I GENERAL COMMENTS

COURSE DESIGN

To provide flexibility in course design, a number of sections in the textbook are designated as "Optional." The instructor wishing to adopt a leisurely pace, or who is designing a course for one quarter, can omit all optional sections and, moreover, give only light coverage to some other sections. The *Comments on Chapters* in Section II of this Manual indicate specifically those sections and parts of sections where light coverage may be appropriate. For an even briefer schedule, Chapters 11 and 12 can be omitted entirely. Also, there is no new material in Chapter 13; rather, this chapter provides a perspective of the preceding chapters. The purpose of Chapter 13 is to help summarize and put into perspective the many inference methods discussed in the text.

Because each chapter builds on the preceding ones, it is not advisable to alter the order of the chapters. However, it would be possible to postpone coverage of Section 6.6 (on confidence intervals for proportions) until covering hypothesis testing for categorical data in Chapter 10. (This was the placement of confidence intervals for proportions in the first edition; the third edition treats confidence intervals for means and for proportions together in Chapter 6.)

EXERCISES

Calculators and Computers

The exercises in the text are designed to minimize numerical drudgery and emphasize understanding. Exercises with simple numbers familiarize the students with the meaning and structure of formulas, after which additional exercises, based on real data, focus primarily on interpretation. Any calculations required for the latter exercises are easily carried out on a hand calculator.

If computing facilities are available, the use of a computer can be easily integrated into the course. To that end, a data disk comes with the textbook. A number of the exercises give raw data and could be done on either a computer or a calculator. Also, certain exercises, labeled *computer exercise*, are especially designed for computer use and would not be suitable for hand computation.

Location of Exercises

Most exercises are located at the ends of sections. At the end of each chapter are Supplementary Exercises, some of which use material from more than one section.

Class Discussion

Many of the exercises can be used as starting points for class discussion. This is especially true of exercises that emphasize interpretation of results or that request a critique of an inappropriate analysis or conclusion.

<u>Sampling Exercises</u>

Scattered throughout the first half of the book are exercises that require students to carry out random sampling. These sampling exercises are discussed in detail in Section III of this Manual.

APPENDICES

At the end of the textbook are appendices that provide more detail on selected topics discussed in the text. These appendices are not intended to form part of the course (except, perhaps, for Appendix 6.1 on significant digits), but rather to provide supplementary material for interested students.

STATISTICAL TABLES

The tables of critical values are all used in essentially the same way. In the t table (Table 4 and inside back cover), the column headings are upper tail areas. Thus, when the alternative hypothesis for a t test is directional, students can easily bracket the P-value by reading the column headings; when the alternative hypothesis is nondirectional, the column headings must be multiplied by 2. The tables for the Wilcoxon-Mann-Whitney test, the sign test, and the Wilcoxon signed-rank test give two sets of column headings: one for use with nondirectional alternatives and one for use with directional alternatives. Tables for the chi-square test and the F test give upper tail areas, as is appropriate for use with the usual nondirectional alternative hypotheses. When the alternative hypothesis in a chi-square test is directional, the column headings must be multiplied by 1/2.

The examples in the textbook and the answers to the exercises in this Manual do not use interpolation in the statistical tables. For instance, in entering Table 4 the nearest value of df is used; in ambiguous cases (e.g., df = 35 or df = 200), either one of the nearest values is considered correct. Students may need some guidance on this point.

II COMMENTS ON CHAPTERS

Chapter 1 Introduction

The study of statistics will seem more inviting to students in the life sciences if they see that statistical questions arise in biologically interesting settings. Chapter 1 consists primarily of a series of examples of such settings. Instructors may choose to discuss these or other examples in the first lecture.

Chapter 2 Description of Samples and Populations

Comments on Sections 2.1 - 2.3

Students are sometimes uneasy when approaching these introductory sections because they are not sure what is expected of them. The instructor may wish to reassure them on the following points: (a) the distinction between Y and y in Section 2.1 is for clarity only; (b) for an exercise requiring a grouped frequency distribution, there are many different "right answers"; (c) the material on histograms with unequal class widths is for "consciousness-raising" purposes -- the exercises do not require students to create such histograms.

Comments on Sections 2.4 - 2.6

This is an opportunity to get students started using their calculators effectively. Two useful general principles are:

(a) Minimize roundoff error by keeping intermediate answers in the calculator rather than writing them down. Round the answers only at the end of the entire computation. (The topic of how far to round when reporting the mean and SD is discussed in Section 6.2; for use in Chapter 2, the instructor may wish to give students a temporary answer, such as "always round to four significant digits.")

(b) Take full advantage of the calculator's memory (and parentheses, if it has them) to keep track of intermediate answers.

Students may need guidance in using the statistics function of the calculator; the following tips are worth mentioning:

(a) Use the "change-sign" key (not the "minus" key) to enter negative data.

(b) Use the "data-removal" key to correct erroneous entries.

(c) Some calculators have two SD keys, one that uses "n-1" as a divisor, and another that uses "n" instead. The former definition is used throughout the text.

Students may wonder whether they are permitted to use the SD function of their calculator in doing homework. Generally, the exercises have been written under the assumption that students would use the SD function of their calculator. However, in the entire text the total number of exercises in which students are expected to calculate a standard deviation from raw data is quite small.

Comments on Optional Section 2.7

This section aims to enhance the student's intuition about the effect of linear and nonlinear transformations of data. The logarithmic transformation is especially important in biology. It may be enlightening to students to point out that choice of scale is rather arbitrary and that there is nothing wrong with choosing a new scale in order to aid presentation and interpretation. For example, acidity is generally measured by pH, which is in log scale.

