Statistics for the Life Sciences 5th Edition Samuels Solutions Manual
Full Download: http://testbanklive.com/download/statistics-for-the-life-sciences-5th-edition-samuels-solutions-manual/

INSTRUCTOR'S SOLUTIONS MANUAL

STATISTICS
FOR THE LIFE SCIENCES

FIFTH EDITION

Myra Samuels
Purdue University

Jeffrey A. Witmer
Oberlin College

Andrew Schaffner
California Polytechnic State University

PEARSON

Boston Columbus Hoboken Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Full download all chapters instantly please go to Solutions Manual, Test Bank site: testbanklive.com

This work is protected by United States copyright laws and is provided solely for the use of instructors in teaching their courses and assessing student learning. Dissemination or sale of any part of this work (including on the World Wide Web) will destroy the integrity of the work and is not permitted. The work and materials from it should never be made available to students except by instructors using the accompanying text in their classes. All recipients of this work are expected to abide by these restrictions and to honor the intended pedagogical purposes and the needs of other instructors who rely on these materials.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Reproduced by Pearson from electronic files supplied by the author.

Copyright © 2016, 2012, 2003 Pearson Education, Inc.
Publishing as Pearson, 501 Boylston Street, Boston, MA 02116.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

ISBN-13: 978-013-397624-3
ISBN-10: 0-133-97624-6

www.pearsonhighered.com

PEARSON

Contents

I. General Comments 1

II. Comments on Chapters 3

III. Comments on Sampling Exercises 19

Chapter 1 Introduction 24

Chapter 2 Description of Samples and Population 27

Chapter 3 Probability and the Binomial Distribution 48

Chapter 4 The Normal Distribution 59

Chapter 5 Sampling Distributions 72

Unit I Summary 91

Chapter 6 Confidence Intervals 95

Chapter 7 Comparison of Two Independent Samples 112

Chapter 8 Comparison of Paired Samples 145

Unit II Summary 164

Chapter 9 Categorical Data: One Sample Distributions 170

Chapter 10 Categorical Data: Relationships 188

Unit III Summary 219

Chapter 11 Comparing the Means of Many Independent Samples 222

Chapter 12 Linear Regression and Correlation 255

Unit IV Summary 295

Chapter 13 A Summary of Inference Methods 299

Copyright © 2016 Pearson Education, Inc.

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.

EXERCISES

Calculators and Computers

The exercises in the text are designed to minimize numerical drudgery and emphasize understanding. Exercises with simple numbers familiarize the student 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, data files are available for many of the examples and exercises in the book. Several exercises present computer output and ask the student to interpret the results. 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. Exercises in the Unit Summary sections often use material from more than one chapter.

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.

Sampling Exercises

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 website for 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 non-directional, 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 selected values of the test statistic in bold type and corresponding non-directional P-values in italics. This is somewhat non-standard, but we believe is more informative than are standard tables. The column headings direct the reader to critical values for which the P-value is less than or equal to the given column heading. Tables for the chi-square test and the F test give upper tail areas, as is appropriate for use with the usual non-directional 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 begins with a series of examples of such settings. Instructors may choose to discuss these or other examples in the first lecture. Section 1.3 then addresses sampling issues, including discussion of sampling errors and nonsampling errors.

Comments on Section 1.2

Section 1.2 discusses data collection, particularly the difference between an observational study and an experiment. By drawing attention to data collection issues at the onset of the course, we hope that students are mindful of need to question where data came from before conducting any analysis.

Comments on Section 1.3

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. 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; and (b) sampling exercises in Chapters 1, 5, 6, 7, and 8 (see Section III of this Manual) require the technique.

Note that in some cases we regard the data as being like a random sample, although they did not arise that way. For example, consider a group of students in a class: We would not think of them as being a random sample of all students on campus, but if the variable of interest is blood type, then we might regard their blood types as being like a random sample, on the assumption that blood type is unrelated to a student enrolling in a given class.

Chapter 2 Description of Samples and Populations

Comments on Sections 2.1 and 2.2

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."

Comments on Sections 2.3 - 2.6

Students will want to use their calculators, or computers, when solving homework problems. Two useful general principles are:

(a) Minimize roundoff error by keeping intermediate answers in a 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 a 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 a 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) Note that 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 or computer software in doing homework. Generally, the exercises have been written under the assumption that students would use the SD function of a calculator or would use software to obtain a sample SD. 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 Probability, and the Binomial Distribution

Comments on Section 3.2

This section introduces 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 3.3. 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 3.3), although Section 3.2 introduces probability trees, which implicitly use conditional probabilities. Rather independence is introduced informally in various settings throughout the text (which is why Section 3.3 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 also introduced in Chapter 6; 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 3.3

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 3.4 and 3.5

Continuous distributions are introduced in Section 3.4. The purpose of Exercises 3.4.1-3 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 3.5 introduces the concept of a random variable and more formally the population mean and variance via a few simple examples.

Comments on Section 3.6

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 3.1; the derivation of the mean and of the standard deviation are found in Appendix 3.2.

Comments on Optional Section 3.7

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 treatment of the normal distribution. Some instructors will want to use technology completely and not make use of Table 3. Even so, it is helpful if students 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.S.1 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 quantile 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. Others will wish to discuss the optional sub-section on the Shapiro-Wilk test, wherein the presentation is necessarily brief and somewhat informal, given that P-values are not fully discussed until later in the text.

Chapter 5 Sampling Distributions

Comments on Sections 5.1 and 5.2

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 1.3 and 2.8).

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 (5.2.1, 5.2.2, and 5.2.3) 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 5.2.7 and 5.S.12), 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 5.3

Section 5.3 illustrates the effect of the Central Limit Theorem on a moderately skewed population (Example 5.3.1) and a violently skewed population (Example 5.3.2). 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 5.3 need not be covered at all. A sketchy coverage of Section 5.3 can be achieved in ten minutes of class time by simply displaying and briefly discussing Figures 5.3.1-5.3.4; a more leisurely coverage permits detailed discussion of Example 5.3.2 and assignment of one of the exercises (5.3.1 or 5.3.2).

Background Information on Examples 5.3.1 and 5.3.2

For the record, the distribution in Example 5.3.1 is a shifted and scaled chi-square distribution fitted to a graph of the data of Zeleny (see source cited in Example 5.3.1). The distributions in Figure 5.3.2 were estimated by the authors and R. B. Becker using computer simulation; for each sampling distribution, we drew 100,000 samples from the shifted chi-square distribution. The distribution in Example 5.3.2 is a 9:1 mixture of two normal distributions with parameters $\mu_1 = 115$, $s_1 = 37.5$, $\mu_2 = 450$, and $s_2 = 75$; the parameters are based on a simplified version of data of Bradley (see sources cited in Example 5.3.2). The curves in Figure 5.3.4 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 = 0.9.

