TABLE 1.1

Some Frequently Used Terms and Their General Meanings

Term	Meaning
95% confidence interval	A range of values (above, below or above and below) the sample (mean, median, mode, etc.) has a 95% chance of containing the true value of the population (mean, median, mode). Also called the fiducial limit; equivalent to Pc 0.05
Bias	Systemic error as opposed to a sampling error. For example, selection bias may occur when each member of the population does not have an equal chance of being selected for the sample
Degrees of freedom	The number of independent deviations and is usually abbreviated df
Independent variables	Also known as predictors or explanatory variables
p-Value	Another name for significance level; usually 0.05
Power	The effect of the experimental conditions on the dependent variable relative to sampling fluctuation. When the effect is maximized, the experiment is more powerful. Power can also be defined as the probability that there will not be a type II error (1-Beta). Conventionally, power should be at least 0.07
Random	Each individual member of the population has the same chance of being selected for the sample
Robust	Having inferences or conclusions little affected by departure from assumptions
Sensitivity	The number of subjects experiencing each experimental condition divided by the variance of scores in the sample.
Significance level	The probability that a difference has been erroneously declared to be significant, typically 0.05 and 0.01 corresponding respectively to a 5% and 1% chance of error.
Type I error	Concluding that there is an effect when there really is not an effect
Type II error	Concluding there is no effect when there really is an effect

(Adapted from Marriott, 1991.)

associated with the method. To help in better understanding the chapters to come, terms frequently used in discussion throughout this book should first be considered. These are presented in Table 1.1.

References

Marriott, F.H.C. (1991) *Dictionary of Statistical Terms*. Longman Scientific and Technical, Essex, England.

2

Basic Principles

Let us start by reviewing a few simple terms and concepts that are fundamental to an understanding of statistics.

Each measurement we make — each individual piece of experimental information we gather — is called a datum. It is extremely unusual, however, to either obtain or attempt to analyze a datum. Rather, we gather and analyze multiple pieces at one time, the resulting collection being called data.

Data are collected on the basis of their association with a treatment (intended or otherwise) as an effect (a property) that is measured in the experimental subjects of a study, such as body weights. These identifiers (that is, treatment and effect) are termed "variables." Our treatment variables (those that the researcher or nature control, and which can be directly controlled) are termed "independent," while our effect variables (such as weight, life span, and number of neoplasms) are termed "dependent" variables — their outcome is believed to depend on the "treatment" being studied.

All possible measures of a given set of variables in all possible subjects that exist are termed the "population" for those variables. Such a population of variables cannot be truly measured; for example, one would have to obtain, treat, and measure the weights of all of the Fischer 344 rats that were, are, or ever will be. Instead, we deal with a representative group — a "sample." If our sample of data is appropriately collected and of sufficient size, it serves to provide good estimates of the characteristics of the parent population from which it was drawn.

Two terms refer to the quality and reproducibility of our measurements of variables. The first, accuracy, is an expression of the closeness of a measured or computed value to its actual or "true" value in nature. The second, precision, reflects the closeness or reproducibility of a series of repeated measurements of the same quantity.

If we arrange all of our measurements of a particular variable in order as a point on an axis marked as to the values of that variable, and if our sample were large enough, the pattern of distribution of the data in the sample would begin to become apparent. This pattern is a representation of the frequency distribution of a given population of data — that is, of the incidence of different measurements, their central tendency, and dispersion.

The most common frequency distribution — and one we will talk about throughout this book — is the normal (or Gaussian) distribution. This distribution is the most common in nature and is such that two thirds of all values are within one standard deviation (to be defined in Chapter 2) of the mean (or average value for the entire population) and 95% are within 1.96 standard deviations of the mean. The mathematical equation for the normal curve is

$$y = \frac{e^{\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

where μ is the mean and σ is the standard deviation.

There are other frequency distributions, such as the binomial, Poisson, and chi-square.

