1 Interpretive Tools

Chapter Objectives

At the conclusion of this chapter the student should be able to:

1. Understand the basic tools required to analyze and interpret data sets from the clinic, laboratory, or literature.
2. Describe the differences between classic dosage forms and modern drug delivery systems.
3. Use dimensional analysis.
4. Understand and apply the concept of significant figures.
5. Define determinant and indeterminant errors, precision, and accuracy.
6. Calculate the mean, median, and mode of a data set.
7. Understand the concept of variability.
8. Calculate standard deviation and coefficient of variation and understand when it is appropriate to use these parameters.
9. Use graphic methods to determine the slope of lines.
10. Interpret slopes of lines and how they relate to absorption and elimination from the body.

Introduction

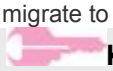
"One of the earmarks of evidence-based medicine is that the practitioner should not just accept the conventional wisdom of his/her mentor. Evidence-based medicine uses the scientific method of using observations and literature searches to form a hypothesis as a basis for appropriate medical therapy. This process necessitates education in basic sciences and an understanding of basic scientific principles."[1,2] Today more than ever before, the pharmacist and the pharmaceutical scientist are called upon to demonstrate a sound knowledge of biopharmaceutics, biochemistry, chemistry, pharmacology, physiology, and toxicology and an intimate understanding of the physical, chemical, and biopharmaceutical properties of medicinal products. Whether engaged in research and development, teaching, manufacturing, the practice of pharmacy, or any of the allied branches of the profession, the pharmacist must recognize the need to rely heavily on 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 boundaries between the various sciences and must keep abreast of advances in the physical, chemical, and biological fields in order to understand and contribute to the rapid developments in his or her profession. You are also expected to provide concise and practical interpretations of highly technical drug information to your patients and colleagues. With the abundance of information and misinformation that is freely and publicly available (e.g., on the Internet), having the tools and ability to provide meaningful interpretations of results is critical.

Historically, *physical pharmacy* has been associated with the area of pharmacy that dealt with the quantitative and theoretical principles of physicochemical science as they applied to the practice of pharmacy. Physical pharmacy attempted to integrate the factual knowledge of pharmacy through the development of broad principles of its own, and it aided the pharmacist and the pharmaceutical scientist in their attempt to predict the solubility, stability, compatibility, and biologic action of drug products. Although this remains true today, the field has become even more highly integrated into the biomedical aspects of the practice of pharmacy. As such, the field is more broadly known today as the *pharmaceutical sciences* and the chapters that follow reflect the high degree of integration of the biological and physical–chemical aspects of the field.

Developing new drugs and delivery systems and improving upon the various modes of administration are still the primary goals of the pharmaceutical scientist. A practicing pharmacist must also possess a thorough understanding of modern drug delivery systems as he or she advises patients on the best use of prescribed medicines. In the past, drug delivery focused nearly exclusively on *pharmaceutical*

technology (in other words, the manufacture and testing of tablets, capsules, creams, ointments, solutions, etc.). This area of study is still very important today. However, the pharmacist needs to understand how these delivery systems perform in and respond to the normal and pathophysiologic states of the patient. The integration of physical–chemical and biological aspects is relatively new in the pharmaceutical sciences. As the field progresses toward the complete integration of these subdisciplines, the impact of the biopharmaceutical sciences and drug delivery will become enormous. The advent and commercialization of molecular, nanoscale, and microscopic drug delivery technologies is a direct result of the integration of the biological and physical–chemical sciences. In the past, a dosage (or dose) form and a drug delivery system were considered to be one and the same. *A dosage form* is the entity that is administered to patients so that they receive an effective dose of a drug. The traditional understanding of how an oral dosage form, such as a tablet, works is that a patient takes it by mouth with some fluid, the tablet disintegrates, and the drug dissolves in the stomach and is then absorbed through the intestines into the bloodstream. If the dose is

too high, a lower-dose tablet may be prescribed. If a lower-dose tablet is not commercially available, the patient may be instructed to divide the tablet. However, a pharmacist who dispenses a nifedipine (Procardia XL) extended-release tablet or an oxybutynin (Ditropan XL) extended-release tablet to a patient would advise the patient not to bite, chew, or divide the “tablet.” The reason for this is that the tablet dosage form is actually an elegant osmotic pump *drug delivery system* that looks like a conventional tablet (see Key Concept Box on Dosage Forms and Drug Delivery Systems). This creative and elegant approach solves numerous challenges to the delivery of pharmaceutical care to patients. On the one hand, it provides a sustained-release drug delivery system to patients so that they take their medication less frequently, thereby enhancing patient compliance and positively influencing the success rate of therapeutic regimens. On the other hand, patients see a familiar dosage form that they can take by a familiar route of administration. In essence, these osmotic pumps are delivery systems packaged into a dosage form that is familiar to the patient. The subtle differences between dose forms and delivery systems will become even more profound in the years to come as drug delivery systems successfully migrate to the molecular scale.



Key Concept

Pharmaceutical Sciences

Pharmacy, like many other applied sciences, has passed through a descriptive and empiric era. Over the past decade a firm scientific foundation has been developed, allowing the “art” of pharmacy to transform itself into a quantitative and mechanistic field of study. The integration of the biological, chemical, and physical sciences remains critical to the continuing evolution of the pharmaceutical sciences. The theoretical links between the diverse scientific disciplines that serve as the foundation for pharmacy are reflected in this book. 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.

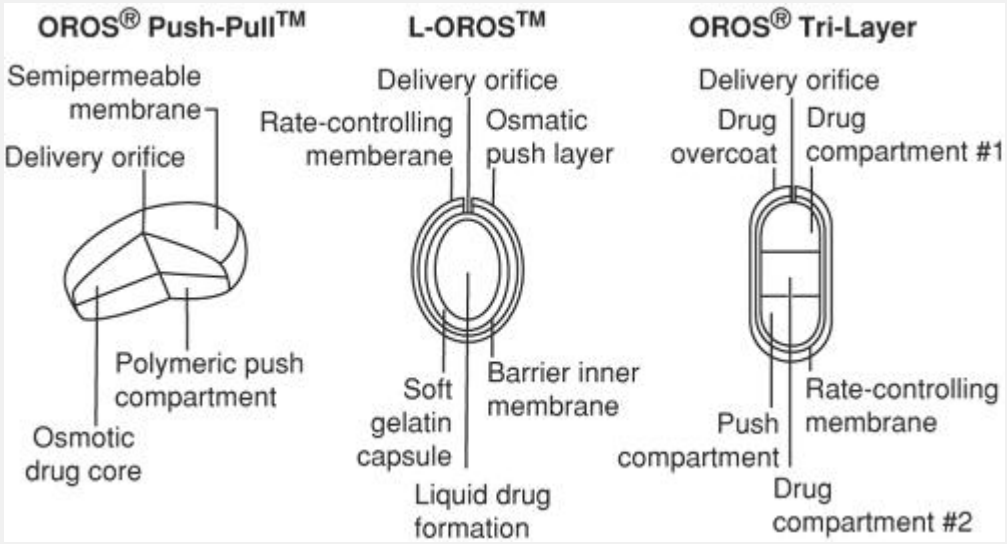
Key Concept

Dosage Forms and Drug Delivery Systems

A Procardia XL extended-release tablet is similar in appearance to a conventional tablet. It consists, however, of a semipermeable membrane surrounding an osmotically active drug core. The core is divided into two layers: an “active” layer containing the drug and a “push” layer containing pharmacologically inert but osmotically active components. As fluid from the gastrointestinal (GI) tract enters the tablet, pressure increases in the osmotic layer and “pushes” against the drug layer, releasing drug through the precision laser-drilled tablet orifice in the active layer. Procardia XL is designed to provide nifedipine at an approximately constant rate over 24 hr. This controlled rate of drug delivery into the GI lumen is independent of pH or GI motility. The nifedepine release profile from Procardia XL depends on the

existence of an osmotic gradient between the contents of the bilayer core and the fluid in the GI tract. Drug delivery is essentially constant as long as the osmotic gradient remains constant, and then gradually falls to zero. Upon being swallowed, the biologically inert components of the tablet remain intact during GI transit and are eliminated in the feces as an insoluble shell. The information that the pharmacist provides to the patient includes “Do not crush, chew, or break the extended-release form of Procardia XL. These tablets are specially formulated to release the medication slowly into the body. Swallow the tablets whole with a glass of water or another liquid. Occasionally, you may find a tablet form in the stool. Do not be alarmed, this is the outer shell of the tablet only, the medication has been absorbed by the body.” On examining the figure, you will notice how the osmotic pump tablet looks identical to a conventional tablet.

