

1

Introduction

Dimensions and Units
Some Elements of Mathematics

Statistical Methods and the Analysis of Errors

The pharmacist today more than ever before is called upon to demonstrate a sound knowledge of pharmacology, organic chemistry, and biochemistry and an intelligent understanding of the physical and chemical properties of the new medicinal products that he or she prepares and dispenses.

Whether engaged in research, teaching, manufacturing, community pharmacy, or any of the allied branches of the profession, the pharmacist must recognize the need to borrow heavily from the basic sciences. This stems from the fact that pharmacy is an applied science, composed of principles and methods that have been culled from other disciplines. The pharmacist engaged in advanced studies must work at the boundary between the various sciences and must keep abreast of advances in the physical, chemical, and biologic fields to understand and contribute to the rapid developments in his own profession.

Pharmacy, like many other applied sciences, has passed through a descriptive and an empiric era and is now entering the quantitative and theoretic stage.

The scientific principles of pharmacy are not as complex as some would believe, and certainly they are not beyond the understanding of the well-educated pharmacist of today. In the following pages, the reader will be directed through fundamental theory and experimental findings to practical conclusions in a manner that should be followed easily by the average pharmacy student.

The name *physical pharmacy* has been associated with the area of pharmacy that deals with the quantitative and theoretic principles of science as they apply to the practice of pharmacy. Physical pharmacy attempts to integrate the factual knowledge of pharmacy through the development of broad principles of its own, and it aids the pharmacist, the pharmacologist, and the pharmaceutical chemist in their attempt to predict the solubility, stability, compatibility, and biologic action of drug products. As a result of this knowledge, the

pharmaceutical scientist is in a better position to develop new drugs and dosage forms and to improve upon the various modes of administration.

This course should mark the turning point in the study pattern of the student, for in the latter part of the pharmacy curriculum, emphasis is placed upon the application of scientific principles to practical professional problems. Although facts must be the foundation upon which any body of knowledge is built, the rote memorization of disjointed "particles" of knowledge does not lead to logical and systematic thought. The student should strive in this course to integrate facts and ideas into a meaningful whole. In the pharmacist's career, he or she frequently will call upon these generalizations to solve practical pharmaceutical problems.

The comprehension of course material is primarily the responsibility of the student. The teacher can guide and direct, explain and clarify, but facility in solving problems in the classroom and the laboratory depends largely on the student's understanding of theory, recall of facts, ability to integrate knowledge, and willingness to devote sufficient time and effort to the task. Each assignment should be read and outlined, and assigned problems should be solved outside the classroom. The teacher's comments then will serve to clarify questionable points and aid the student to improve his or her judgment and reasoning abilities.

DIMENSIONS AND UNITS

The properties of matter are usually expressed by the use of three fundamental dimensions: length, mass, and time. Each of these properties is assigned a definite *unit* and a *reference standard*. In the metric system, the units are the centimeter (cm), the gram (g), and the second (sec); accordingly, it is often called the *cgs* system. A reference standard is a fundamental unit

TABLE 1-1. Fundamental Dimensions and Units

Dimension (Measurable Quantity)	Dimensional Symbol	CGS Unit	SI Unit	Reference Standard
Length (<i>l</i>)	<i>L</i>	Centimeter (cm)	Meter (m)	Meter
Mass (<i>m</i>)	<i>M</i>	Gram (g)	Kilogram (kg)	Kilogram
Time (<i>t</i>)	<i>T</i>	Second (sec)	Second (s)	Atomic frequency of Cesium 133

relating each measurable quantity to some natural or artificial constant in the universe.

Measurable quantities or dimensions such as area, density, pressure, and energy are compounded from the three fundamental dimensions just referred to. In carrying out the operation of measurement, we assign to each property a dimension that is expressed quantitatively in units. Thus the quantities of length, area, and volume are measured in the dimension of length (*L*), length squared (*L*²), and length cubed (*L*³), respectively corresponding to the unit of cm, cm², and cm³ in the cgs system. The fundamental dimensions, units, and reference standards are given in Table 1-1.

The International Union of Pure and Applied Chemistry (IUPAC) has introduced a *Système International* or *SI units* in an attempt to establish an internationally uniform set of units. *Physical Pharmacy* generally uses the cgs or common system of units. However, since SI units appear with increasing frequency in research articles and are found in some textbooks, they are introduced to the student in this chapter. They are also used in Chapter 4 and to some extent elsewhere in the book. SI units are listed in Tables 1-1 and 1-2, and some appear inside the front cover of the book under *Physical Constants*.

Length and Area. The dimension of length serves as a measure of distance and has as its reference standard the *meter*. It is defined as follows:

$$1 \text{ meter} = 1.65076373 \times 10^6 \lambda_{\text{Kr-86}}$$

in which $\lambda_{\text{Kr-86}} = 6.0578021 \times 10^{-7} \text{ m}$ is the wavelength in vacuo of the transition between two specific energy levels of the krypton-86 atom. Prior to this definition, the meter was arbitrarily defined as the distance between two lines on a platinum-iridium bar preserved at the International Bureau of Weights and Measures in Sèvres, France. The unit of length, the centimeter, is

one hundredth of a meter, the common dimensions and multiples of which are found in Table 1-2. In the microscopic range, lengths are expressed as micrometers (μm), nanometers (nm), and angstroms, \AA , sometimes written \AA . Units are often multiplied by positive and negative powers of 10 to indicate their magnitude, the micrometer being 1×10^{-3} millimeters or 10^{-4} cm, the nanometer $0.001 \mu\text{m}$, and the angstrom 0.1 nm or 10^{-8} cm. Although the micrometer (μm) is the preferred term for 0.001 mm in modern textbooks on colloid chemistry, the practice is sometimes to use the older and more familiar term, micron (μ). Similarly, the nanometer has replaced the millimicron ($\text{m}\mu$). The student should be familiar with the prefixes (see Table 1-2) accompanying units such as mass, volume, and time. For example, a nanosecond, or ns, is 10^{-9} second; a megaton (Mton) is 10^6 tons. Area is the square of a length and has the unit of square centimeters (sq. cm or cm^2).

Volume. The measurable quantity, volume, is also derived from length. Its reference standard is the *cubic meter*; its cgs unit is one millionth of this value or 1 cubic centimeter (cc or cm^3). Volume was originally defined in terms of the *liter*, the volume of a kilogram of water at 1 atmosphere pressure and 4° C , and was meant to be equivalent to 1000 cm^3 . Owing to the failure to correct for the dissolved air in the water, however, the two units do not compare exactly. It has since been established that 1 liter actually equals 1000.027 cm^3 . Thus, there is a discrepancy between the milliliter (one thousandth of a liter) and the cubic centimeter, but it is so slight as to be disregarded in general chemical and pharmaceutical practice. Volumes are usually expressed in milliliters in this book, abbreviated ml or mL, in conformity with the U.S. Pharmacopeia and the National Formulary; however, cubic centimeters are used in the book where this notation seems more appropriate.

The pharmacist uses cylindrical and conical graduates, droppers, pipettes, and burettes for the measurement of volume; graduates are used more frequently than the other measuring apparatus in the pharmacy laboratory. The flared conical graduate is less accurate than the cylindrical type, and the use of the flared graduate should be discouraged except for some liquids that need not be measured accurately. The selection of the correct graduate for the volume of liquid to be measured has been determined by Goldstein et al.¹

TABLE 1-2. Fractions and Multiples of Units

Multiple	Prefix	Symbol
10^{12}	tera	T
10^9	giga	G
10^6	mega	M
10^3	kilo	k
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p

Mass. The standard of mass is the kilogram. It is the mass of a platinum-iridium block preserved at the Bureau of Weights and Measures. The practical unit of mass in the cgs system is the gram (g), which is one thousandth of a kilogram. Mass is often expressed as the weight of a body. The balance is said to be used for "weighing," and the standard masses are known as "weights." The proper relationship between mass and weight will be considered under the topic of force.

To weigh drugs precisely and accurately, the pharmacist must understand the errors inherent in operating a balance. A Class A balance, used for the compounding of prescriptions, is serviceable only if kept in good working condition and checked periodically for equality of arm length, beam rider accuracy, and sensitivity. These tests are described in the booklet by Goldstein and Mattocks.² Furthermore, a good balance is of no use unless an accurate set of weights is available.

Density and Specific Gravity. The pharmacist frequently uses these measurable quantities when interconverting between mass and volume. Density is a derived quantity since it combines the units of mass and volume. It is defined as mass per unit volume at a fixed temperature and pressure and is expressed in the cgs system in grams per cubic centimeter (g/cm^3). In SI units, density is expressed as kilograms per cubic meter.

Specific gravity, unlike density, is a pure number without dimension, however, it may be converted to density by the use of appropriate formulas.³ Specific gravity is defined as the ratio of the density of a substance to the density of water, the values for both substances being determined at the same temperature unless otherwise specified. The term *specific gravity*, in light of its definition, is a poor one; it would be more appropriate to refer to it as *relative density*.

Specific gravity is defined more often for practical purposes as the ratio of the mass of a substance to the mass of an equal volume of water at 4° or at some other specified temperature. The following notations are frequently found to accompany specific gravity readings: 25°/25°, 25°/4°, and 4°/4°. The first figure refers to

the temperature of the air in which the substance was weighed; the figure following the slash is the temperature of the water used. The official pharmaceutical compendia use a basis of 25°/25° to express specific gravity.

Specific gravity may be determined by the use of various types of pycnometers, the Mohr-Westphal balance, hydrometers, and other devices. The measurements and calculations are discussed in elementary chemistry, physics, and pharmacy books.

Other Dimensions and Units. The derived dimensions and their cgs and SI units are listed in Table 1-3. Although the units and relations are self-explanatory for most of the derived dimensions, force, pressure, and energy require some elaboration.

Force. One is familiar with force in everyday experience as a push or pull required to set a body in motion. The larger the mass of the body and the greater the required acceleration, the greater the force that one must exert. Hence, the force is directly proportional to the mass (when acceleration is constant) and to the acceleration (when the mass is constant). This may be represented by the relation

$$\text{Force} \propto \text{Mass} \times \text{Acceleration} \quad (1-1)$$

This proportionality is converted to an equality, that is, to an equation or mathematical expression involving an equal sign, according to the laws of algebra, by the introduction of a constant. Accordingly, we write

$$f = k \times m \times a \quad (1-2)$$

in which f is the force, k is the *proportionality constant*, m is the mass, and a is the acceleration. If the units are chosen so that the constant becomes unity (i.e., has the value of 1), the well-known force equation of physics is obtained:

$$f = m \times a \quad (1-3)$$

The cgs unit of force is the *dyne*, defined as the force that imparts to a mass of 1 g an acceleration of 1 cm/sec^2 .

The reader should recall from physics that weight is the force of gravitational attraction that the earth

TABLE 1-3. *Derived Dimensions and Units*

Derived Dimensions	Dimensional Symbol	CGS Unit	SI Unit	Relationship to Other Dimensions
Area (A)	L^2	cm^2	m^2	the square of a length
Volume (V)	L^3	cm^3	m^3 *	the cube of a length
Density (ρ)	ML^{-3}	g/cm^3	kg m^{-3}	mass/unit volume
Velocity (v)	LT^{-1}	cm/sec	m s^{-1}	length/unit time
Acceleration (a)	LT^{-2}	cm/sec^2	m s^{-2}	length/(time) ²
Force (f)	MLT^{-2}	$\text{g cm}/\text{sec}^2$ or dyne	$\text{kg m s}^{-2} = \text{J m}^{-1} = \text{N}$	mass \times acceleration
Pressure (p)	$ML^{-1}T^{-2}$	dyne/cm^2	$\text{N m}^{-2} = \text{kg m}^{-1}\text{s}^{-2} = \text{Pa}$	force/unit area
Energy (E)	ML^2T^{-2}	$\text{g cm}^2/\text{sec}^2$ or erg	$\text{kg m}^2 \text{s}^{-2} = \text{N m} = \text{J}$	force \times length

Key: N = newton, or kilogram \times meter \times second⁻²; Pa = pascal, or newton \times meter⁻²; J = joule; in this table, m = meter, not mass; L = length; T = time; M = mass.

