1

Data Reporting and Analysis

1.1 Introduction

Science uses experimentation, data collection, and data analysis to become informed about the natural world. The data associated with *physical quantities* are expressed in terms of *units of measure*. The physical quantities and units of measure have related *dimensions*. The methods used for measuring these physical quantities can provide a certain level of accuracy, precision, robustness, linearity, and quantitation limit. Data should be reported with appropriate significant figures (SFs), which reflect the precision and robustness of the analytical method. Often data values need to be *rounded off* to reflect the correct number of SFs.

1.1.1 Fundamental/Base and Derived Physical Quantities

Physical quantities are known as *fundamental* or *derived* (Ade 2021). Fundamental quantities provide a standard of measure of independent physical parameters. Derived physical quantities are a function of fundamental quantities. There are seven fundamental physical quantities: amount of substance, like number of moles or particles, time, length, mass, temperature, electric current, and luminous intensity. The first five quantities are the most frequent fundamental physical quantities discussed in this book. Energy, heat, work, force, chemical potential, pressure, velocity, acceleration, momentum, frequency, area, volume, and density are derived quantities found throughout the text.

Fundamental units are standard units of measure associated with fundamental physical quantities. In pharmaceutical sciences, several unit systems are still used today. The globally recognized scientific standard is the Système International d'Unités (SI) system (SI Reference Point 2025). A second unit system is he IUPAC centimeter-gram-second (cgs) system (IUPAC 1993). The English foot-ounce-second system (Purdue University) is also still used in practice and found in textbooks. A brief discussion of the SI standard units used throughout this book follows.

Since 2019, SI has linked SI units to physical constants. This guarantees their stability and universality.

Mole (mol) is the SI unit defining the amount of a substance. One mole consists of exactly 6.022 140 76E23 *elementary entities*. (The *E* notation has been used by the *SHARE Operating System* since 1958. The *E* means "times ten raised to the power of *n* or $\times 10^n$.") In 2019, Avogadro's number was changed to a constant having units of elementary entities mol^{-1} or mol^{-1}. Avogadro's constant, N_A, is 6.022 140 76E23 mol^{-1}. The elementary entities may be any particle like electrons, atoms, molecules, ions, and others.

The SI unit of time is the *second* (s). One second is the time it takes for a cesium atom to vibrate 9.192 631 770E9 times.

Thermodynamics for Drug Product Design: A Guide for Pharmaceutical Scientists, First Edition. William Craig Stagner.

The meter (m) is the SI unit of length and is defined as the distance traveled by light in a vacuum in 1/2.997 924 58E8 of a second. (The speed of light is 2.997 924 58E8 ms^{-1}.)

The kilogram (kg) is now expressed in terms of the Planck constant (h), 6.626 070 15-34 kg m^2 s^{-1}, with meter and second as defined earlier. The older definition of kg relied on the mass of the kilogram.

The SI unit for temperature is the kelvin (K). In 2019, the definition was changed from fixing the temperature scale using the triple point of water to using the Boltzmann's constant (k) to give an *energy* equivalent. The Boltzmann constant is 1.380 649E-23 J K^{-1} or kg m^2 s^{-2} K^{-1}.

Table 1.1 provides a list of the fundamental physical quantities, SI unit names, SI unit symbols, and the fundamental quantity dimensions.

Table 1.2 lists selected derived quantities, some of which are identified by specific names. As noted earlier, derived physical quantities are a function of or composed of the fundamental quantities, which is evident from their units and dimensions.

Table 1.1 List of Fundamental Physical Quantities

Physical quantity	Name of SI[a] unit	SI unit symbol	Fundamental dimension
Amount of substance	Mole	mol	[n]
Time	Second	s	[t]
Length	Meter	m	[L]
Mass	Kilogram	kg	[M]
Temperature	Kelvin	K	[T]
Electric current	Ampere	A	[I]
Luminous intensity	Candela	cd	[I_v]

[a] SI – Système International.

Table 1.2 Selected Derived Physical Quantities

Physical quantity	Formula or name	SI[a] derived unit symbol(s)	Dimensional formula(s)
Area	Length × width	m^2	[L^2]
Volume	Length × width × height	m^3	[L^3]
Velocity	Distance per second	m s^{-1}	[Lt^{-1}]
Concentration	Amount per volume	mol m^{-3}	[nL^{-3}]
Density	Mass per volume	kg m^{-3}	[ML^{-3}]
Force[b]	Newton (N)	m kg s^{-2}	[LMt^{-2}]
Energy, heat, and work[c]	Joule (J)	N m = m^2 kg s^{-2}	[L^2Mt^{-2}]
Heat capacity	Joule per kelvin	N m K^{-1} = m^2 kg s^{-2} K^{-1}	[L^2Mt^{-2} K^{-1}]
Pressure	Pascal (Pa)	N m^{-2} = m^{-1} kg s^{-2}	[$L^{-1}Mt^{-2}$]
Surface tension	Newton per length	N m^{-1}	[Mt^{-2}]
Surface energy	Joule per length × length	J m^{-2}	[Mt^{-2}]
Gas constant[d]	Gas constant (R)	J K^{-1} mol^{-1} = m^2 kg s^{-2} K^{-1} mol^{-1}	[$L^2Mt^{-2}T^{-1}n^{-1}$]
Gas constant	Gas constant (R)	m^3 Pa K^{-1} mol^{-1} = m^2 kg s^{-2} K^{-1} mol^{-1}	[$L^2Mt^{-2}T^{-1}n^{-1}$]
Gas constant	Gas constant (R)	m^2 kg s^{-2} K^{-1} mol^{-1}	[$L^2Mt^{-2}T^{-1}n^{-1}$]

[a] SI – Système International.
[b] Force = mass × acceleration.
[c] Work = force × distance.
[d] The numerical value for each gas constant is 8.314 462 618 153 24.

1.1.2 Basic Mathematics and Statistics

Mathematics is used to describe or define the nature of our universe. Mathematics can also be viewed as the language of science. Thermodynamics and other areas of science require knowledge of mathematics, especially basic operations of arithmetic, algebra, and geometry; transcendental functions such as logarithms; and calculus to describe and interpret the physical nature of the universe.

Mathematics and statistics are routinely employed to analyze experimental data. Mathematical equations can show fundamental relationships between independent and dependent variables. Many important thermodynamic relationships are described by equations or formulas. Most of the thermodynamic data can be described by linear or exponential functions.

Basic statistics shed light on the central tendency of the data and its precision or variation around the central tendency.

In this chapter, important data reporting considerations such as quantity units, dimensional analysis (DA), significant figures (SFs), and rounding data values are reviewed. Key mathematical functions and statistical parameters that are used in reporting and analyzing experimental data are also reviewed.

1.2 Physical Quantity Dimensions and Dimensional Analysis

The nature of a physical quantity is described by its fundamental dimension or its dimensional formula. In terms of a derived physical quantity, a *dimensional formula is defined by the fundamental dimension(s) raised to an appropriate power necessary to fully describe the quantity of interest*. Tables 1.1 and 1.2 show the fundamental physical quantity dimensions and derived quantity dimensional formulas (Ade 2021a).

1.2.1 Application of Dimensional Analysis

Dimensional analysis (DA) is used throughout the book. DA of a physical quantity shows the association and interrelationships between or among two or more fundamental or derived physical quantities. DA can be used to check for the dimensional consistency of an equation and associated property values. Physical properties of interest can be reduced to their fundamental dimensions of amount, time, length, mass, temperature, electrical current, and luminous intensity by employing DA. DA plays a key role in converting units of one system to another. Dimensions can also be used to determine unknown relationships and equations by utilizing fundamental dimensions for known quantities. Likewise, when deriving a physicochemical equation, dimensional consistency requires that the units and dimensions on the left-hand and right-hand sides be the same.

1.2.2 Limitations of Dimensional Analysis

There are limitations to DA. DA does not apply to transcendental exponential, logarithmic, and trigonometric functions. If one side of an equation has a sum, DA cannot be applied. However, the units on each side of the equation should have the same units. That is, when physical quantities are added or subtracted, they must have the same units, but the units are not added or subtracted. Stating the obvious, DA is not applied to dimensionless numbers.

1.2.3 Dimensionless Numbers

Not all quantities have a dimension. These quantities are known as *dimensionless* quantities. They are also called a *scalar quantity* or *quantity of dimension one*. There are numerous quantities that are dimensionless. Notable dimensionless numbers are pi (π), Euler's number "e," and transcendental values such as the logarithm or trigonometric operations of a quantity. We will talk more about logarithmic functions later in this chapter. Many dimensionless numbers are obtained from ratios of quantities. Pi is a good example, which is the ratio of the circumference of any circle to its diameter.

1.2.4 Dimensional Nomenclature

There are several general rules regarding dimensional quantity formatting and nomenclature. Dimensions are placed in closed brackets as shown in Tables 1.1 and 1.2. Simplified or reduced dimensions are enclosed within closed brackets. For example, density is mass [M] per volume [L^3]. The reported dimension would be [ML^{-3}], not [M] [L^{-3}]. Exponential rules are used in dimensional nomenclature. [L^1] is [L] and [L^0] is 1.

1.2.5 Dimensional Quantity Algebra or Dimensional Algebra

A physical quantity value can be expressed as the *product* of a *numerical value* and a *unit*.

$$\text{Physical quantity} = \text{numerical value} \times \text{unit} \tag{1.1}$$

By convention, for convenience and efficiency, the "times sign" is not used. The commonly accepted formatting without the "times sign" is

$$\text{Temperature} = 298\ \text{K}. \tag{1.2}$$

To employ dimensional algebra, it is important to remember that a physical quantity is the *product* of a numerical value and a unit. From Eqs. (1.1) and (1.2), all three terms can be treated as algebraic quantities amenable to algebraic operations. As an example,

$$\text{Temperature}/\text{K} = 298 \tag{1.3}$$

Dimensional algebra is the operational means by which DA is performed. For purposes of this book, dimensional algebra and DA will be used interchangeably.

As stated earlier, dimensional algebra and analysis find several uses. It is used to convert a physical quantity from one unit system to another unit system. Dimensional algebra is also used to reduce units in an equation to confirm that the left-hand and right-hand sides of an equation are the same. This confirmation also denotes dimensional consistency.

In these cases, one uses *conversion factors* like 1 erg = 1E-7 J or 1 m = 1000 mm. Quantity algebra rearranges these conversion factors into quotients that are equal to 1. These quotients are called *unitary conversion quotients* or *conversion ratios*. The conversion ratio for the above mentioned conversion factor is 1 erg/1E-7 J or 1E-7 J/1 erg. Note, for convenience and ease of cancellation of dimensional terms, conversion ratios can also be displayed as 1 erg · 1E7 J^{-1} or 1E-7 J · 1 erg^{-1}.