Remember, most of the time when a patient takes a tablet, it is also the delivery system. It has been optimized so that it can be mass-produced and can release the drug in a reproducible and reliable manner. Complete disintegration and deaggregation occurs and there is little, if any, evidence of the tablet dose form that can be found in the stool. However, with an osmotic pump delivery system, the “tablet” does not disintegrate even though all of the drug will be released. Eventually, the outer shell of the depleted “tablet” passes out of the body in the stool.



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. This chapter provides a foundation for interpreting the observations and results that come from careful

scientific study. At the conclusion of this chapter, you should have the ability to integrate facts and ideas into a meaningful whole and concisely convey a sense of that meaning to a third party. For example, if you are a pharmacy practitioner, you should be able to translate a complex scientific principle to a simple, practical, and useful recommendation for a patient.

Key Concept

“”我听见 我忘记;我看见 我记住;我做 我了解:
I HEAR AND I FORGET. I SEE AND I REMEMBER. I DO AND I UNDERSTAND.”

The ancient Chinese proverb emphasizes the value of active participation in the learning process. Through the illustrative examples and practice problems in this book and on the online companion Web site, the student is encouraged to actively participate.

The comprehension of course material is primarily the responsibility of the student. The teacher can guide and direct, explain, and clarify, but competence 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.

Measurements, Data, Propagation of Uncertainty

The goal of this chapter is to provide a foundation for the *quantitative reasoning* skills that are fundamental to the pharmacy practitioner and pharmaceutical scientist. “As mathematics is the language of science, statistics is the logic of science.”3 Mathematics and statistics are fundamental tools of the pharmaceutical sciences. You need to understand how and when to use these tools, and how to interpret what they tell us. You must also be careful not to overinterpret results. On the one hand, you may ask “do we really need to know how these equations and formulas were derived in order to use them effectively?” Logically, the answer would seem to be no. By analogy, you do not need to know how to build a computer in order to use one to send an e-mail message, do you? On the other hand, graphically represented data convey a sense dynamics that benefit from understanding a bit more about the fundamental equations behind the behavior. These equations are merely tools (that you should not memorize!) that allow for the transformation of a bunch of numbers into a behavior that you can interpret.

The mathematics and statistics covered in this chapter and this book are presented in a format to promote understanding and practical use. Therefore, many of the basic mathematical “tutorial” elements have been removed from the sixth edition, and in particular this chapter, because of the migration of numerous college-level topics to secondary school courses over the years. However, if you believe that you need a refresher in basic mathematical concepts, this information is still available in the online companion to this text (at thePoint.lww.com/Sinko6e). Statistical formulas and graphical method explanations have also been dramatically reduced in this edition. Depending on your personal goals and the philosophy of your program of study, you may well need an in-depth treatment of the subject matter. Additional detailed treatments can be found on the Web site and in the recommended readings.

Data Analysis Tools

Readily available tools such as programmable calculators, computer spreadsheet programs (e.g., Microsoft Excel, Apple Numbers, or OpenOffice.org Calc), and statistical software packages (e.g., Minitab, SAS or SPSS) make the processing of data relatively easy. Spreadsheet programs have two distinct advantages: (1) data collection/entry is simple and can often be automated, reducing the possibility of errors in transcription, and (2) simple data manipulations and elementary statistical calculations are also easy to perform. In addition, many spreadsheet programs seamlessly interface with statistical packages when more robust statistical analysis is required. With very little effort, you can add data sets and generate pages of analysis. The student should appreciate that while it may be possible to automate data entry and have the computer perform calculations, the final interpretation of the results and statistical analysis is your responsibility! As you set out to analyze data keep in mind the simple acronym *GIGO—Garbage In, Garbage Out*. In other words, solid scientific results and sound methods of analysis will yield meaningful interpretations and conclusions. However, if the scientific foundation is weak, there are no known statistical tools that can make bad data significant.

Dimensional Analysis

Dimensional analysis (also called the factor-label method or the unit factor method) is a problem-solving method that uses the fact that any number or expression can be multiplied by 1 without changing its value. For example, since 2.2 kg = 1 lb, you can divide both sides of the relationship by "1 lb," resulting in 2.2 kg/1 lb = 1. This is known as a *conversion factor*,

which is a ratio of like-dimensioned quantities and is equal to the dimensionless unity (in other words, is equal to 1). On the face of it, the concept may seem a bit abstract and not very practical. However, dimensional analysis is very useful for any value that has a "unit of measure" associated with it, which is nearly everything in the pharmaceutical sciences. Simply put, this is a practical method for converting the units of one item to the units of another item.

Example 1-1

Solving problems using dimensional analysis is straightforward. You do not need to worry about the actual numbers until the very end. At first, simply focus on the units. Plug in all of the conversion factors that cancel out the units you do not want until you end up with the units that you do want. Only then do you need to worry about doing the calculation. If the units work out, you will get the right answer every time. In this example, the goal is to illustrate how to use the method for converting one value to another.

Question: How many seconds are there in 1 year?

Conversion Factors:

365 days = 1 year 24 hr = 1 day 60 min = 1 hr 60 sec = 1 min

Rearrange Conversion Factors:

$$\frac{365\ \text{days}}{1\ \text{year}} = 1 \quad \frac{24\ \text{hr}}{1\ \text{day}} = 1 \quad \frac{60\ \text{min}}{1\ \text{hr}} = 1 \quad \frac{60\ \text{sec}}{1\ \text{min}} = 1$$

Solve (arrange conversion factors so that the units that you do not want cancel out):

$$\frac{365\ \text{days}}{1\ \text{year}} \times \frac{24\ \text{hr}}{1\ \text{day}} \times \frac{60\ \text{min}}{1\ \text{hr}} \times \frac{60\ \text{sec}}{1\ \text{min}}$$

as you see the units become seconds.

Calculate: Now, plug the numbers carefully into your calculator and the resulting answer is 31,536,000 sec/year.

Example 1-2

This example will demonstrate the use of dimensional analysis in performing a calculation. How many 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. This is the key conversion factor required to solve this problem. The question can then be asked keeping in

mind the conversion factor: If 1 cal equals 4.184 joules, how many calories are there in 3.00 joules? Write down the conversion factor, being careful to express each quantity in its proper units. For the unknown quantity, use an *X*.

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

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

Table 1-1 Writing or Interpreting Significant figures in Numbers

Rule	Example
All nonzero digits are considered significant	98.513 has five significant figures: 9, 8, 5, 1, and 3
Leading zeros are not significant	0.00361 has three significant figures: 3, 6, and 1
Trailing zeros in a number containing a decimal point are significant	998.100 has six significant figures: 9, 9, 8, 1, 0, and 0
The significance of trailing zeros in a number not containing a decimal point can be ambiguous	The number of significant figures in numbers like 11,000 is uncertain because a decimal point is missing. If the number was written as 11,000, it would be clear that there are five significant figures
Zeros appearing anywhere between two nonzero digits are significant	607.132 has six significant figures: 6, 0, 7, 1, 3, and 2

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 the 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:

$$X \text{ (in gallons)} = 2.0 \text{ liter} \times (1000 \text{ mL/liter}) \times (1 \text{ pint}/473 \text{ mL}) \times (1 \text{ gallon}/8 \text{ pints})$$

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

One may be concerned about the apparent disregard for the rules of significant figures 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.

