work: the error probability distribution in F is more closely normal than are the error probability distributions in the experimental variables. Thus, to a satisfactory approximation, the value of t to be used can be taken to be that for $\nu = \infty$: For $P = 95$ percent, $t = 1.96$.

However, we are usually working in situations where many or most of the uncertainties in the experimental variables are estimated on the basis of judgment and experience. On the assumption that the errors in such variables are distributed in a roughly normal manner, we may resort to the following approximate expression in terms of confidence limits:†

$$\Delta^2(F) = \left(\frac{\partial F}{\partial x}\right)^2 \Delta^2(x) + \left(\frac{\partial F}{\partial y}\right)^2 \Delta^2(y) + \left(\frac{\partial F}{\partial z}\right)^2 \Delta^2(z) + \ldots \quad (54)$$

†The use of Δ as a limit of error, as in $\Delta(x)$, is not to be confused with its use as a difference, as in Δx.

In certain cases, the propagation of random errors can be carried out very simply.

1. For $F = ax \pm by \pm cz$,

$$\Delta^2(F) = a^2\,\Delta^2(x) + b^2\,\Delta^2(y) + c^2\,\Delta^2(z) \tag{55}$$

2. For $F = axyz$ (or axy/z or ax/yz or a/xyz),

$$\frac{\Delta^2(F)}{F^2} = \frac{\Delta^2(x)}{x^2} + \frac{\Delta^2(y)}{y^2} + \frac{\Delta^2(z)}{z^2} \tag{56}$$

3. For $F = ax^n$,

$$\frac{\Delta^2(F)}{F^2} = n^2\,\frac{\Delta^2(x)}{x^2} \rightarrow \frac{\Delta(F)}{F} = n\,\frac{\Delta(x)}{x} \tag{57}$$

4. For $F = ae^x$,

$$\Delta^2(F) = a^2e^{2x}\,\Delta^2(x) \rightarrow \frac{\Delta(F)}{F} = \Delta(x) \tag{58}$$

5. For $F = a \ln x$,

$$\Delta^2(F) = \frac{a^2}{x^2}\,\Delta^2(x) \rightarrow \Delta(F) = a\,\frac{\Delta(x)}{x} \tag{59}$$

Note the differences between Eqs. (55) and (56) for random errors and the comparable expressions in Eqs. (50) and (51) for systematic errors.

Equation (55) includes the important special case $(x - y)$ which is applicable, for example, to differences between two readings of the same variable (e.g., the weight of a liquid plus container and the weight of the empty container, the initial and final readings of a thermometer). In many cases the limits of error in the initial and final readings may be taken as equal, in which case the limit of error in the difference between the two readings is equal to $\sqrt{2}$ times the limit of error in a single reading. Equation (56) can be simplified in the special case that the fractional error in each variable is approximately the same [i.e., $\Delta(x)/x \simeq \Delta(y)/y \simeq \Delta(z)/z$]. In this case, the fractional error in F is roughly equal to the sum of the fractional errors in the individual variables divided by the square root of the number of variables.

Many propagation-of-error treatments are not so simple as those illustrated by Eqs. (55) to (59). The expression for F is often complicated enough or the functional form sufficiently awkward to justify a breakdown of the procedure into steps:

$$F = f_0(A, B, \ldots)$$

where (60)

$$A = f_1(x_1, x_2, \ldots), \qquad B = f_2(y_1, y_2, \ldots), \qquad \text{etc.}$$

We can then write

$$\begin{aligned}
\Delta^2(F) &= \left(\frac{\partial F}{\partial A}\right)^2 \Delta^2(A) + \left(\frac{\partial F}{\partial B}\right)^2 \Delta^2(B) + \ldots \\
\Delta^2(A) &= \left(\frac{\partial A}{\partial x_1}\right)^2 \Delta^2(x_1) + \left(\frac{\partial A}{\partial x_2}\right)^2 \Delta^2(x_2) + \ldots \\
\Delta^2(B) &= \left(\frac{\partial B}{\partial y_1}\right)^2 \Delta^2(y_1) + \left(\frac{\partial B}{\partial y_2}\right)^2 \Delta^2(y_2) + \ldots, \text{ etc.}
\end{aligned} \tag{61}$$

provided that A, B, . . . are *independent.* If the quantities A, B, etc., are not independent (that is, in the event that some of the variables x_i are identical with some of the y_i), the limit of error calculated in this manner will be incorrect. For example, if we let

$$F = a(e^{kx} - 1) + b(y - cx) = A + B$$

it would be improper to write

$$\Delta^2(F) = \Delta^2(A) + \Delta^2(B)$$
$$\Delta^2(A) = (ake^{kx})^2\,\Delta^2(x) \qquad \Delta^2(B) = b^2\,\Delta^2(y) + b^2c^2\,\Delta^2(x)$$

Treatment of the function as a whole gives the correct result

$$\Delta^2(F) = (ake^{kx} - bc)^2\,\Delta^2(x) + b^2\,\Delta^2(y)$$

which differs from the result first given by a term $-2abcke^{kx}\,\Delta^2(x)$.