Example 1.1 Conversion Ratios

a) Given the conversion factor 1 L = 1E3 mL; the conversion ratios are

$$1\ \mathrm{L}/1\mathrm{E3}\ \mathrm{mL} \text{ or } 1\ \mathrm{L}\cdot 1\mathrm{E\text{-}3}\ \mathrm{mL}^{-1} \tag{1.4}$$

and

$$1\mathrm{E3}\ \mathrm{mL}/1\ \mathrm{L} \text{ or } 1\mathrm{E3}\ \mathrm{mL}\cdot 1\ \mathrm{L}^{-1} \tag{1.5}$$

b) Given the conversion factor 1 m = 1E3 mm; the conversion ratios are

$$1\ \mathrm{m}/1\mathrm{E3}\ \mathrm{mm}, \text{ or } 1\ \mathrm{m}\cdot 1\mathrm{E\text{-}3}\ \mathrm{mm}^{-1} \tag{1.6}$$

and

$$1\mathrm{E3}\ \mathrm{mm}/1\ \mathrm{m} \text{ or } 1\mathrm{E3}\cdot 1\ \mathrm{m}^{-1} \tag{1.7}$$

Example 1.2 Converting a Physical Quantity from One Unit System to Another

Convert 600.0 ergs to joules using conversion ratio dimensional algebra.

$$600.0\ \cancel{\mathrm{ergs}}\left(1\mathrm{E\text{-}7}\ \mathrm{J}\cdot 1\ \cancel{\mathrm{erg}}^{-1}\right) = \underline{6.000\mathrm{E\text{-}5}\ \mathrm{J}} \tag{1.8}$$

Example 1.3 Using Dimensional Algebra to Confirm Both Sides of an Equation Have the Same Dimensionality

Using the ideal gas law PV = nRT and fundamental units, show that the left-hand and right-hand sides of the equation have the same units and dimensions.

Given from Table 1.2

P = Pa = $\mathrm{m}^{-1}\ \mathrm{kg}\ \mathrm{s}^{-2}$

V = m^3

n = mole

R = $\mathrm{m}^2\ \mathrm{kg}\ \mathrm{s}^{-2}\ \mathrm{K}^{-1}\ \mathrm{mol}^{-1}$

T = K

J = $\mathrm{m}^2\ \mathrm{kg}\ \mathrm{s}^{-2}$

1) *Left-hand side* of the ideal gas law equation. Substitute the fundamental units $\mathrm{m}^{-1}\ \mathrm{kg}\ \mathrm{s}^{-2}$ for P and m^3 for V. Apply dimensional algebra.

$$\left(\cancel{\mathrm{m}}^{-1}\ \mathrm{kg}\ \mathrm{s}^{-2}\right)\times \mathrm{m}^{3\cancel{2}} = \underline{\mathrm{m}^2\ \mathrm{kg}\ \mathrm{s}^{-2}} \tag{1.9}$$

2) *Right-hand side* of the ideal gas law equation, PV = nRT.

   a) Substitute the fundamental units for n and T. Moles and K are directly substituted for n and T.
   b) Substitute $\mathrm{m}^2\ \mathrm{kg}\ \mathrm{s}^{-2}\ \mathrm{K}^{-1}\ \mathrm{mol}^{-1}$ for R
   c) Perform dimensional algebra.

$$\cancel{\mathrm{mol}}\times\left(\mathrm{m}^2\ \mathrm{kg}\ \mathrm{s}^{-2}\ \cancel{\mathrm{K}}^{-1}\ \cancel{\mathrm{mol}}^{-1}\right)\times\cancel{\mathrm{K}} = \underline{\mathrm{m}^2\ \mathrm{kg}\ \mathrm{s}^{-2}} \tag{1.10}$$

Therefore, the left-hand (Eq. 1.9) and right-hand (Eq. 1.10) sides of the ideal law equation have the same dimensional and unit consistency.

The gas constant, *R,* can be expressed in units of energy per temperature per mole. It is the molar equivalent to the Boltzmann constant. The gas constant value changes depending on the "unit system" employed in making the experimental measurements. As mentioned earlier, the three most used unit systems are the SI, *cgs*, and English systems. Dimensional quantity algebra can be used to convert the gas constant value and its respective units from one unit system to another. Dimensional algebra can also be employed to convert the units from derived units to fundamental units for a particular unit system.

Example 1.4 Use Dimensional Algebra to Convert the Energy Gas Constant, R, from SI Derived Units to Fundamental SI Units
Given

$$R = 8.314\,J\,K^{-1}\,mol^{-1}$$

$$J = m^2\,kg\,s^{-2} \text{ and the conversion ratio is } \left(m^2\,kg\,s^{-2}\right)\cdot J^{-1}$$

To determine the fundamental SI energy gas constant units, use the conversion ratio for J and perform dimensional algebra.

$$\left(8.314\,\cancel{J}\,K^{-1}\,mol^{-1}\right)\cdot\left(m^2\,kg\,s^{-2}\right)\cdot\cancel{J}^{-1} = \underline{8.314\,m^2\,kg\,s^{-2}\,K^{-1}\,mol^{-1}}$$

This is the gas constant in fundamental SI units. It is noted that the "joules" cancel as a result of the quantitative algebra operation.

Example 1.5 Use Dimensional Algebra to Convert *cgs* Energy Gas Constant Units to SI Fundamental Gas Constant Units
Given

$$R = 8.314E7\,erg\,K^{-1}\,mol^{-1}$$

$$1\,erg = 1E-7\,J; \text{conversion ratio is } 1\,erg\cdot 1E+7\,J \text{ or } 1E-7\cdot 1\,erg^{-1}$$

$$J = m^2\,kg\,s^{-2}; \text{conversion ratio is } J\cdot\left(m^{-2}\,kg^{-1}\,s^2\right) \text{ or } \left(m^2\,kg\,s^{-2}\right)\cdot J^{-1}$$

Employing conversion ratios and dimensional algebra, the energy *cgs* gas constant can be converted to the SI energy gas constant fundamental units.

$$8.314\cancel{E7\,erg}\,K^{-1}\,mol^{-1}\times\left(1\cancel{E\text{-}7\,J\cdot erg}^{-1}\right)\left(kg\,m^2\,s^{-2}\right)\cdot\cancel{J}^{-1}) = \underline{8.314\,kg\,m^2\,s^{-2}}\ \underline{K^{-1}mol^{-1}},$$

in SI fundamental units.

The reader is referred to Connors and Mecozzi (2010) for additional discussion of dimensional quantitative algebra.

1.3 Data Reporting Using Decimal, Scientific, Normalized Scientific, and E Scientific Numerical Notation

There are several numerical notations that can be used to report and analyze data in base-10 logarithm, including decimal, scientific, normalized (normalized scientific), engineering, and E notations. This book will use decimal, normalized scientific, and E notation systems.

Decimal notation represents base-10 logarithm fractions or real numbers using digits 0 through 9 and plus and minus signs to represent positive and negative numbers, respectively. A decimal point may be used to indicate SFs for whole numbers or represent the fractional portion of a real number.

Scientific notation has been developed to make arithmetic operations easier and improve the convenience of writing large and small numeric values. Scientific notation takes the form:

$$s \times 10^n \tag{1.11}$$

where s is a nonzero real number called the *significand* and n is a positive or negative integer *exponent*. If the significand value is negative, a minus sign precedes s.

Normalized scientific notation employs scientific notation and, by convention, has one nonzero integer before a decimal point. This notation is useful for entering data into tables and to quickly view the magnitude differences of the quantity value.

$$s. \times 10^n \tag{1.12}$$

E scientific notation is an extension of normalized scientific notation, where E means "times ten raised to the power of n, or $\times 10^n$." In E scientific notation, Eq. (1.12) would be written as

$$\text{s.En} \tag{1.13}$$

E scientific notation minimizes keystrokes and provides easy readability. It is also the most compact notation for text and table data reporting.

Examples of these four numeric notation methods are summarized in Table 1.3.

Table 1.3 Comparison of Decimal, Scientific, Normalized Scientific, and E Scientific Numerical Notation

Decimal	Scientific	Normalized scientific	E scientific
0.0000000095123	0.95123×10^{-8}	9.5123×10^{-9}	9.5123E-9
5	5×10^0	5×10^0	5.E0
6471.998	6471.998×10^0	6.471998×10^3	6.471998E3
−10 122.3	$-10\,122.3 \times 10^0$	-1.01223×10^4	−1.01223E4
−2000[a]	-2×10^3	-2×10^3	−2.E3

[a] The SFs of this number are ambiguous. The numerical value may have one or as many as four SFs. A decimal point (2000.) indicates the physical quantity has four SFs and is unambiguous. In this example, the value is assumed to have one SF.

1.4 Accuracy

Accuracy of a measurement system is the difference between the true value or target value of a quantity and the measured value or mean of measured values. The FDA has set the accuracy acceptance criteria to be 97.0%–103.0% for a known "spiked" concentration (FDA 2019). Accuracy measurements within this acceptance range are considered to have high accuracy. The further a measured value is from the lower and upper ranges of this acceptance range, the lower the accuracy of the measurement. Most of this section's discussion is based on the United States Food and Drug Administration and USP 43-NF 38.

1.5 Precision as Measurement Variability and Relative Uncertainty

Precision deals with the closeness or difference of measured values independent of the measured accuracy values. Precision is related to the variability of the measured quantities. Precision is also related to the *repeatability* and *reproducibility or robustness* of the analytical method. Repeatability is the variation that occurs under rigorous experimental control, such as using the exact same instrument, with the same operator, in the same lab over a relatively short period of time (days or weeks). Reproducibility represents a more "real-life" scenario, where different instruments, but the same model is used, by different operators, in different labs, over a number of months.

There are two primary types of variability: systematic or determinant variability or error and random or nondeterminate variability or error. Both types of variability can be the result of the operator, analytical instruments, chemicals, environment, and more. Systematic variability results when all the causes of the variability remain constant throughout each measurement. Often poor calibration or instrumental setup can result in systematic variability. On the other hand, random variability can occur unpredictably with each measurement. This random variability or error fluctuates around the mean of the physical quantity.

Precision is commonly reported as *standard deviation (SD)* or *percent relative standard deviation (% RSD)* as an indication of measurement variability. Percent RSD is

$$\text{Percent RSD} = (\text{SD}/\text{mean}) \times 100\% \tag{1.14}$$

where SD is the measurement's standard deviation. For a normal distribution, three SDs account for 99.7% of the measured values. The FDA's precision acceptance criteria for repeatability is 2.0% RSD (FDA 2019).

1.6 Valid Measurements

A reported measurement is considered *valid* when it is both accurate and precise. This concept is illustrated in Figure 1.1. A valid measurement is depicted in the upper left quadrant. The other quadrants show possible scenarios that do not provide a valid physical quantity measurement. The lower left-hand quadrant depicts a set of measurements that are not accurate. This may be due to systematic error since all the data points lie outside the center target value but fall fairly close to each other. The upper right-hand corner shows reasonably high accuracy with physical values scattered around the center. This quadrant exemplifies low precision and relatively high accuracy. Finally, the lower right-hand corner displays poor accuracy and precision. Not one value is in the center target range, and the measurements are spread widely across the measurement space. These data exhibit poor accuracy and large random variability.