Significant Figures

A significant figure is any digit used to represent a magnitude or a quantity in the place in which it stands. The rules for interpreting significant figures and some examples are shown in Table 1-1. Significant figures give a sense of the accuracy of a number. They include all digits except leading and trailing zeros where they are used merely to locate the decimal point. Another way to state this is, 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 such as “not less than 99” is given in an official compendium, it means 99.0 and not 98.9.

Example 1-4

How Many Significant Figures in the Number 0.00750?

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 because it is not needed to write the number; if it were not significant, it could be omitted. Thus, the value contains three significant figures.

How Many Significant Figures in the Number 7500?

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. *Hint*: 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.

Significant figures are particularly useful for indicating the precision of a result. The proper interpretation of a value may be questioned specifically in cases when performing calculations (e.g., when spurious digits introduced by calculations carried out to greater accuracy than that of the original data) or when reporting measurements to a greater precision than the equipment supports. It is important to remember that the instrument used to make the measurement limits the precision of the resulting value that is reported. For example, 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. These zeros are not significant, however, 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.

Key Concept

“When Significant Figures do Not Apply”

Since significant figure rules are based upon estimations derived from statistical rules for handling probability distributions, they apply only to *measured* values. The concept of significant figures does not pertain to values that are known to be exact. For example, integer counts (e.g., the number of tablets dispensed in a prescription bottle); legally defined conversions such as 1 pint = 473 mL; constants that are defined arbitrarily (e.g., a centimeter is 0.01 m); scalar operations such as "doubling" or "halving"; and mathematical constants, such as π and *e*. However, physical constants such as Avogadro's number have a limited number of significant figures since the values for these constants are derived from measurements.

Example 1-5

The following example is used to illustrate *excessive precision*. If a faucet is turned on and 100 mL of water flows from the spigot in 31.47 sec, what is the average volumetric flow rate? By dividing the volume by time using a calculator, we get a rate of 3.177629488401652 mL/sec. Directly stating the uncertainty is the simplest way to indicate the precision of any result. Indicating the flow rate as 3.177 ± 0.061 mL/sec is one way to accomplish this. This is particularly appropriate when the uncertainty itself is important and precisely known. If the degree of precision in the answer is not important, it is acceptable to express trailing digits that are not known exactly, for example, 3.1776 mL/sec. If the precision of the result is not known you must be careful in how you report the value. Otherwise, you may overstate the accuracy or diminish the precision of the result.

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.
2. 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.
3. 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:

   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 are discussed in some detail by Brecht.5

4. In multiplication or division, the rule commonly used is to retain the same number of significant figures in the result as appears 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.

5. 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. Because calculations involved in theoretical 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 calculator is more convenient, however, and tables of logarithms are rarely used today.
6. 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.

Remember, significant figures are not meant to be a perfect representation of uncertainty. Instead, they are used to prevent the loss of precision when rounding numbers. They also help you avoid stating more information than you actually know. Error and uncertainty are not the same. For example, if you perform an experiment in triplicate (in other words, you repeat the experiment three times), you will get a value that looks something like 4.351±0.076. This does not mean that you made an error in the experiment or the collection of the data. It simply means that the outcome is naturally statistical.

Data Types

The scientist is continually attempting to relate phenomena and establish generalizations with which to consolidate and interpret experimental data. The problem frequently resolves 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

$$y \propto x \qquad (1\text{–}1)$$

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,

$$\frac{y_1}{x_1} = \frac{y_2}{x_2} = \ldots \qquad (1\text{–}2)$$

Because 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

$$\frac{y}{x} = \text{Constant} \qquad (1\text{–}3)$$

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

$$y = kx \qquad (1\text{–}4)$$

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

$$y = f(x) \qquad (1\text{–}5)$$

which is read "*y* is some function of *x*." That is, *y* may be equal, for example, to $2x$, to $27x^2$, or to $0.0051 + \log(a/x)$. The functional notation in equation **(**1-5**)** merely signifies that *y* and *x* are related in some way without specifying the actual equation by which they are connected.

As we begin to lay the foundation for the interpretation of data using descriptive statistics, some background information about the types of data that you will encounter in the pharmaceutical sciences is needed. In 1946, Stevens defined measurement as "the assignment of numbers to objects or events according to a rule."4 He proposed a classification system that is widely used today to define data types. The first two, *intervals* and *ratios*, are categorized as continuous variables. These would include results of laboratory measurements for nearly all of the data that are normally collected in the laboratory (e.g., concentrations, weights). Only ratio or interval measurements can have units of measurement, and

these variables are quantitative in nature. In other words, if you were given a set of “interval” data you would be able to calculate the exact differences between the different values. This makes this type of data “quantitative.” Since the interval between measurements can be very small, we can also say that the data are “continuous.” Another laboratory example of interval data measures is temperature. Think of the gradations on a common thermometer (in Celsius or Fahrenheit scale)—they are typically spaced apart by 1 degree with minor gradations at the 1/10th degree. The intervals could become even smaller; however, because of the physical limitations of common thermometers, smaller gradations are not possible since they cannot be read accurately. Of course, with digital thermometers the gradations (or intervals) could be much smaller but then the precision of the thermometer may become questionable. Another temperature scale that will be used in various sections of this text is the Kelvin scale, a thermodynamic temperature scale. By international agreement,

the Kelvin and Celsius scales are related through the definition of absolute zero (in other words, 0 K = -273.15°C). Since the thermodynamic temperature is measured relative to absolute zero, the Kelvin scale is considered a ratio measurement. This also holds true for other physical quantities such as length or mass. The third common data type in the pharmaceutical sciences is *ordinal scale* measurements. Ordinal measurements represent the rank order of what is being measured. “Ordinals” are more subjective than interval or ratio measurements.

The final type of measurement is called *nominal* data. In this type of measurement, there is no order or sequence of the observations. They are merely assigned different groupings such as by name, make, or some similar characteristic. For example, you may have three groups of tablets: white tablets, red tablets, and yellow tablets. The only way to associate the various tablets is by their color. In clinical research, variables measured at a nominal level include sex, marital status, or race. There are a variety of ways to classify data types and the student is referred to texts devoted to statistics such as those listed in the recommended readings at the end of this chapter.6,7

Error and Describing Variability

If one is to maintain a high degree of accuracy in the compounding of prescriptions, the manufacture of products on a large scale, or the analysis of clinical or laboratory research results, one must know how to locate and eliminate constant and accidental errors as far 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, can 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, and 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, although not constant, is a systematic error that can also be determined and accounted for once the coefficient of expansion is known.

Determinate errors can be reduced in analytic work by using a calibrated apparatus, using blanks and controls, using several different analytic procedures and apparatus, eliminating impurities, and carrying out the experiment under varying conditions. In pharmaceutical manufacturing, determinate errors can

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, whereas 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, whereas *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.