*The cubic meter is a large volume, so that volume is often expressed in SI units as decimeter cubed (dm^3) which is equal to 1000 cm^3 .

exerts on a body, and it should be expressed properly in force units (dynes) rather than mass units (grams). The relationship between weight and mass can be obtained from equation (1-3). Substituting weight w for force and g for acceleration, the equation becomes

$$w = m \times g \quad (1-4)$$

Although the gravitational acceleration of a body varies from one part of the earth to another, it is approximately constant at 981 cm/sec². Substituting this value for g , the weight of a 1-g mass is calculated from equation (1-4) as follows:

$$w = 1 \text{ g} \times 981 \text{ cm/sec}^2$$

and

$$w = 981 \text{ g cm/sec}^2 \text{ or } 981 \text{ dynes}$$

Therefore, the weight of a body with a mass of 1 gram is actually 981 dynes. It is common practice to express weight in the mass unit, grams, since weight is directly proportional to mass; however, in problems involving these physical quantities, the distinction must be made.

The SI unit of force is the newton (N), which is equal to one kg m s⁻². It is defined as the force that imparts to a mass of 1 kg an acceleration of 1 m/sec² (see Table 1-3).

Pressure. Pressure may be defined as force per unit area; the unit commonly used in science is dyne/cm². Pressure is often given in atmospheres (atm) or in centimeters or millimeters of mercury. This latter unit is derived from a measurement of the height of a column of mercury in a barometer, which is used to measure the atmospheric pressure. The equation from elementary physics used to convert height in a column of mercury or another liquid into pressure units is

$$\text{pressure (dyne/cm}^2\text{)} = \rho \times g \times h \quad (1-5)$$

where ρ is the density of the liquid in g/cm³ at a particular temperature, g is the acceleration of gravity 980.665 cm/sec², and h is the height in cm of the column of liquid. At sea level, the mean pressure of the atmosphere supports a column of mercury 76 cm (760 mm) or 29.9 inches in height. The barometric pressure may be translated into the fundamental pressure unit, dyne/cm², by multiplying the height, $h = 76$ cm, times 1 cm² cross-sectional area by the density ρ of mercury, 13.595 g/cm³, at 0° to give the mass and multiplying this by the acceleration of gravity, $g = 980.7$ cm/sec². The result divided by cm² is 1.0133×10^6 dyne/cm² and is equal to 1 atm. This series of multiplication and division is expressed simply by equation 5.

In the SI system, the unit of pressure (or stress) is the newton divided by the meter squared (Nm⁻²) and is called the pascal (Pa), (see Table 1-3).

Example 1-1. Convert the pressure of a column of ethyl alcohol 76 cm Hg (760 mm Hg) high to a pressure at sea level and 0° C (standard pressure) expressed in (a) dyne/cm² and (b) pascals (Pa). The density (ρ) of ethanol at 0° C is 0.80625 g/cm³.

(a) To obtain the standard pressure in dyne/cm², one uses equation (1-5) with the density $\rho = 0.80625$ g/cm³, the acceleration of gravity g at sea level as 980.665 cm/sec², and the height h of the column of mercury as 76.000 cm Hg.

$$\begin{aligned} \text{Pressure} &= 0.80625 \text{ g/cm}^3 \times 980.665 \text{ cm/sec}^2 \times 76.000 \text{ cm} \\ &= 6.00902 \times 10^4 \text{ dyne/cm}^2 \end{aligned}$$

(b) To obtain the standard pressure in pascals (Pa), we use SI units in equation (1-5):

$$\begin{aligned} \text{Pressure} &= \left(0.80625 \text{ g/cm}^3 \times \frac{\text{kg}}{10^3 \text{ g}} \times \frac{(10^2)^3 \text{ cm}^3}{1 \text{ m}^3} \right)^* \\ &\times \left(980.665 \text{ cm/sec}^2 \times \frac{\text{m}}{100 \text{ cm}} \right) \times \left(76.000 \text{ cm} \times \frac{\text{m}}{100 \text{ cm}} \right) \\ &= 6.00902 \times 10^3 \text{ kg m}^{-1} \cdot \text{s}^{-2} \text{ (or N} \cdot \text{m}^{-2}\text{, or Pa)} \end{aligned}$$

*1 meter = 10² cm; therefore, 1 m³ = (10²)³ cm³ = 10⁶ cm³.

Work and Energy. Energy is frequently defined as the condition of a body that gives it the capacity for doing work. The concept actually is so fundamental that no adequate definition can be given. Energy may be classified as kinetic energy or potential energy.

The idea of energy is best approached by way of the mechanical equivalent of energy known as work and the thermal equivalent of energy or heat. When a constant force is applied to a body in the direction of its movement, the work done on the body equals the force multiplied by the displacement, and the system undergoes an increase in energy. The product of force and distance has the same dimensions as energy, namely ML^2T^{-2} . Other products also having the dimensions of energy are pressure \times volume, surface tension \times area, mass \times velocity², and electric potential difference \times quantity of electricity.

The cgs unit of work, also the unit of kinetic and potential energy, is the erg. It is defined as the work done when a force of 1 dyne acts through a distance of 1 centimeter:

$$1 \text{ erg} = 1 \text{ dyne} \times 1 \text{ cm}$$

The erg is often too small for practical use and is replaced by the joule (J) (pronounced *jewel*), which is equal to 10⁷ ergs:

$$1 \text{ joule} = 1 \times 10^7 \text{ erg}$$

In carrying out calculations in the cgs system involving work and pressure, work must be expressed in ergs and pressure in dynes/cm². When using the SI or any other system, consistent units must also be employed.

Heat and work are equivalent forms of energy and are interchangeable under certain circumstances. The thermal unit of energy in the cgs system is the gram calorie (small calorie). Formerly it was expressed as the amount of heat necessary to raise the temperature of 1 gram of water from 15° to 16° C. The small calorie is now defined as equal to 4.184 joules. The large or kilogram calorie (kcal) equals 1000 small calories. The SI unit for energy or work is the joule (J), which is seen in Table 1-3 to be equivalent to the newton \times meter (N m).

Temperature. Temperature is assigned a unit known as the degree. On the centigrade and the Kelvin or absolute scales, the freezing and boiling points of pure water at 1 atm pressure are separated by 100 degrees. Zero degree on the centigrade scale equals 273.15° on the Kelvin scale.

SOME ELEMENTS OF MATHEMATICS

The student should become familiar with the fundamental concepts of mathematics that are frequently used in the physical sciences and upon which are based many of the equations and graphic representations encountered in this book.

Calculations involving dimensions. Ratio and proportions are frequently used in the physical sciences for conversions from one system to another. The following calculation illustrates the use of proportions.

Example 1-2. How many gram calories are there in 3.00 joules? One should first recall a relationship or ratio that connects calories and joules. The relation 1 cal = 4.184 joules comes to mind. The question is then asked in the form of a proportion: If 1 calorie equals 4.184 joules, how many calories are there in 3.00 joules? The proportion is set down, being careful to express each quantity in its proper units. For the unknown quantity, an "X" is used.

$$\frac{1 \text{ cal}}{4.184 \text{ joules}} = \frac{X}{3.00 \text{ joules}}$$

$$X = \frac{3.00 \text{ joules} \times 1 \text{ cal}}{4.184 \text{ joules}}$$

$$X = 0.717 \text{ cal}$$

A second method, based on the requirement that the units as well as the dimensions be identical on both sides of the equal sign, is sometimes more convenient than the method of proportions.

Example 1-3. How many gallons are equivalent to 2.0 liters? It would be necessary to set up successive proportions to solve this problem. In the method involving identity of units on both sides of the equation, the quantity desired, X (gallons), is placed on the left and its equivalent, 2.0 liters, is set down on the right side of the equation. The right side must then be multiplied by known relations in ratio form, such as 1 pint per 473 ml, to give the units of gallons. Carrying out the indicated operations yields the result with its proper units.

$$X \text{ (in gallons)} = 2.0 \text{ liter} \times (1000 \text{ mL/liter})$$

$$\times (1 \text{ pt}/473 \text{ mL}) \times (1 \text{ gal}/8 \text{ pt})$$

$$X = 0.53 \text{ gal}$$

One may be concerned about the apparent disregard for the rules of significant figures (p. 11) in the equivalents such as 1 pint = 473 mL. The quantity of pints can be measured as accurately as that of milliliters, so that we assume 1.00 pint is meant here. The quantities 1 gallon and 1 liter are also exact by definition, and significant figures need not be considered in such cases.

Exponents. The various operations involving exponents, that is, the powers to which a number is raised, are best reviewed by studying the examples set out in Table 1-4.

TABLE 1-4. *The Rules of Exponents*

$a \times a \times a = a^3$	$a^2/a^4 = a^{2-4} = a^{-2} = \frac{1}{a^2}$
$a^2 \times a^3 = a^{2+3} = a^5$	$a^2/a^2 = a^{2-2} = a^0 = 1$
$(a^2)^3 = a^2 \times a^2 \times a^2 = a^6$	$a^{1/2} = \sqrt{a}$
$\left(\frac{a}{b}\right)^3 = a^3/b^3$	$a^{1/2} \times a^{1/2} = a^{1/2+1/2} = a^1 = a$
$a^5/a^2 = a^{5-2} = a^3$	$a^{2/3} = (a^2)^{1/3} = \sqrt[3]{a^2}$
$a^5/a^4 = a^{5-4} = a^1 = a$	

Logarithms. The equality

$$10^3 = 1000 \quad (1-6)$$

is expressed in logarithmic notation as:

$$\log_{10} 1000 = 3 \quad (1-7)$$

The exponent 3 to which the base 10 is raised to give 1000 in equation (1-6) is referred to as the *logarithm* of the number 1000. The number 1000 is known as the *antilogarithm* of the number 3. In general, if b , raised to the power x , gives the number a , then the logarithm to the base b of a is x :

$$b^x = a \quad (1-8)$$

$$\log_b a = x \quad (1-9)$$

When 10 is used as the base, the logarithm is known as the *common* or *Briggsian* logarithm, whereas the number 2.71828 . . . , designated as e , is used as the base for the *natural* or *Napierian* logarithms. The quantity e is important in the theoretic development of the physical and biochemical sciences and is discussed in some detail by Daniels.⁴ It is the sum of the series $1 + 1 + 1/2! + 1/3! + 1/4! . . .$ in which ! denotes a factorial number that is defined as the product of the positive integers between 1 and the number. Thus, $2! = 1 \times 2$, $3! = 1 \times 2 \times 3 = 6$, and $4! = 1 \times 2 \times 3 \times 4 = 24$. The common logarithms are designated by the symbol \log_{10} or simply as \log , while the natural logarithms are written as \log_e or \ln .

Although one usually has access to a hand calculator for obtaining the logarithms of numbers, it sometimes happens that one has only a table of common logarithms (see the back cover of this book). To convert from one system to another, particularly from the natural to the common logarithm, the following formula is used:

$$\ln a = 2.303 \log a^* \quad (1-10)$$

Equation (10) may be derived as follows. Let

$$\log a = x \quad (1-11)$$

so that

$$a = 10^x \quad (1-12)$$

*The conversion factor, 2.303, is more accurately expressed as 2.302585.

and taking the natural logarithm, equation (1-12) becomes

$$\ln a = \ln 10^x = x \ln 10 \quad (1-13)$$

Now $\ln 10 = 2.303$, and equation (1-13) becomes

$$\ln a = 2.303 x \quad (1-14)$$

Substituting the identity $x = \log a$ from equation (1-11) into equation (1-14) gives the desired formula.

The application of logarithm is best demonstrated by considering several examples. In the expression,

$$\log 60.0 = 1.778$$

the digit 1 to the left of the decimal point in the logarithm is known as the *characteristic* and signifies that the number 60.0 belongs to that class of numbers with a magnitude of 10^1 and thus contains two figures to the left of the decimal point. The quantity 0.778 of the logarithm is known as the *mantissa* and is found in the table of common logarithms. It is often convenient to express the number 60.0 by writing it with one significant figure to the left of the decimal point, 6.00, multiplied by 10 raised to the first power, viz., 6.00×10^1 . The exponent of 10 then gives the characteristic, and the value in the logarithm table gives the mantissa directly.

This method may be used to obtain the logarithm of 6000 from a table as follows. The number is first written as 6.000×10^3 if it is accurate to four significant figures. The characteristic is observed to be 3, and the mantissa is found in the table as 0.778. Hence,

$$\log 6000 = 3.778$$

For decimal fractions that frequently appear in problems involving molar concentration, the following method is used. Suppose one desires to know the logarithm of 0.0600. The number is first written as 6.00×10^{-2} . The characteristic of a number may be positive or negative; the mantissa is always positive. The characteristic in this case is -2 and the mantissa is 0.778. Hence,

$$\log 0.0600 = -2 + 0.778 = -1.222$$

Finding the number in a table when the logarithm is given, that is, obtaining the *antilogarithm*, is shown by the following example. What is the value of a if $\log a = 1.7404$? The characteristic is 1 and the mantissa is 0.7404. From the table of logarithms, one finds that the number corresponding to a mantissa of 0.7404 is 5.50. The characteristic is 1, so the antilogarithm is 5.50×10^1 or 55.0.

