Chapter 1

- Read all sections.

Chapter 2

- Read all sections.

Chapter 3

- Section 3.1: Read all.

- Section 3.2: Read all. The key idea here is that “every subset of \( n \) measurements from the population has an equal chance of being included in the sample.”

- Section 3.3: Read all. Equation (3.2) is the \textit{defining} formula the sample variance; equation (3.3) is a \textit{short-cut} formula for use when doing calculations by hand. Memorize (3.2) because it shows that the variance is measuring squared deviations from the mean and you should remember this. Don’t worry about (3.3). When data sets are large enough that the calculations are difficult, you’ll be using a calculator or a computer. Box plots and stem-and-leaf plots are best left to study with the aid of a statistical software package. Read this material so that you’ll get an idea of what box plots and stem-and-leaf plots are. Of the two, box plots are more useful and I’ll cover what I expect you to know in class.

- Section 3.4: Read all. Example 3.7 is an illustration of simple random sampling without replacement (once an element is selected it is removed from the population and cannot be selected again). The sampling distribution of a statistic is a difficult concept to understand, so don’t be surprised if you are confused by this material initially.

- Section 3.5: Read all. The material in Box 3.2 is important.

- Section 3.6: Read all.

- Section 3.7: Read all. The use of the notation \( SE \) in the text is inconsistent with many other texts and statistical software packages. In the text \( SE \) refers to the standard deviation of a statistic. Elsewhere \( SE \) refers to an estimate of the standard deviation of a statistic (in the text an estimate of a \( SE \) is denoted by \( \bar{SE} \)).

- Section 3.8: Skip from the middle of page 97 to the top of page 99.

- Section 3.9: Skip the subsections \textit{Hypothesis testing with confidence intervals}, and \textit{Confidence Interval versus test statistic approach to hypothesis testing}
• Section 3.10: Skip all.
• Section 3.11: Skip all.
• Section 3.12: Skip all.

Chapter 4

• Section 4.1: Read all.
• Section 4.2: Read all.

• Section 4.3: Read all. Confidence intervals and hypothesis tests for means of one or two populations are described, categorized, and boxed in this section. Boxes 4.1 – 4.5 contain concise summaries of results for specific cases categorized according type of data. It might be helpful to reread Sections 3.8 and 3.9 in which confidence intervals and hypothesis tests are described in general terms. Boxes 4.1 – 4.5

  – Box 4.1: One or two independent samples, large sample sizes, no knowledge about normality or equality of variances. Note: in the first displayed equation $S_1/n_1$ should be $S_1/\sqrt{n_1}$.
  – Box 4.2: One sample from a population known to have a normal distribution. Note: at the top of page 133, “Let $\sigma_\mu = S/\sqrt{n}$” should be “Let $\hat{\sigma}_\mu = S/\sqrt{n}$.”
  – Box 4.3: Two independent samples, from normal populations with unknown means and a common, but unknown variance.
  – Box 4.4: Two independent samples, from normal populations with unknown means and unknown variances not known to be equal.
  – Box 4.5: One sample of paired data with normally distributed differences.

The methods in Boxes 4.2 – 4.5 assume that samples are from normal populations. Normality is not assumed in Box 4.1; however, in Box 4.1 it is assumed that sample sizes are large. The Central Limit Theorem ensures that if samples sizes are large (the text uses “sample size > 30” as a guideline), then sample means have approximately normal distributions, and this is enough to justify the methods in Box 4.1.

Deciding on what procedure is appropriate (or best) for a given situation is often not easy. In the make-believe world of homework and exam questions usually there are keywords in the statement of the problem that indicate the most appropriate method. So on a homework or exam problem read carefully and determine whether you’re told
that populations are normal, or whether variances are equal, or whether data are “paired,” or from independent samples, and so on.

Real-world research problems don’t come bundled with leading keywords and phrases (too bad — maybe your research advisor can supply these; if not, maybe the statistics consulting service can help). Of the possible pitfalls, the two potentially most serious are: mistakenly thinking that you have independent samples when you’ve really got paired data; and mistakenly assuming that variances are equal when they are not (and are substantially different). So make sure that you know the distinction between paired data and independent samples; when in doubt use the method in Box 4.4 instead of that in Box 4.3; and do your best to collect as much data as possible (large samples sizes).

• Section 4.4: Read all. Confidence intervals and hypothesis tests for variances of one or two normal populations are described, categorized, and summarized in Boxes 4.6 – 4.7.

  – Box 4.6: Confidence intervals and hypothesis tests about the variance of a single normal population.
  – Box 4.7: Confidence intervals and hypothesis tests about the variances of two populations with independent samples.

Note that the methods in for comparing variances assume the data are from a normal population. Methods for comparing variances from non-normal populations are more involved and typically not described in introductory text books.

• Section 4.5: Read all. Statistical models are convenient for describing how data from different populations differ. We’ll discuss statistical models in more detail in connection with analysis of variance and regression modeling later in the course.

• Section 4.6: Skip.

• Section 4.7: Read all. This material is of second-order importance for a first-course in statistics. However, it serves to emphasize the point the validity of certain statistical methods depends assumptions about the population(s) sampled.