In all areas of biological research, optimal design and appropriate interpretation of experiments require that the researcher understand both the biological and technological underpinnings of the system being studied and of the data being generated. From the point of view of the statistician, it is vitally important that the experimenter both know and be able to communicate the nature of the data and understand its limitations. One classification of data types is presented in Table 2.1.

The nature of the data collected is determined by three considerations. These are the biological source of the data (the system being studied), the instrumentation and techniques being used to make measurements, and the design of the experiment. The researcher has some degree of control over each of these — least over the biological system (he/she normally has a choice of only one of several models to study) and most over the design of the experiment or study. Such choices, in fact, dictate the type of data generated by a study.

Statistical methods are based on specific assumptions. Parametric statistics — those that are most familiar to the majority of scientists — have

TABLE 2.1

Types of Variables (Data) and Examples of Each Type

Classified By	Type	Example
Scale		
Continuous	Scalar	Body weight
	Ranked	Severity of a lesion
Discontinuous	Scalar	Weeks until the first observation of a tumor in a carcinogenicity study
	Ranked	Clinical observations in animals
	Attribute	Eye colors in fruit flies
	Quantal	Dead/alive or present/absent
Frequency distribution	Normal	Body weights
	Bimodal	Some clinical chemistry parameters
	Others	Measures of time-to-incapacitation

more stringent underlying assumptions than do nonparametric statistics. Among the underlying assumptions for many parametric statistical methods (such as the analysis of variance) is that the data are continuous. The nature of the data associated with a variable (as described above) imparts a "value" to that data, the value being the power of the statistical tests that can be employed.

Continuous variables are those that can at least theoretically assume any of an infinite number of values between any two fixed points (such as measurements of body weight between 2.0 and 3.0 kg). Discontinuous variables, meanwhile, are those that can have only certain fixed values, with no possible intermediate values (such as counts of 5 and 6 dead animals, respectively).

Limitations on our ability to measure constrain the extent to which the real-world situation approaches the theoretical, but many of the variables studied in toxicology are in fact continuous. Examples of these are lengths, weights, concentrations, temperatures, periods of time, and percentages. For these continuous variables, we may describe the character of a sample with measures of central tendency and dispersion that we are most familiar with — the mean, denoted by the symbol $\bar{x}$ and called the arithmetic average, and the standard deviation SD, which is denoted by the symbol σ; and is calculated as being equal to

$$\sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum x)^2}{N}}{N-1}}$$

where X is the individual datum and N is the total number of data in the group.

Contrasted with these continuous data, however, we have discontinuous (or discrete) data, which can only assume certain fixed numerical values. In these cases our choice of statistical tools or tests is, as we will find later, more limited.

Probability

Probability is simply the frequency with which, in a sufficiently large sample, a particular event will occur or a particular value will be found. Hypothesis testing, for example, is generally structured so that the likelihood of a treatment group being the same as a control group (the so-called null hypothesis) can be assessed as being less than a selected low level (very frequently 5%), which implies that we are $1.0 - \alpha$ (that is, $1.0 - 0.05 = 0.95$, or 95%) sure that the groups are *not* equivalent.

Functions of Statistics

Statistical methods may serve to do any combination of three possible tasks. The one we are most familiar with is hypothesis testing — that is, determining if two (or more) groups of data differ from each other at a predetermined level of confidence. A second function is the construction and use of models that may be used to predict future outcomes of chemical–biological interactions. This is most commonly seen in linear regression or in the derivation of some form of correlation coefficient. Model fitting allows us to relate one variable (typically a treatment or independent variable) to another. The third function, reduction of dimensionality, continues to be less commonly utilized than the first two. This final category includes methods for reducing the number of variables in a system while only minimally reducing the amount of information, therefore making a problem easier to visualize and to understand. Examples of such techniques are factor analysis and cluster analysis. A subset of this last function, discussed later under descriptive statistics, is the reduction of raw data to single expressions of central tendency and variability (such as the mean and standard deviation).