Let us find the antilogarithm of a negative number, -2.699 . Recalling that the mantissa must always be positive, we first separate the logarithm into a negative characteristic and positive mantissa:

$$-2.699 = -3.00 + 0.301$$

This transformation is easily seen in Figure 1-1, in

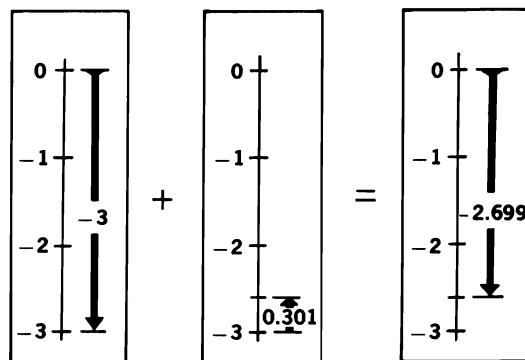


Fig. 1-1. Schematic representation for finding the antilogarithm of a negative number.

which -2.699 corresponds to going down the scale in a negative direction to -3 and coming back up the scale 0.301 units in the positive direction. Actually, by this process, we are subtracting 1 from the characteristic and adding 1 to the mantissa, or to the quantity

$$-2.699 = (-2) + (-0.699)$$

we subtract and add 1 to yield

$$(-2 - 1) + (-0.699 + 1) = -3 + 0.301$$

The result $(-3 + 0.301)$ is sometimes abbreviated to $(\bar{3}.301)$, in which the minus sign above the 3 applies only to the characteristic. $\bar{3}$ is commonly referred to as "bar three." It has been the practice in some fields, such as quantitative analysis, to use the form in which 10 is added and subtracted to give

$$\bar{3}.301 = 7.301 - 10$$

For physical chemical calculations and for plotting logarithms of numbers, it is more convenient to use the form -2.699 than one of the forms having a mixture of negative and positive parts. For use with logarithm tables, however, the mixed form is needed. Thus, in order to obtain the antilogarithm, we write the logarithm as $\bar{3}.301$. The number corresponding to the mantissa is found in the logarithm table to be 2.00. The characteristic is observed to be -3 , and the final result is therefore 2.00×10^{-3} .

We have dealt with logarithms to the base 10 (common logarithms) and to the base $e = 2.71828 \dots$ (natural logarithms). Logarithms to any other positive number as the base, b , may also be obtained. The formula used for this purpose is

$$\log_b(a) = \log_e(a)/\log_e(b) \quad (1-15)$$

To obtain the logarithm of the number $a = 100$ to the base $b = 37$, we substitute in equation (1-15):

$$\log_{37}(100) = \ln(100)/\ln(37) = 1.2753$$

One may also use common logs on the right side of the equation:

$$\log_{37}(100) = \log_{10}(100)/\log_{10}(37) = 1.2753$$

TABLE 1-5. Rules of Logarithms

$\log ab = \log a + \log b$	$\log \frac{1}{a} = \log 1 - \log a = -\log a$
$\log \frac{a}{b} = \log a - \log b$	$\log a^2 = \log a + \log a = 2 \log a$
$\log 1 = 0$ since $10^0 = 1$	$\log \sqrt{a} = \log a^{1/2} = \frac{1}{2} \log a$
$\log a^{-2} = -2 \log a = 2 \log \frac{1}{a}$	

These formulas allow one to obtain a logarithm to a base b for any whole or fractional positive number desired. *Problem 1-12* is an exercise involving the change from one logarithmic base to another.

As seen in the table of exponents (Table 1-4), numbers may be multiplied and divided by adding and subtracting exponents. Since logarithms are exponents, they follow the same rules. Some of the properties of logarithms are exemplified by the identities collected in Table 1-5.

Variation. The scientist is continually attempting to relate phenomena and establish generalizations with which to consolidate and interpret experimental data. The problem frequently resolves itself into a search for the relationship between two quantities that are changing at a certain rate or in a particular manner. The dependence of one property, the *dependent variable* y , on the change or alteration of another measurable quantity, the *independent variable* x , is expressed mathematically as

$$y \propto x \tag{1-16}$$

which is read: “ y varies directly as x ,” or “ y is directly proportional to x .” A proportionality is changed to an equation as follows. If y is proportional to x in general, then all pairs of specific values of y and x , say y_1 and x_1 , y_2 and x_2 , . . . , are proportional. Thus

$$\frac{y_1}{x_1} = \frac{y_2}{x_2} = \dots \tag{1-17}$$

Since the ratio of any y to its corresponding x is equal to any other ratio of y and x , the ratios are constant, or, in general

$$\frac{y}{x} = \text{constant} \tag{1-18}$$

Hence, it is a simple matter to change a proportionality to an equality by introducing a *proportionality constant*, k . To summarize, if

$$y \propto x$$

then

$$y = kx \tag{1-19}$$

It is frequently desirable to show the relationship between x and y by the use of the more general notation,

$$y = f(x) \tag{1-20}$$

which is read: “ y is some function of x .” That is, y may be equal to $2x$, to $27x^2$, or to $0.0051 + \log(a/x)$. The functional notation, equation (1-20), merely signifies that y and x are related in some way without specifying the actual equation by which they are connected. Some well-known formulas illustrating the principle of variation are shown in Table 1-6.

Graphic Methods. Scientists are not usually so fortunate as to begin each problem with an equation at hand relating the variables under study. Instead, the investigator must collect raw data and put them in the form of a table or graph to better observe the relationships. Constructing a graph with the data plotted in a manner so as to form a smooth curve often permits the investigator to observe the relationship more clearly, and perhaps allows expression of the connection in the form of a mathematical equation. The procedure of obtaining an empiric equation from a plot of the data is known as *curve fitting* and is treated in books on statistics and graphic analysis.

The magnitude of the independent variable is customarily measured along the horizontal coordinate scale called the x axis. The dependent variable is measured along the vertical scale or the y axis. The data are plotted on the graph, and a smooth line is drawn through the points. The x value of each point is known as the x coordinate or the *abscissa*, the y value is known as the y coordinate or the *ordinate*. The intersection of the x axis and the y axis is referred to as the *origin*. The x and y values may be either *negative* or *positive*.

The simplest relationship between two variables, where the variables contain no exponents other than one (*first-degree equation*), yields a straight line when plotted on rectangular graph paper. The straight-line or linear relationship is expressed as

$$y = a + bx \tag{1-21}$$

in which y is the dependent variable, x is the independent variable, and a and b are constants. The constant

TABLE 1-6. Formulas Illustrating the Principle of Variation

Measurement	Equation	Dependent Variation	Independent Variable	Proportionality Constant
Circumference of a circle	$C = \pi D$	Circumference, C	Diameter, D	$\pi = 3.14159 \dots$
Density	$M = \rho V$	Mass, M	Volume, V	Density, ρ
Distance of falling body	$s = \frac{1}{2}gt^2$	Distance, s	Time, t^2	Gravity constant, $\frac{1}{2}g$
Freezing point depression	$\Delta T_f = K_f m$	Freezing point depression, ΔT_f	Molality, m	Cryoscopic constant, K_f

b is the *slope* of the line; the greater the value of b , the steeper the slope. It is expressed as the change in y with the change in x or $b = \frac{\Delta y}{\Delta x}$. b is also the tangent of the angle that the line makes with the x axis. The slope may be positive or negative depending on whether the line slants upward or downward to the right. When $b = 1$, the line makes an angle of 45° with the x axis ($\tan 45^\circ = 1$), and the equation of the line may then be written

$$y = a + x \quad (1-22)$$

When $b = 0$, the line is horizontal (i.e., parallel to the x axis), and the equation reduces to

$$y = a \quad (1-23)$$

The constant a is known as the *y intercept* and signifies the point at which the line crosses the y axis. If a is positive, the line crosses the y axis above the x axis; if negative, it intersects the y axis below the x axis. When a is zero, equation (1-21) may be written,

$$y = bx \quad (1-24)$$

and the line passes through the origin.

The results of the determination of the refractive index of a benzene solution containing increasing concentrations of carbon tetrachloride are found in Table 1-7. The data are plotted in Figure 1-2 and are seen to produce a straight line with a negative slope. The equation of the line may be obtained by using the two-point form of the linear equation,

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1} (x - x_1) \quad (1-25)$$

The method involves selecting two widely separated points (x_1, y_1) and (x_2, y_2) on the line and substituting into the two-point equation.

Example 1-4. Referring to Figure 1-2, let 10.0% be x_1 and its corresponding y value 1.497 be y_1 ; let 60.0% be x_2 and 1.477 be y_2 . The equation then becomes

$$y - 1.497 = \frac{1.477 - 1.497}{60.0 - 10.0} (x - 10.0)$$

$$y - 1.497 = -4.00 \times 10^{-4} (x - 10.0)$$

$$y = -4.00 \times 10^{-4} x + 1.501$$

The value -4.00×10^{-4} is the slope of the straight line and corresponds to b in equation (1-21). A negative

TABLE 1-7. Refractive Indices of Mixtures of Benzene and Carbon Tetrachloride

(x) Concentration of CCl ₄ (volume %)	(y) Refractive Index
10.0	1.497
25.0	1.491
33.0	1.488
50.0	1.481
60.0	1.477

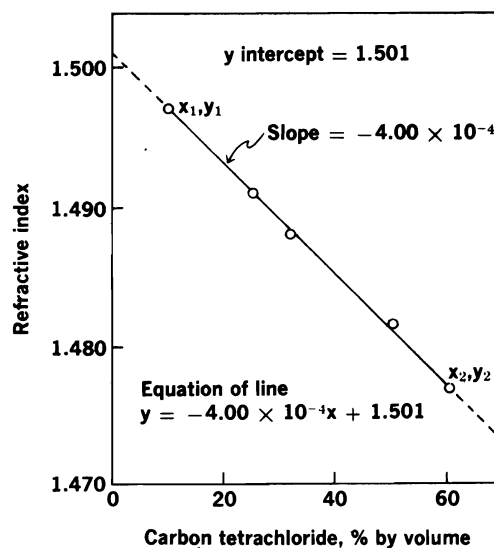


Fig. 1-2. Refractive index of the system benzene-carbon tetrachloride.

value for b indicates that y decreases with increasing values of x as observed in Figure 1-2. The value 1.501 is the y intercept and corresponds to a in equation (1-21).* It may be obtained from the plot in Figure 1-2 by *extrapolating* (extending) the line upwards to the left until it intersects the y axis. It will also be observed that

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x} = b \quad (1-26)$$

and this simple formula allows one to compute the slope of a straight line. The use of *statistics* to determine whether data fit the slope of such a line and its intercept on the y axis is illustrated later in this chapter.

Not all experimental data form straight lines when plotted on ordinary rectangular coordinate paper. Equations containing x^2 or y^2 are known as *second-degree* or *quadratic equations*, and graphs of these equations yield parabolas, hyperbolas, ellipses, and circles. The graphs and their corresponding equations may be found in standard textbooks on analytic geometry.

Logarithmic relationships occur frequently in scientific work. The data relating the amount of oil separating from an emulsion per month (dependent variable, y) as a function of the emulsifier concentration (independent variable, x) are collected in Table 1-8.

The data from this experiment may be plotted in several ways. In Figure 1-3, the oil separation y is plotted as ordinates against the emulsifier concentration x as abscissas on a rectangular coordinate grid. In

*The y -intercept, 1.501, is of course the refractive index of pure benzene at 20°C . For the purpose of this example we assume that we were not able to find this value in a table of refractive indices. In handbooks of chemistry and physics the value is actually found to be 1.5011 at 20°C .

TABLE 1-8. Emulsion Stability as a Function of Emulsifier Concentration

Emulsifier (% Concentration) (x)	Oil Separation (mL/month) (y)	Logarithm of Oil Separation (log y)
0.50	5.10	0.708
1.00	3.60	0.556
1.50	2.60	0.415
2.00	2.00	0.301
2.50	1.40	0.146
3.00	1.00	0.000

Figure 1-4, the logarithm of the oil separation is plotted against the concentration. In Figure 1-5, the data are plotted on semilogarithm paper, consisting of a logarithmic scale on the vertical axis and a linear scale on the horizontal axis.