Comments on Section 2.8

The major goal of this section is to convince students that the population/sample concept is a reasonable one in biological research, and to help them develop some intuition about the relationship between sample statistics and population parameters.

Chapter 3 Random Sampling, Probability, and the Binomial Distribution

Comments on Sections 3.1 and 3.2

In keeping with the nature of most biological data, the text treats the random sampling model as a model -- that is, an idealization -- rather than as a reflection of a physical sampling process. (The relationship between the random sampling model and the physical act of randomization in allocating experimental units to treatment groups will be made explicit in Chapter 8.) To motivate this approach, and to counter the common preconception that a "random" sample and a "representative" sample are the same thing, the instructor can point out that the statistical approach takes a natural but unanswerable question that a biological researcher might ask and translates it into a slightly different question that can be answered:

Researcher's question: "How representative is my sample?"

Statistical translation: "How representative is a random sample likely to be?"

The word "likely" in the translated question is unavoidable, because a random sample can be quite unrepresentative. This motivates the use of probability in statistical analysis.

The technique of drawing a random sample has two applications in the text: (a) the technique is central to randomized allocation, as explained in Chapter 8; and (b) sampling exercises in Chapters 3, 5, 6, 7, and 9 (see Section III of this Manual) require the technique. The instructor who wishes to omit sampling exercises could defer discussion of the technique until Chapter 8.

Comments on Sections 3.3 and 3.4

These sections introduce probability and the use of probability trees. The presentation bypasses formal definitions of sample space, event, etc., and the "addition" and "multiplication" rules for combining probabilities; formal treatment of these topics is taken up in optional Section 3.5. Instead, the emphasis is on a central theme: the interpretation of probability as long-run relative frequency (from which the "addition" rule follows very naturally). The intent of this approach is to spend less time on probability manipulations and more time on later chapters where probability ideas are applied in the analysis of data.

The concept of "independence" is not defined formally in these sections (it is presented in optional Section 3.5), although Section 3.4 introduces probability trees, which implicitly use conditional probabilities. Rather independence is introduced informally in various settings throughout the text (which is why Section 3.5 can be skipped). Independence of trials and independence as part of the definition of a random sample are introduced in Chapter 3; independence of observations in a sample is discussed in Chapter 6; independence of two samples is introduced in Chapter 7; independence of two categorical variables and the related concept of conditional probability are introduced in Chapter 10. Conditional distributions, conditional means, and conditional standard deviations are introduced in Chapter 12.

Comments on Optional Section 3.5

This section introduces formal rules for probability, including the "addition" and "multiplication" rules for combining probabilities. This section is optional; some instructors will omit it, while others who wish to cover probability in a more formal way will include the section.

Comments on Sections 3.6 and 3.7

Continuous distributions are introduced in Section 3.6. The purpose of Exercise 3.16 is to reassure students that the mysterious "area under the curve" will be quite easy for them to handle, and, further, to ward off the common misconception that all continuous distributions are normal. Section 3.7 introduces the concept of a random variable and presents a few simple examples.

Comments on Section 3.8

This section introduces the binomial distribution, which is presented as a tool for computing certain kinds of probabilities that will later be seen to be relevant to statistics. The binomial formula is presented, but details of the derivation of this formula are left to Appendix 3.2; the derivation of the mean and of the standard deviation are found in Appendix 3.3.

Comments on Optional Section 3.9

Many students find this section intriguing because it makes probability ideas very concrete; however, the material is not referred to again in the text.

Chapter 4 The Normal Distribution

Chapter 4 is a straightforward coverage of the normal distribution. Note that the table of normal areas (Table 3) is reproduced inside the front cover of the textbook.

Students should have little difficulty with the exercises in Chapter 4 if they are reminded to draw a sketch for each calculation. In addition to skill in determining normal areas, this chapter gives students experience in visualizing a population distribution for a population whose size is large and unspecified.

Those naturally occurring distributions that are used as examples of normal distributions in Chapter 4 are in fact (approximately) normal, as determined by an examination of the raw data (in Example 4.1 the raw data are shown) or by theory (e.g., in Exercise 4.26 the distribution is Poisson with large mean). However, most population distributions encountered in biology are not approximately normal but are distinctly skewed to the right. Thus, the challenge is to convey to the students the twin messages that (a) it is not true that the "typical" distribution is normal, but (b) methods of data analysis based on normal theory are useful anyway. The simplest example of the latter is the "typical" percentages (68%, 95%, > 99%) rule given in Section 2.6, which is derived from the normal distribution but works rather well for many nonnormal distributions. Deeper examples are the many inferential methods (first encountered in Chapter 6) that are based on normal theory but (because of the Central Limit Theorem) can be validly applied in nonnormal settings.

Comments on Section 4.4

Section 4.4 takes up the topic of assessing normality. Normal probability plots are introduced here and are used to assess normality in many examples later in the text. However, some instructors will choose to spend minimal time on this topic, preferring to rely on other means of assessing normality, such as examining histograms.

Comments on Optional Section 4.5

Section 4.5 on the continuity correction is optional. The instructor who plans to cover the normal approximation to the binomial distribution in optional Section 5.5 may wish to cover Section 4.4, but this is not really necessary because the continuity correction is introduced anew in Section 5.5.

