Part I

Teaching Hints

Suggestions for Using the Text

In writing this text, we have followed the premise that a good textbook must be more than just a repository of knowledge. A good textbook should be an agent interacting with the student to create a working knowledge of the subject. To help achieve this interaction, we have modified the traditional format, to encourage active student participation.

Each chapter begins with Preview Questions, which indicate the topics addressed in each section of the chapter. Next comes a Focus Problem that uses real-world data. The Focus Problems show the students the kinds of questions they can answer when they have mastered the material in the chapter. In fact, students are asked to solve each chapter's Focus Problem as soon as the concepts required for the solution have been introduced.

A special feature of this text are the Guided Exercises built into the reading material. These Guided Exercises, with their completely worked solutions, help the students focus on key concepts in the newly introduced material. The Section Problems reinforce student understanding and sometimes require the student to look at the concepts from a slightly different perspective than that presented in the section. Chapter Review problems are much more comprehensive. They require students to place each problem in the context of all they have learned in the chapter. Data Highlights at the end of each chapter ask students to look at data as presented in newspapers, magazines, and other media and then to apply relevant methods of interpretation. Finally, Linking Concept problems ask students to verbalize their skills and synthesize the material.

We believe that the approach from small-step Guided Exercises to Section Problems, to Chapter Review problems, to Data Highlights, to Linking Concepts will enable the instructor to use his or her class time in a very profitable way, going from specific mastery details to more comprehensive decision-making analysis.

Calculators and statistical computer software take much of the computational burden out of statistics. Many basic scientific calculators provide the mean and standard deviation. Those calculators that support two-variable statistics provide the coefficients of the least-squares line, the value of the correlation coefficient, and the predicted value of y for a given x. Graphing calculators sort the data, and many provide the least-squares line. Statistical software packages give full support for descriptive statistics and inferential statistics. Students benefit from using these technologies. In many examples and exercises in *Understandable Statistics,* we ask students to use calculators to verify answers. Illustrations in the text show TI-83 calculator screens, MINITAB outputs, and ComputerStat outputs, so that students can see the different types of information available to them through the use of technology.

However, it is not enough to enter data and punch a few buttons to get statistical results. The formulas producing the statistics contain a great deal of information about the *meaning* of the statistics. The text breaks down formulas into tabular form so that students can see the information in the formula. We find it useful to take class time to discuss formulas. For instance, an essential part of the standard deviation formula is the comparison of each data value to the mean. When we point this out to students, it gives meaning to the standard deviation. When students understand the content of the formulas, the numbers they get from their calculator or computer begin to make sense.

The seventh edition features Focus Points at the beginning of each section, describing that section's primary learning objectives. Also new to the seventh edition is the change from Calculator Notes to Technology Notes. The Technology Notes briefly describe relevant procedures for using the TI-83 Plus calculator, Microsoft Excel, and MINITAB. In addition, the Using Technology segments have been updated to include Excel material.

For courses in which technologies are strongly incorporated into the curriculum, we provide two separate supplements, the *Technology Guide* (for the TI-83 Plus, MINITAB, and ComputerStat) and the *Excel Guide* (for Microsoft Excel). These guides gives specific hints for using the technologies, and also give Lab Activities to help students explore various statistical concepts.

Alternate Paths Through the Text

Like previous editions, the seventh edition of *Understandable Statistics* is designed to be flexible. In most one-semester courses, it is not possible to cover all the topics. The text provides many topics so you can tailor a course to fit your students' needs. The text also aims to be a *readable reference* for topics not specifically included in your course.