Comments on Optional Section 5.4

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 9.2 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 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 6.2.4, 6.2.5, 6.2.6, 6.2.7, 6.3.8, 6.S.12, and 6.S.13 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.

The last part of Section 6.3 presents one-sided confidence intervals. This can easily be omitted as the material is not used elsewhere in the text.

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 0.80 for a two-tailed t test at $\alpha = 0.05$.

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 6.5.2, 6.S.7, 6.S.9, 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 5.3 in this Manual).

Comments on Sections 6.6 and 6.7

In addition to introducing the notation for two samples, Section 6.6 alerts students to the fact that distributions can differ in dispersion and shape as well as in location, while noting that our attention will be given to comparisons of center. Section 6.6 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, which some instructors may wish to discuss as preparation for later treatment of analysis of variance. (The case $\sigma_1 \neq \sigma_2$ often occurs in biological research.)

The textbook primarily uses the Satterthwaite formula (Formula (6.7.1) on page 206) for degrees of freedom, on the view that technology should be used for number crunching, so that the messy form of Formula (6.7.1) 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 Section 6.8

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 introduces the general principles of hypothesis testing by beginning with the randomization test and then developing the t test. The chapter concludes with the Wilcoxon-Mann-Whitney test, which is the distribution-free competitor to the t test.

Although some introductory tests include the F test for comparison of variances, we have omitted this topic. In some situations comparing variances is more relevant than comparing means. However, when means are being compared 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 more fully discussed in Section 7.4.

Comments on Section 7.1

Section 7.1 introduces the randomization test for comparing two populations, which some would call a permutation test since it is based on considering the distribution of possible permutations of sample data. Although this is not labeled as an optional section, it is possible to skip Section 7.1 and to begin with Section 7.2. However, we believe that demonstrating a simple basis for determining a P-value in Section 7.1 and then using the t test as an approximation, starting with Section 7.2, is a good way to secure the idea of hypothesis testing and the proper interpretation of a P-value.

Students readily understand that computer simulation is needed when the sample sizes are large and listing all possible permutations in not practical, but it is advisable to conduct a physical randomization, using the data of Example 7.1.1 or of some other small two-sample comparison, with 3x5 cards that contain one observation each. Shuffling and dealing the cards into two piles and calculating the difference in means demonstrates the concept underlying this test procedure.

Comments on Sections 7.2 and 7.3

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 7.2 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.

Comments on Section 7.4

Section 7.4 gives explicit attention to the difference between association and causation, alerting students to the existence of both experimental and observational studies. This is arguably the most important section in the entire book.

Comments on Section 7.5

When considering one-tailed tests and directional H_A the textbook uses a two-step procedure to bracket the P-value. 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 – as well as some statisticians – 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 7.6

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 7.6 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 7.7

Section 7.7 gives a detailed discussion of power and introduces Table 5, which gives the sample size required to achieve a prescribed power.

Comments on Sections 7.8 and 7.9

Section 7.8 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 7.9 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.

Section 7.9 explicitly acknowledges an unspoken condition for the t test (and the Wilcoxon-Mann-Whitney test and the randomization 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.)

Comments on Section 7.10

Section 7.10 introduces the Wilcoxon-Mann-Whitney test. This is the students' first exposure to a classical nonparametric test (they will meet the sign test and Wilcoxon Signed-Rank in Chapter 8).

The first part of Section 7.10 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 7.10 gives the conditions for validity of the Wilcoxon-Mann-Whitney test and compares it with the t test and the randomization 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 65 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.

Chapter 8 Comparison of Paired Samples

Comments on Sections 8.1 - 8.2

Section 8.1 contains a brief introduction to the paired design and to a randomization test for this setting. After explaining the basic notion of using pairwise differences, Section 8.2 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 8 (namely, ($\bar{y}_1$ - $\bar{y}_2$) and $SE_{(\bar{Y}_1 - \bar{Y}_2)}$) is the same as in Chapters 6 and 7. This notation emphasizes that the object of inference (namely, μ_1 - μ_2) is the same in Chapters 6 and 7 as it is in Chapter 8, although the df and the formula for $SE_{(\bar{Y}_1 - \bar{Y}_2)}$ are different.

Comments on Section 8.3

Section 8.3 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 as often there is a positive correlation between the observations on members of a pair, which reduces the variance of the difference. The term "correlation" is not used in Chapter 8, but the idea is conveyed by a scatterplot (Figure 8.3.1); class discussion of such a scatterplot can help students develop some intuition about the meaning of effective pairing. Together, sampling exercises 8.3.1 and 8.3.3 (or 8.3.1 and 7.3.2 or 7.3.3) illustrate the increase in power achievable by effective pairing.

Comments on Sections 8.4 and 8.5

Section 8.4 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 8.5 presents the Wilcoxon Signed-Rank test, which is more powerful than the sign test, but not as widely applicable.

Comments on Section 8.6

This section contains 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.

Chapter 9 Categorical Data: One-Sample Distributions.

Comments on Sections 9.1 and 9.2

Section 9.1 develops the sampling distribution for the Wilson-adjusted sample proportion, $\tilde{P}$ Section 9.2 then presents the Wilson confidence interval for a proportion. Note that some authors refer to this as the "plus 2; plus 4" interval. Appendix 9.1 discusses the use of $\tilde{P}$, which is becoming more widely used in statistical practice, although the traditional $\hat{P}$-based confidence interval remains very common.

Students using graphing calculators or standard statistical software will find that the technology produces confidence intervals based on $\hat{P}$ (although it is easy to add 2 successes and 2 failures to the data so as to get the Wilson interval). 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 further discussion, see Appendix 9.1.

The material in these sections is related to the material in Sections 6.2 and 6.3. In Appendix 3.2 it is shown that in the setting of 0-1 data, $\mu = p$ and $s = \sqrt{p(1-p)}$, so the fact that $s_{\bar{Y}} = s/\sqrt{n}$ corresponds directly to the fact that $s_{\hat{P}} = \sqrt{p(1-p)}/\sqrt{n}$; 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 $\hat{P}$, and likewise of $\tilde{P}$, when n is large.

Comments on Section 9.3

Optional Section 9.3 considers confidence levels other than 95%, for which the rule "add 2 successes and 2 failures" is modified.

Comments on Section 9.4

The chi-square goodness-of-fit test, introduced in Section 9.4, 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 9.4. This notation may seem unnecessarily cumbersome, but its extension in Section 10.2 to the conditional probability notation $\hat{\Pr}\{A|B\}$ will be very useful.

The term "compound hypothesis" introduced in Section 9.4 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.)