When the expression for F is a complicated one, it is advisable to watch for opportunities for transforming the expression or parts of the expression so that single symbols representing known quantities can be substituted for groupings of several symbols. The most advantageous simplifications are usually obtained when the quantities substituted are the end results of the calculations, such as F itself, or at least quantities which represent the later stages of the calculation of F. This is advantageous not only because it results in the greatest economy of symbols but also because it presents the most frequent opportunities for cancellation and simplification. For example, the limit of error of

$$F = C\frac{x^2}{y}$$

can be obtained from

$$\Delta^2(F) = \left(\frac{2Cx}{y}\right)^2 \Delta^2(x) + \left(\frac{Cx^2}{y^2}\right)^2 \Delta^2(y)$$

but for computational purposes it is more convenient to divide this equation by F^2, since the value of F will already have been calculated; we obtain

$$\frac{\Delta^2(F)}{F^2} = 4\,\frac{\Delta^2(x)}{x^2} + \frac{\Delta^2(y)}{y^2}$$

Estimation of uncertainties is often more complicated than in the above instances because the functional form of F may be more complicated. Before one begins to differentiate F, which is always in principle possible but often yields masses of unwieldly expressions with ever-present possibility of mistakes, it is well to examine F carefully with two purposes in mind: (1) to see if the expression for F can be simplified to make differentiation easier; and (2) to see if there are any terms that can be neglected because their effect on the uncertainty is small. While F must be evaluated in its precise form to yield the desired numerical result to the maximum possible precision, it should be remembered that the value of the limit of error need not be known to great precision; thus a quantity can often be dropped in a limit-of-error treatment even if it cannot be neglected in the calculation of F itself. A term can ordinarily be dropped in the limit-of-error treatment if the effect produced in $\Delta(F)$ is no more than about 5 percent or even 10 percent. An overgenerous (overconservative) estimate of the limit of error is to be preferred to an insufficient (overoptimistic) estimate; if a term is to be dropped, it is better that dropping it make $\Delta(F)$ a little too large than a little too small. For example, in the cryoscopic determination

of the molecular weight of an unknown substance, the molecular weight is given by Eq. (10-21) as

$$M = \frac{1000gK_f}{G\,\Delta T_f}(1 - k_f\,\Delta T_f)$$

Since k_f is of the order of 0.01 K^{-1}, the term $k_f\,\Delta T_f$ is itself very small compared with unity and will contribute negligibly to the uncertainty in comparison to ΔT_f itself. Therefore, the treatment of uncertainties, but not the calculation of M itself, may be based on the simpler, approximate expression

$$M = \frac{1000gK_f}{G\,\Delta T_f}$$

The value of $\Delta(M)/M$ is then easily obtained by using the form given in Eq. (56).

Another technique that may be used in cases where the formula for F is very complicated and difficult to differentiate is "computational differentiation," which is extremely tedious to do by hand or with a calculator but reasonably easy to do on a programmable calculator or a computer that has already been programmed to compute the result F. This consists of first computing the result F by substituting in the values of the x_i, then calculating the result $F + \Delta_1 F$ obtained by changing x_1 to $x_1 + \Delta x_1$, where Δx_1 is a very small increment, leaving the other x_i unchanged. Then the change $\Delta_1 F$ is obtained by subtraction. This is done successively with x_2, x_3, . . . , yielding $\Delta_2 F$, $\Delta_3 F$, ... Then

$$\frac{\partial F}{\partial x_i} \simeq \frac{\Delta_i F}{\Delta x_i} \tag{62}$$

These values of partial derivatives may then be substituted into Eq. (54) to yield the desired limit-of-error value.

EXAMPLES

Given below are two numerical examples of a propagation of errors treatment.

The first example is related to the optical rotation measurement described as a case study in the next section. The formula for calculating the specific rotation $[\alpha]$ is

$$[\alpha] = \frac{V}{Lm}\alpha \tag{63}$$

where V is the volume of the solution in cubic centimeters, L is the length of the polarimeter tube in decimeters, m is the mass of solute contained in volume V, and α is the rotation angle in degrees. Values of these four independent variables together with their 95 percent confidence limits are as follows:

	V	L	m	$\bar{\alpha}$
	25.00 cm^3	2.000 dm	1.7160 g	+20.950°
Δ =	0.02 cm^3	0.002 dm	0.0003 g	0.016°

The confidence limits for the first three variables were estimated on the basis of judgment and experience. The last was determined from a statistical analysis of data from 10 individual measurements with Eq. (8). From Eq. (63) we calculate

$$F = [\alpha]_D^{25} = \frac{25.00}{2.000 \times 1.7160} \times 20.950 = 152.61 \text{ deg dm}^{-1} \text{ (g cm}^{-3})^{-1}$$

We observe that Eq. (63) is of the type for which Eq. (56) is applicable. Therefore,

$$\Delta^2([\alpha]) = [\alpha]^2\left[\frac{\Delta^2(V)}{V^2} + \frac{\Delta^2(L)}{L^2} + \frac{\Delta^2(m)}{m^2} + \frac{\Delta^2(\alpha)}{\alpha^2}\right]$$

$$= 152.61^2\left(\frac{0.02^2}{25.00^2} + \frac{0.002^2}{2.000^2} + \frac{0.0003^2}{1.7160^2} + \frac{0.016^2}{20.950^2}\right)$$

$$= 2.33 \times 10^4(64 + 100 + 3 + 58) \times 10^{-8} = 0.0524$$

$$\Delta([\alpha]) = 0.23 \text{ deg dm}^{-1} \text{ (g cm}^{-3})^{-1}$$

One would report this result as $[\alpha] = 152.61 \pm 0.23$ deg dm^{-1} (g cm^{-3})$^{-1}$ or 1.5261 ± 0.0023 deg m^2 kg^{-1} in SI units.