Table of Prerequisite Material

Chapter	Prerequisite Sections
1 Getting Started	**none**
2 Organizing Data	**1.1**
3 Averages and Variation	**1.1, 1.2, 2.2**
4 Elementary Probability Theory	**1.1, 1.2, 2.2, 3.1, 3.2**
5 The Binomial Probability Distribution and Related Topics	**1.1, 1.2, 2.2, 3.1, 3.2, 4.1, 4.2, with 4.3 useful but not essential**
6 Normal Distributions	
with 6.4 omitted	**1.1, 1.2, 2.2, 3.1, 3.2, 4.1, 4.2, 5.1**
with 6.4 included	**add 5.2, 5.3**
7 Introduction to Sampling Distributions	**1.1, 1.2, 2.2, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3**
8 Estimation	
with 8.3 and parts of 8.4, 8.5 omitted	**1.1, 1.2, 2.2, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 7.1, 7.2**
with 8.3 and all of 8.4, 8.5 included	**add 5.2, 5.3, 6.4**
9 Hypothesis Testing	
with 9.5 and part of 9.7 omitted	**1.1, 1.2, 2.2, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 7.1, 7.2**
with 9.5 and all of 9.7 included	**add 5.2, 5.3, 6.4**
10 Regression and Correlation	
with part of 10.2 and 10.4, 10.5 omitted	**1.1, 1.2, 3.1, 3.2**
with all of 10.2 and 10.4, 10.5 included	**add 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 7.1, 7.2, 8.1, 9.1, 9.2**
11 Chi-square and *F* Distributions	
with 11.3 omitted	**1.1, 1.2, 2.2, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 7.1, 7.2, 9.1**
with 11.3 included	**add 8.1**
12 Nonparametric Statistics	**1.1, 1.2, 2.2, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 7.1, 7.2, 9.1, 9.5**

Teaching Tips for Each Chapter

CHAPTER 1 GETTING STARTED

Double-Blind Studies (SECTION 1.3)

The double-blind method of data collection, mentioned at the end of Section 1.3, is an important part of standard research practice. A typical use is in testing new medications. Because the researcher does not know which patients are receiving the experimental drug and which are receiving a more familiar drug (or a placebo), the researcher is prevented from subconsciously doing things that might skew the results.

If, for instance, the researcher communicates a more optimistic attitude to patients in the experimental group, this could influence how they respond to diagnostic questions or might actually influence the course of their illness. And if the researcher wants the new drug to prove effective, this could subconsciously influence how he or she handles information related to each patient's case. All such factors are eliminated in double-blind testing.

The following appears in the physician's dosing instructions package insert for the prescription drug QUIXIN™:

> In randomized, double-masked, multicenter controlled clinical trials where patients were dosed for 5 days, QUIXIN™ demonstrated clinical cures in 79% of patients treated for bacterial conjunctivitis on the final study visit day (day 6-10).

Note the phrase "double-masked." This is, apparently, a synonym for "double-blind." Since "double-blind" is widely used in the medical literature and in clinical trials, why do you suppose they chose to use "double-masked" instead?

Perhaps this will provide some insight: QUIXIN™ is a topical antibacterial solution for the treatment of conjunctivitis, i.e., it is an antibacterial eye drop solution used to treat an inflammation of the conjunctiva, the mucous membrane that lines the inner surface of the eyelid and the exposed surface of the eyeball. Perhaps, since QUIXIN™ is a treatment for eye problems, the manufacturer decided the word "blind" shouldn't appear *anywhere* in the discussion.

Source: Package insert. QUIXIN™ is manufactured by Santen Oy, P.O. Box 33, FIN-33721 Tampere, Finland, and marketed by Santen Inc., Napa, CA 94558, under license from Daiichi Pharmaceutical Co., Ltd., Tokyo, Japan.

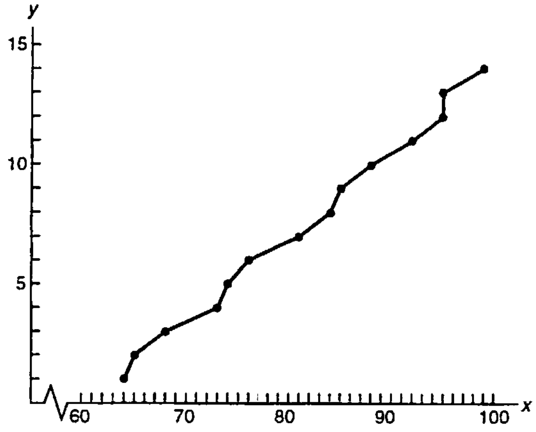
CHAPTER 2 ORGANIZING DATA

Emphasize when to use the various graphs discussed in this chapter: bar graphs when comparing data sets; circle graphs for displaying how a total is dispersed into several categories; time plots to display how data changes over time; histograms or frequency polygons to display relative frequencies of grouped data; stem-and-leaf displays for displaying grouped data in a way that does not lose the detail of the original raw data.