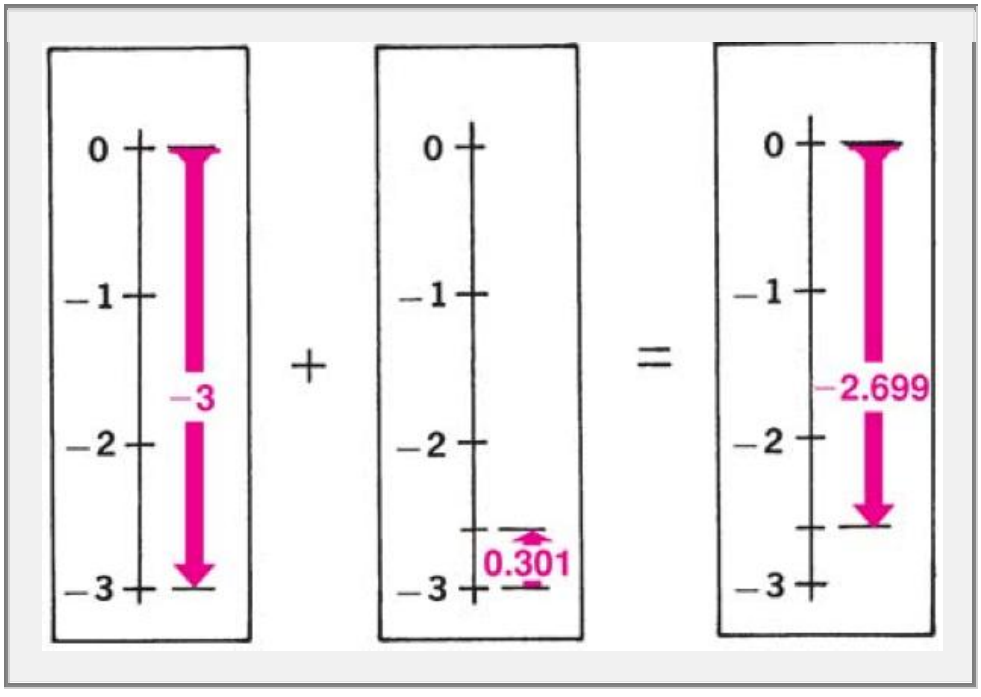


Fig. 1-1. The normal curve for the distribution of indeterminate errors.

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 having various errors along the vertical axis against the magnitude of the errors on the horizontal axis, one obtains a bell-shaped curve, known as a *normal frequency distribution curve*, as shown in Figure 1-1. 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.

Descriptive Statistics

Since the typical pharmacy student has sufficient exposure to descriptive statistics in other courses, this section will focus on introducing (or reintroducing) some of the key concepts that will be used numerous times in later chapters. The student who requires additional background in statistics is advised to seek out one of the many outstanding texts that have been published.[6,7] Descriptive statistics depict the basic features of a data set collected from an experimental study. They give summaries about the sample and the measures. However, viewing the individual data and tables of results alone is not always sufficient to understand the behavior of the data. Typically, a graphic analysis is paired with a tabular description to perform a quantitative analysis of the data set. The third component of descriptive statistics is "summary" statistics. These are single numbers that summarize the data. With interval data (e.g., the dose strength of individual tablets in a batch of 10,000 tablets), summary statistics focus on how big the value is and the variability among the values. The first of these aspects relates to measures of "central tendency" (e.g., what is the average?), while the second refers to "dispersion" (in other words, the "variation" among a group of values).

Central Tendency: Mean, Median, Mode

Central tendency can be described using a summary statistic (the mean, median, or mode) that gives an indication of the average value in the data set. The theoretical 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 [X with bar above] 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

$$\overline{X} = \frac{\sum(X_i)}{N} \qquad (1\text{–}6)$$

in which Σ stands for "the sum of," X_i is the ith individual measurement of the group, and N is the number of values. [X with bar above] is an estimate of μ and approaches it as the number of measurements N is increased. Remember, the "equations" used in all of the calculations are really a shorthand notation describing the various relationships that define some parameter.

Example 1-6

A new student has just joined the lab and is being trained to pipette liquids correctly. She is using a 1-mL pipettor and is asked to withdraw 1 mL of water from a beaker and weigh it on a balance in a weighing boat. To determine her pipetting skill, she is asked to repeat this 10 times and take the average. What is the average volume of water that the student withdraws after 10 repeats? The density of water is 1 g/mL.

Attempt	Weight (g)
1	1.05
2	0.98
3	0.95
4	1.00
5	1.02
6	1.00
7	1.10
8	1.03
9	0.96
10	0.98

If ΣX_i = 9.99 and N = 10, so 9.99/10 = 0.999. Given the number of significant figures, the average would be reported as 1.00 g, which equals 1 mL since the density of water is 1 g/mL.

The *median* is the middle value of a range of values when they are arranged in rank order (e.g., from lowest to highest). So, the median value of the list [1, 2, 3, 4, 5] is the number 3. In this case, the mean is also 3. So, which value is a better indicator of the central tendency of the data? The answer in this case is neither—both indicate central tendency equally well. However, the value of the median as a summary statistic

becomes more obvious when the data set is skewed (in other words, when there are outliers or data points with values that are quite different from most of the others in the data set). For example, in the data set [1, 2, 2, 3, 10] the mean would be 3.6 but the median would be 2. In this case, the median is a better summary statistic than the mean because it gives a better representation of central tendency of the data set. Sometimes the median is referred to as a more “robust” statistic since it gives a reasonable outcome even with outlier results in the data set.

Example 1-7

As you have seen, calculating the median of a data set with an odd number of results is straightforward. But, what do you do when a data set has an even number of members? For example, in the data set [1, 2, 2, 3, 4, 10] you have 6 members to the data set. To calculate

the median you need to find the two middle members (in this case, 2 and 3) then average them. So, the median would be 2.5.

Although it is human nature to want to “throw out” an outlying piece of data from a data set, it is not proper to do so under most circumstance or at least without rigorous statistical analysis. Using median as a summary statistic allows you to use all of the results in a data set and still get an idea of the central tendency of the results.

The *mode* is the value in the data set that occurs most often. It is not as commonly used in the pharmaceutical sciences but it has particular value in describing the most common occurrences of results that tend to center around more than one value (e.g., a bimodal distribution that has two commonly occurring values). For example, in the data set [1, 2, 4, 4, 5, 5, 5, 6, 9, 10] the mode value is equal to 5. However, we sometimes see a data set that has two “clusters” of results rather than one. For example, the data set [1, 2, 4, 4, 5, 5, 5, 6, 9, 10, 11, 11, 11, 11, 13, 14] is bimodal and thus has two modes (one mode is 5 and the other is 11). Taking the arithmetic mean of the data set would not give an indication of the bimodal behavior. Neither would the median.

Variability: Measures of Dispersion

In order to fully understand the properties of the data set that you are analyzing, it is necessary to convey a sense of the dispersion or scatter around the central value. This is done so that an estimate of the variation in the data set can be calculated. This variability is usually expressed as the *range*, the *mean deviation*, or the*standard deviation*. Another useful measure of dispersion commonly used in the pharmaceutical sciences is the *coefficient of variation* (CV), which is a dimensionless parameter. Since much of this will be a review for many of the students using this text, only the most pertinent features will be discussed. The results obtained in the physical, chemical, and biological aspects of pharmacy have different characteristics. In the physical sciences, for example, instrument measurements are often not perfectly reproducible. In other words, variability may result from random measurement errors or may be due to errors in observations. In the biological sciences, however, the source of variation is viewed slightly differently since members of a population differ greatly. In other words, biological variations that we typically observe are intrinsic to the individual, organism, or biological process.

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 ð of the population. The sum of the positive and negative deviations about the mean equals zero; hence, the algebraic signs are disregarded 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 [X with bar above], 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

$$d = \frac{\sum |X_i - \overline{X}|}{N} \qquad (1\text{–}7)$$

in which $\Sigma|X_i$-[X with bar above]| 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 because 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 lowercase 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* and σ is, therefore, called the *population standard deviation*. Population standard deviations are shown in Figure 1-1.

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 letter *s*. The formula is

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{N}} \qquad (1\text{–}8)$$

For a small sample, the equation is written

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{N - 1}} \qquad (1\text{–}9)$$

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 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 **(**1-9**)**. When *N* is large, say *N* > 100, we can use *N* instead of (*N* - 1) to estimate the population standard deviation because the difference between the two is negligible.

Modern statistical methods handle small samples 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 population standard deviation as fewer samples are used to compute the estimate.

However, for many students studying pharmacy there is no compelling reason to view standard deviation in highly technical terms. So, we will simply refer to standard deviation as “SD” from this point forward.