There is also a special subset of statistical techniques that is part of both the second and third functions of statistics. This is data transformation, which includes such things as the conversion of numbers to log or probit values.

As a matter of practicality, the contents of this book are primarily designed to address the first of the three functions of statistical methods that we presented (hypothesis testing). The second function, modeling — especially in the form of risk assessment — is becoming increasingly important as the science continues to evolve from the descriptive phase to a mechanistic phase (i.e., the elucidation of mechanisms of action), and as such is addressed in some detail. Likewise, because the interrelation of multiple factors is becoming a real concern, a discussion of reduction of dimensionality has been included.

Descriptive Statistics

Descriptive statistics are used to convey, in summary, the general nature of the data. As such, the parameters describing any single group of data have two components. One of these describes the location of the data, while the other gives a measure of the dispersion of the data in and about this location. Often overlooked is the fact that the choice of which parameters are used to give these pieces of information implies a particular type of distribution for the data.

Most commonly, location is described by giving the (arithmetic) mean and dispersion by giving the standard deviation (SD) or the standard error of the mean (SEM).

The use of the mean with either the SD or SEM implies, however, that we have reason to believe that the data being summarized are from a population that is at least approximately normally distributed. If this is not the case, then we should rather use a set of statistical quantities that do not require a normal distribution. These are the median, for location, and the semi-quartile distance, for a measure of dispersion.

Other descriptive statistics that are commonly used in toxicology include the geometric mean and the coefficient of variation (CV).

Calculation and discussion of the mean and standard deviation have already been covered. The formulae and notes for the other statistics mentioned follow:

Standard Error of the Mean

If we again denote the total number of data in a group as N, then the SEM would be calculated as

$$\text{SEM} = \frac{\text{SD}}{\sqrt{N}}$$

The standard deviation and the standard error of the mean are related to each other yet are quite different. To compare these two, let us first demonstrate their calculation from the same set of 15 observations.

		Sum (Σ)
Data Points (X_i):	1, 2, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9	78
Squares (X_i^2):	1, 4, 9, 16, 16, 25, 25, 25, 36, 36, 36, 49, 49, 64, 81	472

The standard deviation can then be calculated as:

$$\text{SD} = \sqrt{\frac{472 - \frac{(78)^2}{15}}{15-1}} = \sqrt{\frac{472 - \frac{(6084)}{15}}{14}} = \sqrt{\frac{472 - 405.6}{14}} = \sqrt{4.7428571} = 2.1778$$

with a mean ($\bar{x}$) of 78/15 = 5.2 for the data group. The SEM for the same set of data, however, is

$$\text{SEM} = \frac{2.1778}{\sqrt{15}} = \frac{2.1778}{3.8730} = 0.562303$$

The SEM is quite a bit smaller than the SD, making it very attractive to use in reporting data. This size difference is because the SEM is actually an

estimate of the error or variability involved in measuring the means of samples, and not an estimate of the error (or variability) involved in measuring the data from which means are calculated. This is implied by the Central Limit Theorem, which tells us three major things:

- The distribution of sample means will be approximately normal regardless of the distribution of values in the original population from which the samples were drawn.
- The mean value of a set of samples from the collection will tend toward the mean of the collection with a large number of samples; i.e., the mean of the collection of all possible means of samples of a given size is equal to the population mean.
- The SD of the collection of all possible means of samples of a given size, called the SEM, depends on both the SD of the original population and the size of the sample.

Since the sample means are normally distributed regardless of the population distribution, a probable range for the population mean can be calculated based on the sample mean and the SEM. If the population mean is represented by $\bar{x}_P$ and the sample mean is represented by $\bar{x}_S$, then the range $\bar{x}_S \pm (1.96)(\text{SEM})$ has a 95% probability of containing $\bar{x}_P$, which is comparable to the earlier assertion that a normally distributed population has 95% of its values within $\bar{x}_P \pm (1.96)(\text{SD})$. Put simply, the SD is a measure of the variability of the data, while the SEM is a measure of the variability of the mean of samples of the data.