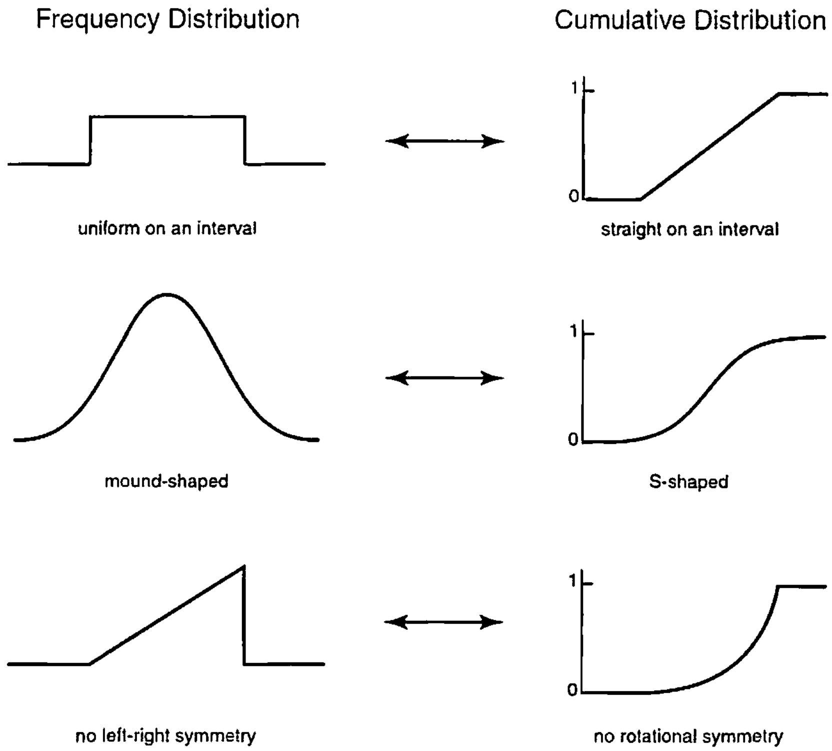
Drawing and Using Ogives (Section 2.2)

The text describes how an ogive, which is a graph displaying a cumulative frequency distribution, can easily be constructed using a frequency table. However, a graph of the same basic sort can be constructed even more quickly than that. Simply arrange the data values in ascending order and then plot one point for each data value, where the x coordinate is the data value and the y coordinate starts at 1 for the first point and increases by 1 for each successive point. Finally, connect adjacent points with line segments. In the resulting graph, for any x, the corresponding y value will be (roughly) the number of data values less than or equal to x.

Here, for example, is the graph for the data set 64, 65, 68, 73, 74, 76, 81, 84, 85, 88, 92, 95, 95, 99:



This graph and others like it are not technically ogives, since for one thing, the possibility of duplicate data values (like 95 in this example) means that the graph will not necessarily be a function. But the graph can be used to get a quick fix on the general shape of the cumulative distribution curve. And by implication, the graph can be used to get a quick idea of the shape of the frequency distribution, as illustrated below.



The pseudo-ogive obtained for the example data set suggests a uniform distribution on the interval 63 to 100 or thereabouts.

CHAPTER 3 AVERAGES AND VARIATION

Students should be instructed in the various ways that sets of numeric data can be represented by a single number. The concepts of this section can be motivated to students by emphasizing the need to represent a set of data by a single number.