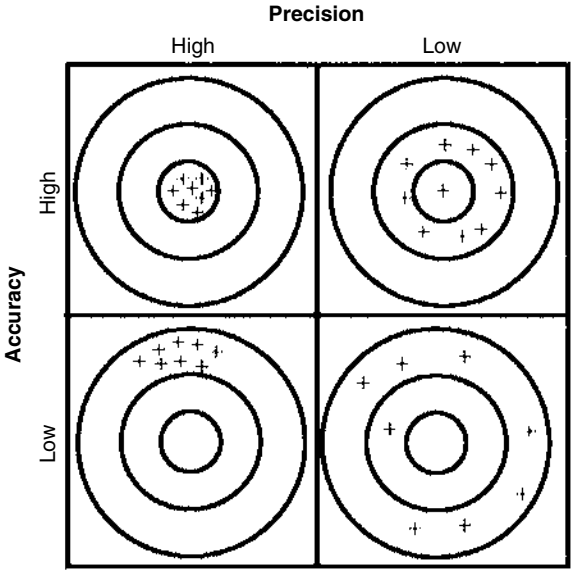

Figure 1.1 Possible accuracy and precision scenarios. The four quadrants represent targets, with the center being the target value or true value. The + sign indicates one of eight individual measured physical quantities.

1.7 Uncertainty

1.7.1 Uncertainty Associated with a Single Measurement

Physical quantity measurement carries a level of *uncertainty or error (σ) or SD*. This level of uncertainty is propagated throughout the calculations that use these measurements. There is uncertainty in the scale of the measuring device, such as a ruler or graduated cylinder. In this case, the uncertainty is equal to the smallest increment divided by 2. The uncertainty for a digital device is the smallest increment or digit position. To acknowledge this uncertainty, the general form of any measurement is written as

$$\text{Measurement or physical quantity value} = x_{best\ estimate\ of\ value} \pm \sigma_{uncertainty\ or\ error}$$

The level of uncertainty determines the number of SFs recorded for measurement estimates. SFs will be discussed in more detail in Section 1.8, Significant Figures (Digits) for Measured Values and Calculations.

There are several rules in reporting uncertainty, $\sigma_{uncertainty\ or\ error}$.

Rule 1- Stating Uncertainties: experimental uncertainties should be expressed to only *one* SF because the uncertainty is an estimate and carries no more significance than one digit.

Example 1.6 Uncertainty Reported to Only One SF

$$\text{Length} = 30.57 \pm 0.034\ \text{cm would be } \underline{30.57 \pm 0.03\ \text{cm}} \tag{1.15}$$

The uncertainty of 0.034 cm should be reported to one SF.

Rule 2- Stating *best estimate* measurements or values: the last SF in a reported quantity should have the same place or decimal place value as the uncertainty.

Example 1.7 Reporting the Last SF for a Reported Quantity

$$\text{Velocity} = 2345.67 \pm 200\ \text{cm s}^{-1} \text{ is correctly written } \underline{2300 \pm 200\ \text{cm s}^{-1}} \tag{1.16}$$

In this example, the uncertainty of the reported estimate value lies in the hundred's place. Therefore, the last SF of the reported estimate is at the hundred's place giving 2300 cm s^{-1}. Note that the uncertainty of 200 cm s^{-1} has one SF according to Rule 1.

1.7.2 Uncertainty Associated with Replicate Measurements

As discussed earlier, the FDA acceptance criteria for instrumental methods repeatability precision is 2.0% RSD. The SD and % RSD are statistical calculations that represent the random variability of the replicate sample measurements.

1.7.3 Propagation of Error of Combined Measurements and Calculations

An in-depth discussion of *the propagation of error* is beyond the scope of this book. It is important to be aware that there are propagation of error equations that properly combine uncertainties from different measurements. Propagation of error determination plays a major role in electronic device design. As an example, carbon resistors are widely used in all sorts of electronic circuits. Resistors manufactured to a "specified" resistance will have some variability or error in their "actual" resistance. Resistors connected in series or in parallel will exhibit different propagation of variability or error. In electronic device design, the overall resistance variation/error has a significant impact on the device's performance.

When a calculation requires more than one variable to provide a solution, propagation of errors is required to reflect the appropriate uncertainty. Table 1.4 provides propagation of error SDs for the various arithmetic operations.

In the case of designing resistors in series, the overall SD or error for the series of resistors is the square root of the sum of the squares of the individual resistor's SD or variance. This propagation of error can add significantly to the overall variability of the design circuitry. It is important to be aware that propagation of error occurs for all physical measurements and can have a significant impact on the reported value accuracy.

1.7.4 Author's Anecdote: A Priori Estimation of Dose Variation Using Propagation of Error Equations

It is extremely important that a patient receives the same dose of medication every time the dosage form is administered. My first industrial project was to improve the dose variability of a capsule that contained an active pharmaceutical ingredient (API) in the form of round beads. Inactive beads

Table 1.4 Propagation of Error Equations for Various Arithmetic Operations

Arithmetic operation	Example	Propagated Standard Deviation (SD or σ_x)
Addition or subtraction	$x = a + b - c$	$SD_x = (SD_a{}^2 + SD_b{}^2 + SD_c{}^2)^{0.5}$
Multiplication or division	$x = (ab)/c$	$SD_x = [(SD_a/a)^2 + (SD_b/b)^2 + (SD_c/c)^2]^{0.5}$

a, b, and c are the measured physical quantity values. SD_a, SD_b, and SD_c are the SDs of the physical quantity values (Propagation of Error. Chemistry Libretexts).

were used as a diluent. Dose uniformity is often measured as % RSD, where the SD is divided by the mean, and this quotient is multiplied by 100. Dose uniformity of 3.0% RSD is generally thought to be a relatively stringent measure of acceptable dose uniformity. This allows room for operator-to-operator, day-to-day, and plant-to-plant variation. The United States Pharmacopeia content uniformity limit is 7.8% RSD.

The content uniformity of the product I was assigned typically showed 6.5% RSD, with some batches closer to 5.5% and some batches closer to 7.5% RSD. From time to time batches would fail the USP content uniformity limit.

As I researched what had been done to improve the product content uniformity, I learned the high dose variability had been a chronic problem from the time the product was launched, and the FDA was scrutinizing it very closely. My research revealed that every aspect of the manufacturing process had been investigated, and all improvement attempts had failed. The analytical methods had also undergone extensive method revalidation and gave acceptable results.

I began to think that the propagation of the variabilities due to weight variation, blend inhomogeneity, and assay variability was the reason for the high observed % RSD content/dose uniformity values.

Capsule dose can be generically defined as

$$D = P\,W$$

where D (dose) is the drug amount/capsule, P is the proportion or ratio of drug amount/formulation blend weight and W is the formulation blend weight/capsule. P is dependent on blend homogeneity, which is related to mixing effectiveness, and assay accuracy and precision.

Using the propagation error equation in Table 1.4 for multiplied terms, the dose, SD_D is

$$SD_D = \left[(SD_P/P)^2 + (SD_W/W)^2\right]^{0.5} \tag{1.17}$$

where P and W have been defined earlier, SD_P is the drug formulation proportion SD, and SD_W is the formulation blend weight/capsule SD. SD_P is function blend heterogeneity and assay variability. This is also true for a capsule blend that is a mixture of active beads and placebo beads. The drug proportion standard deviation, SD_P, can be expressed as:

$$SD_P = \left[(SD_{Ph})^2 + (SD_{Pa})^2\right]^{0.5} \tag{1.18}$$

where Ph is the independent measurement of the bead mixture heterogeneity., SD_{Ph} is the bead mixture standard deviation, Pa is the independent measure of the assay content and SD_{Pa} is the assay content standard deviation. Substituting for SD_P from Eq. (1.18) into Eq. (1.17) gives :

$$SD_D = \left\{\left[\left[(SD_{Ph})^2 + (SD_{Pa})^2\right]^{0.5}/P\right]^2 + (SD_W)^2/W\right\}^{0.5} \tag{1.19}$$

Equation (1.19) shows that dose variability is propagated by blend heterogeneity, assay variability, and capsule fill weight variability.

Independent experiments were undertaken to determine SDs estimates. The binomial statistical theory was used to make an independent assessment of the SD_{Ph}. SD_{Pa} was determined from assay validation data. Independent formulation blend weight/capsule SD was determined by weighing beads emptied from individual capsules.

Using Eq. 1.19 and the independent SD measurements an *a priority* SD_D was calculated. It was shown that SD_{Ph} was responsible for 75.3% of the SD_D. 20.1% of the SD_D resulted from SD_W and the remaining 4.6% SD_D resulted from Pa variability (Stagner et al. 1991). From the propagation of error

analysis, the blend heterogeneity was determined to be the major cause of the relatively high dose variability. As mentioned earlier, numerous attempts to improve blend uniformity had been undertaken by making changes to blender design, using different types of blenders, and studying the effects of mixing time without satisfactory results.

Interestingly, the binomial distribution model can be applied to this problem. The binomial distribution can be used to determine the probability of a discrete event occurring or not. In this case, if you choose one bead, it will either be a drug bead or a placebo bead. The mean of the binomial distribution is $\mu = Np$, where N is the total number of beads in the capsule and p is the proportion of active beads. The binomial SD of the number of active beads is $SD = (Npq)^{0.5}$, where p is the proportion of active beads, q is the proportion of inactive beads $(1-p)$, and N is the total number of beads in the capsule. Employing the equation for the mean and SD shows that one can decrease the drug dose/content % RSD by increasing the total number of beads and increasing the proportion of drug beads (Al-Achi et al. 2023a).

The use of propagation of errors was instrumental in isolating the primary cause of the dose variability to the blend heterogeneity. The binomial distribution theory was used to independently determine the a priori blend heterogeneity. It was also used to resolve the dose content uniformity issue. The binominal distribution showed that blend variability was attributed to two factors, the total number of beads and proportion of active beads. The dose variability % RSD was improved to about 3% RSD from 6.5% RSD by increasing the total number of beads and increasing the proportion of active beads.

1.8 Significant Figures (Digits) for Measured Values and Calculations

As mentioned earlier, *the level of uncertainty determines the number of SFs in a recorded measurement*. The SFs reported for a measurement are equal to the number of digits that are known with a degree of confidence and are considered reliable. The last digit of an SF is an estimate or approximation. As such, 25.14 mg is reported to the 1/100 mg. The digit "4" is an estimate and has the most uncertainty but is still considered reliable. If the exact same amount of material were weighed on a balance capable of measuring the weight to 1/1000 mg, the quantity could be reported to 1/1000 mg, say 25.147 mg. In this case, the "7" would be an estimated measure.

Reporting the last reliable digit depends on the measurement device. For example, if the weight of the material is made on an electronic balance that has digits to the fourth decimal place, the fourth decimal place digit is considered an estimate and recorded as the last SF. This last decimal place is the most uncertain and is still considered reasonably reliable. In the case of reporting the volume using a graduated cylinder that is marked in units of 1 mL, depends on the scientist estimating the last SF to the nearest 0.1 mL.

It is important to note that the category of *exact numbers* is counted, and therefore, exact numbers do not have an uncertainty. For example, a pharmacist counts a specific number of tablets to be dispensed to a patient. Obviously, this example does not account for the possibility of "human error."