The SEM should be used only when the uncertainty of the estimate of the mean is of concern — which is almost never the case in toxicology. Rather, we are concerned with an estimate of the variability of the population — for which the SD is appropriate.

Median

When all the numbers in a group are arranged in a ranked order (that is, from smallest to largest), the median is the middle value. If there is an odd number of values in a group, then the middle value is obvious (in the case of 13 values, for example, the seventh largest is the median). When the number of values in the sample is even, the median is calculated as the midpoint between the $(N/2)$th and the $([N/2] + 1)$th number. For example, in the series of numbers 7, 12, 13, 19 the median value would be the midpoint between 12 and 13, which is 12.5. Note that for $N = 2$ the mean and median will be identical.

Semi-Quartile Distance

When all the data in a group are ranked, a quartile of the data contains one ordered quarter of the values. Typically, we are most interested in the borders

of the middle two quartiles, Q_1 and Q_3, which together represent the semi-quartile distance and which contain the median as their center. Note this is the same as finding the distance between the 25th and 75th percentile. Given that there are N values in an ordered group of data, the upper limit of the jth quartile (Q_j) may be computed as being equal to the $[j(N + 1)/4]$th value. Once we have used this formula to calculate the upper limits of Q_1 and Q_3, we can then compute the semi-quartile distance (which is also called the quartile deviation, and as such is abbreviated as the QD) with the formula $QD = (Q_3 - Q_1)/2$.

For example, for the fifteen-value data set 1, 2, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9, we can calculate the upper limits of Q_1 and Q_3 as

$$\text{Position of } Q_1 = \frac{1(15+1)}{4} = \frac{16}{4} = 4 \qquad \text{Position of } Q_3 = \frac{3(15+1)}{4} = \frac{48}{4} = 12$$

The 4th and 12th values in this data set are 4 and 7, respectively. The semi-quartile distance can then be calculated as

$$QD = (Q_3 - Q_1)/2 = (7 - 4)/2 = 1.5$$

In the case that the calculated lower and upper bounds are not integers, the values of Q_3 and Q_1 should be calculated using interpolation. For example, if we add the value 1 to the above data set, bring N up to 16, the position of $Q_1 = 4.25$ and $Q_3 = 12.25$. The value of Q_1 becomes 3 (the 4th data point) plus 0.25 times the difference between the fourth data point and the fifth, whose value is 4:

$$Q_1 = 3 + (4 - 3)^*(0.25) = 3.25 \quad \text{and} \quad Q_3 = 7 + (7 - 7)^*(0.25) = 7$$

therefore

$$QD = (Q_3 - Q_1)/2 = (7 - 3.25)/2 = 1.875$$

The value $Q_3 - Q_1$ is also known as the quartile distance or the interquartile range.

Geometric Mean

One final sample parameter which sees some use in toxicology (primarily in inhalation studies) is the geometric mean, denoted by the term $\bar{X}_g$. This is calculated as

$$\bar{X}_g = (X_1^* X_2^* X_3^* \ldots X_N)^{1/N} = \sqrt[N]{X_1^* X_2^* X_3^* \ldots X_N}$$

and has the attractive feature that it does not give excessive weight to extreme values (or "outliers"), such as the mass of a single very large particle in a dust sample. In effect, it "folds" extreme values in toward the center of the distribution, decreasing the sensitivity of the parameter to the undue influence of the outlier. This is particularly important in the case of aerosol samples where a few very large particles would cause the arithmetic mean of particle diameters to present a misleading picture of the nature of the "average" particle.