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

Example 1-8

A pharmacist receives a prescription for a patient with rheumatoid arthritis calling for seven divided powders, each of which is to weigh 1.00 g. To check his skill in filling the powders, he removes the contents from each paper after filling the prescription by the block-and-divide method and then weighs the powders carefully. The results of the weighings are given in the first column of Table 1-2; 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 can be expressed as 0.98 ± 0.046 g. The variability of a single powder can also be expressed in terms of percentage deviation by dividing the mean deviation by the arithmetic mean and multiplying

by 100. The result is 0.98% ± 4.6%; of course, it includes errors due to removing the powders from the papers and weighing the powders in the analysis.

Table 1-2 Statistical Analysis of Divided Powder Compounding Technique

	Weight of Powder Contents (g)	Deviation (Sign Ignored), $\lvert X_i - [\bar{X}] \rvert$	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	Σ = 6.84	Σ = 0.32	Σ = 0.0362
Average	0.98	0.046	

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

Statisticians have 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 1-1.

Goldstein7 selected 1.73δ 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 $\pm 2s$ 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 1-8 is calculated as follows:

$$s = \sqrt{\frac{0.0362}{(7-2)}} = 0.078 \text{ g}$$

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

The smaller the standard deviation estimate (or the mean deviation), the more *precise* is the 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—because 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 1-8, the true value is 1.00 g, the amount requested by the physician. The apparent error involved in compounding this prescription is

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

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 1-8 is sufficiently great that no systemic error can be presumed. However, on further analysis it is found that one or several results are questionable. This possibility is considered later. If the arithmetic mean inExample 1-8 were 0.90 instead of 0.98, the difference could be stated with assurance to have statistical significance because the probability that such a result could occur by chance alone would be small.

The mean error in this case is

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

The *relative error* is obtained by dividing the mean error by the true value. It can 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

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

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 1-8 had yielded an average weight of 0.60 g with a mean deviation of 0.5%, the results would have been accepted as precise. The degree of accuracy, however, would have been low because 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 observed, “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 1-8 (Table 1-2), we note one rather discordant value, namely, 0.81 g. If the arithmetic mean is recalculated ignoring this measurement, we obtain a mean of 1.01 g. The mean deviation without the doubtful result is 0.02 g. It is now seen that the divergent result is 0.20 g 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 is

probably caused by some definite error in technique. Statisticians rightly question this rule, but it is a useful though not always reliable criterion for finding discrepant results.

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 become a safe and proficient compounder before entering the practice of pharmacy.

The CV is a dimensionless parameter that is quite useful. The CV relates the standard deviation to the mean and is defined as

$$\mathrm{CV} = \mathrm{SD}/\mathrm{mean} \qquad (1\text{–}10)$$

It is valid only when the mean is nonzero. It is also commonly reported as a percentage (%CV is CV multiplied by 100). For example, if SD = 2 and mean = 3, then the %CV is 67%. The CV is useful because the standard deviation of data must always be understood in the context of the mean of the results. The CV should be used instead of the standard deviation to assess the difference between data sets with dissimilar units or very different means.

Visualizing Results: Graphic Methods, Lines

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 will allow expression of the connection in the form of a mathematical equation. The procedure of obtaining an empirical 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.

We will first go through some of the technical aspects of lines and linear relationships. The simplest relationship between two variables, in which the variables contain no exponents other than 1, yields a straight line when plotted using rectangular coordinates. The straight-line or linear relationship is expressed as

$$y = a + bx \qquad (1\text{–}11)$$

in which y is the dependent variable, x is the independent variable, and a and b are constants. The constant b is the *slope* of the line; the greater the value of b, the steeper is the slope. It is expressed as the change in y with the change in x, or $b = \frac{\Delta y}{\Delta x}$; 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, respectively. When $b = 1$, the line makes an angle of 45° with the x axis and the equation of the line can be written as follows:

$$y = a + x \qquad (1\text{–}12)$$

When $b = 0$, the line is horizontal (in other words, parallel to the x axis), and the equation reduces to

$$y = a \qquad (1\text{–}13)$$

The constant a is known as the *y intercept* and denotes 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 it is negative, the line intersects the y axis below the x axis. When a is zero, equation (1-11) may be written as

$$y = bx \qquad (1\text{–}14)$$

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 shown in Table 1-3 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

Table 1-3 Refractive Indices of Mixtures of Benzene and Carbon Tetrachloride

Concentration of CCl_4 (x) (Volume %)	Refractive Index (y)
10.0	1.497
25.0	1.491
33.0	1.488
50.0	1.481
60.0	1.477

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \qquad (1\text{–}15)$$

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 1-9

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 let 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 (1-11). A negative 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 (1-11). It can be obtained

from the plot in Figure 1-2 by *extrapolating* (extending) the line upward to the left until it intersects the y axis. It will also be observed that

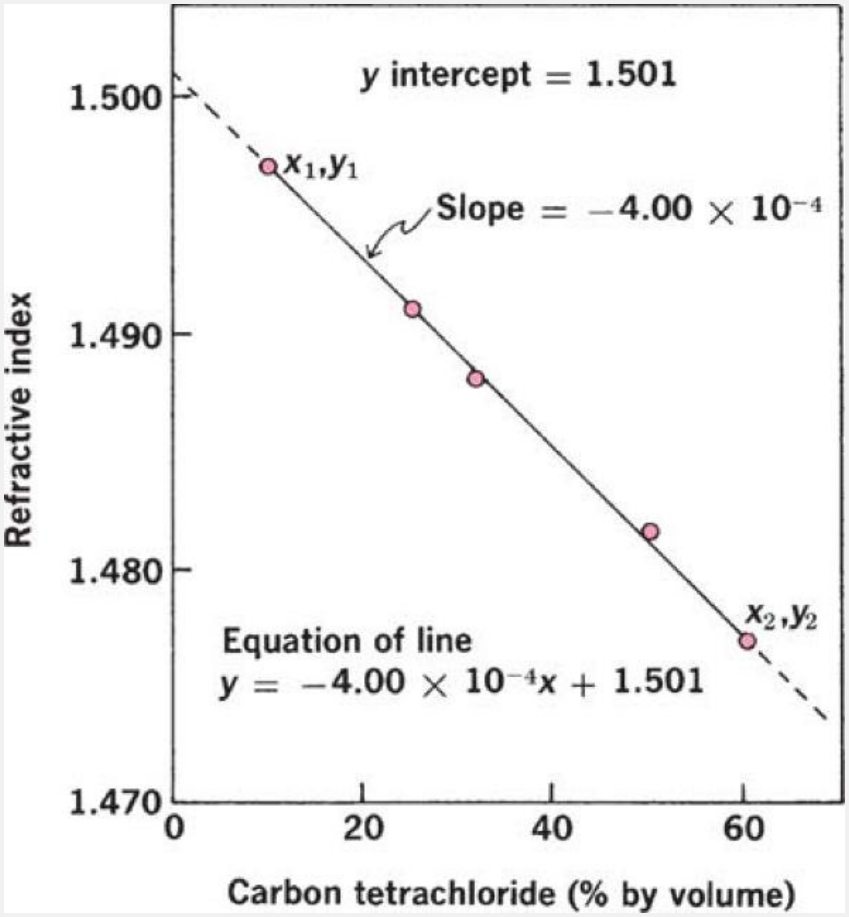


Fig. 1-2. Refractive index of the system benzene–carbon tetrachloride at 20°C.

Table 1-4 Emulsion Stability as a Function of Emulsifier Concentration

Emulsifier (x) (% Concentration)	Oil Separation (y) (mL/month)	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

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x} = b \qquad (1\text{–}16)$$

and this simple formula allows one to compute the slope of a straight line.

Not all experimental data form straight lines. 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 can be found in standard textbooks on analytic geometry.

Logarithmic relationships occur frequently in scientific work. 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 1-4.