Although Figure 1-3 provides a direct reading of oil separation, difficulties arise when one attempts to draw a smooth line through the points or to extrapolate the curve beyond the experimental data. Furthermore, the equation for the curve cannot be obtained readily from Figure 1-3. When the logarithm of oil separation is plotted as the ordinate, as in Figure 1-4, a straight line results, indicating that the phenomenon follows a logarithmic or exponential relationship. The slope and the y intercept are obtained from the graph, and the equation for the line is subsequently found by use of the two-point formula:

$$\log y = 0.85 - 0.28x$$

Figure 1-4 requires that we obtain the logarithms of the oil-separation data before the graph is constructed and, conversely, that we obtain the antilogarithm of the ordinates to read oil separation from the graph. These inconveniences of converting to logarithms and antilogarithms may be overcome by the use of semilogarithm paper. The x and y values of Table 1-8 are plotted directly on the graph to yield a straight line, as seen in

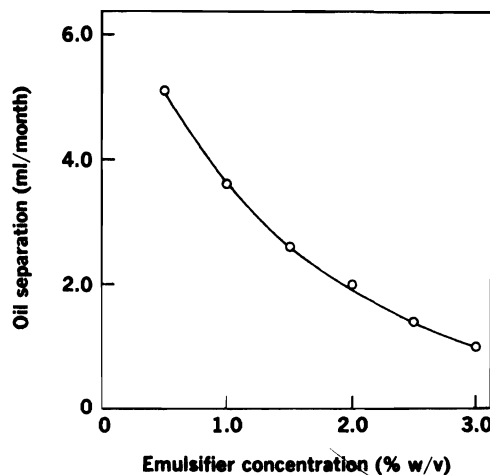


Fig. 1-3. Emulsion stability data plotted on a rectangular coordinate grid.

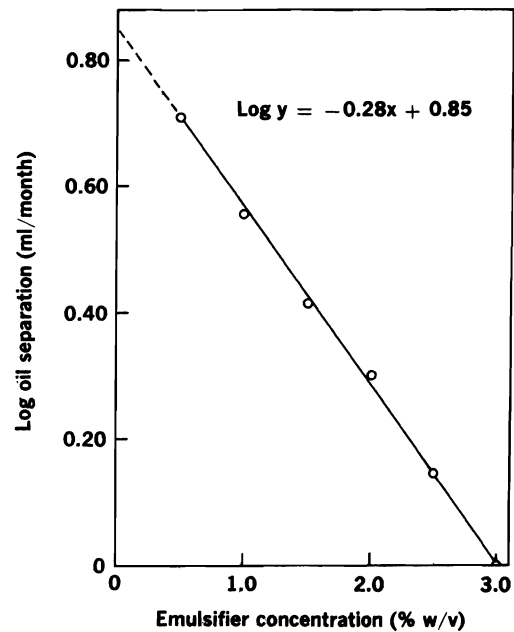


Fig. 1-4. A plot of the logarithm of oil separation of an emulsion vs. concentration on a rectangular grid.

Figure 1-5. Although such a plot ordinarily is not used to obtain the equation of the line, it is convenient for reading the oil separation directly from the graph. It is well to remember that the \ln of a number is simply 2.303 \log of the number. Therefore, logarithmic graph scales may be used for \ln as well as for \log plots.

Computers and Calculators. Computers are used widely in industry, government, business, education, and research and touch the lives of nearly everyone in one way or another. Computers may be divided into analog and digital machines. Digital computers deal

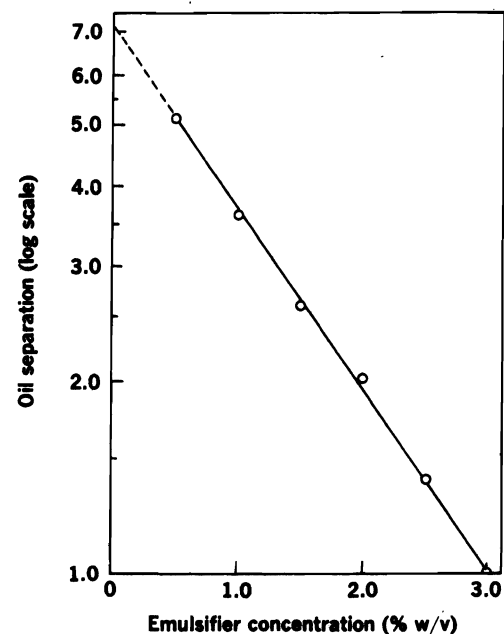


Fig. 1-5. Emulsion stability plotted on a semilogarithmic grid.

with numbers much like desk and hand-held calculators do. The modern calculator is provided with registers for the storage of data and a central core that can be programmed with mathematical instructions to carry out most mathematical functions. The programmable calculator, like the computer, also has a decision-making capacity through its ability to determine whether a number is larger than zero (positive), smaller than zero (negative), or equal to zero. The computer differs from the calculator in that it is faster, capable of greater storage, and more versatile.

The analog computer, unlike the digital computer, handles mathematical problems using voltages to represent variables such as concentration, pressure, time, and temperature. If the problem can be written as a differential equation (Examples A-9, A-12, p. 599), the solution is obtained by expressing the equation in voltages, capacitances, and resistances and then causing the voltage to vary with time as determined by the differential equation. The solution to the problem appears as a graphic plot on a recorder or is displayed as a tracing on an oscilloscope screen. The analog computer is composed of tens or hundreds of amplifiers that are used for the mathematical operations of addition, multiplication, and so on. The amplifiers are connected by the operator into a circuit that represents the differential equation at hand. Each amplifier corresponds to one step in the chemical, physical, or mechanical process under consideration. The analog computer is used in engineering to simulate the spring action on the axles of an automobile or the movement of a skyscraper in a high wind. It has been used to calculate the absorption, distribution, and elimination constants for a drug that is administered to a patient and to plot the curves for uptake and excretion. Today the digital computer can also calculate such values and prepare graphs with facility, and the popularity of the analog computer has diminished in pharmacy in recent years.

Currently, the microcomputer and the hand-held calculator are of great interest in small business and education and for personal use. The microcomputer, at a price within reach of the average individual, is more powerful today than the large institutional computers of the 1960s.

Programs may be written for large electronic computers and microcomputers in a number of languages, the most popular of which are FORTRAN, BASIC, PASCAL, C, and C++. Even some hand-held calculators can now be programmed in BASIC.

It will profit the student to become familiar with BASIC and/or FORTRAN and with the operation of an institutional or personal microcomputer. A hand-held calculator will be useful for working the problems at the ends of the chapters of this book. A programmable calculator is particularly convenient to obtain the slopes and intercepts of lines and for carrying out a repetitive sequence of mathematical operations. For example, in the chapter on solubility, if one desires to calculate the

minimum pH for complete solubility of a series of 10 acidic drugs of known pK_a values, it is simpler and faster to program the calculator than to carry out a number of repetitive steps for each of the 10 drugs.

Significant Figures. A significant figure is any digit used to represent a magnitude or quantity in the place in which it stands. The number zero is considered as a significant figure except when it is used merely to locate the decimal point. The two zeros immediately following the decimal point in the number 0.00750 merely locate the decimal point and are not significant. However, the zero following the 5 is significant since it is not needed to write the number; if it were not significant, it could be omitted. Thus, the value contains three significant figures. The question of significant figures in the number 7500 is ambiguous. One does not know whether any or all of the zeros are meant to be significant, or whether they are simply used to indicate the magnitude of the number. To express the significant figures of such a value in an unambiguous way, it is best to use exponential notation. Thus, the expression 7.5×10^3 signifies that the number contains two significant figures, and the zeros in 7500 are not to be taken as significant. In the value, 7.500×10^3 , both zeros are significant, and the number contains a total of four significant figures. The significant figures of some values are shown in Table 1-9.

The significant figures of a number include all certain digits plus the first uncertain digit. For example, one may use a ruler, the smallest subdivisions of which are centimeters, to measure the length of a piece of glass tubing. If one finds that the tubing measures slightly greater than 27 cm in length, it is proper to estimate the doubtful fraction, say 0.4, and express the number as 27.4 cm. A replicate measurement may yield the value 27.6 or 27.2 cm so that the result is expressed as 27.4 ± 0.2 cm. When a value such as 27.4 cm is encountered in the literature without further qualification, the reader should assume that the final figure is correct to within about ± 1 in the last decimal place, which is meant to signify the mean deviation of a single measurement. However, when a statement is given in the official compendia (U.S. Pharmacopeia and National Formulary) such as "not less than 99," it means 99.0 and not 98.9.

Significant figures are particularly useful for indicating the precision of a result. The precision is limited by

TABLE 1-9. Significant Figures

Number	Number of Significant Figures
53.	2
530.0	4
0.00053	2
5.0030	5
5.3×10^{-2}	2
5.30×10^{-4}	3
53000	indeterminate

the instrument used to make the measurement. A measuring rule marked off in centimeter divisions will not produce as great a precision as one marked off in 0.1 cm or mm. One may obtain a length of 27.4 ± 0.2 cm with the first ruler and a value of, say, 27.46 ± 0.02 cm with the second. The latter ruler, yielding a result with four significant figures, is obviously the more precise one. The number 27.46 implies a precision of about 2 parts in 3000, whereas 27.4 implies a precision of only 2 parts in 300.

The absolute magnitude of a value should not be confused with its precision. We consider the number 0.00053 mole/liter as a relatively small quantity because three zeros immediately follow the decimal point. But these zeros are not significant and tell us nothing about the precision of the measurement. When such a result is expressed as 5.3×10^{-4} mole/liter, or better as $5.3 (\pm 0.1) \times 10^{-4}$ mole/liter, both its precision and its magnitude are readily apparent.

In dealing with experimental data, certain rules pertain to the figures that enter into the computations:

1. In rejecting superfluous figures, increase by 1 the last figure retained if the following figure rejected is 5 or greater. Do not alter the last figure if the rejected figure has a value of less than 5. Thus, if the value 13.2764 is to be rounded off to four significant figures, it is written as 13.28. The value 13.2744 is rounded off to 13.27.

2. In addition or subtraction, include only as many figures to the right of the decimal point as there are present in the number with the least such figures. Thus, in adding 442.78, 58.4, and 2.684, obtain the sum and then round off the result so that it contains only one figure following the decimal point:

$$442.78 + 58.4 + 2.684 = 503.684$$

This figure is rounded off to 503.9.

Rule 2 of course cannot apply to the weights and volumes of ingredients in the monograph of a pharmaceutical preparation. The minimum weight or volume of each ingredient in a pharmaceutical formula or a prescription should be large enough that the error introduced is no greater than, say, 5 in 100 (5%), using the weighing and measuring apparatus at hand. Accuracy and precision in prescription compounding is discussed in some detail by Brecht.⁵

3. In multiplication or division, the rule commonly used is to retain the same number of significant figures in the result as appear in the value with the least number of significant figures. In multiplying 2.67 and 3.2, the result is recorded as 8.5 rather than as 8.544. A better rule here is to retain in the result the number of figures that produces a percentage error no greater than that in the value with the largest percentage uncertainty.

4. In the use of logarithms for multiplication and division, retain the same number of significant figures in the mantissa as there are in the original numbers.

The characteristic signifies only the magnitude of the number and accordingly is not significant. Since calculations involved in theoretic pharmacy usually require no more than three significant figures, a four-place logarithm table yields sufficient precision for our work. Such a table is found on the inside back cover of this book. The hand calculator is more convenient, however, and tables of logarithms are used less frequently today. Logarithms to nine significant figures are obtained by the simple press of a button on the modern hand calculator.

5. If the result is to be used in further calculations, retain at least one digit more than suggested in the rules just given. The final result is then rounded off to the last significant figure.

STATISTICAL METHODS AND THE ANALYSIS OF ERRORS

If one is to maintain a high degree of exactitude in the compounding of prescriptions and the manufacture of products on a large scale, one must know how to locate and eliminate constant and accidental errors insofar as possible. Pharmacists must recognize, however, that just as they cannot hope to produce a perfect pharmaceutical product, neither can they make an absolute measurement. In addition to the inescapable imperfections in mechanical apparatus and the slight impurities that are always present in chemicals, perfect accuracy is impossible because of the inability of the operator to make a measurement or estimate a quantity to a degree finer than the smallest division of the instrument scale.

Error may be defined as a deviation from the absolute value or from the true average of a large number of results. Two types of errors are recognized: *determinate* (constant) and *indeterminate* (random or accidental).

Determinate Errors. Determinate or constant errors are those that, although sometimes unsuspected, may be avoided or determined and corrected once they are uncovered. They are usually present in each measurement and affect all observations of a series in the same way. Examples of determinate errors are those inherent in the particular method used, errors in the calibration and the operation of the measuring instruments, impurities in the reagents and drugs, biased personal errors that, for example, might recur consistently in the reading of a meniscus, in pouring and mixing, in weighing operations, in matching colors, and in making calculations. The change of volume of solutions with temperature, while not constant, is, however, a systematic error that also may be determined and accounted for once the coefficient of expansion is known.