Coefficient of Variation

There are times when it is desired to describe the relative variability of one or more sets of data. The most common way of doing this is to compute the coefficient of variation (CV), which is calculated simply as the ratio of the SD to the mean, or

$$CV = SD/\overline{X}$$

A CV of 0.2 or 20% thus means that the SD is 20% of the mean. In toxicology the CV is frequently between 20 and 50% and may at times exceed 100%.

Outliers and Rounding of Numbers

These two considerations in the handling of numerical data can be, on occasion, of major concern to the toxicologist because of their pivotal nature in borderline cases. Outliers should also be of concern for other reasons, however. On the principle that one should always have a plan to deal with all reasonably likely contingencies in advance of their happening, early decisions should be made to select a policy for handling both outliers and the rounding of numbers.

Outliers are extreme (high or low) values that are widely divergent from the main body of a group of data and from our common experience. They may arise from an instrument (such as a balance) being faulty, the apparently natural urge of some animals to frustrate research, or be indicative of a "real" value. Outlying values can be detected by visual inspection of the data, use of a scattergram (described later), or (if the data set is small enough, which is usually the case in toxicology) by a large increase in the parameter estimating the dispersion of data, such as the SD.

When we can solidly tie one of the above error-producing processes (such as a balance being faulty) to an outlier, we can safely delete it from consideration. But if we cannot solidly tie such a cause to an outlier (even if we have strong suspicions), we have a much more complicated problem, for

then such a value may be one of several other things. It could be the result of a particular case that is the grounds for the entire study — that is, the very "effect" that we are looking for — or it could be because of the collection of legitimate effects that constitute sample error. As will be discussed later (under exploratory data analysis), and is now more widely appreciated, outliers can be an indication of a biologically significant effect that is not yet statistically significant. Variance inflation can be the result of such outliers and can be used to detect them. Outliers, in fact, by increasing the variability within a small group, decrease the sensitivity of our statistical tests and actually preclude our having a statistically significant result (Beckman and Cook, 1983).

Alternatively, the outlier may be the result of, for example, an unobserved technician error, and may be such as to change the decisions made from a set of data. In this case we want to reject the data point — to exclude it from consideration with the rest of the data. But how can one identify these legitimate statistical rejection cases?

There are a wide variety of techniques for data rejection. Their proper use depends on one's having an understanding of the nature of the distribution of the data. For normally distributed data with a single extreme value, a simple method such as Chauvenet's Criterion (Meyer, 1975) may legitimately be employed. This states that if the probability of a value deviating from the mean is greater than $\frac{1}{2}N$, one should consider that there are adequate grounds for its rejections.

In practice, this approach is demonstrated below.

Use of Chauvenet's Criterion

Having collected 20 values as a data set, we find they include the following values: 1, 6, 7, 8, 8, 9, 9, 9, 10, 10, 10, 10, 10, 11, 11, 11, 12, 12, 13, and 14. Was the lowest value (1) erroneous and should it be rejected as an outlier? Some simple calculations are performed, as

$$\text{Mean} = 9.55$$

$$\text{SD} = 2.80$$

$$\text{Chauvenet's Criterion Value} = \tfrac{1}{2}N = 20/2 = 10$$

So we would reject the value of "1" if its probability of occurrence were less than 10%. Going to a table of Z scores (such as Table H in Appendix 1), we see that 10% of the values in a normal distribution are beyond ±1.645 SD of the mean. Multiplying this by the SD for the sample, we get (1.645)(2.80) = 4.606. This means we would reject values beyond this range from the mean; that is, less than (9.55 − 4.606) = 4.944 or greater than (9.55 + 4.606) = 14.156. We therefore reject the value of "1."

One should note that as the sample size gets bigger, the rejection zone for Chauvenet's Criterion will also increase. Indeed, an N of 20 is about as large as this method is useful for.