1.8.1 Rules for Determining Significant Figures for a Measured or Reported Value

Zeros serve as place holders and set the decimal point, like 0.1, which adds a degree of complexity to determining SFs. The rules to determine SFs are given as follows.

1) All nonzero numbers are significant.
2) A zero within a number is significant.
3) Zeros that do not "set" the decimal point are significant. 5.00 has three SFs.
4) Zeros that "set" the decimal point are not significant. 0.00057 and 0.57 only have two SFs.
5) Zeros that follow digits may or may not be significant. A decimal point notation or E scientific notation must be used to unambiguously indicate significance. For example, 3200 mg has ambiguous SFs compared to 3.200E3 mg (four SFs) or 3200. mg (four SFs).

1.8.2 Determination of Significant Figures for Addition and Subtraction Calculations

The reported value should have the same number of decimal places as the term with the least number of decimal places. To add numbers written in scientific or E scientific notation, the number needs to be written to the same power.

Example 1.8 Addition SFs Using Decimals

$$\begin{aligned} &135.210 \text{ mg} \\ &+\underline{10.2 \text{ mg}} \\ &145.410 \text{ mg} \end{aligned}$$

10.2 mg contains one decimal place, so the calculated value is reported as $\underline{145.4 \text{ mg or } 1.454\text{E}2 \text{ mg}}$ (1.20)

Example 1.9 Subtraction SFs Using E Scientific Notation

$$\begin{aligned} &1.35210\text{E}2 \text{ mg} \\ &-\underline{0.102\ \text{E}2 \text{ mg}} \\ &1.25010\text{E}2 \text{ mg} \end{aligned}$$

10.2 mg contains one decimal place, so the calculated value is reported as $\underline{1.250\text{E}2 \text{ mg,}}$ which has one decimal place. (1.21)

Note: Dimensional algebra is not performed for addition and subtraction. The units are not added or subtracted. In the case of Examples 1.8 and 1.9, "mg" units are attached to the calculated physical quantity regardless of the math operation.

1.8.3 Determination of Significant Figures for Multiplication and Division Calculations

The reported value should have the same number of SFs as the term with the least number of SFs.

Example 1.10 Multiplication SFs

$$21.2 \text{ cm} \times 705.06 \text{ cm} = 14,947.272 \text{ cm}^2$$

21.2 cm has the least number of SFs at three SFs. So the calculated value is reported as $\underline{14{,}900 \text{ cm}^2}$ $\underline{\text{or } 1.49\text{E}4 \text{ cm}^2}$. (1.22)

Reporting 14,900 leaves some ambiguity regarding the number of SFs this number has. Reporting the answer as "14,900." incorrectly implies that the number has five SFs. The E scientific notation makes it clear that the answer had three SFs.

Example 1.11 Division SFs

$$705.06\ \text{cm}/21.2\ \text{cm} = 33.257$$

$$\text{Since 21.2 cm has the fewest SFs at 3, the reported calculation is } \underline{33.3}. \tag{1.23}$$

Note: Dimensional algebra is performed for multiplication and division. In the case of multiplication, the units are multiplied as shown in Example 1.10. In the division Example 1.11, the units are divided, resulting in an answer having no reported units.

1.8.4 Logarithm Significant Figures

A logarithm is composed of a *characteristic* and *mantissa*. The characteristic part is an integer, and the mantissa is the decimal portion of the logarithm. The base-10 logarithm ($\log_{10}$) of 339 is 2.530. The "2" is the characteristic, and ".530" is the mantissa. The number of digits in the reported mantissa equals the number of SFs in 339, which is three SFs.

Example 1.12 Logarithm SFs (base-10)
The base-10 logarithm of 0.001237, which has four SFs is

$$\text{Log}_{10}\ 0.001237 = \underline{-2.9076} \tag{1.24}$$

The four digits in mantissa indicate that the original number had four SFs.

Example 1.13 Logarithm SFs (base-e)

$$\text{The base-e natural logarithm, } \ln e, \text{ of } 0.001237 \text{ is } \underline{-6.6951}. \tag{1.25}$$

The number of digits in the mantissa indicates that the logarithm has four SFs.

Note: Logarithms have no associated units because they are a transcendental value.

1.8.5 Antilogarithm Significant Figures

An antilogarithm has *the same number of SFs in the antilogarithm as there are digits in the mantissa of the logarithm value*.

Example 1.14 Antilogarithm SFs (Base-10)
The base-10 antilogarithm (antilog_{10}) of the $\log_{10}$ value −4.470 has three SFs. This is the number of digits −4.470 has in the mantissa.

$$\text{Antilog}_{10} - 4.470 = 10^{-4.470} = \underline{3.39\text{E-}5.} \tag{1.26}$$

The antilogarithm has three SFs, which are highlighted.

As stated earlier, the antilogarithm's SFs equals the number of digits in the mantissa of the log value.

Example 1.15 Antilogarithm SFs (base-e)
The base-e logarithm value −4.470 has three SFs. This is the number of digits it has in its mantissa.

$$\text{Antilog}_{e} - 4.470 = e^{-4.470} = \underline{0.0114.} \tag{1.27}$$

The antilogarithm's three SFs are highlighted. This equals the number of digits in the mantissa of the log value −4.470.

1.9 Rounding Numbers

Rounding should be done on the final calculation to avoid accumulating round-off errors that can occur if intermediate values are rounded. Likewise, in performing the calculations, it is acceptable to carry more SFs than are needed for the SFs in the final calculation.

The digit that is considered for rounding is the last SF. The rounding rules used in this book are as follows: the last significant digit of a number is maintained if the digit is followed by 0, 1, 2, 3, or 4; and the last SF is rounded up one unit if the digit is followed by 5, 6, 7, 8, and 9.

Example 1.16 Rounding 2.4488E4 to Four SFs

Consider the number 2.4488E4. It needs to be rounded to four SFs. In this example, the underlined 8 is the digit that will be considered for rounding since it is the last SF. Since "8" follows the last SF, the reported value would be rounded up one unit to 2.449E4. (1.28)

Example 1.17 Rounding 2.4484E4 to Four SFs

Consider the number 2.4484E4. It needs to be rounded to four SFs. In this instance, the underlined 8 is the digit that will be rounded. Since "4" follows the last SF, the "8" would be maintained, and the reported value would be 2.448E4. (1.29)

Example 1.18 Rounding 2.4485E4 to Four SFs

Consider the number 2.4485E4. It needs to be rounded to four SFs. In this case, the underlined 8 is the digit that will be rounded since it is the last SF. Since "5" follows the last SF, the "8" would be rounded up one unit, and the reported value would be 2.449E4. (1.30)

1.10 How Many Digits and Decimals to Use in Research Reports, Tables, and Presentations

Once precision, uncertainty, and SFs have been determined for a measured or calculated value, it would seem straightforward to know how many digits to report. Unfortunately, there are a numerous "rules" that have been suggested for how to best represent data in reports, tables, and presentations (Coghill and Garson 2006; Cole 2015). In general, it is important that the presented number fairly represents the uncertainty of the measurement technique, the data variability, and the significance of the numerical value. As Cole (2015) states, "rules that specify the number of decimal places at the expense of reporting significant figures or vice versa are often found to be

unsatisfactory." He suggests that reporting data has a degree of "art" and "common sense." Presenting a large number of digits can overwhelm the reader and the importance of the data lost.

Cole makes the following suggestions:

1) Keep the number of reported digits to no more than 4.
2) The final decimal place should represent the uncertainty in the measurement.
3) The SD and RSD final decimal place should be the same as the reported measured value.
4) The *p value* should be reported to three or four digits.
5) The coefficient of variation should be reported to three or four digits.
6) Scientific or E scientific notation should represent the correct number of SFs and decimal uncertainty while balancing suggestion point number 1.

Table 1.5 follows these suggestions.

Table 1.5 Acyclovir Activation Energy and Arrhenius Frequency Constant at Three pH Values

pH ± SD[a]	*Ea*[b] (Kcal mol^{-1})	*A*[c] (h^{-1})	*R*[b]
1.19 ± 0.02	29.7	40.9	0.939
1.69 ± 0.02	32.9	44.3	0.995
2.19 ± 0.02	31.2	40.6	1.000
Mean	31.3	41.9	
SD[a]	1.6	2.0	

Source: Begum et al. (2018)/with permission of Elsevier.
[a] SD – Standard deviation ($n = 3$).
[b] *Ea* – Arrhenius activation energy.
[c] *A* – Arrhenius frequency constant.

1.11 Exponents

Exponential terms are used extensively in the sciences and especially in thermodynamics. A brief review of exponents is provided with examples to prepare you for the following chapters. The exponent operational rules need to be memorized so they can easily be applied to the reporting and analysis of data.

1.11.1 Exponent Properties and Operations

Exponential properties and operational rules (Anton 1980; Stein 1977):

1. $x^0 = 1$ (1.31)
2. $x^1 = x$ (1.32)
3. $x^m \times x^n = x^{n+n}$ (1.33)
4. $x^{-m} = 1/x^m$ (1.34)
5. $x^m/x^n = x^{m-n}$ (1.35)
6. $(x^m)^2 = x^{2m}$ (1.36)

7. $(x^m)^n = x^{mn}$ (1.37)
8. $(xy)^m = x^m y^m$ (1.38)

In general, x and y can be numbers like 10 or Euler's number, e.

1.11.2 Examples of Scientific Exponent Phenomena

Example 1.19 Richter Scale
The Richter Scale (RS) value is a measure of earthquake power or magnitude. It is a dimensionless number. In this example, let us compare the magnitudes of two earthquakes registered on the RS 8 and 4. The values 8 and 4 represent the exponents of 10. The magnitude of RS values RS_8 and RS_4 are

$$RS_8 = 10^8 = 100,000,000 \tag{1.39}$$

$$RS_4 = 10^4 = 10,000 \tag{1.40}$$

To determine how much more powerful RS_8 is compared to RS_4, we divide RS_8 by RS_4 or $RS_8/RS_4 = 10^8/10^4$. From Eq. (1.35), this quotient equals 10^{8-4}, which equals 10^4 or 10,000. This means that the RS_8 earthquake was 10,000-fold more powerful than the RS_4 earthquake.

Example 1.20 Drug Degradation
It is common for a drug (D) to undergo degradation to form by-products (BP). This can be illustrated by the following degradation equation.

$$D \rightarrow BP \tag{1.41}$$

D is a drug and BP is the degradation by-product. This scheme represents a first-order process that can be described by an exponential function

$$D(t) = D_0 e^{-kt} \tag{1.42}$$

where D is the drug concentration at time t, D_0 is the initial drug concentration, e is the Euler's number, and $k_{(\text{temperature in Kelvin})}$ is the degradation rate constant in units of reciprocal time and a function of Kelvin temperature. The rate constant can be used to predict the drug concentration at any given time. Note the negative sign of the rate constant. This indicates an exponential loss of drug over time. Table 1.6 illustrates this exponential decay, where k at 298.15 K, is 1.05E-2 h^{-1}. The starting concentration is set at 100%.