Chapter 5 Sampling Distributions

Comments on Sections 5.1 - 5.3

Chapter 5 introduces the very important concept of a sampling distribution. As motivation, the students can be reminded that the question "How representative is a random sample likely to be?" is the foundation of statistical inference (see Comments on Sections 2.8, 3.1, and 3.2).

The sampling distribution of a proportion $\hat{p}$ is covered first (Section 5.2) because the students have already worked with the binomial distribution. However, none of the data-analytic methods in the book use the exact sampling distribution of $\hat{p}$ (Section 6.6 uses the normal approximation to the binomial), whereas Sections 6.2 and 6.3 use the sampling distribution of $\bar{Y}$ (Section 5.3) . Consequently, the instructor may wish to cover Section 5.2 only very lightly.

Many students find Chapter 5 difficult. To motivate them to make a special effort, the instructor can stress that the chapter lays an important foundation, because many concepts used in the analysis of real data (two examples are standard errors and P-values) can be understood only if sampling distributions are first understood. Students should be encouraged to read the material in Chapter 5 more than once. Doing the sampling exercises (5.11, 5.12, and 5.13) and then discussing them in class are very helpful to the students.

Many students will need to be reminded several times that the sampling distribution of $\bar{Y}$ is not the same as the distribution of observations in the sample (nor in the population).

To help the student put Chapter 5 in perspective, the instructor can explain that, whereas in Chapter 5 we are assuming that we know the population parameters and we are predicting the behavior of samples, in real inferential data analysis (which starts in Chapter 6) we are in the reverse position: we know the characteristics of the sample and we are trying to learn something about the unknown characteristics of the population. Moreover, although computations using the sampling distribution of $\bar{Y}$ may appear to require the knowledge of μ and σ, actually, useful computations can be made in ignorance of μ (as shown by Exercises 5.17 and 5.56), and furthermore in Chapter 6 it will be seen that very useful similar computations can be made in ignorance of σ.

Comments on Optional Section 5.4

Section 5.4 illustrates the effect of the Central Limit Theorem on a moderately skewed population (Example 5.13) and a violently skewed population (Example 5.14). The two populations are used again in Section 6.5 to show the impact of the Central Limit effect on the coverage probability of a Student's t confidence interval. However, Section 6.5 can be used independently; Section 5.4 need not be covered at all. A sketchy coverage of Section 5.4 can be achieved in ten minutes of class time by simply displaying and briefly discussing transparencies of Figures 5.13-5.16; a more leisurely coverage permits detailed discussion of Example 5.14 and assignment of one of the exercises (5.29 or 5.30).

Background Information on Examples 5.13 and 5.14

For the record, the distribution in Example 5.13 is a lognormal distribution fitted to data of Zeleny (see source cited in Example 5.13). The mean and SD of the natural logarithm of Y are 4.1 and .33, respectively. The distributions in Figure 5.14 were estimated by the authors and R. B. Becker using computer simulation; for each sampling distribution, we drew 40,000 samples from the lognormal distribution (20,000 for n = 64). The distribution in Example 5.14 is a 9:1 mixture of two normal distributions with parameters $\mu_1 = 115$, $\sigma_1 = 17.7$, $\mu_2 = 450$, and $\sigma_2 = 83.4$; the parameters are based on a simplified version of data of Bradley (see sources cited in Example 5.14). The curves in Figure 5.16 were determined analytically; each sampling distribution is a mixture of normal distributions with mixing proportions determined from the binomial distribution with parameters n and p = .9.

Comments on Optional Section 5.5

This section presents the normal approximation to the binomial distribution, both unadorned and in its guise as the sampling distribution of $\hat{p}$. The material is never used again. (A brief mention of the normal approximation in Section 6.6 is entirely self-contained.)

Chapter 6 Confidence Intervals

Comments on Sections 6.1 - 6.3

These sections introduce the idea of a standard error and its use in constructing a confidence interval. The confidence interval for μ based on Student's t (σ unknown) is the only one presented; the interval based on Z (σ known) is used as motivation, but is not presented as a technique for analyzing data. Of course, the two intervals are nearly identical if n is large; the instructor can clarify this by explaining the relationship between the normal table (Table 3 and front cover) and the bottom row of the t table (Table 4 and back cover). Confidence levels are given at the bottom of each column in Table 4, for use with confidence intervals in Section 6.3 and later.

The explicit comparison between the SD and the SE in Section 6.2 helps students to feel more comfortable with the different interpretations of these two statistics -- a difference that is important but subtle. Exercises 6.4, 6.5, 6.6, 6.7, 6.14, 6.62, and 6.63 can serve as reinforcement.

Students generally have difficulty interpreting confidence statements; this difficulty is natural. Several of the exercises ask students to interpret, in context, confidence intervals that they have constructed. Students should not be allowed to simply refer to "the mean," lest they confuse in their minds the sample mean and the population mean. If they balk at being held accountable for precise usage of words, they may benefit from hearing the saying that "The temple of reason is entered through the courtyard of habit." A major goal in asking students to interpret confidence intervals is to clarify their reasoning by instilling good habits of English usage.

The "Rule for Rounding" given in Section 6.2 is intended as a guide for reporting summary statistics in research articles. It is not rigidly adhered to in the answers to exercises given in this Manual.

Students who are uncomfortable with the concept of significant digits (used in Section 6.2) can be referred to Appendix 6.1.

Comments on Section 6.4

This section shows students that (a) there is a rational way to decide how large n should be, but (b) the decision requires input from the researcher. While these two principles apply quite generally, the only other explicit sample size computation given in the text is for the two-sample t test (optional Section 7.8 on power).