Determinate errors may be combated in analytic work by the use of calibrated apparatus, by the use of blanks and controls, by using several different analytic

procedures and apparatus, by eliminating impurities, and by carrying out the experiment under varying conditions. In pharmaceutical manufacturing, determinate errors may be eliminated by calibrating the weights and other apparatus and by checking calculations and results with other workers. Adequate corrections for determinate errors must be made before the estimation of indeterminate errors can have any significance.

Indeterminate Errors. Indeterminate errors occur by accident or chance, and they vary from one measurement to the next. When one fires a number of bullets at a target, some may hit the "bull's eye," while others will be scattered around this central point. The greater the skill of the marksman, the less scattered will be the pattern on the target. Likewise, in a chemical analysis, the results of a series of tests will yield a random pattern around an average or central value, known as the *mean*. Random errors will also occur in filling a number of capsules with a drug, and the finished products will show a definite variation in weight.

Indeterminate errors cannot be allowed for or corrected because of the natural fluctuations that occur in all measurements.

Those errors that arise from random fluctuations in temperature or other external factors and from the variations involved in reading instruments are not to be considered accidental or random. Instead, they belong to the class of determinate errors and are often called *pseudoaccidental* or *variable determinate* errors. These errors may be reduced by controlling conditions through the use of constant-temperature baths and ovens, the use of buffers, and the maintenance of constant humidity and pressure where indicated. Care in reading fractions of units on graduates, balances, and other apparatus can also reduce pseudoaccidental errors. Variable determinate errors, although seemingly indeterminate, can thus be determined and corrected by careful analysis and refinement of technique on the part of the worker. Only errors that result from pure random fluctuations in nature are considered truly indeterminate.

Precision and Accuracy. *Precision* is a measure of the agreement among the values in a group of data, while *accuracy* is the agreement between the data and the true value. Indeterminate or chance errors influence the precision of the results, and the measurement of the precision is accomplished best by statistical means. Determinate or constant errors affect the accuracy of data. The techniques used in analyzing the precision of results, which in turn supply a measure of the indeterminate errors, will be considered first, and the detection and elimination of determinate errors or inaccuracies will be discussed later.

Indeterminate or chance errors obey the laws of probability, both positive and negative errors being equally probable, and larger errors being less probable than smaller ones. If one plots a large number of results

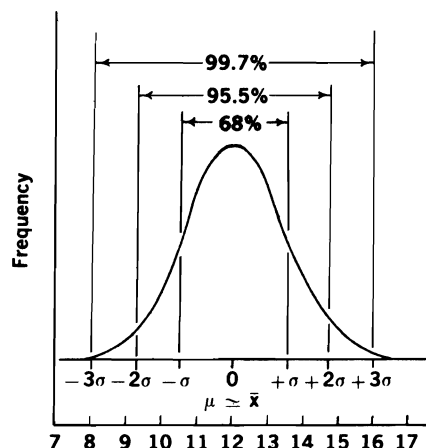


Fig. 1-6. The normal curve for the distribution of indeterminate errors.

having various errors along the vertical axis against the magnitude of the errors on the horizontal axis, a bell-shaped curve, known as a *normal frequency distribution curve*, is often obtained, as shown in Figure 1-6. If the distribution of results follows the normal probability law, the deviations will be represented exactly by the curve for an infinite number of observations, which constitute the *universe* or *population*. Whereas the population is the whole of the category under consideration, the *sample* is that portion of the population used in the analysis.

The Arithmetic Mean. When a normal distribution is obtained, it follows that the arithmetic mean is the best measure of the central value of a distribution; that is, the mean represents a point corresponding closest to the "bull's eye." The theoretic mean for a large number of measurements (the universe or population) is known as the *universe* or *population mean* and is given the symbol μ (mu).

The arithmetic mean \bar{X} is obtained by adding together the results of the various measurements and dividing the total by the number N of the measurements. In mathematical notation, the arithmetic mean for a small group of values is expressed as

$$\bar{X} = \frac{\Sigma(X_i)}{N} \quad (1-27)$$

in which Σ is the Greek capital letter sigma standing for "the sum of," X_i is the i th individual measurement of the group, and N is the number of values. \bar{X} is an estimate of μ and approaches it as the number of measurements N is increased.

Measures of Dispersion. After having chosen the arithmetic mean as the central tendency of the data, it is necessary to express the dispersion or scatter about the central value in a quantitative fashion so as to establish an estimate of variation among the results. This variability is usually expressed as the *range*, the *mean deviation*, or the *standard deviation*.

The range is the difference between the largest and the smallest value in a group of data and gives a rough idea of the dispersion. It sometimes leads to ambiguous results, however, when the maximum and minimum values are not in line with the rest of the data. The range will not be considered further.

The average distance of all the hits from the "bull's eye" would serve as a convenient measure of the scatter on the target. The average spread about the arithmetic mean of a large series of weighings or analyses is the mean deviation δ of the population.* The sum of the positive and negative deviations about the mean equals zero; hence, the algebraic signs are disregarded in order to obtain a measure of the dispersion.

The *mean deviation* d for a sample, that is, the deviation of an individual observation from the arithmetic mean of the sample, is obtained by taking the difference between each individual value X_i and the arithmetic mean \bar{X} , adding the differences without regard to the algebraic signs, and dividing the sum by the number of values to obtain the average. The mean deviation of a sample is expressed as

$$d = \frac{\sum |X_i - \bar{X}|}{N} \quad (1-28)$$

in which $\sum |X_i - \bar{X}|$ is the sum of the absolute deviations from the mean. The vertical lines on either side of the term in the numerator indicate that the algebraic sign of the deviation should be disregarded.

Youden⁶ discourages the use of the mean deviation since it gives a biased estimate that suggests a greater precision than actually exists when a small number of values are used in the computation. Furthermore, the mean deviation of small subsets may be widely scattered around the average of the estimates, and accordingly, d is not particularly efficient as a measure of precision.

The standard deviation σ (the Greek lower case letter sigma) is the square root of the mean of the squares of the deviations. This parameter is used to measure the dispersion or variability of a large number of measurements; for example, the weights of the contents of several million capsules. This set of items or measurements approximates the *population* or the *universe*, and σ is therefore called the *standard deviation of the universe*.** Universe standard deviations are shown in Figure 1-6.

*The population mean deviation is written as

$$\delta = \frac{\sum |X_i - \mu|}{N}$$

where X_i is an individual measurement, μ is the population mean, and N is the number of measurements.

**The equation for the universe standard deviation is

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

As previously noted, any finite group of experimental data may be considered as a subset or sample of the population; the statistic or characteristic of a sample from the universe used to express the variability of a subset and supply an estimate of the standard deviation of the population is known as the *sample standard deviation* and is designated by the small letter s . The formula is

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}} \quad (1-29)$$

For a small sample the equation is written

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}} \quad (1-30)$$

The term $(N - 1)$ is known as the *number of degrees of freedom*. It replaces N to reduce the bias of the standard deviation s , which on the average is lower than the universe standard deviation.

The reason for introducing $(N - 1)$ is explained as follows. When a statistician selects a sample and makes a single measurement or observation, he or she obtains at least a rough estimate of the mean of the parent population. This single observation, however, can give no hint as to the degree of variability in the population. When a second measurement is taken, however, a first basis for estimating the population variability is obtained. The statistician states this fact by saying that two observations supply one *degree of freedom* for estimating variations in the universe. Three values provide two degrees of freedom, four values provide three degrees of freedom, and so on. Therefore, we do not have access to all N values of a sample for obtaining an estimate of the standard deviation of the population. Instead, we must use 1 less than N or $(N - 1)$, as shown in equation (1-30). When N is large, say $N > 100$, one may use N instead of $(N - 1)$ to estimate the population standard deviation, since the difference between the two is negligible.

Modern statistical methods handle the small sample quite well; however, the investigator should recognize that the estimate of the standard deviation becomes less reproducible and, on the average, becomes lower than the universe standard deviation as fewer samples are used to compute the estimate.

A sample calculation involving the arithmetic mean, the mean deviation, and the estimate of the standard deviation follows.

Example 1-5. A pharmacist received a prescription for a patient with rheumatoid arthritis calling for seven divided powders, the contents of which were each to weigh 1.00 gram. To check his skill in filling the powders, he removed the contents from each paper after filling the prescription by the block-and-divide method and then weighed the powders carefully. The results of the weighings are given in the first column of Table 1-10; the deviations of each value from the arithmetic mean, disregarding the sign, are given in column 2; and the squares of the deviations are shown in the last column. Based on the use of the mean deviation, the weight of the powders may be expressed as 0.98 ± 0.046 gram. The variability of a single

TABLE 1-10. Statistical Analysis of Divided Powder Compounding Technique

Weight of Powder Contents (grams)	Deviation, (Sign Ignored) $ X_i - \bar{X} $	Square of the Deviation $(X_i - \bar{X})^2$
1.00	0.02	0.0004
0.98	0.00	0.0000
1.00	0.02	0.0004
1.05	0.07	0.0049
0.81	0.17	0.0289
0.98	0.00	0.0000
1.02	0.04	0.0016
Total $\Sigma = 6.84$	$\Sigma = 0.32$	$\Sigma = 0.0362$
Average 0.98	0.046	

powder may also be expressed in terms of percent deviation by dividing the mean deviation by the arithmetic mean and multiplying by 100. The result is $0.98 \pm 4.6\%$; of course, it includes errors due to removing the powders from the papers and weighing the powders in the analysis.

The standard deviation is used more frequently than the mean deviation in research work. For large sets of data, it is approximately 25% larger than the mean deviation, that is, $\sigma = 1.25 \delta$.

The statistician has estimated that owing to chance errors, about 68% of all results in a large set will fall within one standard deviation on either side of the arithmetic mean, 95.5% within ± 2 standard deviations, and 99.7% within ± 3 standard deviations, as seen in Figure 1-6.

Goldstein⁷ selected 1.73σ as an equitable tolerance standard for prescription products, whereas Saunders and Fleming⁸ advocated the use of $\pm 3 \sigma$ as approximate limits of error for a single result.

In pharmaceutical work, it should be considered permissible to accept $\pm 2 s$ as a measure of the variability or "spread" of the data in small samples. Then, roughly 5 to 10% of the individual results will be expected to fall outside this range if only chance errors occur.

The estimate of the standard deviation in *Example 1-4* is calculated as follows:

$$s = \sqrt{\frac{0.0362}{(7-1)}} = 0.078 \text{ gram}$$

and $\pm 2 s$ is equal to ± 0.156 gram. That is to say, based upon the analysis of this experiment, the pharmacist should expect that roughly 90 to 95% of the sample values will fall within ± 0.156 gram of the sample mean.

The smaller the standard deviation estimate (or the mean deviation) the more *precise* is the compounding operation. In the filling of capsules, precision is a measure of the ability of the pharmacist to put the same amount of drug in each capsule and to reproduce the result in subsequent operations. Statistical techniques for predicting the probability of occurrence of a specific deviation in future operations, although important in

pharmacy, require methods that are outside the scope of this book. The interested reader is referred to treatises on statistical analysis.

Whereas the average deviation and the standard deviation can be used as measures of the *precision* of a method, the difference between the arithmetic mean and the *true* or *absolute* value expresses the error that can often be used as a measure of the *accuracy* of the method.

The true or absolute value is ordinarily regarded as the universe mean μ —that is, the mean for an infinitely large set—since it is assumed that the true value is approached as the sample size becomes progressively larger. The universe mean does not, however, coincide with the true value of the quantity measured in those cases in which determinate errors are inherent in the measurements.

The difference between the sample arithmetic mean and the true value gives a measure of the accuracy of an operation; it is known as the *mean error*.

In *Example 1-5*, the true value is 1.00 gram, the amount requested by the physician. The apparent error involved in compounding this prescription is

$$E = 1.0 - 0.98 = +0.02 \text{ gram}$$

in which the positive sign signifies that the true value is greater than the mean value. An analysis of these results shows, however, that this difference is not statistically significant, but rather is most likely due to accidental errors.* Hence, the accuracy of the operation in *Example 1-5* is sufficiently great that no systemic error can be presumed. We may find on further analysis, however, that one or several results are questionable. This possibility is considered later. If the

*The deviation of the arithmetic mean from the true value or the mean of the parent population can be tested by use of the following expression:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N-1}}$$

In this equation, t is a statistic known as Student's t value, after W. S. Gosset, who wrote under the pseudonym of "Student." The other terms in the equation have the meaning previously assigned to them.