The topic of multiple comparisons arises implicitly for the first time in Section 9.4; 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 9 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.

Chapter 10 Categorical Data: Relationships.

Comments on Sections 10.1 and 10.2

Section 10.1 introduces the 2 x 2 contingency table and conditional probability, followed by a randomization test for the 2 x 2 setting. Section 10.2 then presents the chi-square test for comparing two proportions.

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 10.2 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.)

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 sample proportions 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 10.1.2 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 2.5 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 9.4, 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 10.6.1 and 10.6.2 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 8 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 Many 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. A randomization test for comparing several means is presented.

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. The computational formulas presented are based on familiar summary statistics (the sample mean and standard deviation) and are not emphasized: 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 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 could be omitted, although the one-way ANOVA model is extended in optional Section 11.7 when two-way ANOVA is presented, so instructors who intend to cover Section 11.7 should cover Section 11.3 as well.

Comments on Sections 11.4 and 11.5

Section 11.4 and 11.5 describe the global F test and the conditions for its validity. Section 11.5 also introduces residual analysis as a pooled technique to assess the validity conditions.

Comments on Section 11.6

Section 11.6 presents ANOVA for randomized (complete) blocks designs. The section begins with a discussion of the powerful idea of blocking and how it can be used in a variety of settings to remove variability and increase power. The latter portion of this section, which provides computational details, could be omitted by those only wishing to provide a conceptual introduction to blocking.

Comments on Section 11.7

Section 11.7 presents two-way ANOVA for factorial designs. Interactions, main effects, and simple effects are presented here.

Comments on Optional Section 11.8

Section 11.8 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 11.9

Section 11.9 presents three multiple comparison procedures: Fisher's Least Significant Difference, the Bonferroni method, and Tukey's Honest Significant Difference. Of course, the Bonferroni idea is very general, so instructors who choose to skip this section may wish to discuss Bonferroni briefly. Exercises 11.9.7 and 11.9.8 present uses of the Bonferroni adjustment outside of ANOVA settings.

Comments on Section 11.10

Section 11.10 reviews the topics of Chapter 11 and also briefly mentions some other approaches to comparison of several samples. The perspective provided by Section 11.10 may be especially helpful if optional Sections 11.8 and 11.9 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 Example 12.1.1 the values of X are specified by the experimenter (that is, they are constants), and in Example 12.1.2 they are observed (that is, they are values of random variables).

Comments on Section 12.2

Section 12.2 introduces the correlation coefficient as a descriptive statistic that measures the strength of a linear association. The bivariate sampling model is presented, which provides the foundation for testing the hypothesis that the population correlation is zero. A randomization test is presented along with the t test. An optional sub-section that can easily be omitted uses the Fisher transformation to develop a confidence interval.

Comments on Section 12.3

Section 12.3 presents the regression line as a smoothed version of the Graph of Averages and notes that the slope is equal to the correlation scaled by the ratio of two standard deviations. Algebraic detail is provided in Appendix 12.1. Later in the section the least-squares criterion and the residual SS are presented, along with the residual standard deviation. In the exercises, quantities such as SS(resid) are provided for the student.

Comments on Sections 12.4 and 12.5

Section 12.3 is concerned entirely with descriptive aspects of regression. Sections 12.4 and 12.5 develop the inferential aspects. For abbreviated coverage of Chapter 12, the instructor could omit Sections 12.4 and 12.5; this approach would require also omitting those parts of Sections 12.6 - 12.9 that deal with inference.

Section 12.4 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 regression and correlation contexts presented in Section 12.1.

Section 12.5 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 regression and correlation contexts presented in Section 12.1 (for instance, because the conditional coverage probability of the 95% confidence interval is 0.95 for any given X_1, X_2, .., X_n, it follows that the unconditional coverage probability is also 0.95). A randomization test is mentioned for settings in which the normality condition is violated.

Comments on Section 12.6

Section 12.6 presents guidelines for interpreting linear regression and correlation and distinguishes the effects on the regression line due to outliers and influential points. Several common pitfalls are discussed. The use of the logarithmic transformation is illustrated.

Comments on Optional Section 12.7

Section 12.7 presents prediction intervals, which are contrasted with confidence intervals. It is worth noting that prediction intervals can be constructed in many other statistical settings and that a common confusion among users of statistics is to confuse these two types of intervals (e.g., when analyzing a mean with a single sample, as in Chapter 6).

Comments on Sections 12.8 and 12.9

Section 12.8 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 12.8 can be omitted without loss of continuity.

Section 12.9 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.

III COMMENTS ON SAMPLING EXERCISES

To illustrate various aspects of probability, the text includes sampling exercises in Chapters 1, 5, 6, 7, and 8. For most effective use, these exercises should be carried out independently by each student, and the results then aggregated and discussed in class. Briefly, the sampling exercises cover the following topics.

Exercise 1.3.6: The difference between random sampling and judgment sampling.

Exercise 1.3.7: The binomial distribution.

Exercise 5.2.1: The inadequacy of judgment sampling.

Exercises 5.2.2 and 5.2.3: The sampling distribution of $\bar{Y}$ and (optionally) the sampling distribution of s.

Exercises 6.3.1 and 6.3.2: Confidence intervals.

Exercises 7.3.1, 7.3.2, 7.3.3: Significance level and power.

Exercises 8.3.1, 8.3.2, 8.3.3: Effect of matching on power.

The exercises in different chapters can be used independently, except for Exercises 6.3.1 and 6.3.2, which refer to results of 5.2.1 and 5.2.3.

Below are comments and background information on specific sampling exercises.

Exercises 1.3.6 and 1.3.7

> There are 39 ellipses with tail bristles, so that in fact p = 0.39 for this population. By accumulating the results from the entire class the instructor can illustrate two key ideas: (1) the outcome of any given sample is uncertain, but (2) we can predict with reasonable certainty the aggregate pattern formed by many samples. For a deeper understanding, the instructor can point out that in fact the trials are not really independent -- for instance, the (conditional) probability that the second ellipse is a mutant is either 38/99 or 39/99, depending on whether the first ellipse chosen was mutant or not -- but that the binomial approximation is reasonably good here because 38/99 and 39/99 are approximately equal. For the record, the following table shows binomial probabilities and the exact hypergeometric probabilities based on sampling without replacement.

Number of		Probability	
Mutants	Nonmutants	Binomial	Exact
0	5	0.0845	0.0790
1	4	0.2700	0.2703
2	3	0.3452	0.3542
3	2	0.2207	0.2221
4	1	0.0706	0.0666
5	0	0.0090	0.0076

The two sets of probabilities are barley distinguishable. Of course, the relative frequencies from a class sampling exercise may be expected to differ from the probabilities, both because of sampling error and because not all students carry out the assignment correctly.