Table 1.6 Percent *w/v* Drug Degradation Over Time

Time (h)	Drug concentration (% *w/v*)
1.00	98.96[a]
2.00	97.92
4.00	95.89
8.00	91.94
16.00	84.54
32.00	71.15

[a] $D(t) = (100\%\ w/v)e^{-0.0105\ (1/h)(1\ h)} = 98.96\%\ w/v$.

We may be interested in the percent of the drug that remains after 12 hours. This can be calculated by

$$D(t) = (100.\%\ w/v)e^{-0.0105\ (1.00/h)(12.00\ h)} = \underline{88.2\%\ w/v\ \text{drug}} \tag{1.43}$$

1.12 Logarithms

There is an interesting quote by an unknown author regarding the use of logarithms in thermodynamics. It states, "thermodynamics is research on everything for which the negative logarithm is linear with 1/T." As we will see in examples in this chapter and later chapters, this uncanny tongue-in-cheek quip does highlight the importance of becoming fluent in logarithmic properties and operations.

1.12.1 Logarithmic Properties and Operations

When handling logarithmic terms in an equation, the following logarithmic properties and mathematical rules apply (Anton 1980; Stein 1977):

$$1.\ \log\ 10 = 1 \quad \ln\ e = 1 \tag{1.44}$$

$$2.\ \log\ 1 = 0 \quad \ln\ 1 = 0 \tag{1.45}$$

$$3.1\log\ (xy) = \log\ x + \log\ y \quad \ln\ (xy) = \ln\ x + \ln\ y \tag{1.46}$$

$$4.\ \log(x/y) = \log\ x - \log\ y \quad \ln(x/y) = \ln\ x - \ln\ y \tag{1.47}$$

$$5.\ \log\ x^m = m\ \log\ x \quad \ln\ x^m = m\ \ln\ x \tag{1.48}$$

$$6.\ \log\ x = (\ln\ x)/2.303 \quad \ln\ x = 2.303\ \log\ x \tag{1.49}$$

Logarithmic operations can be used to linearize exponential functions. The two most common logarithmic bases used in math and science are the *base-10 logarithm* and the *base-e logarithm*, also known as a *natural logarithm*. The Euler's number e has a nonrepeating decimal with a value of 2.71828. The base-10 logarithm operation symbol is "log," which designates it as $\log_{10}$. The base-e logarithm operational symbol is "ln," which indicates $\log_e$.

Example 1.21 Derivation of a Conversion Factor Used to Convert a Base-e logarithm to a Base-10 Logarithm and Vice Versa

We can derive the conversion factor, 2.303, by solving Eq. (1.49) for the conversion factor c. That is, we will replace 2.302, with c in Eq. (1.49) to give Eq. (1.50).

$$\ln x = c \log x \tag{1.50}$$

Solving for c gives

$$c = \ln x / \log x \tag{1.51}$$

The conversion factor, c, can now be determined by setting x to any value. For convenience, let us set $x = 10$. Then,

$$c = \ln 10 / \log 10 = 2.303/1 \text{ or } 2.303 \tag{1.52}$$

Substituting 2.303 for c in Eq. (1.50) gives Eq. (1.49) in Rule 6.

1.12.2 Scientific Examples of Base-10 Logarithms (Log_{10} or Log)

Base-10 logarithm or $\log_{10}$ or log conveniently indicates a 10-fold magnitude difference or change at a glance. There are a number of examples in science that use the base-10 logarithm scale such as the RS, microbial growth enumeration, acoustical power, decibel scale, star brightness, and pH scale. An example of the pH scale is shown in the following text.

Example 1.22 pH Is Inversely Proportional to the Molar Hydrogen Ion Concentration.

$$\text{By definition } \mathbf{pH} = -\log[\mathbf{H}^+] \tag{1.53}$$

where $[H^+]$ is the molar concentration of the hydrogen ion expressed in moles/liter. The "p" indicates the mathematical operation of "taking" the negative log of a number. That is,

$$p\,(x) = -\log\,(\mathrm{x}) \tag{1.54}$$

where x is a number, and $-\log\,(x)$ represents "taking" the negative log of the number, x. From Eq. (1.48) (logarithmic Rule 5), Eq. (1.53) can be rewritten to give

$$\mathrm{pH} = \log[\mathrm{H}^+]^{-1} \tag{1.55}$$

Using Eq. (1.34), exponential Rule 4, Eq. (1.55) can now be rewritten as

$$\mathrm{pH} = \log 1/[\mathrm{H}^+] \tag{1.56}$$

which shows the important property of pH. Equation (1.56) shows that pH is inversely proportional to the molar hydrogen ion concentration. That is, as the $[H^+]$ increases, the pH decreases. This is consistent with the pH scale, which ranges from 0 to 14. A pH of 0 is very acidic, and pH 14 is very alkaline or basic. A pH of 7 is considered a neutral pH, where the $[H^+]$ and [OH] molar concentrations are equal.

Example 1.23 Derivation Showing Neutral pH Is 7

Neutral pH is associated with the autoionization of water shown as follows.

$$\mathrm{H_2O\,(l)} \rightleftharpoons \mathrm{H^+\,(aq)} + \mathrm{OH^-\,(aq)} \tag{1.57}$$

Equation (1.57) shows that water auto dissociates equally into a hydronium ion and a hydroxide ion. Let us assume that the molar concentration of both ions is 1×10^{-7} M. From Eq. (1.53), the pH at this hydronium ion concentration is

$$\mathrm{pH} = -\log\left[1 \times 10^{-7}\right] \tag{1.58}$$

The steps to determine the pH at this concentration are shown as follows.
Step 1: Employing Eq. (1.46), logarithmic Rule 3, Eq. (1.58) becomes

$$\mathrm{pH} = -\log 1 + -\log\left[10^{-7}\right] \tag{1.59}$$

Step 2: Assessing the first term on the right-hand side of Eq. (1.59)
From Eq. (1.45), logarithm Rule 2, the first term $-\log 1$ becomes

$$-\log 1 = 0 \tag{1.60}$$

Steps 3 to 7: Assessing Second Term of Eq. (1.59).
Step 3: Applying Eq. (1.48), logarithm Rule 5, the second term $-\log\,[10^{-7}]$ results in:

$$\log\left[10^{-7}\right]^{-1} \tag{1.61}$$

Step 4: Utilizing Eq. (1.36), exponential Rule 6, Eq. (1.61) gives

$$\log \left[10^{+7}\right] \tag{1.62}$$

Step 5: Using Eq. (1.48), logarithm Rule 5, Eq. (1.62) yields

$$7\,(\log 10) \tag{1.63}$$

Step 6: From Eq. (1.44), logarithmic Rule 1, the second term in Eq. (1.63) converts to

$$\log 10 = 1$$

Step 7: Substituting "1" into Eq. (1.63) for log 10 gives

$$7 \times 1 = 7 \tag{1.64}$$

Step 8: The calculated pH for a hydronium concentration of 1×10^{-7} (1E-7) M combines the values of the first term (Eq. 1.60) and the second term (Eq. 1.64) of Eq. (1.59) to give

$$\text{pH} = 0 + 7 = 7 \tag{1.65}$$

The pH scale ranges from 0 to 14. Zero to 6.99 pH is the acidic range and has an abundance of hydronium ion. pH 7.00 is a neutral pH and has equal molar concentrations of hydronium ion and hydroxide ion. pH 7.01–14.00 is designated the basic range which exhibits a higher portion of hydroxide ion. This derivation confirms that when the hydronium ion and hydroxide ion are in equal concentration, the acid and base neutralize each other giving a neutral pH of 7.

Similar calculations can be undertaken to calculate the pH at any other hydrogen ion molar concentration. Such calculations would confirm that pH is inversely related to hydrogen ion concentration (see Eq. 1.56, $\text{pH} = \log 1/[\text{H}^+]$).

1.12.3 Scientific Examples of Base-*e* Logarithms (Log_e or Ln)

There are many natural phenomena that can be described as a function of *e*. Examples are radioactive decay, heat sterilization rates, numerous thermodynamic concepts, and many others. The use of logarithms to linearize drug degradation rate and drug solubility expressions are provided as follows. Linearization of the drug degradation rate equation is an extension to Example 1.20.

Example 1.24 Linearization of the First-Order Drug Degradation Rate Expression
It is common for a drug (*D*) to undergo degradation to form by-products (*BP*). This can be illustrated by the following degradation equation.

$$D \rightarrow BP \tag{1.66}$$

where *D* is the drug and *BP* is the degradation by-product. This scheme represents a first-order process that can described by an exponential function:

$$D(t) = D_0 e^{-kt} \tag{1.67}$$

where *D* is the drug concentration at time *t*, D_0 is the initial drug concentration, *e* is the Euler's number, and *k* is the first-order degradation rate constant. The rate constant can be used to predict how much drug will be left at any given time. Equation (1.67) can be linearized by taking the natural logarithm of both sides.

$$\ln D(t) = \ln \left[D_0 e^{-kt}\right] \tag{1.68}$$

Exponential and logarithmic operations can now be used to derive a linear form of this expression.

Concentrating on the right-hand side of Eq. (1.68) and using Eq. (1.46) (Rule 3, logarithm of a "product term") gives

$$\ln D_0 + \ln e^{-kt} \tag{1.69}$$

Applying Eq. (1.48), Rule 5 to $\ln e^{-kt}$ gives

$$-kt \ \ln e \tag{1.70}$$

Employing Eq. (1.44) (Rule 1), $-kt \ln e$ becomes

$$-kt\,(1) \text{ where } \ln e = 1 \tag{1.71}$$

The right-hand side of Eq. (1.68), $\ln D(t) = \ln [D_0 e^{-kt}]$, produces

$$\ln D_0 - kt$$

Now, $\ln D(t) = \ln [D_0 e^{-kt}]$ has been converted into the linear form:

$$\ln D(t) = \ln D_0 - kt \tag{1.72}$$

where a plot of drug concentration versus time gives a straight line with a slope of $-k$, the degradation rate constant.

Example 1.25 Linearization of Exponential Ideal Drug Solubility Equation

Equation (1.73) shows that a crystalline drug's solubility is an exponential function of the drug's melting point, the temperature of the solvent, and the molar heat of fusion of the crystalline material.

$$x_2 = e^{\frac{-\Delta H_f (T_m - T)}{R \quad (T_m T)}} \tag{1.73}$$

Here x_2 is the mole fraction drug solubility, ΔH_f is the molar heat of fusion, R is the gas constant, T_m is the melting point of the drug in kelvin, and T is the temperature of the solvent/solution in kelvin. This solubility equation can be linearized by taking the natural logarithm of both sides and implementing the logarithmic rules and operations provided earlier.