Student's t Values for Six Degrees of Freedom

Probability of a plus or minus deviation greater than t	0.8	0.6	0.4	0.2	0.02	0.001
t value	0.27	0.55	0.91	1.44	3.14	5.96

Substituting the results of the analysis of the divided powders into the equation just given, we have

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N-1}} = \frac{0.98 - 1.00}{0.08/\sqrt{6}} = \frac{-0.02}{0.033} = -0.61$$

Entering the table with a t value of 0.61, we see that the probability of finding a \pm deviation greater than t is roughly equal to 0.6. This means that in the long run there are about 60 out of 100 chances of finding a t value greater than -0.61 by chance alone. This probability is sufficiently large to suggest that the difference between the mean and the true value may be taken as due to chance.

arithmetic mean in *Example 1-5* were 0.90 instead of 0.98, the difference could be stated with assurance to have statistical significance, since the probability that such a result could occur by chance alone would be small.* The mean error in this case is

$$1.00 - 0.90 = 0.10 \text{ gram}$$

The *relative error* is obtained by dividing the mean error by the true value. It may be expressed as a percentage by multiplying by 100, or in parts per thousand by multiplying by 1000. It is easier to compare several sets of results by using the relative error rather than the absolute mean error. The relative error in the case just cited is

$$\frac{0.10 \text{ gram}}{1.00 \text{ gram}} \times 100 = 10\%$$

The reader should recognize that it is possible for a result to be precise without being accurate, that is, a constant error is present. If the capsule contents in *Example 1-5* had yielded an average weight of 0.60 gram with a mean deviation of 0.5%, the results would have been accepted as precise. The degree of accuracy, however, would have been low since the average weight would have differed from the true value by 40%. Conversely, the fact that the result may be accurate does not necessarily mean that it is also precise. The situation can arise in which the mean value is close to the true value, but the scatter due to chance is large. Saunders and Fleming⁸ observe that "it is better to be roughly accurate than precisely wrong."

A study of the individual values of a set often throws additional light on the exactitude of the compounding operations. Returning to the data of *Example 1-5*, we note one rather discordant value, namely, 0.81 gram. If the arithmetic mean is recalculated ignoring this measurement, one obtains a mean of 1.01 grams. The mean deviation without the doubtful result was 0.02 gram. It is now seen that the divergent result is 0.20 gram smaller than the new average or, in other words, its deviation is 10 times greater than the mean deviation. A deviation greater than four times the mean deviation will occur purely by chance only about once or twice in 1000 measurements; hence, the discrepancy in this case was probably caused by some definite error in technique. This rule is rightly questioned by statisticians, but it is a useful though not always reliable criterion for finding discrepant results.

*When the mean is 0.90 gram, the *t* value is

$$t = \frac{0.90 - 1.00}{0.08/\sqrt{6}} = \frac{-0.10}{0.033} = -3$$

and the probability of finding a *t* value greater than -3 as a result of chance alone is found in the *t* table to be about 0.02, or 2 chances in 100. This probability is sufficiently small to suggest that the difference between the mean and the true value is real, and the error may be computed accordingly.

TABLE 1-11. *Refractive Indices of Mixtures of Benzene and Carbon Tetrachloride*

(x) Concentration of CCl ₄ (volume %)	(y) Refractive Index
10.0	1.497
26.0	1.493
33.0	1.485
50.0	1.478
61.0	1.477

Having uncovered the variable weight among the units, one can proceed to investigate the cause of the determinate error. The pharmacist may find that some of the powder was left on the sides of the mortar or on the weighing paper or possibly was lost during trituration. If several of the powder weights deviated widely from the mean, a serious deficiency in the compounder's technique would be suspected. Such appraisals as these in the college laboratory will aid the student in locating and correcting errors and will help the pharmacist to become a safe and proficient compounder before entering the practice of pharmacy. Bingenheimer⁹ has described such a program for students in the dispensing laboratory.

Linear Regression Analysis. The data given in Table 1-7 and plotted in Figure 1-2 clearly indicate the existence of a linear relationship between the refractive index and volume percent of carbon tetrachloride in benzene. The straight line that joins virtually all the points can be drawn readily on the figure by sighting the points along the edge of a ruler and drawing a line that can be extrapolated to the *y* axis with confidence.

Let us suppose, however, that the person who prepared the solutions and carried out the refractive index measurements was not skilled and, as a result of poor technique, allowed indeterminate errors to appear. We might then be presented with the data given in Table 1-11. When this is plotted on graph paper, an appreciable scatter is observed (Fig. 1-7) and we are unable, with any degree of confidence, to draw the line that expresses the relation between refractive index and concentration. It is here that we must employ better means of analyzing the available data.

The first step is to determine whether or not the data in Table 1-11 should fit a straight line, and for this we calculate the *correlation coefficient*, *r*, using the following equation:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (1-31)$$

When there is perfect correlation between the two variables (i.e., a perfect linear relationship), *r* = 1. When the two variables are completely independent, *r* = 0. Depending on the degrees of freedom and the

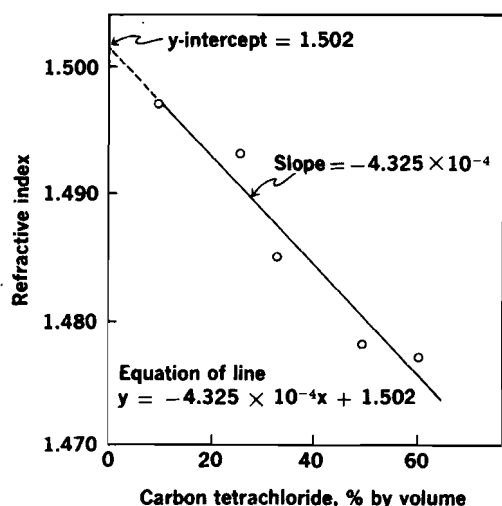


Fig. 1-7. Slope, intercept, and equation of line for data in Table 1-11 calculated by regression analysis.

chosen probability level, it is possible to calculate values of r above which there is significant correlation and below which there is no significant correlation. Obviously, in the latter case, it is not profitable to proceed further with the analysis unless the data can be plotted in some other way that will yield a linear relation. An example of this is shown in Figure 1-4, in which a linear plot is obtained by plotting the *logarithm* of oil separation from an emulsion against emulsifier concentration, as opposed to Figure 1-3, in which the raw data are plotted in the conventional manner.

Assuming that the calculated value of r shows a significant correlation between x and y , it is then necessary to calculate the slope and intercept of the line using the equation:

$$b = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} \quad (1-32)$$

in which b is the *regression coefficient*, or slope. By substituting the value for b in equation (1-33), we can then obtain the y intercept:

$$\bar{y} = y + b(x - \bar{x}) \quad (1-33)$$

The following series of calculations, based on the data in Table 1-11, will illustrate the use of these equations.

Example 1-6. Using the data in Table 1-11, calculate the correlation coefficient, the regression coefficient, and the intercept on the y axis.

Examination of equations (1-31), (1-32), and (1-33) shows the various values we must calculate, and these are set up as shown:

x	$(x - \bar{x})$	$(x - \bar{x})^2$
10.0	-26.0	676.0
26.0	-10.0	100.0
33.0	-3.0	9.0
50.0	+14.0	196.0
61.0	+25.0	625.0
$\Sigma = 180.0$	$\Sigma = 0$	$\Sigma = 1606.0$
$\bar{x} = 36.0$		

y	$(y - \bar{y})$	$(y - \bar{y})^2$
1.497	+0.011	0.000121
1.493	+0.007	0.000049
1.485	-0.001	0.000001
1.478	-0.008	0.000064
1.477	-0.009	0.000081
$\Sigma = 7.430$	$\Sigma = 0$	$\Sigma = 0.000316$
$\bar{y} = 1.486$		

$(x - \bar{x})(y - \bar{y})$
-0.286
-0.070
+0.003
-0.112
-0.225
$\Sigma = -0.693$

Substituting the relevant values into equation (1-31) gives

$$r = \frac{-0.693}{\sqrt{1606.0 \times 0.000316}} = -0.97^*$$

From equation (1-32)

$$b = \frac{-0.693}{1606.0} = -4.315 \times 10^{-4}$$

and finally, from equation (1-33)

$$\begin{aligned} \text{the intercept on the } y \text{ axis} &= 1.486 \\ &- 4.315 \times 10^{-4} (0 - 36) \\ &= +1.502 \end{aligned}$$

Note that for the intercept, we place x equal to zero in equation (1-31). By inserting an actual value of x into equation (1-33), we obtain the value of y that should be found at that particular value of x . Thus, when $x = 10$,

$$\begin{aligned} y &= 1.486 - 4.315 \times 10^{-4} (10 - 36) \\ &= 1.486 - 4.315 \times 10^{-4} (-26) \\ &= 1.497 \end{aligned}$$

The value agrees with the experimental value, and hence this point lies on the statistically calculated slope drawn in Figure 1-7.

Multiple Linear and Polynomial Regression. Regression for two, three, or more independent variables may be performed, using a linear equation:

$$y = a + bx_1 + cx_2 + dx_3 + \dots \quad (1-34)$$

*The theoretic values of r when the probability level is set at 0.05 are:

Degrees of freedom ($N - 2$):	2	3	5	10	20	50
Correlation coefficient, r :	0.95	0.88	0.75	0.58	0.42	0.27

In *Example 1-6*, ($N - 2$) = 3 and hence the theoretic value of r = 0.88. The calculated value was found to be 0.97, and the correlation between x and y is therefore significant.

For a power series in x , a polynomial form is employed:

$$y = a + bx + cx^2 + dx^3 + \dots \quad (1-35)$$

In equations (1-34) and (1-35), y is the dependent variable, and a , b , c , and d are regression coefficients obtained by solving the regression equation. Computers are used to handle these more complex equations, but some hand-held calculators, such as the Hewlett Packard HP41C, are programmed to solve multiple regression equations containing two or more independent variables. The r^2 used in multiple regression is called the square of the multiple correlation coefficient and is given the symbol R^2 in some texts to distinguish it from r^2 , the square of the linear correlation coefficient. Multiple regression analysis is treated by Draper and Smith.¹⁰

References and Notes

1. S. W. Goldstein, A. M. Mattocks, and U. Beirmacher, *J. Am. Pharm. Assoc., Pract. Ed.* 12, 421, 1951.
2. S. W. Goldstein and A. M. Mattocks, *Professional Equilibrium and Compounding Precision*, a bound booklet reprinted from the *J. Am. Pharm. Assoc., Pract. Ed.* April-August 1951.
3. *CRC Handbook of Chemistry and Physics*, 53rd Edition, CRC Press, Boca Raton, FL., 1982-83, p. D-227.
4. F. Daniels, *Mathematical Preparation for Physical Chemistry*, McGraw-Hill, New York, 1928.
5. E. A. Brecht, in *Sprolows' American Pharmacy*, L. W. Dittert, Ed., 7th Edition, Lippincott, Philadelphia, 1974, Chapter 2.
6. W. J. Youden, *Statistical Methods for Chemists*, R. Krieger, Huntington, N.Y., Reprint of original edition; 1977, p. 9.
7. S. W. Goldstein, *J. Am. Pharm. Assoc., Sci. Ed.* 38, 18, 131, 1949; *ibid.*, 39, 505, 1950.
8. L. Saunders and R. Fleming, *Mathematics and Statistics*, Pharmaceutical Press, London, 1957.
9. L. E. Bingenheimer, *Am. J. Pharm. Educ.* 17, 236, 1953.
10. N. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1980.
11. N. C. Harris and E. M. Hemmerling, *Introductory Applied Physics*, McGraw-Hill, New York, 1972, p. 310.
12. M. J. Kamlet, R. M. Doherty, V. Fiserova-Bergerova, P. W. Carr, M. H. Abraham and R. W. Taft, *J. Pharm. Sci.* 76, 14, 1987.
13. W. Lowenthal, "Methodology and Calculations," in *Remington's Pharmaceutical Science*, A. R. Gennaro, Editor, Mack Publishing, Easton, PA., 1985, p. 82.

Problems

1-1. The density ρ of a plastic latex particle is 2.23 g cm^{-3} . Convert this value into SI units.

Answer: $2.23 \times 10^3 \text{ kg m}^{-3}$. (See Table 1-3 and the physical constants on the inside front cover of the book for cgs and SI units.)

1-2. Convert 2.736 nm to cm.

Answer: $2.736 \times 10^{-7} \text{ cm}$

1-3. Convert $1.99 \times 10^4 \text{ \AA}^3$ into (nanometers)³.

Answer: 19.9 nm^3

1-4. The surface tension (γ) of a new synthetic oil has a value of $27.32 \text{ dyne cm}^{-1}$. Calculate the corresponding γ value in SI units.