The data from this experiment may be plotted in several ways. In Figure 1-3, the oil separation y is plotted as ordinate against the emulsifier concentration x as abscissa on a rectangular coordinate grid. In Figure 1-4, the logarithm of the oil separation is plotted against the concentration. In Figure 1-5, the data are plotted using semilogarithmic scale, consisting of a logarithmic scale on the vertical axis and a linear scale on the horizontal axis.

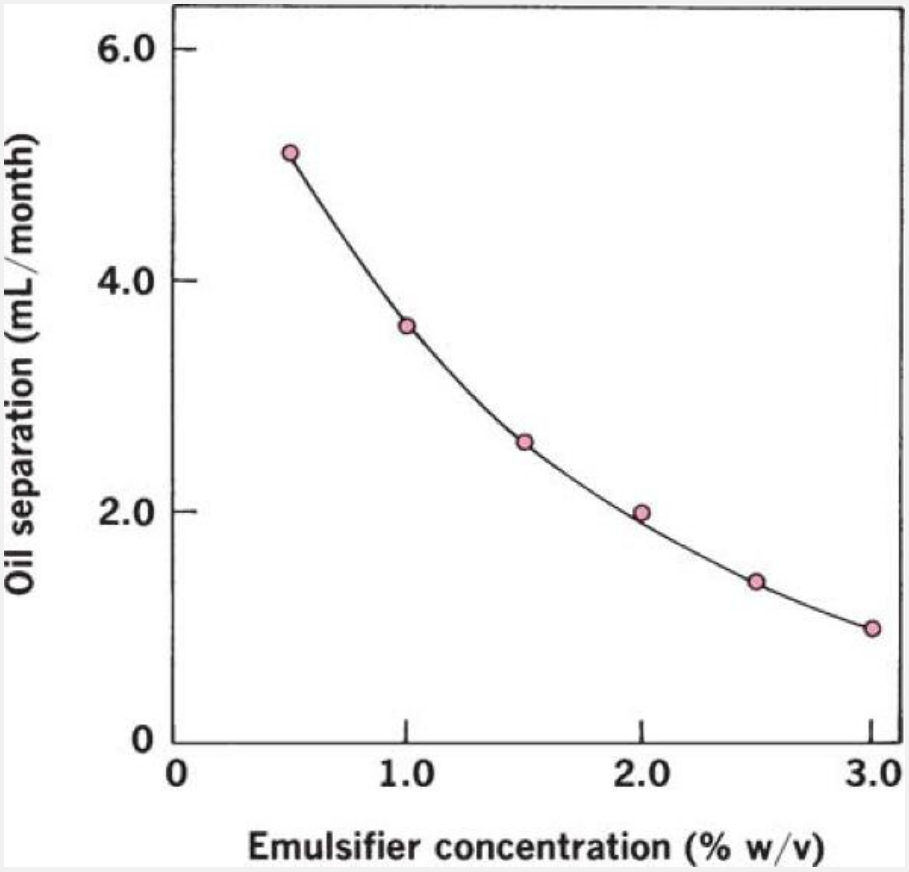


Fig. 1-3. Emulsion stability data plotted on a rectangular coordinate grid.

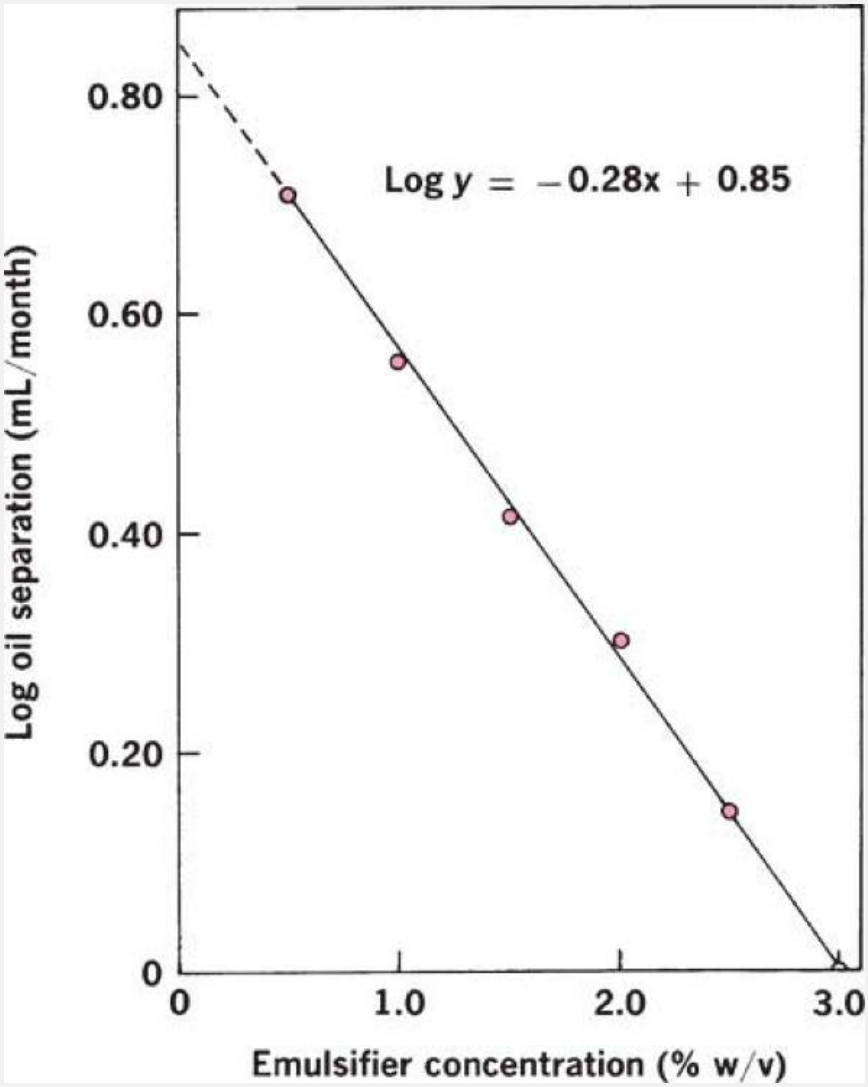


Fig. 1-4. A plot of the logarithm of oil separation of an emulsion versus concentration on a rectangular grid.

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:

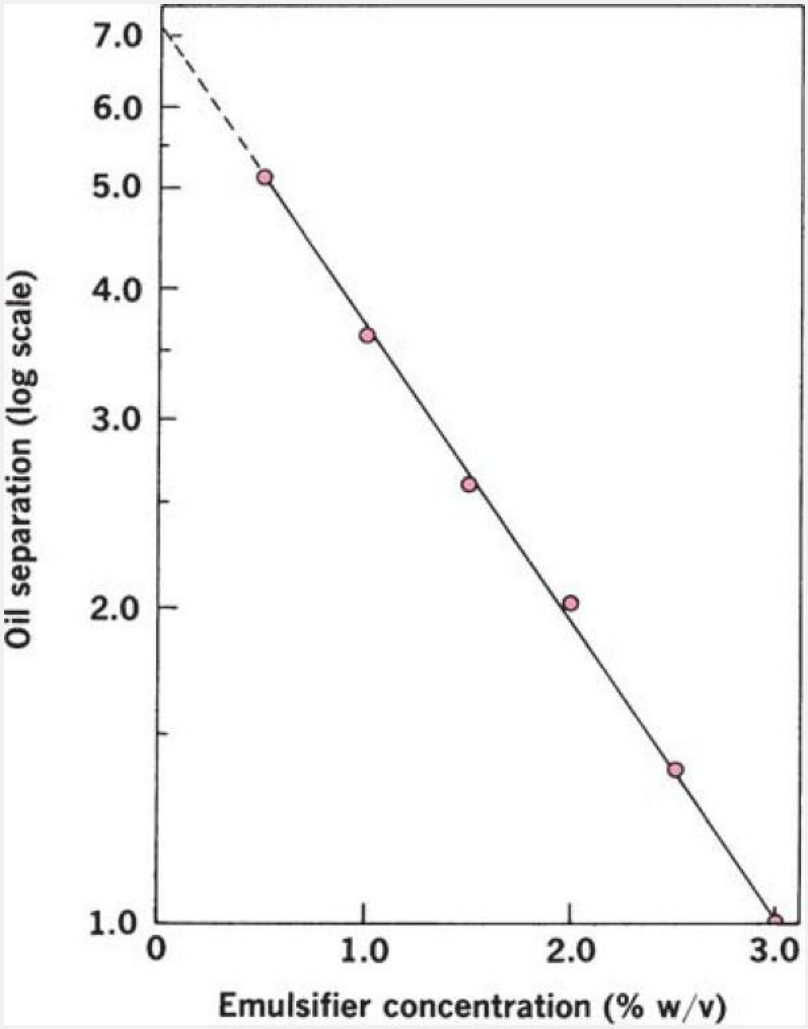


Fig. 1-5. Emulsion stability plotted on a semilogarithmic grid.