Step1: Take the natural logarithm, ln, of both sides of the equation.

$$\ln x_2 = \ln e^{\frac{-\Delta H_f (T_m - T)}{R \quad (T_m T)}} \tag{1.74}$$

Step 2: Use Eq. (1.48), logarithmic Rule 5, on the right-hand side of the Eq. (1.74) to give

$$\ln x_2 = -\left(\Delta H_f / R\right)\left[(T_m - T)/(T_m T)\right] \ \ln e \tag{1.75}$$

Step 3: Use Eq. (1.44), logarithmic Rule 1, $\ln e = 1$, to achieve

$$\ln x_2 = -\left(\Delta H_f / R\right)\left[(T_m - T)/(T_m T)\right] \times 1 \tag{1.76}$$

Step 4: From Eq. (1.76), divide $(T_m - T)$ by (TmT) to get

$$\ln x_2 = -\left(\Delta H_f / R\right)(1/\mathrm{T} - 1/T_m) \tag{1.77}$$

Step 5: The linearized form of $x_2 = e^{\frac{-\Delta H_f (T_m - T)}{R \quad (T_m T)}}$ (Eq. 1.73) is obtained by multiplying $(1/T - 1/T_m)$ by $-(\Delta H_f/R)$ from Eq. 1.77 to give

$$\ln x_2 = -\left(\Delta H_f / R\right)/\mathrm{T} + \left(\Delta H_f / R\right)/T_m \tag{1.78}$$

This is the linear form of the ideal solubility law, Eq. (1.73). A plot of ln x_2 versus $1/T$ gives a straight line with a slope of $-(\Delta H_f/R)$. The intercept is $(\Delta H_f/R)/T_m$, where $\Delta H_f/T_m$ is ΔS_f, the entropy of fusion.

1.13 Differential and Partial Differential Equations

Differential equations involve *derivatives* of a function that tells how the function changes with respect to changes associated with the function variable or variables. In essence, differential equations are about quantifying change. Differential equations are used to explain how a quantity changes as an independent variable or variables change. For example, pharmaceutical scientists are keen to know how the solubility of a drug changes with temperature, how an aerosol's pressure changes with temperature, or how solubility changes with surface area.

Incremental and *infinitesimal* variable changes are distinguished by the magnitude change of the variables. The slope, m, of a straight line, is defined as the incremental change along the y-axis (Δy), divided by the incremental change along the x-axis (Δx). This slope function is written as

$$m = (y_2 - y_1)/(x_2 - y_1) = \Delta y/\Delta x \tag{1.79}$$

where $\Delta y/\Delta x$ is the slope of the line. Δy and Δx are the incremental changes in y and x, respectively.

As the change in y and x approaches zero, the slope function becomes

$$m = dy/dx \tag{1.80}$$

where dy and dx indicate infinitesimal changes in y and x.

To summarize, a derivative expresses the instantaneous rate change of a function with respect to a change in a function's variable. For example, the expression, $y = f(x)$ describes the relationship between a dependent variable y and an independent variable x. This expression can be stated several ways such as "y is a function of x" or "y depends on x." The derivative of $y = f(x)$ is denoted dy/dx or f′(x). As mentioned in the discussion of Eq. (1.80), the differentials dy and dx represent infinitesimal change in y and x. The differential equation form of a derivative can be integrated.

There are several ways to classify differential equations. Differential equations can be classified by *type* and *by order*. There are two types of differential equations: *ordinary* and *partial* differential equations. These will be discussed after the next section on derivative order.

1.13.1 Differential Order of Ordinary Differential Equations

The derivative order is the number of times a function has been differentiated. Similarly, for a differential equation, its order is determined by the highest order derivative of the equation. The order of a differential equation is defined by the highest order of differentiation present in the equation, not by the total count of derivatives or the number of derivative terms. The generalized differential equation order function is $d^n f(x)/dx^n$ or d $f^{(n)}(x)$, where n is the order of the ordinary differential.

For example, the first derivative of position with respect to time gives velocity. The magnitude of velocity is speed. The second derivative of position with respect to time gives acceleration which describes the how velocity/speed is changing. A second-order derivative is the derivative of the first derivative which represents the rate of change of the rate of change. In the case of speed, the second-order derivative of speed is acceleration given in units of distance per time squared. That is, we are interested in the change of speed with change in time.

1.13.2 First-Order Ordinary Differential Equations

An ordinary differential equation (ODE) is a mathematical equation that involves an unknown function of a single independent variable and one or more of its derivatives with respect to that variable. An example of a first-order ordinary differential equation is the average kinetic energy of an ideal monoatomic gas. In this case, $KE_{ave} = (3/2)kT$, where KE_{ave} is the average kinetic energy of a gas in J, k is the Boltzmann constant in J K^{-1}, and T is the temperature in kelvin. The average kinetic energy is only a function of temperature in kelvin. This relationship between the average kinetic energy function and its single-independent variable is indicated as $KE_{ave}(T)$. It is said that KE_{ave} is a function of T (temperature in Kelvin). The derivative of KE_{ave} can only be taken with respect to the function's single variable, T.

Like addition and subtraction, derivatives follow a set of operating rules. Table 1.7 lists some of the more frequently used first-order ordinary derivative operations and provides derivatives and their differentials. A **derivative** is the rate of change of a function (dy/dx), while a **differential** (dx,dy) represents an infinitesimal change in a variable. The derivative is the ratio of two differentials.

Table 1.7 First-Order Ordinary Derivatives and Their Corresonding Differentials[a]

Derivative	Corresponding Differential	Equation number
$dc/dx = 0$	$dc = 0\ dx = 0$	(1.81)
$dx/dx = 1$	$dx = 1\ dx = dx$	(1.82)
$d(cu)/dx = c\ du/dx$	$d(cu) = c\ (du/dx)\ (dx) = c\ du$	(1.83)
$du^n/dx = nu^{n-1}\ du/dx$	$du^n = (nu^{n-1}\ du/dx)\ dx = nu^{n-1}\ du$	(1.84)
$d\ (ln\ u)/dx = (1/u)\ du/dx$	$d(ln\ u) = [(1/u)\ du/dx]\ dx = (1/u)\ du$	(1.85)
$d(e^u)/dx = e^u\ du/dx$	$d(e^u) = (e^u\ du/dx)dx = e^u\ du$	(1.86)
$d(u + v - w)/dx = du/dx + dv/dx - dw/dx$	$d(u + v) = [(du + dv - dw)/dx]\ dx = du + dv - dw$	(1.87)
$d(uv)/dx = u\ dv/dx + v\ d/dx$	$d(uv) = [(u\ dv + u\ dv)/dx]\ dx = u\ dv + u\ du$	(1.88)
$d(uvw)/dx = uv\ dw/dx + vw\ du/dx + uw\ dv/dx$	$d(uvw) = [(uv\ dw + vw\ du + vw\ du)/dx]\ dx = uv\ dw + vw\ du + vw\ du$	(1.89)
$d(u/v)/dx = (v\ du/dx - u\ dv/dx)/v^2$	$d(u/v) = [(v\ du - u\ dv)/dx)/v^2]\ dx = du/v - u/v^2 dv$	(1.90)

Source: From CRC Standard Mathematical Tables.
[a] c and n are constants; u, v, x, and w are variables.

Example 1.26 Temperature Dependence of Monoatomic Gas Kinetic Energy
The average kinetic energy of a monoatomic gas is

$$KE_{ave}(T) = (3/2)kT \tag{1.91}$$

where the average kinetic energy, KE_{ave}, is a function of temperature in Kelvin, indicated by (T). The Boltzmann constant, k, and 3/2 are constants represented by c, in Eq. (1.83). Equation 1.82 and Eq. (1.83) can be used to take the first-order ordinary derivative of Eq. (1.91). The derivative of (1.91) with respect to T is

$$d[KE_{ave}(T)]/\mathrm{dT} = [3/2(k)]\ (\mathrm{d}T)/(\mathrm{d}T) = [3/2(k)] \tag{1.92}$$

1.13.3 First-Order Partial Derivatives

If a mathematical expression is a function of more than one independent variable it is called a multivariate function. The *partial derivative* delineates the rate of change of the multivariate function with respect to one variable while the other variables are held constant. The partial derivative allows one to understand what effect a single variable has on the change in the function value. In other words, the partial derivative examines the sensitivity of the function to changes in a single variable. The *partial differential* represents a *portion or a part* of the *total function differential value* that results from an infinitesimal change a single variable. Simply put, an individual partial derivative provides only a "partial" view of the function's behavior. The sum of all partial derivatives describes the overall change in the function. The ideal gas law is an example of a multivariate function. The ideal gas law is $PV = nRT$. Solving for P gives $P = (nRT)/V$, where P is the pressure in pascals, n is the number moles of gas, R is the gas constant in J mol^{-1} K^{-1}, T is in K, and V in L. In this case, the number of moles is constant and pressure is a function of temperature and volume. We might be interested in knowing how the pressure changes with a change in the temperature if the volume is held constant. We might also be interested in knowing how the pressure changes with a change in the volume when the temperature is held constant. Finally, we might be interested to know how the pressure changes as both the temperature and volume change. In this case, we gain a partial picture of the total differential by taking the derivative of P with respect to T at constant V. We gain another portion of the picture by taking the derivative of P with respect to V at constant T. To gain a full picture of the effect of temperature and volume changes on pressure, the partial derivatives are added to give the total differential equation.

The ∂ *is* the operator symbol for the partial derivative. This symbol alerts the reader that the function has more than one independent variable. It also denotes that the derivative is being taken with respect to one of the independent variables while the other variables are held constant. Let us say z is a function of w, x, and y. The shorthand version of this statement is $z\,(w,x,y)$. The total differential of z is the sum of the partial derivatives of z, where two of the three variables are held constant. The total differential of z is shown in Eq. (1.93).

$$dz\,(w,x,y). = (\partial z/\partial w)_{x,y} + (\partial z/\partial x)_{w,y} + (\partial z/\partial y)_{w,x} \tag{1.93}$$

where the subscripts indicate which independent variables that are held constant.

Example 1.27 Using the Ideal Gas Law, Determine the Differential Change in the Pressure Caused by Changes in the Temperature and Volume

The differential of P is the sum of the partial derivatives of the pressure with respect to T at constant volume and V at constant temperature.

$$P = (nRT)/V \text{ and } P\,(T,V) = nRT/V \tag{1.94}$$

where nR is a constant.

$$(\partial P/\partial T)_V = nR/V \quad (\partial T/\partial T) = nR/V \tag{1.95}$$

where $(\partial T/\partial T) = 1$ from Eq. (1.82).

$$(\partial P/\partial V)_T = nRT\left(\partial V^{-1}/\partial V\right) = nRT \quad \left(-V^{-2}\right) = -nRT/V^2 \tag{1.96}$$

where $(\partial V^{-1}/\partial V)$ is $-V^{-2}$ from Eq. (1.84).

The sum of the partial derivatives from Eqs. (1.95) and (1.96), gives the *total differential change in pressure* with changes in temperature and pressure.

$$dP = (\partial P/\partial T)_V + (\partial P/\partial V)_T = nR/V - nRT/V^2 \tag{1.97}$$

1.14 Integral Equations

Integrals deal with accumulation and net change, while differentials are associated with infinitesimal change. Integrals combine small changes and can be viewed as an antiderivative. Such thinking is like our discussion of logarithms and antilogarithms. The reverse process of determining a differential is finding the integral, which is called integration.