Answer: $0.02732 \text{ N m}^{-1} = 0.02732 \text{ J m}^{-2}$

1-5. The work done by the kidneys in transforming 0.1 mole of urea from the plasma to the urine is 259 cal. Convert this quantity into SI units.

Answer: 1084 J

1-6. The body excretes HCl into the stomach in the concentration of 0.14 M at 37° C . The work done in this process is $3.8 \times 10^{11} \text{ erg}$. Convert this energy into the fundamental SI units of $\text{kg m}^2 \text{ s}^{-2}$. M stands for molarity.

Answer: $3.8 \times 10^4 \text{ kg m}^2 \text{ s}^{-2}$

1-7. The gas constant R is given in SI units as $8.3143 \text{ J }^\circ\text{K}^{-1} \text{ mole}^{-1}$. Convert this value into calories.

Answer: $1.9872 \text{ cal }^\circ\text{K}^{-1} \text{ mole}^{-1}$

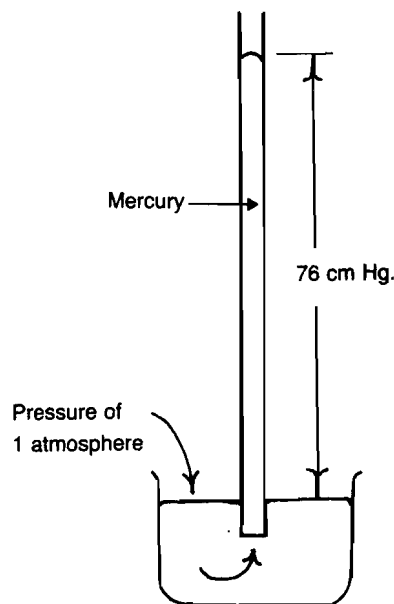


Fig. 1-8. Owing to atmospheric pressure, mercury rises to a height of 76 cm, as demonstrated here.

1-8. Convert 50,237 Pa to torrs or mm Hg, where 1 atm = 760 mm Hg = 760 torrs.

Answer: 376.8 torrs

1-9. Convert an energy of $4.379 \times 10^6 \text{ erg}$ into SI units.

Answer: 0.4379 J

1-10. A pressure of 1 atmosphere will support a column of mercury 760 mm high at 0° C (Fig. 1-8). To what height will a pressure of 1 atm support a column of mineral spirits at 25° C ? The density of mineral spirits (mineral oil fraction) at 25° C is 0.860 g/cm^3 . Express the results both in feet and in millimeters of mineral spirits. *Hint:* See Example 1-1.

Answer: 12010 mm of mineral spirits (or 39 ft)

1-11. How high can an ordinary hand-operated water pump (Fig. 1-9) lift a column of water at 25° C above its surface from a water well? How high could it lift a column of mercury at 25° C from a well filled with mercury? The density of mercury at 25° C is $13.5340 \text{ g cm}^{-3}$.

Answer: According to Harris and Hemmerling,¹¹ "since atmospheric pressure at the sea level is approximately 34 ft of water the most perfect pump of this type could not lift water more than 34 ft from the water level in the well." The same argument applies in the case of mercury.

1-12. Derive equation (1-15), shown on page 6 and used in this problem. Choose a base b for a logarithmic system. For the fun of it, you may pick a base $b = 5.9$ just because your height is 5.9 feet. Set up a log table with $b = 5.9$ for the numbers 1000, 100, 10, 0.1, and 0.001.

Answer: For the number 0.001 you would obtain $\log_{5.9}(0.001) = -3.8916$

1-13. According to Boyle's law of ideal gases, the pressure and volume of a definite mass of gas at a constant temperature are given by the equation $PV = k$ or $P = k(1/V)$, in which P is the pressure in atmospheres and V is the volume in liters. Plot the tabulated data so as to obtain a straight line and find the value of the constant k from the graph. Express the constant in ergs, joules, and calories.

Data for Problem 1-13

P (atm)	0.25	0.50	1.0	2.0	4.0
V (liters)	89.6	44.8	22.4	11.12	5.60

Answer: $k = 22.4 \text{ liter atm}$, $2.27 \times 10^{10} \text{ ergs}$, $2.27 \times 10^3 \text{ joules}$, $5.43 \times 10^2 \text{ cal}$

1-14. The distance traveled by a free-falling body released from rest is given by the equation $s = (1/2)gt^2$. Plot the accompanying data so as to obtain a straight line and determine the value of g , the

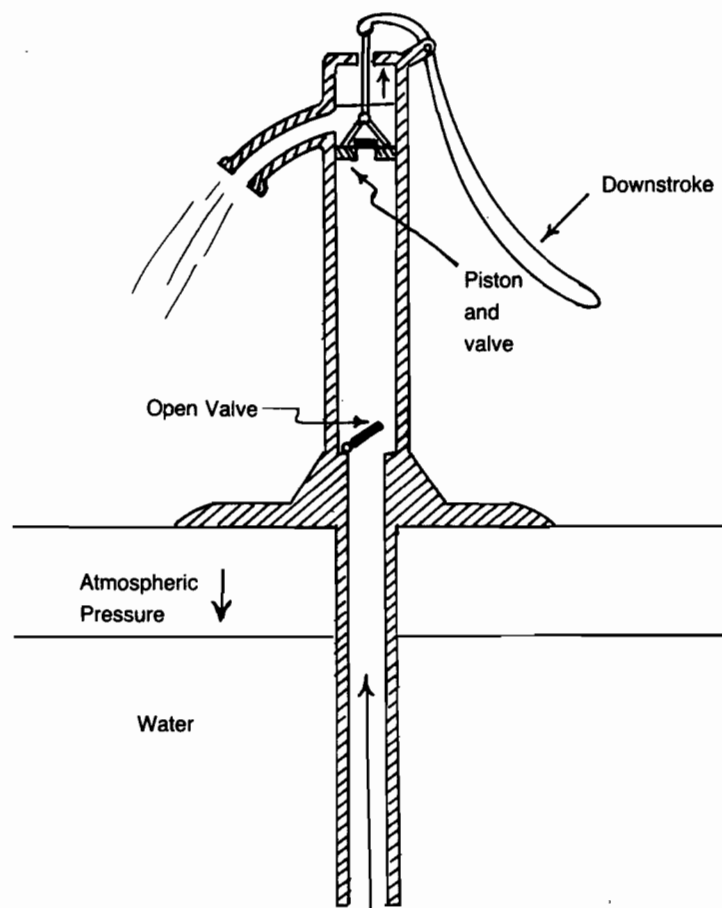


Fig. 1-9. An old-fashioned hand-operated water pump, showing the piston and valves required to lift the water from the well.

acceleration due to gravity. From the graph, obtain the time that has elapsed when the body has fallen 450 feet.

Data for Problem 1-14

s (ft)	0	16	64	144	256	400
t (sec)	0	1	2	3	4	5

Answer: $g = 32$ ft/sec; $t = 5.3$ sec

1-15. The amount of acetic acid adsorbed from solution by charcoal is expressed by the Freundlich equation, $x/m = kc^n$, in which x is the millimoles of acetic acid adsorbed by m grams of charcoal when the concentration of the acid in the solution at adsorption equilibrium is c mole/liter, and k and n are empirical constants. Convert the equation to the logarithmic form, plot $\log(x/m)$ vs. $\log c$, and obtain k and n from the graph.

Data for Problem 1-15

Millimoles of acetic acid per gram of charcoal (x/m)	Concentration of acetic acid (mole/liter) (c)
0.45	0.018
0.60	0.03
0.80	0.06
1.10	0.13
1.50	0.27
2.00	0.50
2.30	0.75
2.65	1.00

Answer: $k = 2.65$; $n = 0.42$

1-16. The equation describing the effect of temperature on the rate of a reaction is the Arrhenius equation,

$$k = Ae^{\frac{E_a}{RT}} \quad (1-36)$$

in which k is the reaction rate constant at temperature T (in degrees absolute), A is a constant known as the Arrhenius factor, R is the gas constant (1.987 cal mole⁻¹ deg⁻¹), and E_a is known as the energy of activation in cal mole⁻¹. Rearrange the equation so as to form an equation for a straight line (see equation (1-19)) and then calculate E_a and A from the following data:

Data for Problem 1-16

t (°C)	40	50	60
k	0.10	0.25	0.70

Hint: Take the natural logarithm of both sides of the Arrhenius equation.

Answer: $E_a = 20,152$ cal mole⁻¹; $A = 1.13 \times 10^{13}$ sec⁻¹

1-17. (a) Convert the equation, $F^N = \eta' G$ into logarithmic form, where F is the shear stress or force per unit area, G is rate of shear, and η' is a viscosity coefficient. N is an exponent that expresses the deviation of some solutions and pastes from the Newtonian viscosity equation (see Chapter 17, in particular pp. 454 and 456). Plot $\log G$ versus $\log F$ (common logarithms) from the table of experimental data given below, and solve for N , the slope of the line. On the same graph plot $\ln G$ vs. $\ln F$ (natural logarithms) and again obtain the slope, N . Does the value of N differ when obtained from these two lines?

(b) Directly plot the F and G values of the table on 1-cycle by 2-cycle log-log paper (i.e., graph paper with a logarithmic scale on both the horizontal and the vertical axes). Does the slope of this line yield the coefficient N obtained from the slope of the previous two plots? How can one obtain N using log-log paper?

(c) Regress $\ln G$ vs. $\ln F$ and $\log G$ vs. $\log F$ and obtain N from the slope of the two regression equations. Which method, (a), (b), or (c), provides more accurate results?

Data for Problem 1-17

G (sec ⁻¹)	22.70	45.40	68.0	106.0	140.0	181.0	272.0
F (dyne/cm ²)	1423	1804	2082	2498	2811	3088	3500

Hint: Convert G and F to \ln and to \log and enter these in a table of G and F values. Carry out the \log and \ln values to four decimal places.

Partial Answer: (c) $\log G = 2.69 \log F - 7.1018$; $r^2 = 0.9982$

$$N = 2.69, \eta' = (\text{antilog } 7.1018) = 1.26 \times 10^7$$

$$\ln G = 2.69 \ln F - 16.3556; r^2 = 0.9982$$

$$N = 2.69; \eta' = \text{antiln}(16.3556) = 1.27 \times 10^7$$

1-18. In the equation

$$-\frac{ds}{dt} = ks \quad (1-37)$$

s is the distance a car travels in time t and k is a proportionality constant. In this problem the velocity ds/dt is proportional to the distance s to be covered at any moment. Thus the speed is not constant, but rather is decreasing as the car reaches its destination, probably because the traffic is becoming heavier as the car approaches the city. The equation may be integrated to solve for k , knowing the distance s at time t and the total distance, s_0 . Separating the variables and integrating between the limits of $s = s_0$ at $t = 0$ and $s = s$ at $t = t$ yields

$$-\int_{s_0}^s \frac{ds}{s} = k \int_0^t dt \quad (1-38)$$

$(-\ln s) - (-\ln s_0) = kt$ and $\ln s = \ln s_0 - kt$, or in terms of common logs, $\log s = \log s_0 - \frac{kt}{2.303}$. The quantity 2.303 must be used because $\ln s = 2.303 \log s$, as shown on p. 5.

Knowing the remaining distance, s , at several times, one obtains k from the slope, and the total distance s_0 from the intercept

Data for Problem 1-18

s (km)	259.3	192.0	142.3	105.4
t (hr)	1	2	3	4

Compute s_0 and k using \ln and \log values with the data given in the table above. Discuss the advantages and disadvantages in using natural logarithms (\ln) and common logarithms (\log) in a problem of this kind. Why would one change from \ln to \log when plotting data, as some workers do?

Answer: Using \ln , $k = 0.300 \text{ hr}^{-1}$; $\ln s_0 = 5.8579$; $s_0 = 350 \text{ km}$. Using \log , $k = 0.300 \text{ hr}^{-1}$; $\log s_0 = 2.5441$; $s_0 = 350 \text{ km}$.

1-19. After preparing a prescription calling for six capsules each containing three grains of aspirin, you remove the contents completely and weigh each. The weights are 2.85, 2.80, 3.02, 3.05, 2.95, and 3.15 grains. Compute the average weight of the contents of the capsules, the average deviation, and the standard deviation. One gram is equal to 15.432 grains (gr.)

Answer: Av. wt. = 2.97 grains; Av. dev. = 0.103 grain; stand. dev. = 0.13 grain.

1-20. (a) Using the data in Table 1-7 and the least-squares method, calculate the slope and the intercept for the linear relationship between refractive index and percent by volume of carbon tetrachloride. Calculate the correlation coefficient, r .

(b) Use the data in Table 1-8 and the least-squares method to obtain the equation of the line plotted in Figure 1-4. Calculate the correlation coefficient, r . Compare your results with the equation shown in Figure 1-4. Explain why your results using the statistical least-squares method might differ from the equation shown in Figure 1-4.