Exercise 5.2.1

Exercise 5.2.1 can be used to demonstrate the inadequacy of human judgment as a substitute for random sampling. Experience has shown that judgment sampling of the ellipse population is biased in the predictable direction: students tend to over-sample the larger ellipses, so that the sampling distribution of $\bar{Y}$ is shifted toward higher values. The true value of the population average length is $\mu = 10.7$ mm. Students tend to get larger values from their judgment samples.

The bias can also be illustrated by comparing the empirical sampling distribution for Exercise 5.2.1 with that for Exercise 5.2.2.

(Note: Exercise 5.2.4 is related to Exercise 5.2.1. In Exercise 5.2.4, an explicit sampling method is proposed that favors larger ellipses.)

Exercises 5.2.2 and 5.2.3

These exercises can be used to illustrate the meaning of sampling distributions. The following classes are convenient for accumulating student results into sampling distributions:

For the sampling distribution of $\bar{Y}$, class limits [5,6), [6,7),…,[22,23) are suitable both for $n = 5$ (Exercise 5.2.2) and $n = 20$ (Exercise 5.2.3); using these class limits is very easy, since only the integer part of each student's answer need be recorded. Optionally, the instructor may also choose to discuss the sampling distribution of the sample standard deviation s. For the sampling distribution of s, class limits of [0,1), [1,2),…,[10,11) are suitable for both $n = 5$ and for $n = 20$; again, only the integer part of each student's answer need be recorded.

The true values of the population parameters are:

$\mu = 10.7$ mm
$s = 4.9$ mm

The following table shows a grouped frequency distribution of the ellipse lengths.

Length (mm)	Frequency
5-6	21
7-8	24
9-10	11
11-12	13
13-14	12
15-16	4
17-18	1
19-20	8
21-22	6
Total	100

An effective class presentation is to display the histograms of the preceding population distribution and of the empirically determined sampling distribution of $\bar{Y}$ for n = 5 and n = 20; if desired, normal curves can be superimposed on the sampling distributions (curves with mean 10.7 and SD $4.9/\sqrt{5}$ = 2.2 and $4.9/\sqrt{20}$ = 1.1). The instructor can point out the following features: (1) the SD of the sampling distribution decreases as n increases; (2) the sampling distributions are less skewed than the population distribution (showing the Central Limit Theorem at work). The first feature can be illustrated again for the sampling distribution of s.

For the record, the following is a list of the lengths (to the nearest mm) of all 100 ellipses in the population.

Ellipse #	Length	Ellipse #	Length	Ellipse #	Length
00	10	34	13	68	13
01	7	35	11	69	22
02	14	36	7	70	12
03	20	37	20	71	8
04	6	38	8	72	6
05	11	39	5	73	5
06	9	40	13	74	14
07	6	41	9	75	11
08	13	42	15	76	8
09	21	43	20	77	21
10	12	44	6	78	7
11	8	45	6	79	8
12	6	46	9	80	13
13	6	47	5	81	7
14	15	48	15	82	12
15	10	49	21	83	20
16	7	50	10	84	7
17	18	51	8	85	11
18	7	52	7	86	9
19	5	53	5	87	7
20	12	54	13	88	14
21	6	55	11	89	20
22	14	56	7	90	10
23	21	57	21	91	7
24	8	58	7	92	6
25	11	59	7	93	6
26	7	60	12	94	14
27	6	61	7	95	11
28	14	62	15	96	7
29	20	63	19	97	19
30	10	64	6	98	5
31	9	65	12	99	8
32	6	66	9		
33	5	67	6		

Exercises 6.3.1 and 6.3.2

These exercises require the results of Exercises 5.2.1 and 5.2.3. As noted above, the true population mean is $\mu = 10.7$ mm. After constructing their confidence intervals, students can be told the value of μ and asked to determine whether each of their confidence intervals covers μ. A show of hands then illustrates the point that about 80% of the confidence intervals cover μ and that this is true regardless of n. (Of course, the 80% is an approximation, because, as seen above, the population shape is not normal.) A brief discussion can then lead the students to recognize that the intervals based on $n = 20$ tend to be narrower than those based on $n = 5$.

Exercise 7.3.1

This exercise illustrates Type I error. After the assignment is completed, a count can be made in class of those students who made Type I errors. The instructor can point out that only about 5% of the students are expected to make Type I errors and that no one could make a Type II error because the null hypothesis is true.

Exercise 7.3.2

This exercise illustrates Type II error and is a natural companion to Exercise 7.3.1. A count in class can be used to illustrate the fact that the risk of Type II error is quite high because n is so small. (Pretending the population distribution is normal gives a power of 0.27.)

Exercise 7.3.3

This exercise can be used as a substitute for Exercises 7.3.1 and 7.3.2. It is a bit more trouble, but also more fun, since each data set is analyzed by a student who does not know whether H_0 is true. The Instructor Copy can be used to give credit for doing the first part of the assignment. When the t tests have been completed, each student should tell his or her partner whether he or she modified one of the samples. Counts of Type I and Type II errors can then be made as indicated in the above comments on Exercises 7.3.1 and 7.3.2. (Note that, if one of the samples was modified, then the chance of rejecting H_0 and deciding that the *other* sample was modified is negligible.)

Exercises 8.3.1, 8.3.2, and 8.3.3

These exercises illustrate the gain in power achievable by a paired design. The correlation between the first and second members of a pair is .94; the paired design is therefore much more powerful than the unpaired design. (The SD of the d's is 1.97; pretending that their distribution is normal gives a power of 0.99 for Exercise 8.3.1; by contrast, the analogous assumption in Exercise 7.3.2 gave a power of only 0.27.)

For fuller understanding, the class results for Exercise 8.3.1 should be compared with an unpaired analysis. If Exercise 7.3.2 or 7.3.3 was assigned, it can be used for comparison; if not, Exercise 8.3.3 can be used instead. Another option is Exercise 8.3.2, which permits comparison of the paired with the unpaired *analysis* of data generated by a paired design.

IV SOLUTIONS TO EXERCISES

Note: Exercises whose answers are given in the back of the textbook are denoted by the symbol •.

CHAPTER 1
Introduction

1.2.1 It appears that the digestive problems were caused by the nocebo effect. People feared that fluoridation of their drinking water would cause health problems and this fear led to digestive problems when, in fact, fluoride was not yet being added to the water.

1.2.2 People eating potato chips made with olestra might expect to have gastrointestinal problems. Thus, the expectation of problems might lead to those problems occurring (a nocebo effect). It is important that the subjects don't know which group they are in, so that any nocebo effect is seen evenly across the two groups. Likewise, the persons evaluating the subjects should be blinded, so that there is no bias in recording any gastrointestinal symptoms that arise.