As a second and more complicated example, consider the measurement of the temperature of a boiling organic liquid with the gas thermometer described in Exp. 1. The absolute temperature is given by

$$T = T_0\frac{p(1 + pv/p_rV + 3\alpha t - B/\tilde{V})}{p_0(1 + p_0v/p_rV - B_0/\tilde{V})} \tag{64}$$

This equation has been obtained by combining Eq. (1-8) with a similar equation in which T is replaced by T_0, which is the temperature at the ice point (273.15 K). For definitions of the quantities in this expression, see Exp. 1; quantities with subscript zero are those corresponding to T_0. The volume V of the thermometer bulb is not to be confused with $\tilde{V} \simeq RT_0/p_0$, which is the molar volume of the gas employed. The direct partial differentiation of this equation with respect to p or p_0 would yield a very messy expression, so we are motivated to look for ways to simplify the equation. Since the last two terms in parentheses in the denominator are small compared to unity, we can apply the series expansion expression for the reciprocal of the binomial $1 + x$,

$$\frac{1}{1 + x} = 1 - x + x^2 - \ldots$$

keeping only the first two terms on the right. Thus we may write the expression for T, ignoring terms of second order in small quantities, as [cf. Eqs. (1-12), (1-13)]

$$T = T_0\frac{p}{p_0}\left[1 + \frac{(p - p_0)}{p_r}\frac{v}{V} + 3\alpha t - \frac{(B - B_0)}{\tilde{V}}\right]$$

$$= T_0\left[\frac{p}{p_0}\left(1 + 3\alpha t - \frac{p_0}{RT_0}(B - B_0)\right) + \frac{(p^2 - pp_0)}{p_0p_r}\frac{v}{V}\right]$$

This may be differentiated with respect to p, p_0, p_r, and v/V without much difficulty. Before going to the trouble of differentiating, we should examine the terms for magnitude. Using values of α, B, and B_0 given in Exp. 1, and noting that v/V is of the order of 0.01, we may surmise that for the purpose of deriving a limit-of-error estimate all terms except the leading one may be neglected;

$$T = T_0\frac{p}{p_0}$$

Let us put in some numbers. Assume that the gas in the bulb is helium; then $B - B_0$ is certainly negligible (see Table 1-1). For $p = 510.1 \pm 1.0$ Torr, $p_0 = 418.2 \pm 1.0$ Torr, $T_0 = 273.15$ K (regarded as exact), T by the above simple equation is 333.2 K; $t = 60.0$°C. With, in addition, $p_r = 456.5 \pm 1.0$ Torr, $3\alpha t = 3(3.2 \times 10^{-6})60 = 0.00058$,

and $v/V = 0.011 \pm 0.003$, the original precise expression gives $T = 333.94\ K$. Using the simple expression for our limit-of-error treatment we have

$$\frac{\Delta^2(T)}{333.94^2} = \frac{1.0^2}{510.1^2} + \frac{1.0^2}{418.2^2}$$

or $\Delta(T) = 1.06$ K. Thus

$$T = 333.9 \pm 1.1 \text{ K} \qquad \text{or} \qquad t = 60.8 \pm 1.1°\text{C}$$

Suppose, however, that we are slightly unsure about the effect of the neglect of the v/V term on the limit of error. Let us disregard the extremely small terms involving α and $(B - B_0)$ and otherwise carry out a full-fledged limit-of-error treatment. The partial derivatives and their values are:

$$\frac{\partial T}{\partial p} = \frac{T_0}{p_0}\left[1 + \frac{(2p - p_0)}{p_r}\frac{v}{V}\right] = 0.662 \text{ K Torr}^{-1}$$

(instead of 0.653 K Torr^{-1} from the simple approximate expression),

$$\frac{\partial T}{\partial p_0} = -T_0\left(\frac{p}{p_0^2} + \frac{p^2}{p_0^2 p_r}\frac{v}{V}\right) = -0.807 \text{ K Torr}^{-1}$$

(instead of -0.798 K Torr^{-1} from the simple approximate expression),

$$\frac{\partial T}{\partial p_r} = -T_0\frac{p(p - p_0)}{p_0 p_r^2}\frac{v}{V} = -0.0016 \text{ K Torr}^{-1}$$

(instead of being neglected altogether), and

$$\frac{\partial T}{\partial (v/V)} = T_0\frac{p(p - p_0)}{p_0 p_r} = 67.1 \text{ K}$$

(instead of being neglected altogether). Then, we obtain from Eq. (54)

$$\begin{aligned}\Delta(T) &= [1.0^2(0.662^2 + 0.807^2 + 0.0016^2) + 0.003^2(67.1)^2]^{1/2} \\ &= (1.0895 + 0.0405)^{1/2} \\ &= 1.063 \text{ K}\end{aligned}$$

which is not significantly different from the result of the earlier calculation.

None of the experiments described in this book will involve mathematical treatments for limit-of-error estimation more complicated than this example. With careful judgment, most limit-of-error treatments can be made much simpler.