Answers: (a) Compare your least-square results with those found in Figure 1-2, which were obtained by use of equation (1-23). $r^2 = 0.9998$.

$$(b) r^2 = 0.9986; \log y = 0.843 - 0.279 x$$

1-21. According to a principle known as Trouton's rule the molar heat of vaporization, ΔH_V (cal/mole) of a liquid divided by its boiling point (T_b) on the Kelvin scale at atmospheric pressure should equal a constant, approximately 23. If this rule holds, a plot of ΔH_V of a number of liquids against their absolute boiling points, T_b , should fall on a straight line with a slope of 23 and an intercept of 0:

$$\Delta H_V = 0 + 23T_b$$

(a) Plot ΔH_V versus T_b on rectangular coordinate paper using all the data points given in the table. With a least-squares linear regression program, obtain the slope and the intercept. Draw a line on the graph corresponding to the equation $\Delta H_V = 23T_b$.

Data for Problem 1-21

Compound	T_b (°K)	ΔH_V (cal mole ⁻¹)	$\Delta H_V/T_b$ (cal °K ⁻¹ mole ⁻¹)
Propane	231	4,812	20.8
Ethyl ether	308	6,946	22.6
Carbon disulfide	320	6,787	21.2
Hexane	342	7,627	22.3
Carbon tetrachloride	350	8,272	23.6
Cyclohexane	354	7,831	22.1
Nitrobenzene	483	12,168	25.2

(b) Repeat the regression, removing nitrobenzene from the data.

(c) Repeat the analysis using the following combinations of data (nitrobenzene is not used in any of these):

- | | | |
|------------------|------------------|----------------------|
| (1) Propane | (2) Propane | (3) Propane |
| Carbon disulfide | Ethyl ether | Carbon tetrachloride |
| Hexane | Carbon disulfide | Carbon disulfide |
| Cyclohexane | Hexane | Hexane |
| | Cyclohexane | |

(d) Compare the slopes you obtain in (a), (b), and (c1), (c2), and (c3). Which one compares best with the Trouton value of 23? Why did you get a slope in (a) quite different from the others?

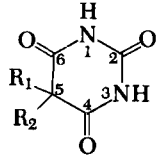
(e) Does this approach we have used, employing the equation $\Delta H_V = 23 T_b$ and linear regression analysis, appear to be a convincing proof of the Trouton rule? Can you suggest another approach to test the validity of the Trouton principle in a more convincing way? (Hint: What result would you expect if you plotted $\Delta H_V/T_b$ on the vertical axis against T_b on the horizontal axis?) Regardless of the method used, the results would probably have been much improved if a large number of organic liquids had been chosen to test the Trouton principle, but in a student problem this approach is not practical.

Answers: (using a Casio hand calculator):

- (a) $r^2 = 0.9843$; slope = 29.46; intercept = -2272
 (b) $r^2 = 0.9599$; slope = 26.14; intercept = -1255
 (c) (1) $r^2 = 0.9914$; slope = 24.65; intercept = -921.5
 (2) $r^2 = 0.9810$; slope = 24.57; intercept = -839
 (3) $r^2 = 0.9639$; slope = 27.0; intercept = -1517

1-22. A series of barbituric acids, disubstituted at the 5,5 (R_a, R_b) position, is tested for hypnotic action against rats. The relative activity required to produce hypnosis is measured for each derivative. It is presumed that this hypnotic activity, dependent variable y , may be linearly related to the logarithm of the partition coefficient, $\log K$ (see Chapter 10, p. 237 for a definition of K) as the independent variable x for each barbituric acid derivative. The observed activity

Data for Problem 1-22

Derivative		K	log K	Relative Activity (RA)		
				R_a	R_b	Observed
Ethyl,	ethyl	4.47			2.79	
Ethyl,	phenyl	26.3			3.12	
Ethyl,	amyl	141.3			3.45	
Ethyl,	butyl	44.7			3.33	
Ethyl	isobutyl	28.2			3.28	
Allyl,	cyclopentyl	97.7			3.67	
Ethyl,	1-methyl-amyl	281.8			3.60	
Ethyl,	isoamyl	89.1			3.50	
Ethyl,	cyclohexyl	61.7			3.45	
Allyl,	1-methyl-butyl	143.3			3.83	

and partition coefficient for each disubstituted barbiturate derivative are as in table above. Calculate the log K values and enter them in the space under the log K heading. Plot the observed relative activities vs. log K values and obtain the slope of the line, using two widely spaced points. Determine the intercept. Then, using linear regression, determine the slope and the intercept by the least-squares method and calculate r , the correlation coefficient. Obtain the least-squares equation of the line and use it to obtain the calculated relative activity for each compound. Enter these calculated activities in the table and record the percent difference between observed and calculated relative activities for each compound.

Answer: $r = 0.9089$; $r^2 = 0.8262$; $RA = 2.472 + 0.5268 \log K$. For the (ethyl, ethyl) derivative, the calculated activity using the least-squares equation is 2.81, which is -0.7% different from the experimental value. Incidentally, the linear relationship found between activity and log K signifies that the more nonpolar the barbiturate derivative (as measured by the partition coefficient), the more active it is as a hypnotic agent in rats. The term r^2 has more significance than the correlation coefficient r ; and r^2 of 0.8262 means that 82.62% of the barbiturate data are explained by the linear equation obtained in this problem.

(This problem came from C. Hansch, *Biological Correlations—The Hansch Approach*, American Chemical Society, 1972, pp. 30, 33. The data are from H. A. Shonle, A. K. Keltch and E. E. Swanson, *J. Am. Chem. Soc.* 52, 2440, 1930. Calculated values are given by C. Hansch, A. R. Steward, S. M. Anderson and D. Bentley, *J. Med. Chem.* 11, 1, 1968. The regression equation calculated here is slightly different from the result found in *Biological Correlations* because we have used only 10 of the 16 data points.)

1-23. The anesthetic activity of nine aliphatic ethers was plotted against the logarithm of the partition coefficient, log K . Log $1/C$ is used as a measure of anesthetic action, C being the molar concentration of each drug. A plot of the data was not linear but rather appeared to be quadratic, suggesting the need for a parabolic equation of the form

$$\log 1/C = a + b(\log K) + c(\log K)^2$$

The observed activity (log $1/C$) and the log partition coefficient for each of the substituted aliphatic ethers are found in the table. Plot log $1/C$ on the vertical axis and log K on the horizontal axis of rectangular coordinate paper to observe the parabolic nature of the curve.

Using a polynomial regression program, available on a personal computer, fit the data points with a parabolic polynomial equation,

$$y = a + bx + cx^2$$

to obtain r^2 , the y -intercept, and the regression coefficients b and c . Substitute the values of log K and $(\log K)^2$ in the parabolic equation

Data for Problem 1-23: Anesthetics in Mice*

Aliphatic Ether	Observed Activity (log $1/C$) [†]	Log Partition Coefficient (log K)
Methyl cyclopropyl	2.85	0.75
Methyl isobutyl	3.00	1.00
Methyl butyl	3.15	1.27
Ethyl tert-butyl	3.25	1.50
Propyl isobutyl	3.33	1.75
Methyl amyl	3.40	2.03
Ethyl isoamyl	3.45	2.35
Di-sec-butyl	3.43	2.57
Diisobutyl	3.35	2.90

*These data are not real but rather were arbitrarily chosen to show an example of a quadratic (parabolic) relationship. See W. Glave and C. Hansch, *J. Pharm. Sci.* 61, 589, 1972, Table I, for the actual data.

[†]The observed activity is recorded as the ED_{50} in mice.

to back-calculate the nine values of log $1/C$. Compare these calculated values with the observed anesthetic activities found in the table for each substituted ether. If the value of r^2 , called the multiple correlation coefficient when associated with multiple regression, is nearly 1.000 and the percent difference between observed and calculated log $1/C$ is small, you can assume that the polynomial equation you have chosen provides a satisfactory fit of the data.

Answer: $r^2 = 0.9964$

$$\log 1/C = 2.170 + 1.058(\log K) - 0.223(\log K)^2$$

1-24. Kamlet et al.¹² found that the logarithmic solubility (log S) of solutes in brain tissue was related to several physical properties according to the model

$$\log S = a + b(V/100) + c\pi + d\beta$$

where V is the intrinsic (van der Waals) molar volume, π is a parameter that measures solute polarity and polarizability, and β expresses the hydrogen bond acceptor basicity character of the solutes. The equation above is treated by multiple linear regression, where the dependent variable is log S and the three independent variables are $V/100$, π , and β .

(a) Using the data below and a computer program or a hand calculator (Hewlett-Packard 41V, for example) that provides the calculations for multiple linear regression of three independent variables, compute the square of the correlation coefficient, r^2 , the y -intercept, a , and the regression coefficients b , c , and d .

Data for Problem 1-24

Solutes*	log S	V/100	π	β
Methanol	1.13	0.405	0.40	0.42
Ethanol	0.69	0.584	0.40	0.45
2-Propanol	0.33	0.765	0.40	0.51
1-Propanol	0.12	0.748	0.40	0.45
Isobutyl alcohol	-0.31	0.920	0.40	0.45
Acetone	0.36	0.734	0.71	0.48
2-Butanone	-0.06	0.895	0.67	0.48
(C ₂ H ₅) ₂ O	-0.31	1.046	0.27	0.47
Benzene	-0.93	0.989	0.19	0.10
CHCl ₃	-0.53	0.805	0.38	0.10

*Selected values from Table I of Kamlet et al.¹²

(b) Use the equation you obtained in part (a) to back-calculate the log S values for the 10 cases, and compare them with the experimentally determined log S values (those found in the table). Give the percent error

$$\frac{\log S_{(\text{exper.})} - \log S_{(\text{calc.})}}{\log S_{(\text{exper.})}} \times 100$$

in the 10 calculated log S values. Do these percentage errors appear to be reasonable for a multiple linear regression? Discuss this point with the instructor and with your colleagues. (See Y. C. Martin, *Quantitative Drug Design*, Marcel Dekker, New York, 1978, pp. 194-198, to determine how well your multiple linear equation fits the data.)

Partial Answer: (a) Using a personal computer or a hand calculator capable of multiple linear-regression analysis, the square of the correlation coefficient is found to be $r^2 = 0.9811$ and the equation is

$$\log S = 1.3793 - 2.5201(V/100) - 0.1216\pi + 1.8148\beta$$

1-25. The specific gravity of alcohol is determined by measuring the mass (weight) of alcohol at 15.56° C and comparing it to the mass (weight) of an equal volume of water, taken as the standard at 15.56° C. The temperature 15.56° C is used because many years ago the United States government settled on a temperature of 60° F (15.56° C) for its testing of alcoholic products.¹³

To obtain the mass of an equal volume of water, one must know the density of water at the standard temperature, 15.56° C. The density of water at various temperatures, as found in handbooks of chemistry, is tabulated below.

Data for Problem 1-25

t (°C)	Density (g/cm ³)*
10	0.9997026
12	0.9995004
14	0.9992474
16	0.9989460
18	0.9985986
20	0.9982071

*The reader is referred to the latest handbooks for tables of values for the density of water. The values above were obtained from the *CRC Handbook of Chemistry and Physics*, 63rd Edition, pp. F5 and F6.

Plot the data and obtain an equation that will reproduce the points on the curve most accurately. If the curve is not linear, it may require the use of a quadratic or a cubic equation to represent the data:

$$\text{Density} = a + bt + ct^2$$

or

$$\text{Density} = a + bt + ct^2 + dt^3$$

Some scientific calculators (such as HP41, TI56, and Casio) and personal computers are provided with multiple regression programs.

Using the equation that best fits the data, calculate the density of water at 15.56° C. Attempt to read the density at 15.56° C directly from the graph. Which method of obtaining the density of water appears to be more accurate, calculation or direct reading?

Using your equation and direct reading from the graph, obtain the density of water at 25° C and at 37° C. Compare your results with those from a chemistry handbook. Is it safe to extrapolate your results obtained from the range of 10° to 20° C to obtain values at 25° and 37° C?

Partial Answer: The cubic equation gives the density at 25° C = 0.9970524 g/cm³; the *CRC Handbook*, p. F5, gives the density at 25° C = 0.9970479 g/cm³.

1-26. Using the data in *Problem 1-15*, compute the correlation coefficient r and the regression coefficient b (n in *Problem 1-15*).

Answer: $r = 0.9995$; $b = 0.432$. The use of a programmed hand calculator or a personal computer will provide these results. The problem may also be done by hand, following the instructions on pages 11 through 16.