• Section 4.8: Read all.

Chapter 6

• Section 6.1: Read all.
• Section 6.2: Skip from 2/3 down p. 208 to 2/3 down p. 211 (this is the material associated with Box 6.2, The binomial test for a binomial proportion). Skip 3/4 down p. 215 to 1/2 down p. 219 (this is the material associated with Box 6.4, The $\chi^2$ goodness-of-fit test of a partially specified set of proportions).

• Section 6.3: Skip from 1/2 down p. 228 to 4/5 down p. 234 (this is the material on Fisher’s Exact Test and McNemar’s test).

Chapter 7

• Section 7.1: Read all.
• Section 7.2: Read all.
• Section 7.3: Read all.
• Section 7.4: Read all.
• Section 7.5: Read all. This section provides guidance to the problem of determining how much data to collect. Of course the more data the better for statistical inference. However, research budgets usually limit the amount of data that can be collected, and the problem is to determine how much data are necessary to meet the objective(s) of the study. When the objectives can be distilled down to a statement about the power of a hypothesis test or the length of a confidence interval, there are some formulas that can help guide the decision. Don’t be intimidated by the formulas in this section on first reading, as some of the formulas will be explained in class. The important point to remember is that there are strategies, based on statistical theory, for deciding how much data to collect. In practice the decision is seldom quite as simple as using one of the formulas in this section, and experience and a little educated guess work are often involved.

• Section 7.6: Skip all.

Chapter 8

• Section 8.1: Read all.
• Section 8.2: Read all. The text uses a subscript ‘+’ to indicate a sum, that is,

$$ Y_{i+} = \sum_{j=1}^{n_i} Y_{i,j}; $$
and a ‘+’ and an overbar to indicate an average, that is,

\[ \bar{Y}_{i,+} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{i,j}. \]

The ‘+’ notation is not very common. Many texts use a subscript dot (or period) ‘.’ in place of the ‘+’ so that sums and averages are

\[ Y_{i,} = \sum_{j=1}^{n_i} Y_{i,j}; \]

and a ‘+’ and an overbar to indicate an average, that is,

\[ \bar{Y}_{i,} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{i,j}. \]

The symbol \( s^2 \) appears in Display 8.1 (page 282) and in Exercise 8.4d, but is not defined in the section. In this case \( s^2 \) is a pooled estimate of variance

\[ s^2 = \frac{(n_1 - 1)s_1^2 + \cdots + (n_t - 1)s_t^2}{N - t}. \]

In the case that \( n_1 = n_2 = \cdots = n_t = n, N - t = t(n - 1) \) and the formula for \( s^2 \) simplifies to

\[ s^2 = \frac{(s_1^2 + \cdots + s_t^2)}{t}. \]

- Section 8.3: Read all. In Box 8.1 Assumption 1 is the least critical, and Assumption 3 is most critical.
- Section 8.4: Read all.
- Section 8.5: Read all.
- Section 8.6: Skip all.
- Section 8.7: Skip all.
- Section 8.8: Skip all.
- Section 8.9: Skip all.
- Section 8.10: Read all.
Chapter 9

This chapter and elsewhere in the book, contains statements about procedures being conservative. This terminology is used in connection with confidence intervals and hypothesis tests. Most applications of statistics involve some degree of approximation (data are never exactly normally distributed). Thus a nominal 95% confidence interval (the confidence coefficient is exactly 95% under certain assumptions, e.g., normality) seldom has an actual coverage probability of 95% (because the assumptions are not satisfied exactly). The confidence interval is conservative if its actual confidence coefficient is greater than 95%. The same terminology is used for hypothesis tests. If a test has a nominal level of \( \alpha = .05 \), so that the \( \Pr(\text{Type I Error}) = .05 \) when the required assumptions hold exactly, then the test is said to be conservative if actual probability of a Type I error is less than .05. With regard to statistical procedures it is generally better to be conservative than liberal.

- Section 9.1: Read all.
- Section 9.2: Read all.
- Section 9.3: Read all.
- Section 9.4: Read all.
- Section 9.5: Read all.
- Section 9.6: Skip all.
- Section 9.7: Skip all.
- Section 9.8: Skip all.
- Section 9.9: Read all.

Chapter 10

- Section 10.1: Read all. This introduction (and much of the chapter) is a little too theoretical and abstract, so don’t get frustrated if it doesn’t sink in on first one reading. After lectures (and before the final exam) reread this section and more of the material should make sense.
- Section 10.2: Read all.
- Section 10.3: Read all.
- Section 10.4: Read all.
- Section 10.5: Read all.
- Section 10.6: Skip all.
- Section 10.7: Skip all.
- Section 10.8: Read all.
- Section 10.9: Skip all.
- Section 10.10: Skip all.
- Section 10.11: Read all.

Chapter 15

- Section 15.1: Read all. We’ll study only the randomized complete block design (RCBD).
- Section 15.2: Read page 701.
- No other reading in this chapter. You can do the homework problems from the lecture notes and the RCBD Summary handout.