A second, relatively straightforward approach for use when the data are normally distributed but contain several extreme values is to Winsorize the data. Although there are a number of variations to this approach, the simplest (called the G-1 method) calls for replacing the highest and lowest values in a set of data. In a group of data consisting of the values 54, 22, 18, 15, 14, 13, 11, and 4, we would replace 54 with a second 22, and 4 with a replicate 11. This would give us a group consisting of 22, 22, 18, 15, 14, 13, 11, and 11, which we would then treat as our original data. Winsorizing should not be performed, however, if the extreme values constitute more than a small minority of the entire data set.

Another approach is to use Dixon's Test (Dixon and Massey, 1969) to determine if extreme values should be rejected. In Dixon's test, the set of observations is first ordered according to their magnitude (as we did earlier for the data set used to demonstrate Chauvenet's Criterion, although there this step was simply to make the case clearer). The ratio of the difference of an extreme value from one of its nearest neighbor values in the range of values in the sample is then calculated, using a formula that varies with sample size. This ratio is then compared to a table value, and, if found to be equal or greater, is considered to be an outlier at the $p \leq 0.05$ level. The formula for the ratio varies with sample size and according to whether it is the smallest or largest value that is suspect.

If we have more information as to the nature of the data or the type of analysis to be performed, there are yet better techniques to handle outliers. Extensive discussions of these may be found elsewhere (Barnett and Lewis, 1994; Grubbs, 1969; Beckman and Cook, 1983; Snedecor and Cochran, 1989).

When the number of digits in a number is to be reduced (due to limitations of space or to reflect the extent of significance of a number) we must carry out the process of rounding off a number. Failure to have a rule for performing this operation can lead to both confusion and embarrassment for a facility (during such times as study audits). One common rule follows.

A digit to be rounded is not changed if it is followed by a digit less than 5 — the digits following it are simply dropped off ("truncated"). If the number is followed by a digit greater than 5 or by a 5 followed by other nonzero digits, it is increased to the next highest number. When the digit to be rounded is followed by 5 alone or by 5 followed by zeros, it is unchanged if it is even but increased by 1 if it is odd. Examples of this rule in effect are (in a case where we must reduce to whole digits):

137.4	becomes	137
137.6	becomes	138
138.52	becomes	139
137.5	becomes	138
138.5	becomes	138

The rationale behind this procedure is that over a period of time, the results should even out — as many digits will be increased as are decreased.

For sets of data, whatever rounding rule is used must generate consistent results. For example, if the values 0.233, 0.034, and 0.746 are rounded to two digits as separate values, the result is 0.23, 0.034, and 0.75 but if they are considered a set the values round to 0.23, 0.03, and 0.75 for consistency.

Sampling

Sampling — the selection of which individual data points will be collected, whether in the form of selecting which animals to collect blood from or to remove a portion of a diet mix for analysis — is an essential step upon which all other efforts toward a good experiment or study are based.

There are three assumptions about sampling that are common to most of the statistical analysis techniques that are used in toxicology. These are that the sample is collected without bias, that each member of a sample is collected independently of the others, and that members of a sample are collected with replacements. Precluding bias, both intentional and unintentional, means that at the time of selection of a sample to measure, each portion of the population from which that selection is to be made has an equal chance of being selected. Ways of precluding bias are discussed in detail in the chapter on experimental design.

Independence means that the selection of any portion of the sample is not affected by and does not affect the selection or measurement of any other portion.

Finally, sampling with replacement means that, in theory, after each portion is selected and measured, it is returned to the total sample pool and thus has the opportunity to be selected again. This is a corollary of the assumption of independence. Violation of this assumption (which is almost always the case in toxicology and all the life sciences) does not have serious consequences if the total pool from which samples are sufficiently large (say 20 or greater) is such that the chance of reselecting that portion is small anyway.

There are four major types of sampling methods — random, stratified, systematic, and cluster. Random is by far the most commonly employed method in toxicology. It stresses the fulfillment of the assumption of avoiding bias. When the entire pool of possibilities is mixed or randomized (procedures for randomization are presented in a later chapter), then the members of the group are selected in the order they are drawn from the pool.