Section 6.4 includes an informal guideline (anticipated difference between two groups should be at least 4 standard errors) that the instructor may wish to mention again when discussing the two-sample t test. (However, there are no exercises using the guideline). The basis of the guideline is that it achieves a power of roughly .80 for a two-tailed t test at $\alpha = .05$.

Comments on Section 6.5

Here, for the first of many times throughout the text, the students encounter the idea that a statistical calculation can give an answer that is numerically correct, but is misleading or meaningless because of either (a) the way the data were obtained, or (b) some feature of the population distribution.

The requirement of independence of the observations is often violated in biological studies; Exercises 6.32, 6.57, 6.59, and many exercises in later chapters address this point.

The numerical results in Table 6.4 on coverage probability when sampling from nonnormal populations were obtained by computer simulation (see Comments on Optional Section 5.4 in this Manual).

Comments on Section 6.6

Confidence intervals for proportions are treated here although, as mentioned earlier, this material could be postponed until Chapter 10, when hypothesis testing with categorical data is discussed. The confidence interval for a proportion is essentially an extension of the material in Section 6.3 to the case in which all observations are 0 or 1. In Appendix 3.3 it is shown that in this setting $\mu = p$ and $\sigma = \sqrt{p(1 - p)}$, so the fact that $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$ corresponds directly to the fact that

$\sigma_{\hat{p}} = \sqrt{p(1 - p)}/\sqrt{n}$; this is discussed in Appendix 5.1. However, the normality condition that is the basis of a confidence interval for μ in Section 6.3 is clearly violated when we have 0-1 data. Thus, we must appeal (via the Central Limit Theorem) to the approximate normality of the sampling distribution of $\hat{p}$ when n is large.

Many instructors will be surprised to find the familiar formula

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

replaced by

$$\tilde{p} \pm 1.96\sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

Appendix 6.2 discusses the use of $\tilde{p}$. If n is large, the familiar $\hat{p}$ -based confidence interval and the $\tilde{p}$ -based confidence interval are virtually identical. However, the $\tilde{p}$ -based confidence interval has superior coverage properties when n is not large. Moreover, using the $\tilde{p}$ -based confidence interval means that it is not necessary to construct tables or rules for how large n must be in order for the confidence interval to have good coverage properties. For furthur discussion, see Appendix 6.2.

Comments on Section 6.7

The material in this section is not specifically reinforced in the exercises, but rather serves to "open a window" between the narrow coverage of Chapter 6 and the real world of research.

Chapter 7 Comparison of Two Independent Samples

Chapter 7 deals with the comparison of two independent samples when the observed variable is quantitative; analogous developments for a categorical observed variable are given in Chapter 10.

After presenting the Student's t confidence interval, Chapter 7 introduces the general principles of hypothesis testing while developing the t test. The chapter concludes with the Wilcoxon-Mann-Whitney test, which is the distribution-free competitor to the t test.

The F test for comparison of variances is omitted. The use of the F test as a preliminary to the pooled t test is strongly discouraged by many statisticians because it is highly sensitive to nonnormality: in the words of George Box, such use is like "putting out to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!" [*Biometrika 40* (1953), p. 333.]

Chapter 7 introduces design issues -- for instance, the term "independent" in the chapter title and the distinction between observational and experimental studies -- that are not fully discussed until Chapter 8. Chapter 7 is placed first so that students with limited experience in research can have considerable exposure to concrete examples before tackling the general principles of design as expounded in Chapter 8.

Comments on Section 7.1

In addition to introducing the notation for two samples, Section 7.1 alerts students to the existence of both experimental and observational studies and to the fact that distributions can differ in dispersion and shape as well as in location.

Comments on Sections 7.2 and 7.3

Section 7.2 introduces the standard error of a difference between two means. This notion may be rather unnatural for life science students, who are generally accustomed to comparing two quantities in terms of their ratio rather than their difference. Class discussion of this point can be helpful.

The unpooled method for computing the standard error is emphasized, although pooling of standard deviations is discussed in an optional sub-section. (The case $\sigma_1 \neq \sigma_2$ often occurs in biological research.)

The textbook primarily uses the Satterthwaite formula (Formula (7.1) on page 227) for degrees of freedom, on the view that technology should be used for number crunching, so that the messy form of Formula (7.1) should provide no obstacle in practice. However, the formulas
$df = n_1 + n_2 - 2$ and $df =$ smaller of $n_1 - 1$ and $n_2 - 1$ are given as liberal and conservative values. The philosophy behind presenting three df formulas is that statistical methods are generally approximate. It may be useful to quote George Box here: "All models are wrong, some models are useful." If different df choices lead to qualitatively different conclusions, then great care should be taken in interpreting the results.

In constructing confidence intervals for $(\mu_1 - \mu_2)$ a choice must be made, implicitly or explicitly, of which sample to denote by 1 and which by 2. Some students find this indeterminacy unsettling; the instructor can help by explaining that either choice is acceptable, and that the apparently different confidence intervals obtained are actually equivalent.

Comments on Sections 7.4, 7.5, and 7.6

These sections introduce the basic ideas of hypothesis testing and the specific technique of the t test. The approach to hypothesis testing is two-pronged, developing both the use of the P-value as a descriptive statistic and also the decision-oriented framework which gives meaning to Type I and Type II error. However, the use of P-values is emphasized and encouraged.