The different ways this can be done that are discussed in Section 3.1: mean, median, and mode vary in appropriateness according to the situation. In many cases of numeric data, the mean is the most appropriate measure of central tendency. If the mean is larger or smaller than most of the data values, however, then the median may be the number that best represents a set of data. The median is most appropriate usually if the set of data is annual salaries, costs of houses, or any set of data which contains one or a few very large or very small values. The mode would be the most appropriate if the population was the votes in an election or Nielsen television ratings, for example. Students should get aquainted with these concepts by calculating the mean, median, and mode for sets of data, and interpreting their meanings, or which is the most appropriate.

Range, variance, and standard deviation can be represented to students as other numbers that aid in representing a set of data in that they measure how data is dispersed. Students will begin to have a better understanding of these measures of dispersion, like mean, median, and mode, by calculating these numbers for given sets of data, and interpreting their respective meanings. These concepts of central tendency and dispersion of data can also be applied to grouped data, and students should get aquainted with interpreting these measures for given realistic situations in which data have been collected.

Chebyshev's theorem is an important theorem to discuss with students that relates to the mean and standard deviation of *any* data set.

Finally, the mean, median, first and third quartiles, and range of a set of data can be easily viewed in a box-and-whisker plot.

CHAPTER 4 ELEMENTARY PROBABILITY THEORY

Ways To Think About Probability (Section 4.1)

As the text describes, there are several methods for assigning a probability to an event. Probability based on intuition is often called *subjective* probability. Thus understood, probability is a numerical measure of a person's confidence about some event. Subjective probability is assumed to be reflected in a person's decisions: the higher an event's probability, the more the person would be willing to bet on its occurring.

Probability based on relative frequency is often called *experimental* probability, because relative frequency is calculated from an observed history of experiment outcomes. But we are already using the word experiment in a way that is neutral among the different treatments of probability—namely, as the name for the activity that produces various possible outcomes. So when we are talking about probability based on relative frequency, we will call this *observed* probability.

Probability based on equally likely outcomes is often called *theoretical* probability, because it is ultimately derived from a theoretical model of the experiment's structure. The experiment may be conducted only once, or not at all, and need not be repeatable.

These three ways of treating probability are compatible and complementary. For a reasonable, well-informed person, the subjective probability of an event should match the theoretical probability, and the theoretical probability in turn predicts the observed probability as the experiment is repeated many times.

Also, it should be noted that although in statistics, probability is officially a property of *events,* it can be thought of as a property of *statements,* as well. The probability of a statement equals the probability of the event that makes the statement true.

Probability and statistics are overlapping fields of study; if they weren't, there would be no need for a chapter on probability in a book on statistics. So the general statement, in the text, that probability deals with known populations while statistics deals with unknown populations is, necessarily, a simplification. However, the statement does express an important truth: if we confront an experiment we initially know absolutely nothing about, then we can collect data, but we cannot calculate probabilities. In other words, we can only calculate probabilities after we have formed some idea of, or acquaintance with, the experiment. To find the theoretical probability of an event, we have to know how the experiment is set up. To find the observed probability, we have to have a record of previous outcomes. And as reasonable people, we need some combination of those same two kinds of information to set our subjective probability.

This may seem obvious, but it has important implications for how we understand technical concepts encountered later in the course. There will be times when we would like to make a statement, say, about the mean of a population, and then give the probability that this statement is true—that is, the probability that the event described by the statement occurs (or has occurred). What we discover when we look closely, however, is that often this can't be done. Often we have to settle for some other conclusion instead. The Teaching Tips for Sections 8.1 and 9.1 describe two instances of this problem.

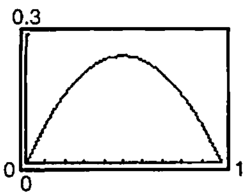
CHAPTER 5 THE BINOMIAL PROBABILITY DISTRIBUTION AND RELATED TOPICS

Binomial Probabilities (Section 5.2)

Students should be able to show that $pq = p(1-p)$ has its maximum value at $p = 0.5$. There are at least three ways to demonstrate this: graphically, algebraically, and using calculus.