Stratified sampling is performed by first dividing the entire pool into subsets or strata, then doing randomized sampling from each strata. This method is employed when the total pool contains subsets that are distinctly different but in which each subset contains similar members. An example is a large batch of a powdered pesticide in which it is desired to determine the nature of the particle size distribution. Larger pieces or particles are on the top, while progressively smaller particles have settled lower in the

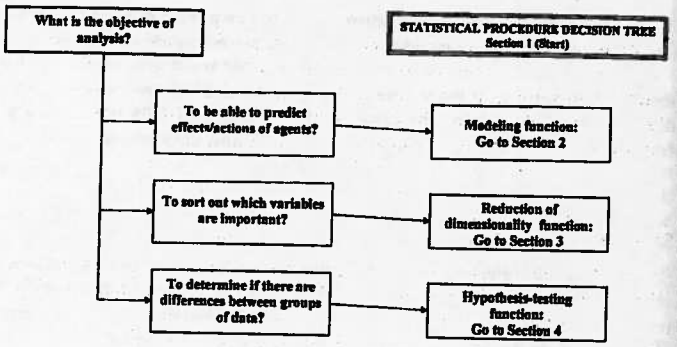

FIGURE 2.1
Overall decision tree for selecting statistical procedures.

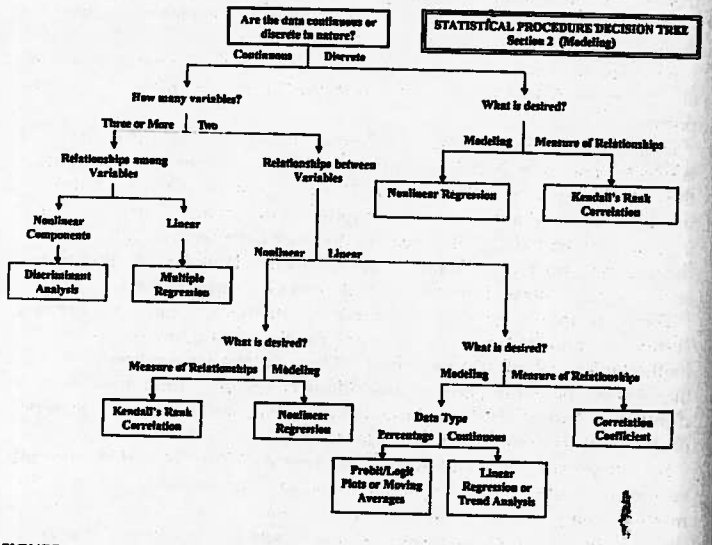

FIGURE 2.2
Decision tree for selecting hypothesis-testing procedures.

container, and, at the very bottom, the material has been packed and compressed into aggregates. To determine a timely representative answer, proportionally sized subsets from each layer or strata should be selected, mixed, and randomly sampled. This method is used more commonly in diet studies.

In systematic sampling, a sample is taken at set intervals (such as every fifth container of reagent or taking a sample of water from a fixed sample point in a flowing stream every hour). This is most commonly employed in quality assurance or (in the clinical chemistry lab) in quality control.

In cluster sampling, the pool is already divided into numerous separate groups (such as bottles of tablets), and we select small sets of groups (such as several bottles of tablets), then select a few members from each set. What one gets then is a cluster of measures. Again, this is a method most commonly used in quality control or in environmental studies when the effort and expense of physically collecting a small group of units is significant.

In classical toxicology studies sampling arises in a practical sense in a limited number of situations. The most common of these are as follows:

- Selecting a subset of animals or test systems from a study to make some measurement (which either destroys or stresses the measured system, or is expensive) at an interval during a study. This may include such cases as doing interim necropsies in a chronic study or collecting and analyzing blood samples from some animals during a subchronic study.
- Analyzing inhalation chamber atmospheres to characterize aerosol distributions with a new generation system.
- Analyzing diet in which test material has been incorporated.
- Performing quality control on an analytical chemistry operation by having duplicate analyses performed on some materials.
- Selecting data to audit for quality assurance purposes.