The text places strong emphasis on verbal statements of hypotheses and conclusions, and the solutions given in this Manual reflect that emphasis. The verbal statements help the student appreciate the biological purpose of each statistical test; without them it is all too easy for the student to look only at the numbers in an example or exercise, while virtually ignoring the descriptive paragraph which gives meaning to the numbers. A potential difficulty is that the verbal statements must be, in the interest of brevity, considerably oversimplified. The instructor can call the students' attention to the demurrer in the Answers to Selected Exercises for Chapter 7, which explicitly recognizes this oversimplification. To further emphasize the point, a distinction is made in Section 7.4 between the "formal" hypotheses H_0 and H_A and the "informal" hypotheses H_0^* and H_A^*; but this cumbersome notation is abandoned in later sections.

The verbal conclusions in this Manual usually use the phrases "sufficient evidence" and "insufficient evidence" (for instance, "There is sufficient evidence to conclude that Drug A is a more effective pain reliever than Drug B"). Some students are more comfortable with "enough" and "not enough," but they may mistakenly believe that "sufficient" and "insufficient" are technical terms that they are required to use. The instructor may wish to use "enough" in class discussion, or to encourage students to use more descriptive phrases such as "little or no evidence," "very strong evidence," etc.

The textbook uses a two-step procedure to bracket the P-value when H_A is directional. The advantage of this rather unusual approach is that it extends readily to tests (such as the chi-square test for a 2 x 2 contingency table) for which H_0 is rejected in only one tail of the null distribution.

The issue of directional versus nondirectional alternative hypotheses is difficult for many students. One difficulty is that the rule that a directional H_A must be formulated before seeing the data is somewhat remote for those students whose only exposure to data is in the textbook itself.

A second difficulty with directional versus nondirectional alternatives concerns the nature of the possible conclusions. The textbook indicates that rejecting H_0 in favor of a nondirectional H_A should lead to a directional conclusion; for instance, "There is sufficient evidence to conclude that Diet I gives a *higher* mean weight gain than Diet 2." However, some students tend to believe that a nondirectional alternative requires a nondirectional conclusion; for instance, "... Diet I gives a *different* mean weight gain than Diet 2." It is worth noting that a biological researcher, having carried out an expensive and time-consuming experiment to find out which diet gives the higher mean, might reasonably be quite dissatisfied with such a noncommittal conclusion.

(Remark: Strictly speaking, the formal machinery of the Neyman-Pearson theory leads to only two possible decisions -- reject H_0 or do not reject H_0. The procedure recommended in this textbook can be formally justified as follows. The two-tailed t test yields *three* possible decisions; namely, D_0: do not reject H_0; D_1: reject H_0 and conclude $\mu_1 < \mu_2$; and D_2: reject H_0 and conclude $\mu_1 > \mu_2$. With this approach one may consider three risks of error; namely, $Pr\{D_1 \text{ or } D_2 | \mu_1 = \mu_2\}$, $Pr\{D_1 | \mu_1 > \mu_2\}$, and $Pr\{D_2 | \mu_1 < \mu_2\}$. It is easy to show that, using the recommended procedure, all three of these risks are bounded by α; indeed, the latter two are bounded by $\alpha/2$. Related ideas are discussed in Bohrer, R. (1979), "Multiple Three-Decision Rules for Parametric Signs," *Journal of the American Statistical Association 74,* 432-437.)

Comments on Section 7.7

The most common abuse of statistical testing is to base scientific conclusions solely on the P-value, which leads inevitably to a troublesome confusion between statistical significance and practical importance. Section 7.7 attempts to forestall this confusion in two ways: by showing how confidence intervals can supplement tests and by introducing the concept of effect size.

Comments on Optional Section 7.8

Section 7.8 gives a detailed discussion of power and introduces Table 5, which gives the sample size required to achieve a prescribed power.

Comments on Sections 7.9 and 7.10

Section 7.9 discusses the conditions on which the Student's t methods are based, and gives some guidelines for informally checking the conditions. Note that the word "conditions" is used in place of the commonly used "assumptions." This is because students tend to think of an assumption as something that one simply assumes. This is not at all the case in statistics; these are conditions that should be verified whenever possible.

Section 7.10 places hypothesis testing in a general setting, thus preparing the students for tests other than the t test. The topics of P-value and Type I and Type II error are revisited in this general setting.

Comments on Section 7.11

Section 7.11 introduces the Wilcoxon-Mann-Whitney test. This is the students' first exposure to a nonparametric test (they will meet the sign test and Wilcoxon Signed-Rank in Chapter 9).

The first part of Section 7.11 describes the Wilcoxon-Mann-Whitney test procedure. The second part gives the rationale for the test; this material is somewhat difficult and can be omitted without loss of continuity. The last part of Section 7.11 gives the conditions for validity of the Wilcoxon-Mann-Whitney test and compares it with the t test.

Some textbooks incorrectly describe the Wilcoxon-Mann-Whitney test as a comparison of medians, and give much stronger conditions for validity of the test than are necessary. In fact, the Wilcoxon-Mann-Whitney procedure tests the null hypothesis that two continuous population distributions are identical against the alternative that one is stochastically larger than the other. (This matter is further discussed in the textbook in Note 46 to Chapter 7.) The confusion is probably due to the fact that many power calculations, and other developments such as the Hodges-Lehmann estimator, assume that the distributions differ only by a shift.

Comment on Section 7.12

Section 7.12 explicitly acknowledges an unspoken condition for the t test and the Wilcoxon-Mann-Whitney test -- that the population distributions are stochastically ordered. (Although, strictly speaking, this condition cannot be satisfied by two normal distributions with different standard deviations, it clearly can be satisfied by the real-world -- and therefore finite-tailed -- distributions for which the normal distributions are always only approximations.)