Graphical method

Recall that $0 \le p \le 1$. So, for $q = 1 - p$, $0 \le q \le 1$ and $0 \le pq \le 1$. Plot $y = pq = p(1-p)$ using MINITAB, a graphing calculator, or other technology. The graph is a parabola. Observe which value of p maximizes pq. (Many graphing calculators can find the maximum value and where it occurs.)



So pq has a maximum value of 0.25, when $p = 0.5$.

Algebraic method

From the definition of q, it follows that $pq = p(1-p) = p - p^2 = -p^2 + p + 0$. Recognize that this is a quadratic function of the form $ax^2 + bx + c$, where p is used instead of x, and $a = -1$, $b = 1$, and $c = 0$.

The graph of a quadratic function is a parabola, and the general form of a parabola is $y = a(x - h)^2 + k$. The parabola opens up if $a > 0$, opens down if $a < 0$, and has a vertex at (h, k). If the parabola opens up, it has its minimum at $x = h$, and the minimum value of the function is $y = k$. Similarly, if the parabola opens down, it has its maximum value of $y = k$ when $x = h$.

Using the method of completing the square, we can rewrite $y = ax^2 + bx + c$ in the form $y = a(x - h)^2 + k$ to show that $h = -b/2a$ and $k = c - b^2/4a$. When $a = -1$, $b = 1$, and $c = 0$, it follows that $h = 1/2$ and $k = 1/4$. So the value of p that maximizes pq is $p = 1/2$, and then $pq = 1/4$. This confirms the results of the graphical solution.

Calculus-based method
This method is shown on page 31.

This result has implications for confidence intervals for p; see the Teaching Tips for Chapter 8.

CHAPTER 6 NORMAL DISTRIBUTIONS

Emphasize the differences between discrete and continuous random variables with examples of each.

Emphasize how normal curves can be used to approximate the probabilities of both continuous and discrete random variables, and in the cases when the distribution of a set of data can be approximated by a normal curve, such a curve is defined by 2 quantities: the mean and standard deviation of the data. In such a case, the normal curve is defined by the equation $y = \dfrac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$.

Review Chebyshev's Theorem from Chapter 3. Emphasize that this theorem implies that for *any* set of data

at *least* 75% of the data lie within 2 standard deviations on each side of the mean; at *least* 88.9% of the data lie within 3 standard deviations on each side of the mean, and at *least* 93.8% of the data lie within 4 standard deviations on each side of the mean.

In comparison, a set of data that has a distribution which is symmetrical and bell-shaped, in particular has a normal distribution, is more restrictive in that

approximately 68% of the data values lie within 1 standard deviation on each side of the mean;
approximately 95% of the data values lie within 2 standard deviations on each side of the mean; and
approximately 99.7% of the data values lie within 3 standard deviations on each side of the mean.

Remind students regularly that a z-value equals the number of standard deviations from the mean for data values of *any* distribution approximated by a normal curve.

Emphasize the connection between the area under a normal curve and probability values of the random variable. That is, emphasize that the area under any normal curve equals 1, and the percentage of area under the curve between given values of the random variable equals the probability that the random variable will be between these values. The values in a z-table are areas *and* probability values.

Emphasize the conditions whereby a binomial probability distribution (discussed in Chapter 5) can be approximated by a normal distribution: $np > 5$ and $n(1-p) > 5$, where n is the number of trials and p is the probability of success in a single trial.

When a normal distribution is used to approximate a discrete random variable (such as the random variable of a binomial probability experiment), the *continuity correction* is an important concept to emphasize to students. A discussion of this important adjustment can be a good opportunity to compare discrete and continuous random variables.

CHAPTER 7 INTRODUCTION TO SAMPLING DISTRIBUTIONS

Emphasize the differences between population parameters and sample statistics. Point out that when knowledge of the population is unavailable, then knowledge of a corresponding sample statistic must be used to make inferences about the population.