• 1.2.3 The acupuncturist expects acupuncture to work better than aspirin, so she or he is apt to "see" more improvement in someone given acupuncture than in someone given aspirin -- even if the two groups are truly equivalent to each other in their response to treatment.

1.2.4 The surgeons selected patients, rather than having patients chosen at random. It is not likely that a patient on the verge of death would have been enrolled in the Vitamin C study, since the surgeons would not have expected that person to live long enough to take much Vitamin C. It is more likely that relatively healthy patients (among those with terminal cancer) would be selected. Such selection would result in a treatment group that, even without the Vitamin C, would have been expected to live longer than did historical controls. Moreover, recently diagnosed cancer patients -- those in the study -- may have been diagnosed earlier in the progression of disease than the historical controls were diagnosed (due to improved medical testing). These factors would make Vitamin C look better than it really is.

1.2.5 This is an anecdote; it is not convincing evidence that the procedure is effective. For example, it might be that rigorous, daily adherence to a treatment regimen made the difference, but the particular treatment was not important. Or it may be that the cumulative effect of previous treatments provided the cure, so that the fungus would have vanished over 100 days even if the person had done nothing.

1.2.6 (a) (I) This should be an experiment; (II) it should be double-blind (i.e., neither the subjects nor the evaluating physicians should know who is in which group).

(b) (I) This should be an observational study; (II) there is no need for blinding here.

1.2.7 (a) (I) This should be an observational study; (II) it should be single blind: The people who measure the brains should be blinded to the sexual orientations of the men.

(b) (I) This should be an experiment, with people randomly assigned to one of two groups: high (more than 1 liter per day) vs low water intake; (II) it should be single blind: The people who measure the weights of the subjects should be blinded to the group assignments.

1.2.8 (a) This was an experiment, since the treatment was applied by the researchers.

(b) The experiment confounds two effects: the effect of the fertilizer and the effect of being on the west side of the garden. It might be that the fertilizer has no effect, but that plants on the west side of the garden grow better than do plants on the east side.

(c) The persons weighing the tomatoes could be blinded as to whether the plants came from the east side or the west side of the garden. Double-blinding, however, does not make sense, since this would involve the tomato plants somehow being unaware of what treatment they were getting.

1.2.9 No. This is an observational study, which means that confounding variables are a concern. In this case, one might expect that relatively healthy persons are more likely to attend religious services than are those with weak immune systems, so it may be that the immune system affects attendance, rather than the other way around. It could also be that attendance at services is beneficial to the immune system, but that the benefit is due to social interaction, not to the religious nature of the services, per se. Other explanations are also possible.

1.2.10 No. It might be that within a given age, sex, and socioeconomic group those people who are healthiest are most likely to play golf, while those in poorer health may not play golf or may have quite playing golf due to poor health.

• **1.3.1 (a)** Cluster sampling. The three clinics are the three clusters.

(b) Simple random sampling.

(c) Stratified random sampling. The strata are the altitudes.

(d) Simple random sampling.

(e) Stratified random sampling. The three breed sizes are the strata.

• **1.3.2 (a)** The sample is nonrandom and likely nonrepresentative of the general population because it consists of (1) volunteers from (2) nightclubs. (i) The social anxiety level of people who attend nightclubs is likely lower than the social anxiety level of the general public. (ii) A better sampling strategy would be to recruit subjects from across the population.

(b) Bias arises from the data being collected only on rainy days. (i) Water pollution readings in the stream might be lower when rain water is mixed in with regular stream water. (ii) A better method would be to sample during all types of weather.

(c) Recording observations only when random coordinates are within a tree canopy will induce bias. (i) Trees with large canopies are more likely to be included in the sample than are other trees, but canopy size is probably related to tree radius, resulting in a sample average that is too large. (ii) A better sampling method would be to measure whichever tree trunk is closest to randomly choose coordinates, but even this produces bias in favor of large trees. In order to avoid the bias, number all of the trees within a region and drawn random numbers to select trees to measure. Geographical regions could be used as strata within a stratified sampling plan of this type.

(d) Fish caught by a single vessel on one day are not a random sample. (i) If the vessel is in a region that has not been fished recently and thus contains large fish, for example, then the sample average will be too large. (ii) To avoid this bias, use randomly chosen fishing vessels on randomly chosen days.

1.3.3 (a) Fish caught by a single vessel on one day are not a random sample. (i) If the vessel is in a region that has not been fished recently and thus contains large fish, for example, then the sample average will be too large. (ii) To avoid this bias, use randomly chosen fishing vessels on randomly chosen days.

(b) Students who eat breakfast, particularly those who eat before 8:30 am, are not a random sample. (i) Those who eat breakfast before 8:30 am in a dining hall might eat a healthier diet than other students, particularly since they don't have to cook breakfast for themselves. (ii) To avoid bias, use a random sample of students including those who do not eat breakfast at all.

(c) Residents who complain of headaches might be highly susceptible to the placebo effect. (i) Residents might experience rapid pain relief due to a placebo effect, meaning that if they were given a sugar tablet but told that it was a pain killer, then they might experience rapid pain relief for psychological reasons. (ii) To control for the placebo effect, give the pain killer to some residents but a placebo to others, in a randomized and double-blind experiment.

1.3.4 (a) Random selection would mean that each digit should be chosen 1/4th of the time.

(b) Results will vary. It is common for people to select the digit 3 more frequently than expected and the digits 1 and 4 less frequently. Based on the experience of the authors, typical proportions for the digits 1, 2, 3, and 4 are around 0.15, 0.25, 0.45, and 0.15, respectively.

(c) These sample proportions suggest that people do not chose numbers randomly, but instead favor certain numbers, particularly 3.

1.3.5 Divide the random digits into groups of size three and use these to select IDs:

7 2 8 || 1 2 1 || 8 7 6 || 4 4 2 || 1 2 1 || 5 9 3 || 7 8 7 || 8 0 3 || 5 4 7 || 2 1 6 || 5 9 6 || 8 5 1

The first potential ID is 728, which is greater than 600 and thus is omitted. The second ID of 121 is included, then 876 is omitted because it is greater than 600, then 442 is included, but the second 121 is skipped since 121 is already in the sample. Continuing, we include 593, 547, and 216. The individuals selected are thus 121, 442, 593, 547, and 216.

1.3.6 – 1.3.7 See Section III of this Manual.