CASE HISTORY OF AN ERROR EVALUATION

A professor once received from a chemical supply house a specimen of a crystalline organic compound needed in her research. This compound was optically active, with a positive rotation (i.e., it was dextrorotatory). The professor wanted to check that the sample was not contaminated by the other (levorotatory) optical isomer. Such contamination would decrease the magnitude of the positive rotation. Therefore she gave a small amount of the material to a graduate student with instructions to determine its *specific rotation* at 25°C with the highly precise polarimeter available in the laboratory (see Exp. 29 and Chapter XX).

The next day, the student came back to the professor bearing his laboratory notebook. This showed that he had first adjusted the thermostat bath that supplies water to the jacket of the polarimeter so that it would cycle nicely between 24.8 and 25.3°C. He then weighed out 1.5220 g of the material (which he had stored overnight in a desiccator) and dissolved it in distilled water in a volumetric flask to make 25.00 mL of solution, being careful to equilibrate the flask in the thermostat bath before making the solution up to the mark. He then filled a 20.00-cm (2.000-dm) polarimeter tube with the solution and obtained the following values for the optical rotation with sodium-D light, taking care to approach the extinction point from alternate sides:

α (deg)	dev
+20.04	−0.004
+20.07	+0.026
+20.05	+0.006
(+20.09)	rejected
+20.04	−0.004
+20.02	−0.024
+20.04	−0.004
+20.03	−0.014
+20.06	+0.016
+20.05	+0.006

$$\bar{\alpha} = +20.044 \qquad \Sigma \text{ dev}^2 = 0.001824$$

$$S = \left(\frac{0.001824}{9-1}\right)^{1/2} = 0.015° \qquad S_m = \frac{0.015°}{\sqrt{9}} = 0.005°$$

In order to evaluate a 95 percent confidence limit Δ for this mean value $\bar{\alpha}$, the Student t distribution can be used. In the present case, we have $N = 9$, $\nu = 8$ and should look in Table 3 under the $P = 0.95$ column to the row corresponding to $\nu = 8$. This yields a $t_{0.95}$ value of 2.31. Equation (34) then gives

$$\Delta = t_{0.95} S_m = 2.31 \times 0.005° = 0.012°$$

Thus

$$\bar{\alpha} = +20.044° \pm 0.012° \qquad (95 \text{ percent}, \ \nu = 8)$$

which means that if the student were to repeat this procedure many times he would find that the mean value of α would differ from 20.044° by no more than 0.012° in 95 percent of the cases.

Using the formula for specific rotation given by Eq. (63), which is equivalent to Eq. (29-14), the student calculated for the specific rotation at 25°C with sodium-D light

$$[\alpha]_D^{25} = \frac{25.00}{2.0000 \times 1.5220} \times (+20.044) = +164.62 \text{ deg dm}^{-1} \text{ (g cm}^{-3})^{-1}$$

Estimating from his general experience 95 percent confidence error limits of 0.02 cm^3 in V, 0.002 dm in L, and 0.0003 g in m, and using his 95 percent confidence limits of 0.012° in the mean value of α, the student calculated by propagation-of-errors techniques 95 percent confidence limits of ±0.22 in the $[\alpha]_D^{25}$ result.

When the professor saw the final result for the specific rotation, +164.62 ± 0.22, her jaw dropped in bewilderment. The reported value *exceeded* the literature value of the pure

dextrarotatory compound, +152.70, by 11.92, about 50 times the claimed limit of error! This ridiculous result could mean only that a serious error or mistake had been made in this determination, unless the literature value itself were seriously in error, which seemed unlikely. She looked for, but could not find, a mistake in arithmetic. She examined the data and asked the student why the reading of 20.09° had been rejected. The student replied that the reading seemed too far out of line. In order to check on this, the professor suggested that they carry out a Q test. This test showed that the reading had been wrongfully rejected (calculated $Q = 0.29$; for $N = 10$, $Q_c = 0.41$; see Table 1 and accompanying discussion). With a revised average of +20.049° ($S = 0.020°$, $S_m = 0.006°$), the professor quickly recalculated the specific rotation, obtaining

$$[\alpha]_D^{25} = +164.66 \pm 0.23 \text{ deg dm}^{-1} \text{ (g cm}^{-3})^{-1}$$

Clearly, the real problem had not yet been found.

She asked to be shown the experimental setup. On seeing the balance on which the student had weighed out the material she noticed that the numerical scale on one of the weight dials was smudged in a few places; one of the 6s looked a little like a 5. Could the student have recorded 1.5220 g when the true weight might have been 1.6220 g? That would be enough to account for most of the difference between the experimental and literature values. Under pressure, the student conceded that this was possible, even likely. The professor calculated, tentatively,

$$[\alpha]_D^{25} = +164.66 \times \frac{1.5220}{1.6220} = +154.51$$

This was now within shooting distance of the literature value, although the occurrence of the weighing mistake could not be considered a certainty. (The professor did point out that had the error gone the other way, she might have thrown out the sample and tried to find new material.)

The professor then inspected the polarimeter setup. She noticed that the circulating pump, which was to supply constant-temperature water from the thermostat bath to the water jacket, was not running. ''Did you have the circulating pump on during your measurements?'' she asked. ''I thought that it went on automatically,'' replied the student, who had used the prominent on–off switch for the sodium-vapor lamp but had not noticed the rather inconspicuous pump switch. Unfortunately, he had also not bothered to read the thermometer that was attached to the polarimeter jacket, which now read 18.0°C rather than 25.0°C. The professor was momentarily stumped, for there were no available data on the temperature coefficient of rotation for this substance. At length she reasoned that the principal effect of temperature on the rotation would be through the temperature dependence of the density of the solution. Using density values for water at the two temperatures, the professor estimated

$$[\alpha]_D^{25} = +154.51 \times \frac{0.998625(@18°C)}{0.997075(@25°C)} = +154.75$$

''Wait a minute,'' said the student. ''If the polarimeter tube were at 18°C instead of 25°C, the solution density would be greater than the value used; thus the corrected specific rotation should be lower, not higher.'' With a red face the professor inverted the ratio and obtained +154.27. (''Scarcely significant,'' mumbled the professor.) The result was still more than a degree higher than the literature value.