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

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 can be overcome by plotting on a semilogarithmic scale. The *x* and *y* values of Table 1-4 are plotted directly on the graph to yield a straight line, as seen in 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 times the log of the number. Therefore, logarithmic graph scales may be used for ln as well as for log plots. In fact, today the natural logarithm is more commonly used than the base 10 log.

Not every pharmacy student will have the need to calculate slopes and intercepts of lines. In fact, once the basics are understood many of these operations can be performed quite easily with modern calculators. However, every pharmacy student should at least be able to look at a visual representation of data and get some sense of what it is telling you and why it is important. For example, the slope of a plasma drug concentration versus time curve is an approximation of the ratio of the input rate and the output rate of the drug in the body at that particular point in time (Fig. 1-6). At any given time point on that curve, the rate of change of drug in the body is equal to the rate of absorption (input) minus the rate of elimination (output or removal from the body). When the two rate processes are equal, the overall

slope of the curve is zero. This is a very important (x,y) point in pharmacokinetics because it is the time point where the peak blood levels occur (C_{max}, t_{max}). The rate of absorption is greater than the rate of elimination to left of the vertical line in Figure 1-6. This is called the *absorption phase*. When the rate of elimination is greater than the rate of absorption, it is called the *elimination phase*. This region falls to the right of the vertical line. The steepness of the slope is an indicator of the rate. For example, in Figure 1-7, three hypothetical products of the same drug are shown. As you can easily see, the rate of absorption of the drug into the bloodstream occurs most quickly from Product 1 and most slowly from Product 2 since the slope of the absorption phase is steepest for Product 1.

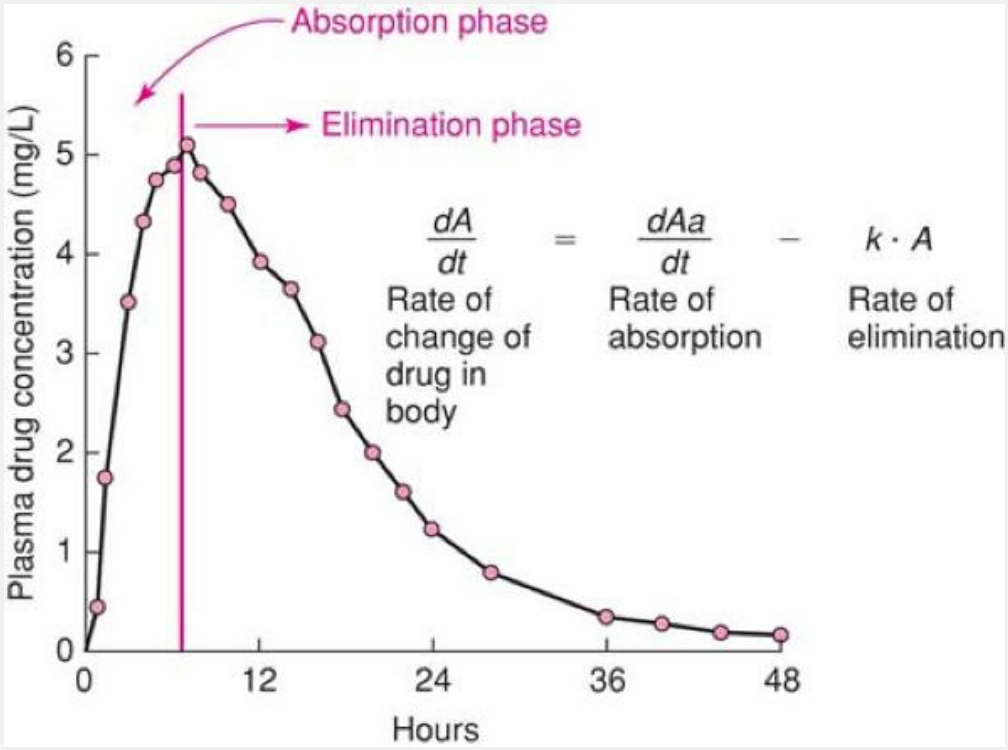


Fig. 1-6. A plot of the absorption and elimination phases of a typical plasma drug concentration versus time curve. During the absorption phase, the rate of input of drug into the body is greater than the rate of output (or elimination) of drug from the body. The opposite is true during the elimination phase. At the peak point (T_{max}, C_{max}), the rates of absorption and elimination are equal.

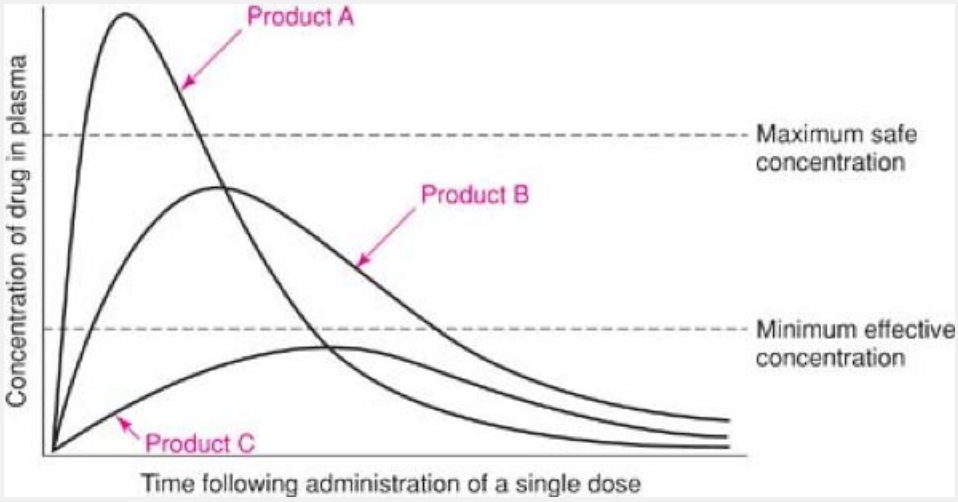


Fig. 1-7. A plot of plasma drug concentration versus time for three different products containing the same dose of a drug. Differences in the profiles are due to differences in the rate of absorption resulting from the three types of formulations. The slope of the absorption phase is equivalent to the rate of drug absorption. The steeper slope (Product A) equals a faster rate of absorption, whereas a less steep slope (Product C) has a slower absorption rate.

Linear Regression Analysis

The data given in Table 1-3 and plotted in Figure 1-2 clearly indicate the existence of a linear relationship between the refractive index and the 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.