If the differential is *dx,* then the *indefinite integral* of *dx*, or the antiderivative of *dx* is

$$\int dx = x + c \tag{1.98}$$

where $\int$ *is* the *integration or integral* symbol, *dx* is part of the *integrand* 1 dx, *x* is the integration variable, and *c* is any constant called the integration constant. Indefinite integrals represent a family of functions that differ by a constant because the limits of the accumulation we are interested in are not defined. For *definite integrals,* the accumulation range is specified. In this case, the determined integral value is specific to the specified *accumulation range* or *low* and *high* integration limits.

Several common integration operation rules are shown in Table 1.8. More complex integrals can be found in books covering integral calculus or standard mathematical tables.

Table 1.8 Common Integration Rules for Indefinite Integrals

Differential function to be integrated	Antiderivative or Indefinite integral	Equation number
$\int a\ dx$	$ax + c$	(1.99)
$\int x^n\ dx$; except -1	$(x^{(n+1)})/(n+1) + c$	(1.100)
$\int x^{-1}\ dx$	Ln $\lvert x\rvert + c$	(1.101)
$\int (1/x)\ dx$	Ln $\lvert x\rvert + c$ (same as 1.101)	(1.102)
$\int e^x\ dx$	$e^x + c$	(1.103)
$\int a^x\ dx$	$a^x/ln\ (a) + c$	(1.104)
$\int \ln (x)\ dx$	$x \ln (x) - x + c$	(1.105)
$\int c\, f(x)\ dx$	$c \int f(x)\ dx + c$	(1.106)
$\int e^{ax}\ dx$	$e^{ax}/a + c$	(1.107)
$\int (f+g)\ dx$	$\int f(x)\ dx + \int g(x)\ dx + c$	(1.108)
$\int (f-g)\ dx$	$\int f(x)\ dx - \int g(x)\ dx + c$	(1.109)

Source: From CRC Standard Mathematical Tables.
Where *a* and *c* are constants. *f* and *g* are functions of *x*.

Example 1.28 Determining the Amount of Reversible Work Obtained during the Isothermal Expansion of an Ideal Gas

Isothermal reversible expansion of an ideal gas is known as $P\Delta V$ work at a constant temperature. $P\Delta V$ work will be discussed in more detail in Chapter 2. The reversible work is given by

$$w_{rev} = -\int_{v1}^{v2} P\,dV \tag{1.110}$$

where P is the ideal gas pressure in pascals and V is the gas volume in liters. Solving the ideal gas law (see Section 1.13.3) for P is $P = nRT/V$. Substituting the expression for P into Eq. (1.110) gives

$$w_{rev} = -\int_{V1}^{V2} \frac{nRT}{V}\,dV \tag{1.111}$$

where n, R, and T are constants, and Eq. (1.111) becomes

$$w_{rev} = -nRT\int_{V1}^{V2} dV/V \tag{1.112}$$

Integrating Eq. (1.112) using Eq. (1.102) from Table 1.8 and evaluating over the limits of V_1 and V_2 give the definite integral, Eq. (1.113).

$$w_{rev} = -nRT(\ln V_2 - \ln V_1) = -nRT \ \ln V_2/V_1 \tag{1.113}$$

From Eq. (1.47), logarithm Rule 4, $\ln V_2 - \ln V_1 = \ln V_2/V_1$. Equation (1.113) tells us that reversible work for an ideal gas is directly proportional to the difference in $\ln V_2 - \ln V_1$. That is, the larger the change in volume, the larger the reversible work.

1.15 Basic Descriptive Statistics

Statistics provides a way to organize and summarize data and provides mathematical techniques to analyze and reach decisions regarding the data. These two aspects of statistics are known as descriptive and inferential statistics.

There are four types of variable data: *nominal, ordinal, interval,* and *ratio*. Nominal data has no order or ranking but can be placed in a specific category like male/female, color, or left-handed/right-handed. Ordinal data exhibits order or hierarchy, but there is no requirement that the distance between each ranking or subcategory is the same. For example, taste can be ranked as "tastes bad" to "tastes great" without intermediate assessment values having equal distance. Ordinal data often uses Hedonic scales of 5–9 rankings or subcategories, which allow for statistical analysis. Interval data has an equal distance between each value but does not have a "true zero." In other words, the data values can be a negative number, as is the case with the Celsius temperature scale. Ratio data displays the equal distance between values and has a true zero, like the Kelvin temperature scale. Nominal and ordinal data variables are called *categorial variables*. Interval and ratio data variables are known as *continuous variables*.

Descriptive statistics summarize the *central tendency* and *dispersion or variation* of values about the central tendency for a set of data. There are several ways to describe the central tendency, which will be discussed further in this section. The *central tendency* is an attempt to determine the *best estimate* of the quantity or function being studied. *Dispersion or data value variation,* on the other hand, provides information about the *variation, uncertainty, and reliability* of the *best estimate*.

Table 1.9 Particulate Counts in Particles per Liter in a Hood on Different Days in Rank Order

Particle Counts per Liter
10
21
33
53
54

1.15.1 Central Tendency of Single Values of Sample Data

There are four main measures of central tendency: arithmetic mean, median, mode, and percentile. The arithmetic mean is the most reported measure of central tendency. The use of the term "*arithmetic mean*" distinguishes it from other means that are also computed, such as the geometric mean and harmonic mean. "*Mean*" will be used for the arithmetic mean for the remainder of this section and book. The mean is sensitive to extreme values in a data set. These extreme values will pull the mean value in their direction. The median, mode, and percentiles are less sensitive to extreme values.

Table 1.9 can be used to illustrate the calculation of mean, median, mode, and percentile of individual data.

Example 1.29 Calculation of Arithmetic Mean

The mean is obtained by adding all the sample values and dividing by the total number of values added.

$$\overline{x} = \frac{\sum x_i}{n} \tag{1.114}$$

where x-bar is the mean, x_i is the ith observation, and n is the total number of observations.

In this case, the mean is

$$(10 + 21 + 33 + 53 + 54)/5 = 34.2 \text{ particles per liter} \tag{1.115}$$

The number of SFs for this data set is 2, so the answer is rounded to two SFs.

The mean is 34 particles per liter

Example 1.30 Determination of the Median

The median of a data set is the value that divides the data set into two equal parts. The number of values equal to and greater than the median is equal to the number of values equal to or less than the median. If the number of values is *odd,* the median value is the middle number when the values are arranged in rank order. In the case, where there are an even number of values, the mean of the two middle observations is considered the median value. In this example, there is an odd number of values, so the third particle count/liter is the median value.

The median is 33 particles/liter.

Example 1.31 Determination of the Mode or Modes

The mode is the value that occurs most frequently. In Table 1.9, each number is different, so there is no mode. It is also possible for a set of data to have multiple modes. Say that the melting points of 10 lots of a compound from the same supplier reported in rank order are 210.4, 211.0, 211.0, 211.2, 211.7, 211.9, 212.1, 212.1, 212.2, and 212.5. Here, 211.0 and 212.1 occur twice in the data set, which has these two mode values.

Example 1.32 Determination of Percentile

Percentile is another measure of central tendency that is used to understand and interpret data. Percentile is the relative standing of the *k*th percentile that splits the data set into two pieces, one equal to and above the *k*th percentile and one below the *k*th percentile. The following steps are used to determine the *k*th percentile:

1) The values are ordered from smallest to largest.
2) For the *k*th percentile of interest, multiply *k*th percentile (say 20th percentile (0.2) times the number of data points. This number is called the *index number*. In this case, $0.2 \times 5 = 1$. The index value is used to locate the *k*th percentile. The index value notes the number of data points needed to be *counted* to determine the *k*th percentile of interest.
3) From Table 1.9, starting with the smallest number and using the index number of 1, the first number counted is 10, which is the 20th percentile of the five values reported. The number 33 would be the 60th percentile data point.
4) If the index number is not a whole number, the number is rounded up. The 50th percentile index value for Table 1.9 is $0.5 \times 5 = 2.5$. So, 2.5 is rounded up to 3. The third number in Table 1.9 is 33, which is the 50th percentile, where half the values are less than 33 and half the values are greater than 33.

Steps 3 and 4 illustrate the shortcomings of reporting percentile data for small data sets. In the case of Table 1.9, 33 counts per liter is considered the 60th percentile and the 50th percentile.

1.15.2 Dispersion of Single Values of Sample Data

Dispersion is also known as the scatter, variation, and spread of the data set. If all the data points in a set of observations are the same, then there is no dispersion, and the data measurement is highly precise (see Section 1.6 Valid Measurements). Data dispersion is small if the values are close together and greater when the observations are scattered widely. Several often-used statistical measures of dispersion are range, variance, SD, coefficient of variation, and % RSD of the mean or standard error.

Example 1.33 Range

The *range* is the difference between the largest and smallest value in a data set. The range of the data found in Table 1.9 is

$$54 \text{ counts/liter} - 10 \text{ counts/liter} = \underline{34 \text{ counts/liter}} \tag{1.116}$$

The range is the least informative of the dispersion measures. It only uses two data points of the data, which limits usefulness, especially for exceptionally large data sets of hundreds to tens of hundreds of observations.

Example 1.34 Variance

The *variance* measures the dispersion of value around the mean. Computationally, the variance is calculated by subtracting the mean for each observed value. This difference is squared to avoid having negative values. The squared differences are summed or added together and divided by the number of observations minus 1. The computational expression is

$$s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2/(n-1) \tag{1.117}$$

where s^2 is the sample variance, x_i is the ith observation, and $n-1$ is the number of degrees of freedom. Here, degrees of freedom allude to the idea that if we know the sum of the $n-1$ "observations minus mean differences," the nth difference value is constrained to be a set value. This is because the differences of all the values must equal zero. In this case, there is only 1 value that will give a "zero" sum of the mean differences. The variance for the values given in Table 1.9 leads to

$$s^2 = \left[(10 - 34.2)^2 + (21 - 34.2)^2 + (33 - 34.2)^2 + (53 - 34.2)^2 + (54 - 34.2)^2\right]/4 = 1506.8/4 = \underline{3.8\text{E}2 \text{ counts}^2 \text{ L}^{-2}} \quad (1.118)$$

where the mean of 34.2 counts L^{-1} for the sample data set was determined for Eq. (1.115) before rounding to 34 counts L^{-1}. The variance in this case has been rounded to 2 SFs.

Since the variance does not have the same units as the measured values, it is difficult to understand its physical significance except to say that the larger the variance, the larger the dispersion or variation around the mean. The variance is an important measure of dispersion for inferential statistics.

Example 1.35 Standard Deviation, *s* or SD

As discussed earlier, it is not meaningful to discuss the variance in terms of the original measurement units because the variance is given in terms of squared units. To measure dispersion in the original units, we take the square root of the variance, which is the *standard deviation*, s or SD, for the sample.

$$s = \text{SD} = \sqrt{s^2} \quad (1.119)$$

The SD of the data in Table 1.9 is 19.4 counts /liter L^{-1}. The SD is reported as 19. to have two SFs.

The SD is used in conjunction with the mean to describe the dispersion about the mean. For this example, the mean and standard deviation (SD) are noted as 34 ± 19 counts L^{-1}.

The SD is a meaningful measure of data dispersion. It has been shown that most large data sets exhibit a *normal distribution* or *bell-shaped curve* that is symmetrical about its mean. For normal distributions, one SD on both sides of the mean includes approximately 68% of the total area under the bell-shaped curve. Extending the boundaries on both sides of the mean by two SDs covers about 95% of the total area of the curve, and three SDs encompasses 99.7% of the total area.