One more thought occurred to her. Something her experience told her should be in the notebook was not there. She took the polarimeter tube, rinsed it thoroughly and filled it with distilled water, and placed it in the instrument (after turning the circulating pump on). Instead of a rotation reading of 0.00° she obtained a reading of +0.26°. The student had

failed to check the zero adjustment on the polarimeter scale! The average of 10 readings gave $+0.258°$ ($S_m = 0.006°$) Making the needed correction, the professor obtained $[\alpha]_D^{25} = +152.52$ deg dm^{-1} (g cm^{-3})$^{-1}$, which was now slightly *below* the literature value of 152.70.

Because of doubt about the weighing error and the temperature correction, it was necessary to repeat the determination; see pp. 59–60 for the result. The new value of $[\alpha]_D^{25}$ was $+152.61 \pm 0.23$ deg dm^{-1} (g cm^{-3})$^{-1}$. Evidently the optical purity of the specimen was quite high; the difference from the literature value of $+152.70$ is not significant at the 95 percent confidence level.

Final Caution. The Student t distribution is to be used when the numerical value to which it attaches is a mean of a definite number of direct observations, or is a numerical result calculated from such a mean by a procedure that introduces no uncertainties comparable in magnitude to the random errors of the direct observations. This is *not* the case for the specific rotation $[\alpha]_D^{25}$ discussed above. The uncertainty contribution to the final result that is due to random errors in the raw data on optical rotations α_i is *not* large compared to the contributions due to the *estimated* uncertainties in the other variables (particularly V and L) that are required for calculating the final result $[\alpha]_D^{25}$. The number of degrees of freedom applicable to the final result cannot be known precisely, or at least it would be an extremely difficult matter to determine. In addition, we must reiterate that the statistical treatment can mean little or nothing in the face of large systematic errors, except to indicate the *precision* of the measurement method.

FUNDAMENTAL LIMITATIONS ON INSTRUMENTAL PRECISION

In all experiments given in this book the precision of the measurements is governed by practical limitations of apparatus construction and operation; the precision is presumably capable of being increased by further refinement of the apparatus or the technique for using it, and it might well be imagined that there is no limit to the possible improvement. However, there are certain definite limitations imposed by physics. Although these limitations are not likely to be encountered by the beginning laboratory student, it is well to be aware of them.

LIMITATIONS DUE TO THERMAL AGITATION

The classical law of equipartition of energy states that every degree of freedom of a system has an average kinetic energy equal to $kT/2$, where k is Boltzmann's constant. The student is used to thinking of this amount of energy (about 2×10^{-21} J at room temperature) as being of significance only for individual atoms and molecules, or at most for microscopic particles undergoing Brownian motion. However, Brownian motion may be observed in systems as large as laboratory instruments. An example of a phenomenon analogous to this is observed in electrical circuits in the form of thermal noise (called Johnson noise). This may, for example, impose an important limitation on the sensitivity of a very sensitive voltmeter or electrometer. It constitutes a very severe limitation on the reception of extremely faint radio waves, especially in radio astronomy, where increasing the amplification is futile because the electrical noise is amplified to the same degree as the desired signal.

An electrical measuring instrument contains electrical circuits incorporating capacitance, inductance, and resistance. In the absence of resistance, a circuit tends to oscillate with a definite frequency f when disturbed. For optimum performance an amount of

resistance is incorporated that is barely sufficient to damp the oscillations resulting from transient inputs; the circuit is then said to be "critically damped." For a critically damped circuit it can be shown[13] that the root-mean-square (rms) fluctuations in voltage V and in current I are given by

$$V_{\text{rms}} \equiv (\overline{V^2})^{1/2} = (\pi kTRf)^{1/2} \tag{65}$$

$$I_{\text{rms}} \equiv (\overline{I^2})^{1/2} = (\pi kTf/R)^{1/2} \tag{66}$$

where R is the total resistance of the circuit, including damping resistance. If the units used in the above equations are volts, amperes, ohms, and hertz (cycles per second), k must be given in joules per kelvin (1.38×10^{-23}).

The rms values given above effectively constitute estimated standard deviations S for *individual* measurements of V or I. These limitations on precision can be overcome to some degree by making many measurements and taking the mean, provided the conditions of measurement do not change during the time required. If a voltage V is automatically and continuously recorded on a chart recorder, a straight line or a smooth curve can be drawn through the wiggly curve traced by the recorder pen. However, as in cases discussed earlier, the improvement in precision varies as the square root of the number of measurements and is therefore severely limited in practice. Since the fluctuations are thermal in origin, they can in principle, and also in practice in some cases, be reduced by reducing the temperature of the detection device.