Table 1-5 Refractive Indices of Mixtures of Benzene and Carbon Tetrachloride

Concentration of CCl_4 (*x*) (Volume %)	Refractive Index (*y*)
10.0	1.497
26.0	1.493
33.0	1.485
50.0	1.478
61.0	1.477

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 1-5. When these data are plotted on graph paper, an appreciable scatter is observed (Fig. 1-8) 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 the data in Table 1-5 should fit a straight line, and for this we calculate the *correlation coefficient, r*, using the following equation:

$$r = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sqrt{\sum(x - \overline{x})^2 \sum(y - \overline{y})^2}} \qquad (1\text{–}17)$$

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

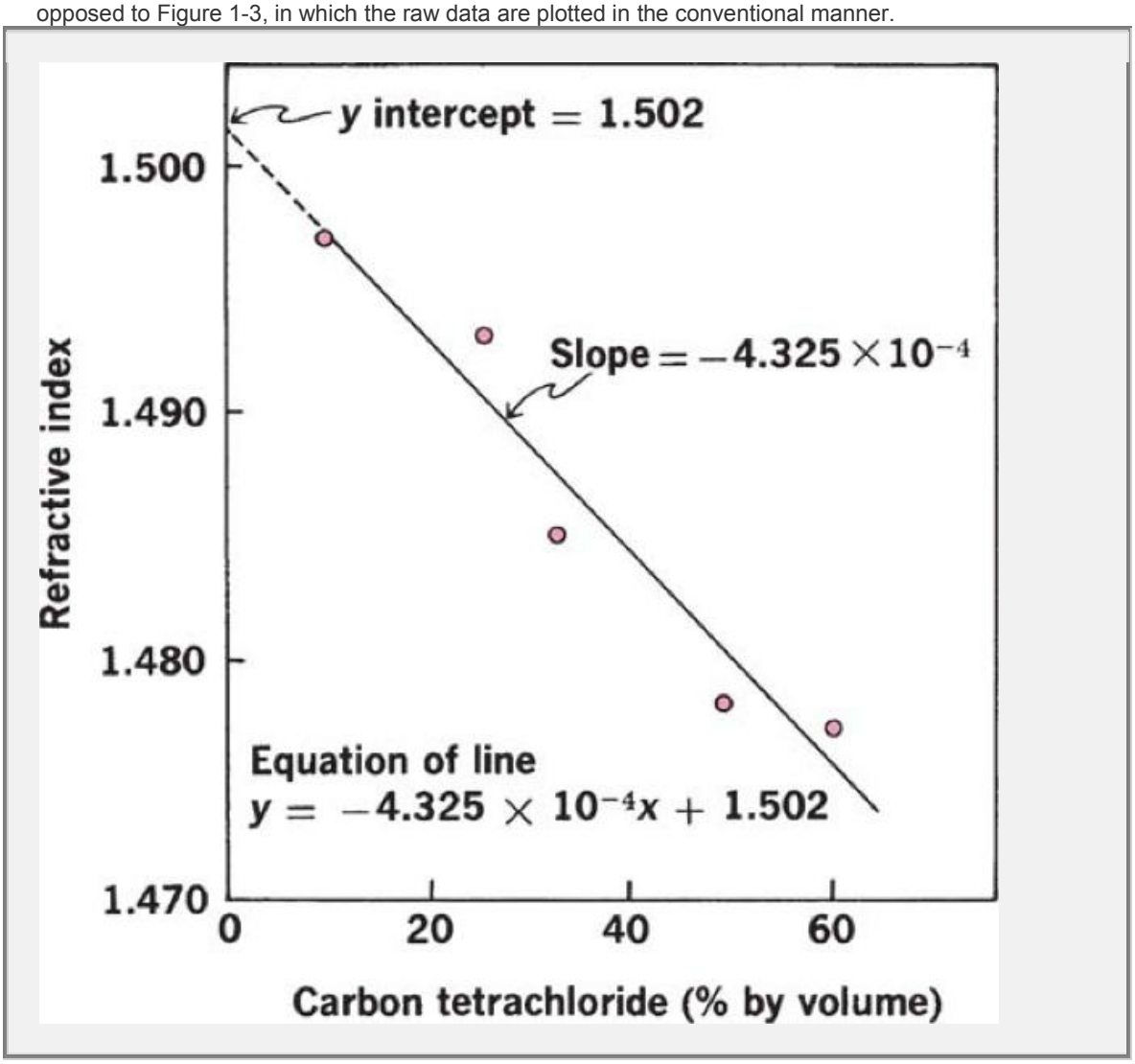


Fig. 1-8. Slope, intercept, and equation of line for data in Table 1-5 calculated by regression analysis.

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

$$b = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sum(x - \overline{x})^2} \qquad (1\text{–}18)$$

in which $\overline{b}$ is the *regression coefficient*, or slope. By substituting the value for **b** in equation **(1-18)**, we can obtain the *y*intercept:

$$\overline{y} = y + b(x - \overline{x}) \qquad (1\text{–}19)$$

The following series of calculations, based on the data in Table 1-5, will illustrate the use of these equations.

Example 1-10

Using the data in Table 1-5, calculate the correlation coefficient, the regression coefficient, and the intercept on the *y* axis.

Examination of equations (1-17) through (1-19) shows the various values we must calculate, and these are set up as follows:

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
$\sum = 180.0$	$\sum = 0$	$\sum = 1606.0$
$\bar{x} = 36.0$		

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
$\sum = 7.430$	$\sum = 0$	$\sum = 0.000316$
$\bar{y} \div 1.486$		

$(x - \bar{x})(y - \bar{y})$
−0.286
0.070
+0.003
−0.112
−0.225
$\sum = -0.690$

Substituting the relevant values into equation (1-17) gives

$$r = \frac{-0.690}{\sqrt{1606.0 \times 0.000316}} = -0.97$$

From equation (1-18)

$$b = \frac{-0.690}{1606.0} = -4.296 \times 10^4$$

and finally, from equation (1-19)

Intercept on the y axis = 1.486

$$-4.315 \times 10^{-4}(0 - 36) = +1.502$$

Note that for the intercept, we place x equal to zero in equation (1-17). By inserting an actual value of x into equation (1-19), 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 1-8.

Chapter Summary

Most of the statistical calculations reviewed in this chapter will be performed using a calculator or computer. The objective of this chapter was not to inundate you with statistical

formulas or complex equations but rather to give the student a perspective on analyzing data as well as providing a foundation for the interpretation of results. Numbers alone are not dynamic and do not give a sense of the behavior of the results. In some situations, equations or graphic representations were used to give the more advanced student a sense of the dynamic behavior of the results.

Practice problems for this chapter can be found at thePoint.lww.com/Sinko6e

References

1. D. L. Sackett, S. E. Straus, W. S. Richardson, W. Rosenberg, and R. B. Haynes, *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd Ed., Churchill Livingstone, Edinburgh, New York, 2000.
2. K. Skau, Am. J. Pharm. Educ. **71**, 11, 2007.
3. S. J. Ruberg, Teaching statistics for understanding and practical use. *Biopharmaceutical Report*, American Statistical Association, 1, 1992, pp. 14.
4. S. S. Stevens, Science, **103**, 677–680, 1946.
5. E. A. Brecht, in *Sprowls' American Pharmacy*, L. W. Dittert, Ed., 7th Ed., Lippincott Williams & Wilkins, Philadelphia, 1974, Chapter 2.
6. W. J. Youden, *Statistical Methods for Chemists*, R. Krieger, Huntington, New York, 1977, p. 9.
7. P. Rowe, in *Essential Statistics for the Pharmaceutical Sciences*, John Wiley & Sons, Ltd, West Sussex, England, 2007.
8. S. Bolton, *Pharmaceutical Statistics: Preclinical and Clinical Applications*, 3rd Ed., Marcel Dekker, Inc., New York, 1997.

Recommended Readings

S. Bolton, *Pharmaceutical Statistics: Preclinical and Clinical Applications*, 3rd Ed., Marcel Dekker, Inc., New York, 1997.

P. Rowe, *Essential Statistics for the Pharmaceutical Sciences*, John Wiley & Sons, Inc., West Sussex, England, 2007.

Chapter Legacy

Fifth Edition: published as Chapter 1 (Introduction). Updated by Patrick Sinko.

Sixth Edition: published as Chapter 1 (Interpretive Tools). Updated by Patrick Sinko.