CHAPTER 2

Description of Samples and Populations

2.1.1 (a) i) Molar width
  ii) Continuous variable
  iii) A molar
  iv) 36

(b) i) Birthweight, date of birth, and race
  ii) Birthweight is continuous, date of birth is discrete (although one might say categorical and ordinal), and race is categorical
  iii) A baby
  iv) 65

• **2.1.2 (a)** i) Height and weight
  ii) Continuous variables
  iii) A child
  iv) 37

(b) i) Blood type and cholesterol level
  ii) Blood type is categorical, cholesterol level is continuous
  iii) A person
  iv) 129

2.1.3 (a) i) Number of leaves
  ii) Discrete variable
  iii) A plant
  iv) 25

(b) i) Number of seizures
  ii) Discrete variable
  iii) A patient
  iv) 20

2.1.4 (a) i) Type of weather and number of parked cars
  ii) Weather is categorical and ordinal, cars is discrete
  iii) A day
  iv) 18

(b) i) pH and sugar content
  ii) Both variables are continuous
  iii) A barrel of wine
  iv) 7

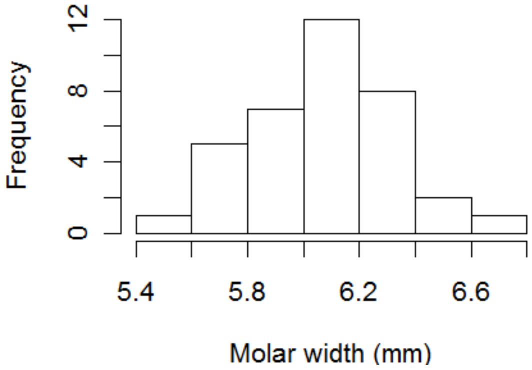
2.1.5 (a) i) body mass and sex
  ii) body mass is continuous, sex is categorical
  iii) A blue jay
  iv) 123

(b) i) lifespan, thorax length, and percent of time spent sleeping

ii) Lifespan is discrete, thorax length and sleeping time are continuous
iii) A fruit fly
iv) 125

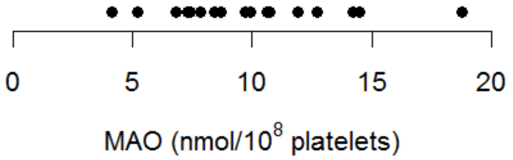
• **2.2.1 (a)**There is no single correct answer. One possibility is:

Molar width	Frequency (no. specimens)
[5.4, 5.6)	1
[5.6, 5.8)	5
[5.8, 6.0)	7
[6.0, 6.2)	12
[6.2, 6.4)	8
[6.4, 6.6)	2
[6.6, 6.8)	1
Total	36



(b) The distribution is fairly symmetric.

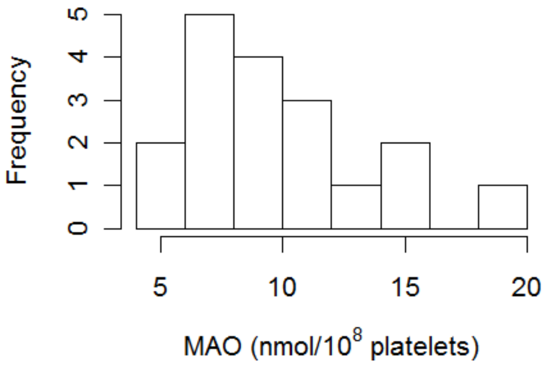
2.2.2



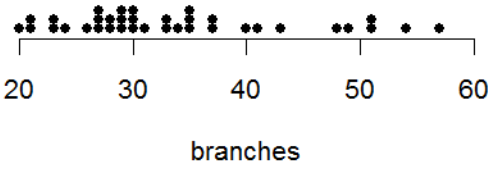
2.2.3 There is no single correct answer. One possibility is

MAO	Frequency (no. patients)
4.0-5.9	2
6.0-7.9	5
8.0-9.9	4
10.0-11.9	3
12.0-13.9	1
14.0-15.9	2

16.0-17.9	0
18.0-19.9	1
Total	18

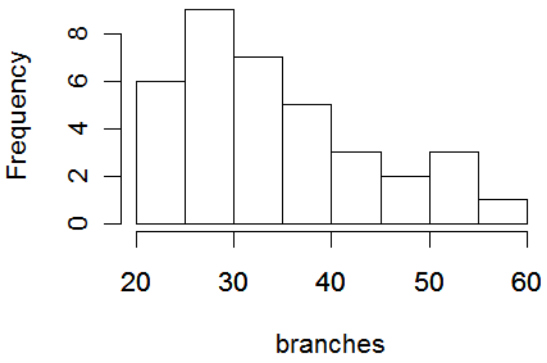


2.2.4



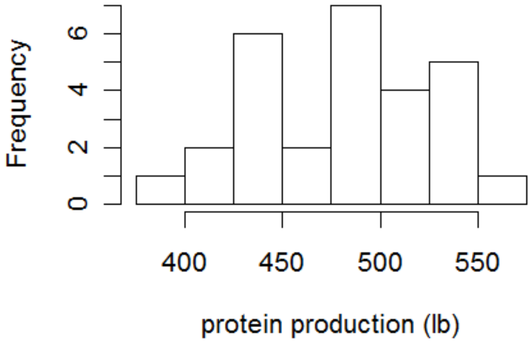
2.2.5 There is no single correct answer. One possibility is

Branches	Frequency (no. cells)
20-24	6
25-29	9
39-34	7
35-39	5
40-44	3
45-49	2
50-54	3
55-59	1
Total	36



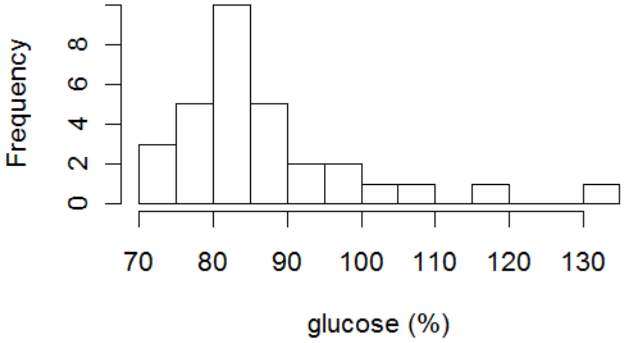
2.2.6 There is no single correct answer. One possibility is

Protein production	Frequency (no. cows)
375-399	1
400-424	2
425-449	6
450-474	2
475-499	7
500-524	4
525-549	5
550-574	1
Total	28

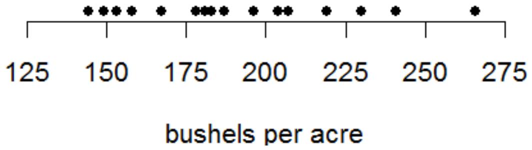


• **2.2.7** There is no single correct answer. One possibility is

Glucose (%)	Frequency (no. of dogs)
70-74	3
75-79	5
80-84	10
85-89	5
90-94	2
95-99	2
100-104	1
105-109	1
110-114	0
115-119	1
120-124	0
125-129	0
130-134	1
Total	31