Chapter 8 Statistical Principles of Design

Comments on Sections 8.1 - 8.4

Section 8.1 introduces students to the need for considering extraneous variation and the important distinction between observational studies and experiments. Section 8.2 discusses observational studies, confounding, spurious association (i.e., the effects of "lurking variables"), and case-control studies. Section 8.3 discusses experiments and the purpose of randomization and explains the technique for carrying out a completely randomized allocation, using the technique for random sampling that was first introduced in Section 3.2. Section 8.4 defines the randomized blocks design, explains how to carry out blocked and stratified randomization, and discusses principles of constructing good blocks.

A particular challenge in discussing this material is to convey the great value of randomization without suggesting that randomization can *guarantee* balance with respect to extraneous variables. (Indeed, there seems to be widespread confusion on this point. Many medical research articles include tables showing the results of random allocation to treatment and control groups, with P-values given for each of several covariates, as if the randomization is meant to guarantee balance, with large P-values indicating that balance was achieved.) Randomization controls bias in a stochastic sense, rather than by creating exactly equivalent groups of experimental units.

Some statisticians prefer to think of randomization as literally providing the basis for statistical inference. Strictly speaking, this view implies the use of permutation tests, in which the null distribution for a test statistic is generated by the randomization process itself. Under certain conditions, a parametric test such as the t test can be viewed as an approximation to a permutation test.

This textbook takes a different view, regarding the randomization as justification for a population mode. In this model, the experimental units are viewed as a random sample from some population (which may, however, be very narrowly defined) and the randomization as defining several conceptual populations that differ only in the treatment received. There are several reasons why we prefer to present the population model in the textbook. First, it is consistent with the random sampling model used in earlier chapters. Second, it is consistent with the model for studies that do not involve randomized allocation (for example, heights of males versus heights of females). Third, without it simple definitions for parameters such as μ_1 and μ_2 are not available, and topics such as power and confidence intervals become much more difficult to discuss.

Comments on Section 8.5

Section 8.5 aims to help the student recognize a common source of nonindependence of observations, namely, the nesting of observations within a random factor.

Comments on Optional Section 8.6

Section 8.6 discusses the difference between sampling errors and non-sampling errors. The randomized response technique is also presented. Students find this technique interesting and enjoyable, but this is clearly a topic that can be skipped.

Comments on Section 8.7

Section 8.7 indicates the relationship between the design and the analysis of a study, and thus prepares the student for the study of paired samples in Chapter 9.

Chapter 9 Comparison of Paired Samples

Comments on Sections 9.1 - 9.2

Section 9.1 contains a brief introduction to the paired design. After explaining the basic notion of using pairwise differences, Section 9.2 describes the paired t test and confidence interval. Although the basis for these techniques is the single-sample standard error introduced in Chapter 6, the notation used in Chapter 9 (namely, $(\bar{y}_1 - \bar{y}_2)$ and $SE_{(\bar{y}_1 - \bar{y}_2)}$) is the same as in Chapter 7. This notation emphasizes that the object of inference (namely, $\mu_1 - \mu_2$) is the same in both Chapter 7 and Chapter 9, although the df and the formula for $SE_{(\bar{y}_1 - \bar{y}_2)}$ are different.

Comments on Section 9.3

Section 9.3 gives several examples of paired designs and explains that pairing can serve two purposes: (1) control of bias, especially in nonrandomized studies, and (2) increased precision. Pairing can increase precision by inducing a positive correlation between the observations on members of a pair, thus reducing the variance of the difference. The term "correlation" is not used in Chapter 9, but the idea is conveyed by a scatterplot (Figure 9.3); class discussion of such a scatterplot can help students develop some intuition about the meaning of effective pairing. Together, sampling exercises 9.10 and 9.12 (or 9.10 and 7.40 or 7.41) illustrate the increase in power achievable by effective pairing.

Comments on Sections 9.4 and 9.5

Section 9.4 introduces the sign test, which is worthwhile for beginning students for two reasons. First, it is widely applicable, even in many nonstandard situations where a parametric analysis may be complicated or unsuitable. Second, because students are familiar with the binomial distribution they can fully understand how P-values for the sign test are calculated, and this enhances their understanding of P-values in general. Section 9.5 presents the Wilcoxon Signed-Rank test, which is more powerful than the sign test, but not as widely applicable.

Comments on Sections 9.6 and 9.7

These sections contain no new techniques, but rather some deeper discussions of earlier ideas and methods. The discussions illustrate (a) the importance of a control group in studying change; (b) suitable reporting of paired data; (c) the folly of using a paired t analysis to compare measurement methods; and (d) the inability of standard designs and analyses to detect interactions of treatments with experimental units. Sections 9.6 and 9.7 can be omitted without loss of continuity.

Chapter 10 Analysis of Categorical Data.

Comments on Section 10.1

Chapter 10 treats categorical data. Since this topic has not been mentioned since Chapter 6, Section 10.1 reminds students of the distinction between categorical and quantitative observed variables.

The chi-square goodness-of-fit test, introduced in Section 10.1, is already familiar to some biology students from their study of genetics, and these students enjoy seeing this formerly mysterious subject in a new light.

In this textbook, the one-sample binomial test is subsumed under the topic of goodness-of-fit tests. This approach minimizes the number of formulas that the student must master, and nothing is lost, since the commonly used Z test based on a standardized binomial deviate is exactly equivalent to the chi-square test.