LIMITATIONS DUE TO PARTICULATE NATURE

An important limitation that is sometimes encountered is due to the particulate nature of electricity (electrons, ions) and of radiation (photons). The measurement of radiation intensities is in certain cases (e.g., X rays) performed by counting particles or photons one at a time. The number N counted in a time interval of given magnitude is subject to statistical fluctuations; a count of N is subject to an estimated standard error given by

$$S = \sqrt{N} \tag{67}$$

The only way to overcome limitations imposed by the particulate nature of electricity and radiation is to scale up the experiment in physical size, time span, or intensity, all of which are subject to severe practical limitations. Doubling the source intensity in an X-ray experiment or doubling the counting time will (on the average) result in the number of counts being doubled, while the standard error is increased by a factor of $\sqrt{2}$ and the relative error is decreased in the same proportion. To improve the relative precision by a factor of 10 requires a scale-up factor of 100; thus, the square-root relationship imposes a severe limitation to improvement of precision.

LIMITATIONS DUE TO UNCERTAINTY PRINCIPLE

According to quantum mechanics, one cannot measure simultaneously both of two conjugate variables (e.g., position and momentum, or energy and time) to infinite precision; the product of the standard errors in the two variables must be greater than $h/2\pi$, where h is Planck's constant (6.64×10^{-34} J s). For instance, the energy of an excited state of an atom relative to the ground state, as determined by the wavelength of an emitted photon, is somewhat uncertain because the lifetime of the excited state is limited by collisions or other perturbations if not by the finite probability of spontaneous emission. The uncertainty principle applies to the results of any *one* experiment comprising simultaneous measurements of a pair of conjugate variables. Suppose that a single photon emitted by a system under study (say, a gas undergoing an electrical discharge) were to

pass between the slits of a very-high-resolution spectrometer, pass through the dispersing element (prism or grating), pass between the very narrow slits of a detector, and then evoke from a photomultiplier tube a response which is recorded as a single event. The wavelength λ of the photon is determined by the angular position of the detector slits with respect to the dispersing element, and the energy of the photon is therefore determined ($E = h\nu = hc/\lambda$, where c is the velocity of light). The conjugate variable t is known, say, from the pressure of the light-emitting gas, which determines the rms time between deactivating molecular collisions. That is just the one photon; many others from the same source possess slightly different wavelengths and therefore miss the detector slit opening. If the experiment were to stop right there, the determination of the wavelength that characterizes the spectral transition under study would be truly limited in precision due to the uncertainty principle.

However, that is not the way we would conduct the wavelength determination in practice. Let the detector slit scan slowly over a narrow range bracketing the position corresponding to the expected wavelength; photons will begin to enter the slit opening at a very low rate at first, and then at a gradually increasing rate which goes through a maximum and declines; if the counting rate is automatically recorded on a chart, a more-or-less Gaussian peak will be obtained. This constitutes in effect a large number (perhaps several thousand) of "individual experiments." The width of the peak is related to the estimated standard deviation S; the center of the peak can be measured with a precision corresponding to S_m. In any practical spectroscopic instrument, a peak of finite width is obtained even with spectral lines of essentially zero width, owing to the finite width of the detector slit and other geometrical factors; in fact,

$$S = (S_{\text{qm}}^2 + S_{\text{inst}}^2)^{1/2} \tag{68}$$

where S_{qm} characterizes the quantum mechanical estimated standard deviation and S_{inst} characterizes the instrumental estimated standard deviation. The same experiment can be done with a photographic plate as the detector. Let us suppose that each incoming photon exposes a single grain in the photographic emulsion; this is an "individual experiment." The resulting "line" on the plate is of finite width, and on microscopic examination is seen to be grainy. The density profile of the "line," which results from many thousands of "individual experiments," can be measured with a scanning microphotometer, and the width and center can be determined. Either with the scanning spectrophotometer or with the photographic spectrograph, the number of "individual experiments" is so large that the square-root relationship between precision and number of measurements almost never provides a severe limitation; in the experiments described above, the precision of the determination of the center of the spectral line is limited more by uncertainties in the geometric dimensions of the apparatus than by quantum mechanics.

Although it is somewhat hard to find real examples in which the precision of a physical chemistry measurement is limited materially and primarily by the uncertainty principle, this principle should be kept in mind as a potential limiting factor in future experiments.

SUMMARY

The principal equations in Part B of this chapter and key references to the text are cited here as a practical aid in the treatment of experimental data. This summary is not intended to be a substitute for thoughtful judgment. Use the equations with care; if there is any doubt concerning the applicability of an equation, review the associated text material.

Rejection of a Discordant Value with Q Test. For $3 \leq N \leq 10$, where N is the number of measurements of the same quantity, calculate

$$Q = \frac{|(\text{suspect value}) - (\text{value closest to it})|}{(\text{highest value}) - (\text{lowest value})} \tag{9}$$

and compare it with Q_c from Table 1 (p. 41). If $Q > Q_c$ then reject the suspect value; otherwise retain it. If $N > 10$, reject the suspect value if and only if it differs from the mean of the others by more than $2.6S_m$, where S_m is the estimated standard deviation of the mean of the others.