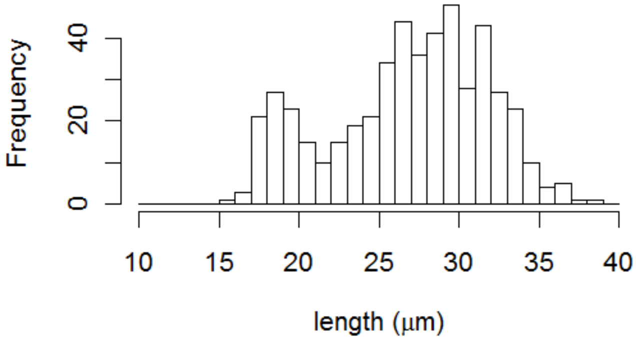


2.2.8 (a)



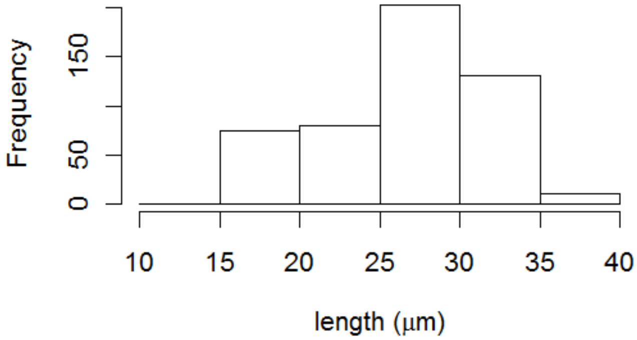
(b)
The distribution is very slightly skewed to the right.

2.2.9 (a)



(b) The distribution is bimodal.

(c) The histogram with only 6 classes obscures the bimodal nature of the distribution.



• **2.3.1** Any sample with $\Sigma y_i = 100$ would be a correct answer. For example: 18, 19, 20, 21, 22.

2.3.2 Any sample with $\Sigma y_i = 100$ and median 15 would be a correct answer. For example: 13, 14, 15, 28, 30.

2.3.3 $\bar{y} = \Sigma y_i/n = \dfrac{6.3 + 5.9 + 7.06.9 + 5.9}{5}$ = 6.40 nmol/gm. The median is the 3rd largest value (i.e., the third observation in the *ordered* array of 5.9 5.9 6.3 6.9 7.0), so the median is 6.3 nmol/gm.

2.3.4 Yes, the data are consistent with the claim that the typical liver tissue concentration is 6.3 nmol/gm. The value of 6.3 fits comfortably near the center of the sample data.

• **2.3.5** $\bar{y}$ = 293.8 mg/dl; median = 283 mg/dl.

• **2.3.6** $\bar{y}$ = 309 mg/dl; median = 292 mg/dl.

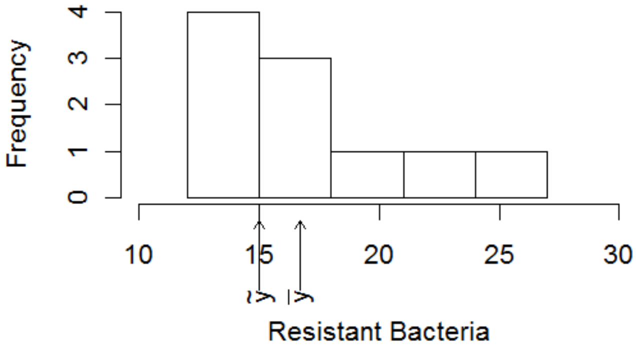
2.3.7 $\bar{y}$ = 3.492 lb; median = 3.36 lb.

2.3.8 Yes, the data are consistent with the claim that, in general, steers gain 3.5 lb/day; the value of 3.5 fits comfortably near the center of the sample data. However, the data do not support the claim that 4.0 lb/day is the typical amount that steers gain. Both the mean and the median are less than 4.0; indeed, the maximum in the sample is less than 4.0.

2.3.9 $\bar{y}$ = 3.389 lb; median = 3.335 lb.

2.3.10 There is no single correct answer. One possibility is

Resistant bacteria	Frequency (no. aliquots)
12-14	4
15-17	3
18-20	1
21-23	1
24-26	1
Total	10



(b) $\bar{y}$ = 16.7, median = $\dfrac{15+15}{2}$ = 15.

• **2.3.11** The median is the average of the 18th and 19th largest values. There are 18 values less than or equal to 10 and 18 values that are greater than or equal to 11. Thus, the median is

$\dfrac{10+11}{2}$ = 10.5 piglets.

2.3.12 $\bar{y}$ = 375/36 = 10.4.

• **2.3.13** The distribution is fairly symmetric so the mean and median are roughly equal. It appears that half of the distribution is below 50 and half is above 50. Thus, mean ≈ median ≈ 50.

2.3.14 Mean ≈ 35, median ≈ 40

2.4.1 (a) Putting the data in order, we have

13 13 14 14 15 15 16 20 21 26

The median is the average of observations 5 and 6 in the ordered list. Thus, the median is $\dfrac{15+15}{2}$ = 15. The lower half of the distribution is

13 13 14 14 15

The median of this list is the 3rd largest value, which is 14. Thus, the first quartile of the distribution is Q_1 = 14. Likewise, the upper half of the distribution is

15 16 20 21 26

The median of this list is the 3rd largest value, which is 20. Thus, the third quartile of the distribution is Q_3 = 20.

(b) IQR = Q_3 - Q_1 = 20 - 14 = 6

(c) To be an outlier at the upper end of the distribution, an observation would have to be larger than $Q_3 + 1.5(\text{IQR}) = 20 + 1.5(6) = 20 + 9 = 29$, which is the upper fence.

- **2.4.2 (a)** The median is the average of the 9th and 10th largest observations. The ordered list of the data is

4.1 5.2 6.8 7.3 7.4 7.8 7.8 8.4 8.7 9.7 9.9 10.6 10.7 11.9 12.7 14.2 14.5 18.8

Thus, the median is $\frac{8.7+9.7}{2} = 9.2$.
To find Q_1 we consider only the lower half of the data set:

4.1 5.2 6.8 7.3 7.4 7.8 7.8 8.4 8.7 9.7

Q_1 is the median of this half (i.e., the 5th largest value), which is 7.4.
To find Q_3 we consider only the upper half of the data set:

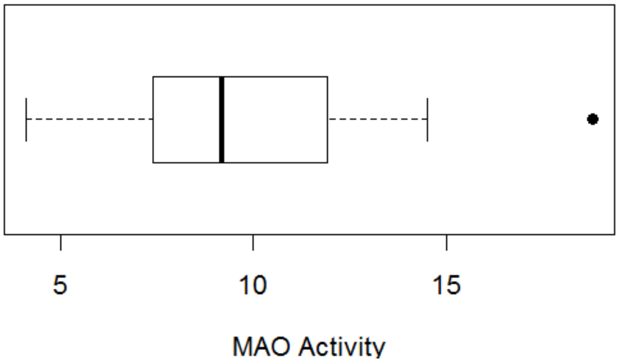
9.7 9.9 10.6 10.7 11.9 12.7 14.2 14.5 18.8.