1.16 Application: Linear Least Squares and Coefficient of Determination

Many natural processes are described by nonlinear functions. Rate processes like chemical degradation, microbial growth, and heat sterilization kill rates are described by nonlinear functions. Equilibrium processes such as solubility and complexation are also described by nonlinear functions. In the past, it was common to derive *linear* forms of these functions by transforming variables. This is done so the data could be plotted using x–y coordinates. Such linear forms and graphic representation of the data are often easier to interpret how specific variables affect the observed values compared to nonlinear data presentation. Today, with ready access to computers, which can easily plot nonlinear data and perform nonlinear regression analysis, the need to derive a linear form of nonlinear equations is not as important.

However, derived linear forms of nonlinear equations will be used throughout this book because the data represented in a linear form is often easier to interpret, and linear least-squares analysis is easy to understand and perform. The well-known linear equation form is

$$y = mx + b \tag{1.120}$$

where y is the dependent variable, x is the independent variable, m is the calculated slope of the line, and b is the calculated y-intercept of the line.

The purpose of linear least squares is to minimize the deviation sum of squares of the individual points from the line. The linear least-squares line is known as the *best-fit line*. The *coefficient of determination,* R^2, is used to measure the *goodness of fit* of the least-squares line and the data points. R^2 determines the strength of the linear association of x and y. R^2 values range from 0 to 1. A value of "0" indicates that the linear model does not explain any of the variation observed between the least-squares line and the data. A "1" denotes that the linear function perfectly explains the data and there is no variation between the least-squares line and the data points.

Figures 1.2 and 1.3 plot the mean of 5 time points (each time point has 144 values) as a zero order and first order dissolution process, respectively. The line in both figures is the least-squares, best-fit line. The coefficient of determination was determined and noted in the graphic.

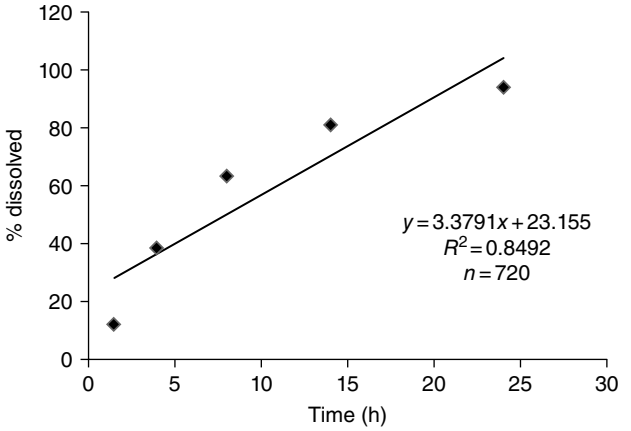

Figure 1.2 Zero-order dissolution model. *Source:* Al-Achi et al. (2023b).

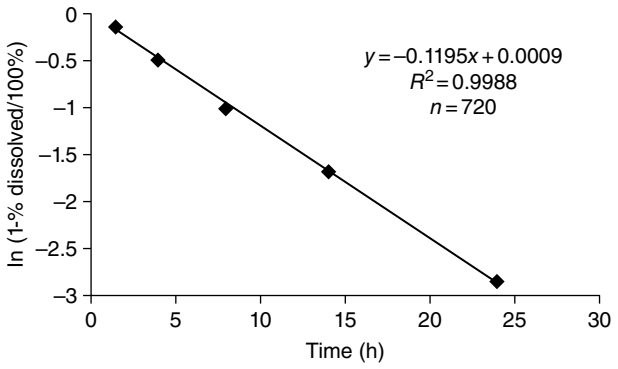

Figure 1.3 First-order dissolution model (fraction of drug remaining to be dissolved). *Source:* Al-Achi et al. (2023b).

Figure 1.2 clearly shows a significant deviation of the data points from the least-squares line. On careful inspection, the zero-order plot shows curvature in the data. The coefficient of determination of 0.8492 is consistent with the data variation seen about the best-fit line. On the other hand, Figure 1.3 exhibits minimal deviation between the data points and the least-squares line. The R^2 of 0.9988 denotes a strong linear association between x and y, indicating that the first-order linear function explains the data and the mechanism of the dissolution process.

References

Ade, S. (2021). *Dimensional Analysis for Dummies: A Simple Approach to Analyzing Dimensional Formulas and Equations Using Physical Quantities*. Made in the USA. Columbia, SC.

Al-Achi, A., Gupta, M.R., and Stagner, W.C. (2023a). *Integrated Pharmaceutics: Applied Preformulation, Product Design, and Regulatory Science*, 2e, 231–232. Hoboken, NJ: John Wiley & Sons.

Al-Achi, A., Gupta, M.R., and Stagner, W.C. (2023b). *Integrated Pharmaceutics: Applied Preformulation, Product Design, and Regulatory Science*, 2e, 467. Hoboken, NJ: John Wiley & Sons.

Anton, W.A. (1980). *Calculus*. New York: Wiley.

Begum, B.C., Johnson, P.R., and Stagner, W.C. (2018). Acyclovir chemical kinetics with the discovery and identification of newly reported degradants and degradation pathways involving formaldehyde as a degradant and reactant intermediate. *Int. J. Pharm.* 535: 172–179.

Coghill, A.M. and Garson, L.R. (ed.) (2006). *The ACS Style Guide: Effective Communications of Scientific Information*, 3e. New York: Oxford University Press.

Cole, T.J. (2015). Too many digits: the presentation of numerical data. *Arch. Dis. Child.* 100: 608–609.

Connors, K.A. and Mecozzi, S. (2010). *Thermodynamics of Pharmaceutical Systems: An Introduction to Theory and Applications*, 2e. Hoboken, NJ: John Wiley & Sons.

FDA (2019). Food and Drug Administration Office Regulatory Affairs Laboratory Manual Volume III Section 4 https://www.fda.gov/media/73535/download (accessed June 2021).

IUPAC (1993). *Mills I, Homann K, Cvitaš T, Kuchitsu K, Quantities, Units, and Symbols in Physical Chemistry*, 2e. Oxford: Blackwell.

IUPAC (2019). https://iupac.org/100/stories/the-new-si/ (accessed May 2021).

NIST (National Institute of Standards and Technology) (2019). SI Redefinition. https://www.nist.gov/si-redefinition (accessed June 2021).

Propagation of Error. Chemistry Libretextx. https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Supplemental_Modules_(Analytical_Chemistry)/Quantifying_Nature/Significant_Digits/Propagation_of_Error. (assessed Jan 2022).

Purdue University. http://chemed.chem.purdue.edu/genchem/topicreview/bp/ch1/units.html (accessed May 2021).

Selby, M. and Girling, B. (ed.) (1965a). *CRC Standard Mathematical Tables*, 303–304. Clevland: Chemical Rubber Company.

Selby, M. and Girling, B. (ed.) (1965b). *CRC Standard Mathematical Tables*, 308. Clevland: Chemical Rubber Company.

SHARE Operating System. https://en.wikipedia.org/wiki/SHARE_Operating_System (accessed April 2023).

SI Reference Point. Version 1.0 Beta. Bureau International des Poids et Mesures. https://si-digital-framework.org/SI/units?lang=en. 2025.

Stagner, W.C., Cerimele, B.J., and Fricke, G.H. (1991). Content uniformity: Separation and quantitation of sources of dose variation. *Drug Dev. Ind. Pharm.* 17 (2): 233–244.

Stein, K.S. (1977). *Calculus and Analytical Geometry*, 2e. New York: McGraw-Hill.
US Metric Association (2019). https://usma.org/international-system-of-units-si-reference-publications#locale-notification (accessed May 2021).
USP-NF. Rounding: https://www.uspnf.com/sites/default/files/usp_pdf/EN/USPNF/generalNoticesandRequirementsFinal.pdf. Significant Figures: http://ftp.uspbpep.com/v29240/usp29nf24s0_uspgn109.html (accessed April 2023).

Chapter 1 Problem Set

1.1 List the seven fundamental physical quantities.

1.2 What are the fundamental dimensions associated with the fundamental physical quantities?

1.3 What are the formulas for the derived physical quantities: area, volume, velocity, concentration, and density?

1.4 Define dimensional analysis and list three examples where dimensional analysis is useful.

1.5 Explain dimensional quantity algebra.

1.6 Explain conversion ratio. What is another name for conversion ratio?

1.7 Using dimensional analysis, determine the units of the drug complexation equilibrium constant, K_{11}, for the following equation when S_t, S_o, and L_t have units of moles/liter

$$S_t = S_o + (K_{11}S_oL_t)/(1 + K_{11}S_o)$$

S_t is the total molar concentration of the drug after complexation, S_o is the solubility of the drug before a complexing agent (ligand) is added, L_t is the total ligand molar concentration added to the drug liquid, and K_{11} is the 1 : 1 drug: ligand equilibrium constant. Note: to be dimensionally correct each side of the equation needs to have the same units and dimensionality.

1.8 Write the following numbers in E scientific notation

$$-200.12, 0.00347, \text{and } 69201748.$$

1.9 Discuss accuracy and precision and how they relate to a valid measurement.

1.10 How does uncertainty or measurement error relate to how data is reported?

1.11 How many significant figures are in
a) 109.0
b) 0.0190
c) 1090

1.12 Report the correct number of significant figures.
a) $3.40 * 208.0$
b) Antilogarithm of 10.800

1.13 Report to the correct number of significant figures.

$$21.1 + 2.037 + 6.13 =$$

1.14 What is the rounded number?
If the final number is reported to three significant figures, and the final number was 89.87, then this value would be rounded off to _____.
If the final number is reported to three significant figures, 89.83 is round off to _____.

1.15 Using exponent rules, calculate the buffer capacity for a buffer having a total buffer concentration of 1 M, $[H_3O^+] = 1.0 \times 10^{-4}$ M, and $K_a = 1.47 \times 10^{-4}$, using the general buffer capacity equation: buffer capacity $= 2.303C\dfrac{K_a[H_3O^+]}{\{K_a + [H_3O^+]\}^2}$

1.16 Convert the nature log form of ln(concentration) $= 1.85 - 0.45 \times$ time (h) to its base-10 logarithm form.

1.17 What is the derivative of the zero-order reaction, $[A] = [A_0] - kt$?

1.18 The diffusion coefficient (D) is a function of temperature (T), solution viscosity (η), and particle (r). A is Avogadro's number, and R is the gas constant:

$$D = \frac{RT}{\pi\eta rA}$$

When all variables change simultaneously, what is the equation for the diffusion coefficient? When the temperature is held constant what is the equation for the diffusion coefficient?

1.19 What is the definite integral of zero-order rate expression $\dfrac{dA}{dt} = -k_0A_0$?

1.20 Find the range, mean, median, mode, standard deviation, and % relative standard deviation for this data set.

90, 76, 53, 78, 88, 80, 81, 91, 99, 68, 62, 78, 67, 82, 88, 89, 78, 72, 77, 96, 93, 88, 88