Mean, Variance, Estimated Standard Deviation *S*. For a data sample of N values of x, denoted x_i where $i = 1, 2, \ldots, N$, the unweighted mean is

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

The variance S^2 is given by

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{N}{N-1}(\overline{x^2} - \bar{x}^2) \tag{5}$$

The estimated standard deviation S is the square root of the variance:

$$S = \frac{1}{\sqrt{N-1}}\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^{1/2} \tag{7}$$

Estimated Standard Deviation of the Mean S_m; Confidence Limit for the Mean. The value of S_m is given by

$$S_m = \frac{S}{\sqrt{N}} = \frac{1}{\sqrt{N(N-1)}}\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^{1/2} \tag{8}$$

The 95 percent confidence limit of error Δ is

$$\Delta = t_{0.95}S_m = t_{0.95}\frac{S}{\sqrt{N}} \tag{34}$$

The value of $t_{0.95}$ is given in Table 3 (p. 49) under $P = 0.95$ for the applicable number of degrees of freedom

$$\nu = N - 1$$

Estimates Based on the Range. See Table 4 and Eq. (36) on p. 49.

Presentation of Results with Uncertainties. See Table 5, p. 50.

Significance Testing. Determine the difference D between the two values being compared: $\bar{x}$ and x_0 (where x_0 is exact) in Case 1 (p. 52), or $\bar{x}_1$ and $\bar{x}_2$ in Case 2 (p. 53). For Case 1, $S_D = S_m$ for the mean involved, and the number of degrees of freedom is $\nu = N - 1$. For Case 2, determine the *pooled* S_D from S_{m1} and S_{m2} using Eqs. (38) and (39) and use $\nu = N_1 + N_2 - 2$.

For the appropriate ν, find t in Table 3 (p. 49) under $P = 0.95$ (or other selected level of confidence) if the question being asked is "Is the difference D significant?" Find t under $P' = 0.95$ if the question being asked is "Is $\bar{x}$ (or $\bar{x}_1$) significantly *greater than* (alternatively, *less than*) x_0 (or $\bar{x}_2$)?" Calculate the quantity Δ_D, where

$$\Delta_D = tS_D \tag{37}$$

If $D > \Delta_D$, then the answer to the question is "Yes."

If one (or both) of the values being compared is not a mean or an exact value, and its uncertainty has been estimated on the basis of judgment and experience, calculate Δ_D with Eq. (42) or Eq. (43).

Propagation of Random Errors. Let x, y, z, . . . be independent, directly measured quantities; let F be a result calculated from them with an equation represented symbolically by

$$F = f(x, y, z, \ldots) \tag{44}$$

$$\Delta^2(F) = \left(\frac{\partial F}{\partial x}\right)^2 \Delta^2(x) + \left(\frac{\partial F}{\partial y}\right)^2 \Delta^2(y) + \left(\frac{\partial F}{\partial z}\right)^2 \Delta^2(z) + \ldots \tag{54}$$

where $\Delta(x)$ represents the 95 percent confidence limit for x, etc. In many cases, the expression for $\Delta(F)$ or $\Delta^2(F)$ can be greatly simplified. See pp. 57–59 for five common simple examples and for a discussion of how to handle complicated cases.

EXERCISES

1. For each of the following series of measurement values, determine whether any measurements should be rejected at the 90 percent confidence level. After rejecting values where appropriate, determine S, S_m, and Δ (95 percent confidence limit) for each series.

(*a*)	2.8	2.7	2.7	2.5	2.9	2.6	3.0	2.6	
(*b*)	97.13	97.10	97.20	97.35	97.10	97.19			
(*c*)	0.134	0.120	0.109	0.124	0.131	0.119	0.135	0.132	0.132

2. A chemist dispenses a titrant, 0.1000 *M* HCl, from a burette at 25.0°C. The initial reading is taken six times; the values (in milliliters) are

   6.79 6.78 6.79 6.77 6.76 6.78

   The final reading is also taken six times; the values are

   28.02 28.03 28.01 28.02 28.02 28.03

   Calculate the number n of moles of HCl in the solution dispensed. Give S and S_m for the initial and final volumes, and give a limit of error (95 percent confidence) for n.

3. A crystallographic society, in a study of systematic errors in X-ray determinations of unit cell sizes of crystals by the powder diffraction method (see Exp. 45), sends a specimen of a standard cubic crystalline material to each of several laboratories. Each is to determine the unit cell parameter a_0 by its own technique in six independent experiments at 25.0°C, and report the mean value together with the estimated standard deviation of the mean, S_m. The first two laboratories to report gave the following results:

   7.4132(18) Å 7.4065(20) Å

   Are the two results significantly different at the 90 percent confidence level? At 95 percent? At 97.5 percent? At 99 percent?

4. The heat of vaporization of a liquid may be obtained from the approximate integrated form of the Clausius-Clapeyron equation,

$$\Delta\tilde{H}_v = -R\frac{\ln(p_2/p_1)}{1/T_2 - 1/T_1} = R\frac{T_1T_2}{T_2 - T_1}\ln\frac{p_2}{p_1}$$

(See Exp. 13 for the exact form of the Clapeyron equation. The integrated form given here involves certain reasonable approximations.) For water, the vapor pressure is measured to be 9.2 Torr at 10.0°C (283.15 K) and 55.3 Torr at 40.0°C (313.15 K). Taking R to be 8.3145 J K^{-1} mol^{-1}, calculate $\Delta\tilde{H}_v$. Taking the limit of error (95 percent confidence) in a temperature measurement to be 0.1 K and that in a pressure measurement to be 0.1 Torr, calculate the uncertainty in $\Delta\tilde{H}_v$. A handbook gives 43,893 J mol^{-1} for $\Delta\tilde{H}_v$ at 25°C. Is the difference between this value and your value significant at the 95 percent confidence level?