The notation $\hat{Pr}\{A\}$ for the estimated probability of a category A is presented in Section 10.1. This notation may seem unnecessarily cumbersome, but its extension in Section 10.3 to the conditional probability notation $\hat{Pr}\{A|B\}$ will be very useful.

The topic of multiple comparisons arises for the first time in Section 10.1; this topic will reappear in the r x k contingency table in Section 10.5 and in analysis of variance in Chapter 11. In Chapter 10 no specific multiple comparison techniques are given; rather, the text simply acknowledges that the chi-square test with more than 1 df is often only the first phase of a statistical analysis.

The term "compound hypothesis" introduced in Section 10.1 is not standard, but was coined by the authors. ("Composite hypothesis" would have been more apt, but this term has a different meaning in statistical theory.)

Comments on Section 10.2

Section 10.2 begins with an extensive development of the 2 x 2 contingency table. The analysis is applied in the context of the comparison of two binomial probabilities. (Note that the commonly used 2-sample Z test for comparing two proportions, based on a standardized difference between estimated probabilities, is exactly equivalent to the chi-square test. The 2-sample Z test for equality of proportions is subsumed in Section 10.3 under the topic of the chi-square test for association in 2 x 2 tables. The chi-square test implicitly "pools" the two sample proportions. By using the chi-square test, rather than presenting the Z test, one avoids the confusion voiced by students who ask "Why do we pool data when calculating the SE here, but not when doing a t-test of $\mu_1 = \mu_2$?" (The answer to this question is that the null distribution of the test statistic (χ^2 or Z) involves a common value of $p_1 = p_2 = p$ and knowing that $p_1 = p_2$ implies that $\sqrt{p_1(1 - p_1)} = \sqrt{p_2(1 - p_2)}$, whereas knowing that $\mu_1 = \mu_2$ does not imply that $\sigma_1 = \sigma_2$ when we have quantitative data.)

Section 10.2 introduces the chi-square test for 2 x 2 contingency table in the setting of comparing two proportions. The notation p_1 and p_2 for the success probabilities is a simple extension of the notation p used in Chapter 6. The expected frequencies and the χ^2 statistic are straightforward to compute, but their relationship to the null hypothesis is not at all obvious to students. Consequently, it is important for students to calculate and compare the estimates $\hat{p}_1$ and $\hat{p}_2$ whenever they carry out a chi-square test.

Comments on Section 10.3

Section 10.3 discusses the interpretation of the chi-square test for the 2 x 2 contingency table in a different context: that in which we have a single bivariate sample and wish to test for independence. (Of course, some applications can be viewed either as a comparison of two binomial proportions (Context 1) or as a test of independence for a single bivariate sample (Context 2). For example, the cross-sectional (Context 2) study of HIV testing in Example 10.19 becomes Context 1 if we condition on treatment.)

Here conditional probability notation is introduced and H_0 is reinterpreted as the hypothesis of independence of the row variable and the column variable. Formal attention is given to the non-obvious fact that the relationship of independence, and moreover the direction of dependence, are invariant under interchange of rows and columns. Examples and exercises emphasize the interpretation and verbal description of dependence relationships (which can be quite difficult for some students); calculation and comparison of estimated conditional probabilities play an important role.

Comments on Optional Section 10.4

Fisher's exact test is presented in Section 10.4. Some instructors will choose to present this material lightly, downplaying the use of combinations to find P-values. Others will choose to skip this section entirely. Some instructors, however, will choose to present the exact test in some detail. If this choice is made, the first three pages of Section 10.4 can be presented before the chi-square test of Section 10.2. The chi-square test can then be presented as an approximation to the exact test, with the point made that the exact test is cumbersome when n is large and the chi-square test is a good approximation in this setting.

Comments on Section 10.5

Section 10.5 extends the ideas of the previous sections to the r x k contingency table. The issue of compound null hypotheses, introduced in Section 10.1, arises again here.

Arguably, the major benefit from studying Section 10.5 is a better intuitive understanding of contingency tables and the relationships they display, rather than the specific technique of the chi-square test. In fact, many r x k tables met in practice involve at least one ordinal variable, and therefore are better analyzed by some other method.

Comments on Section 10.6

Section 10.6 includes conditions for validity of the chi-square goodness-of-fit test and the chi-square test of association in contingency tables. Examples 10.34 and 10.35 illustrate the pitfalls of misapplying these techniques. Explicit attention is given to the fact (too often ignored in practice) that the chi-square test lacks power to detect ordered alternatives (unless r = k = 2).

Comments on Optional Section 10.7

Section 10.7 presents the large-sample approximate confidence interval for $p_1 - p_2$. This interval can be applied in either Context 1 or Context 2.

It is noted that the chi-square test for a 2 x 2 table is approximately equivalent to checking whether the confidence interval includes zero. The equivalence is only approximate because the confidence interval uses an unpooled variance estimate, whereas the chi-square test (which is equivalent to a Z test) uses a pooled variance estimate.

Comments on Optional Section 10.8

Section 10.8 discusses paired categorical data. Perhaps the greatest benefit to students is simply to be made aware that categorical data can be paired (just as Chapter 9 treats paired samples of quantitative data). This fact can be mentioned even if Section 10.8 is skipped.

Comments on Optional Section 10.9

The comparison of samples by calculating *differences*, such as $p_1 - p_2$, can be supplemented by a calculation of *ratios*. Section 10.9 discusses a natural estimate: the relative risk. Odds ratios and confidence intervals for odds ratios are also presented here.