Generalized Methodology Selection

One approach for the selection of appropriate techniques to employ in a particular situation is to use a decision-tree method. Figure 2.1 is a decision tree that leads to the choice of one of three other trees to assist in technique selection, with each of the subsequent trees addressing one of the three functions of statistics that was defined earlier in this chapter. Figure 2.2 is for the selection of hypothesis-testing procedures, Figure 2.3 for modeling procedures, and Figure 2.4 for reduction of dimensionality procedures. For the vast majority of situations, these trees will guide the user into the choice of the proper technique. The tests and terms in these trees will be explained subsequently.

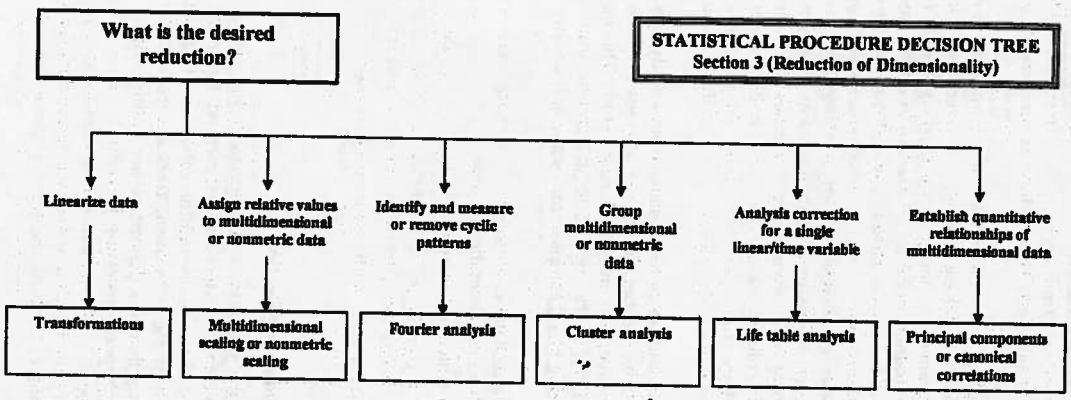

FIGURE 2.3
Decision tree for selecting modeling procedures.

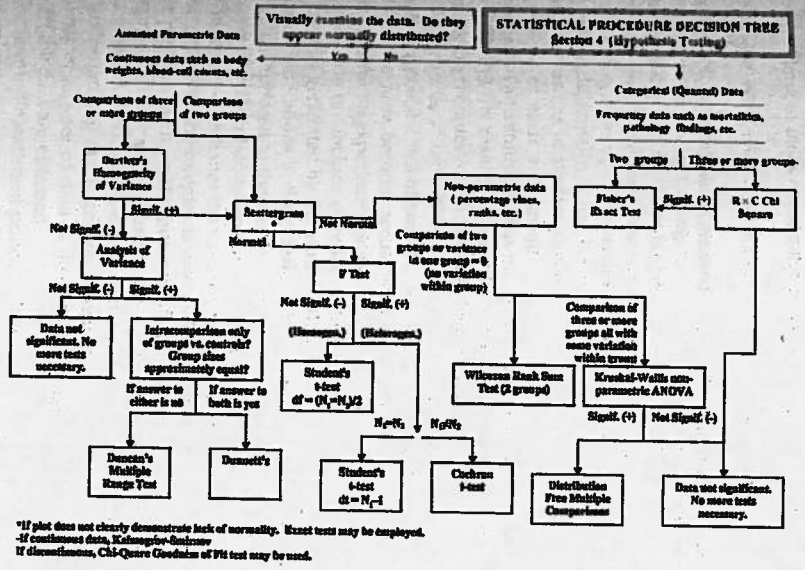

FIGURE 2.4
Decision tree for selecting reduction of dimensionality procedures.