5. A student determines the density of a solid with a pycnometer (see Sample Report, Chapter I) at 25.0°C. The weight of the empty, dry pycnometer is 6.2330 g. Some of the solid material is introduced; the weight of pycnometer plus solid is 39.4156 g. Water (density $\rho = 0.997048$ g cm^{-3} at 25.0°C) is added to fill all space not occupied by the solid; the weight of pycnometer plus solid plus water is 46.1357 g. The pycnometer is emptied and filled with water only; the weight is 19.7531 g. The student's estimated 95 percent confidence limit of error in any single weighing is 0.0020 g. Calculate the density and its 95 percent confidence limit of error, and present it and its limit of error in proper form with the correct number of significant figures and proper units.

REFERENCES

1. P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences,* 2d ed., pp. 42–49, 256–259, McGraw-Hill, New York (1992).
2. R. Annino and R. D. Driver, *Scientific and Engineering Applications with Personal Computers,* pp. 341–350, Wiley-Interscience, New York (1986).
3. *ASYST, a Scientific System,* Macmillan Software Company, New York (1985).
4. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes—The Art of Scientific Computing,* Cambridge University Press, New York (1988). [Available in C, Pascal, and FORTRAN versions. Paperback supplements, *Numerical Recipes Example Book,* accompanied by disks containing the programs given in the book in either C or Pascal, are also available.]
5. W. R. Hamming, *Numerical Methods for Scientists and Engineers,* 2d ed., pp. 567–573, 597–599, Dover, New York (1987).
6. D. A. Skoog, D. M. West, and F. J. Holler, *Fundamentals of Analytical Chemistry,* 6th ed., Saunders, Philadelphia (1992).
7. E. B. Wilson, Jr., *An Introduction to Scientific Research,* pp. 232–242, McGraw-Hill, New York (1952); Dover, New York (1990).
8. W. J. Dixon, *Ann. Math. Stat.* **22,** 68 (1951); R. B. Dean and W. J. Dixon, *Anal. Chem.* **23,** 636 (1951).
9. L. G. Parratt, *Probability and Experimental Errors in Science,* Wiley, New York (1961).
10. F. Reif, *Fundamentals of Statistical and Thermal Physics,* pp. 37–39, McGraw-Hill, New York (1965).

11. A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics,* p. 166, Dover, New York (1949).

12. *CRC Handbook of Probability and Statistics,* 2d ed., pp. 296, 306, CRC Press, Cleveland, Ohio (1976).

13. H. J. J. Braddick, *The Physics of the Experimental Method,* chap. IX, Chapman & Hall, London (1956).

GENERAL READING

D. C. Baird, *Experimentation: An Introduction to Measurement Theory and Experimental Design,* Prentice-Hall, Englewood Cliffs, N.J. (1988).

N. C. Barford, *Experimental Measurements: Precision, Error, and Truth,* 2d ed., Wiley, New York (1985).

P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences,* 2d ed., McGraw-Hill, New York (1992).

W. H. Beyer (ed.), *Standard Mathematical Tables,* 29th ed., CRC Press, Boca Raton, Fla. (1991).

E. O. Doebelin, *Measurement Systems,* 4th ed., McGraw-Hill, New York (1990).

W. R. Hamming, *Numerical Methods for Scientists and Engineers,* 2d ed., Dover, New York (1987).

R. G. Mortimer, *Mathematics for Physical Chemistry,* Macmillan, New York (1981).

E. Whittaker and G. Robinson, *The Calculus of Observations,* 4th ed., Blackie, Glasgow (1944).

E. B. Wilson, Jr., *An Introduction to Scientific Research,* McGraw-Hill, New York (1952); Dover, New York (1990).

EXPERIMENTS IN PHYSICAL CHEMISTRY

SIXTH EDITION

DAVID P. SHOEMAKER
The Late Emeritus Professor of Chemistry
Oregon State University

CARL W. GARLAND
Professor of Chemistry
Massachusetts Institute of Technology

JOSEPH W. NIBLER
Professor of Chemistry
Oregon State University

THE MCGRAW-HILL COMPANIES, INC.
New York St. Louis San Francisco Auckland Bogotá Caracas
Lisbon London Madrid Mexico City Milan Montreal New Delhi
San Juan Singapore Sydney Tokyo Toronto

McGraw-Hill
A Division of The McGraw-Hill Companies

Experiments in Physical Chemistry

Copyright © 1996, 1989, 1981, 1974, 1967, 1962 by The McGraw-Hill Companies, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOW DOW 9 0 9 8 7 6 5

ISBN 0-07-057074-4

This book was set in Times Roman by York Graphic Services, Inc.
The editors were Karen Allanson and Ty McConnell;
the production supervisors were Phil Galea and Annette Mayeski.
The designer was Edward Butler.
R. R. Donnelley & Sons Company was printer and binder.

Library of Congress Cataloging-in-Publication Data

Shoemaker, David P.
Experiments in physical chemistry / David P. Shoemaker, Carl W. Garland, Joseph W. Nibler.—6th ed.
p. cm.
ISBN 0-07-057074-4
1. Chemistry, Physical and theoretical—Laboratory manuals.
I. Garland, Carl W. II. Nibler, Joseph W. III. Title.
QD457.S56 1996
541.3′078—dc20 95-44102