Q_3 is the median of this half (i.e., the 5th largest value in this list), which is 11.9.

(b) $\text{IQR} = Q_3 - Q_1 = 11.9 - 7.4 = 4.5$.

(c) Upper fence = $Q_3 + 1.5 \times \text{IQR} = 11.9 + 6.75 = 18.65$.

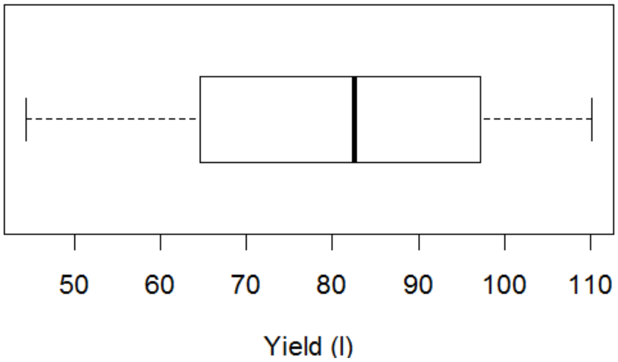
(d)



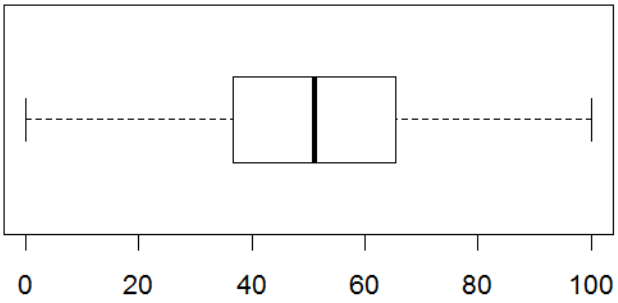
2.4.3 **(a)** Median = 82.6, $Q_1 = 63.7$, $Q_3 = 102.9$.

(b) IQR = 39.2.

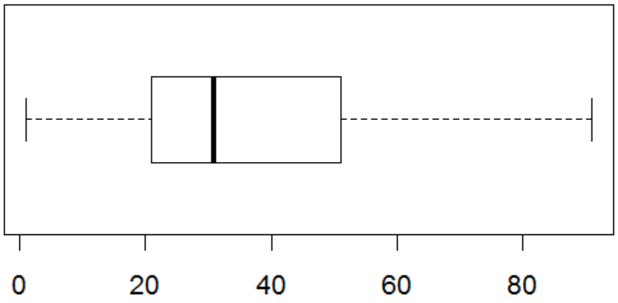
(c)



2.4.4 **(a)** $Q_1 = 35$, median = 50, $Q_3 = 65$.



(b) $Q_1 = 20$, median = 35, $Q_3 = 50$.



2.4.5 The histogram is centered at 40. The minimum of the distribution is near 25 and the maximum is near 65. Thus, boxplot (d) is the right choice.

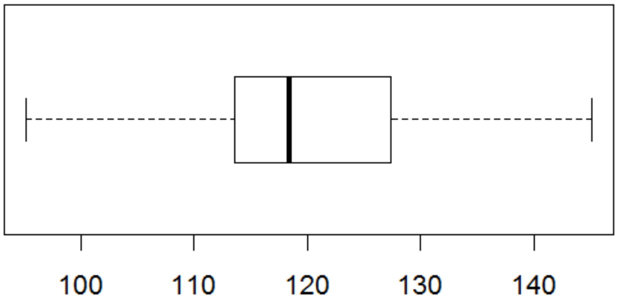
2.4.6 Yes, it is possible that there is no data value of exactly 42. The first quartile does not need to equal a data value. For example, it could be that the first quartile is the average of two points that have the values 41 and 43.

2.4.7 **(a)** The IQR is 127.42 – 113.59 = 13.83.

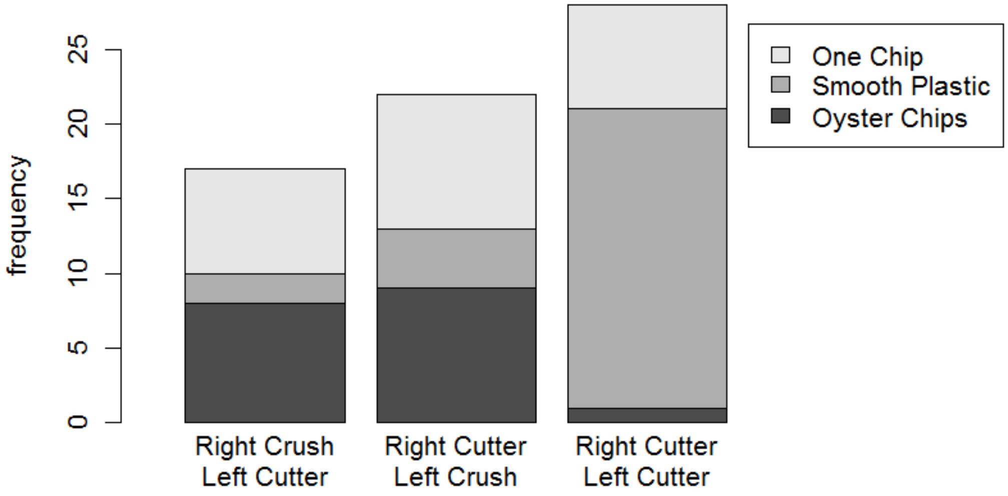
Statistics for the Life Sciences 5th Edition Samuels Solutions Manual
Full Download: http://testbanklive.com/download/statistics-for-the-life-sciences-5th-edition-samuels-solutions-manual/

(b) For a point to be an outlier it would have to be less than 113.59 – 1.5*13.83 = 92.845 or else greater than 127.42 + 1.5*13.83 = 148.165. But the minimum is 95.16 and the maximum is 145.11, so there are no outliers present.

2.4.8



2.5.1 (a)



(b)

Full download all chapters instantly please go to Solutions Manual, Test Bank site: testbanklive.com