Comments on Section 10.10

From the viewpoint of theoretical statistics, the two chi-square tests -- the goodness-of-fit test and the contingency-table test -- are merely special cases of a more general chi-square test. It is more helpful for students, however, to think of these two tests as different and to learn to recognize when each is applicable. For this purpose, Section 10.10 summarizes and contrasts the two tests.

Chapter 11 Comparing the Means of I Independent Samples

Comments on Sections 11.1 and 11.2

Section 11.1 sets the stage for analysis of variance with an example and a brief preview of the topics to be discussed.

Section 11.2 introduces the basic computations of one-way analysis of variance. A graphical understanding of variability within groups and variability between groups is emphasized. Computational formulas are not given: modern computing and graphing calculators make these unnecessary; their inclusion can cloud understanding and lead students to think that statistics is primarily about calculation.

Comments on Optional Section 11.3

The analysis of variance model is presented in Section 11.3. This model foreshadows the linear regression model presented in Chapter 12, but can easily be omitted.

Comments on Sections 11.4 and 11.5

Sections 11.4 and 11.5 describe the global F test and the conditions for its validity.

Comments on Optional Section 11.6

Section 11.6 presents two-way ANOVA, both for factorial designs and for randomized blocks designs. Interactions, main effects, and simple effects are presented here.

Comments on Optional Section 11.7

Section 11.7 presents linear combinations of treatment means in two settings. The first, adjustment for an uncontrolled covariate, is only briefly discussed and can be omitted without loss of continuity. The second, linear contrasts, is more fully developed and serves as reinforcement of the basic ideas of a two-way factorial design. Many students find linear contrasts quite difficult to grasp. They can be helped by leisurely class discussion of a few simple examples.

Comments on Optional Section 11.8

Section 11.8 presents the Newman-Keuls procedure, which students find straightforward, if somewhat tedious to carry out, and the Bonferroni method. Of course, the Bonferroni idea is very general, so instructors who choose to skip this section may wish to discuss Bonferroni briefly.

The fact that the Newman-Keuls and Bonferroni methods are only two of several competing procedures is briefly indicated in Section 11.7; this is further discussed in Appendix 11.1.

Comments on Section 11.9

Section 11.9 reviews the topics of Chapter 11 and also briefly mentions some other approaches to comparison of several samples. The perspective provided by Section 11.9 may be especially helpful if optional Sections 11.7 and 11.8 have been omitted.

Chapter 12 Linear Regression and Correlation

Comments on Section 12.1

Chapter 12 presents linear regression and correlation in a unified framework rather than treating them as separate topics. Accordingly, Section 12.1 describes two contexts for regression and correlation. In Context 1 the values of X are specified by the experimenter (that is, they are constants), and in Context 2 they are observed (that is, they are values of random variables). Each context is illustrated by an example; these two examples are referred to repeatedly throughout Chapter 12.

Comments on Section 12.2

Section 12.2 describes the least-squares criterion and the computation of the fitted regression line, the residual SS and the residual standard deviation. As in Section 11.2, computational formulas are not given. In the exercises, quantities such as SS(resid) are provided for the student.

Comments on Sections 12.3 and 12.4

Section 12.2 is concerned entirely with descriptive aspects of regression. Sections 12.3 and 12.4 develop the inferential aspects. For abbreviated coverage of Chapter 12, the instructor could omit Sections 12.3 and 12.4 and proceed directly to the descriptive aspects of correlation in Section 12.5; this approach would require also omitting those parts of Section 12.5 - 12.8 that deal with inference.

Section 12.3 develops the parametric interpretation of regression. The linear model is presented in terms of the conditional mean and standard deviation of Y given X; this presentation is appropriate for both Context 1 and Context 2.

Section 12.4 introduces the confidence interval for the slope of the regression line and the t test of the hypothesis of zero slope. These procedures are valid in both Context 1 and Context 2 (for instance, because the conditional coverage probability of the 95% confidence interval is .95 for any given X_1, X_2, .., X_n, it follows that the unconditional coverage probability is also .95).

Comments on Section 12.5

Section 12.5 introduces the correlation coefficient first as a descriptive statistic, and shows how the magnitude of r is related to the scatter of the data points about the fitted regression line.

Next, Section 12.5 turns to inference, emphasizing that the sample correlation r is an estimate of a population parameter ρ only if the data can be viewed as a bivariate random sample; this specifically excludes Context 1. In Context 2 the hypothesis H_0: $\beta_1 = 0$ can be reinterpreted as H_0: $\rho = 0$. The confidence interval for ρ is presented in an optional sub-section that can easily be omitted.

Comments on Section 12.6

Section 12.6 presents guidelines for interpreting linear regression and correlation. Several common pitfalls are discussed. The use of the logarithmic transformation is illustrated.

Comments on Sections 12.7 and 12.8

Section 12.7 contains brief and informal descriptions of several topics, including multiple regression, analysis of covariance, and logistic regression. These are intended to widen students' perspective and perhaps to entice some of them to further study of statistics. In addition, the discussion of regression and the 2-sample t test serves as a partial review of Chapters 7 and 12. Section 12.7 can be omitted without loss of continuity.

Section 12.8 contains a summary of all the formulas introduced in Chapter 12.

Chapter 13 A Summary of Inference Methods

Chapter 13 reviews the various inference methods in the text. This chapter, which provides students an opportunity to think about the many topics presented and to consider how they are related, can be useful as students review for the final exam. Warning: Do not be surprised if students find it difficult to choose an appropriate inference method when doing the exercises.