Emphasize the main two facts discussed from the Central Limit Theorem:

1) If x is a random variable with a normal distribution whose mean is μ and standard deviation is σ, then the means of random samples for any fixed size n taken from the x distribution is a random variable $\bar{x}$ that has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

2) If x is a random variable with *any* distribution whose mean is μ and standard deviation is σ, then the mean of random samples of a fixed size n taken from the x distribution is a random variable $\bar{x}$ that has a distribution that approaches a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ as n increases without limit.

Choosing sample sizes greater than 30 is an important point to emphasize in the situation mentioned in part 2 of the Central Limit Theorem above. This commonly-accepted convention insures that the $\bar{x}$ distribution of Part 2 will have a normal distribution regardless of the distribution of the population from which these samples are drawn.

Emphasize that the Central Limit Theorem allows us to infer facts about populations from sample means having normal distributions.

Emphasize that facts about sampling distributions for proportions relating to binomial experiments can be inferred if the same conditions satisfied by a binomial experiment that can be approximated by a normal distribution are satisfied: that is, $np > 5$ and $n(1-p) > 5$, where n is the number of trials and p is the probability of success in a single trial.

Emphasize the difference in the continuity correction that must be taken into account in a sampling distribution for proportions and the continuity correction for a normal distribution used to approximate the probability distribution of the discrete random variable in a binomial probability experiment. That is, instead of subtracting 0.5 from the left endpoint and adding 0.5 to the right endpoint of an interval involved in a normal distribution approximating a binomial probability distribution, $0.5/n$ must be subtracted from the left endpoint and $0.5/n$ added to the right endpoint of such an interval, where n is the number of trials, when a normal distribution is used to approximate a sampling distribution for proportions.

CHAPTER 8 ESTIMATION

Understanding Confidence Intervals (Section 8.1)

As the text says, nontrivial probability statements involve variables, not constants. And if the mean of a population is considered a constant, then the event that this mean falls in a certain range with known numerical bounds has either probability 1 or probability 0.

However, we might instead think of the population mean as itself a variable, since, after all, the value of the mean is initially unknown. In other words, we may think of the population we are sampling from as one of many populations—a population of populations, if you like. One of these populations has been randomly selected for us to work with, and we are trying to figure out which population it is, or, at least, what its mean is.

If we think of our sampling activity in this way, we can then think of the event "The mean lies between *a* and *b*" as having a non-trivial probability of being true. Can we, now, create a 0.90 confidence interval and then say that the mean has a 90% probability of being in that interval? It might seem so, but in general the answer is no. Even though a procedure might have exactly a 90% success rate at creating confidence intervals that contain the mean, a confidence interval created by such a procedure will not, in general, have exactly a 90% chance of containing the mean.

How is this possible? To understand this paradox, let us turn from mean-finding to a simpler task: guessing the color of a randomly-drawn marble. Suppose a sack contains some red marbles and some blue marbles. And suppose we have a friend who will reach in, draw out a marble, and announce its color while we have our backs turned. The friend can be counted on to announce the correct color *exactly 90% of the time* (that is, with a probability of 90%) and the wrong color the other 10% of the time. So if the marble drawn is blue, the friend will say "blue" 9 times out of 10 and "red" the remaining time. And conversely for a red marble. This is like creating an 0.90 confidence interval.

Now the friend reaches in, pulls out a marble, and announces, "blue." Does this mean that we are 90% sure the friend is holding a blue marble? *It depends on what we think about the mix of marbles in the bag.* Suppose we think that the bag contains three red marbles and two blue ones. Then we expect the friend to draw a red marble 3/5 of the time and announce "blue" 10% of those times, or 3/50 of all draws. And we expect the friend to draw a blue marble 2/5 of the time and announce "blue" 90% of those times, or 18/50 of all draws. This means that the ratio of true "blue" announcements to false ones is 18 to 3, or 6 to 1. And thus we attach a probability of 6/7 = 85.7%, not 90%, to our friend's announcement that the marble drawn is blue, even though we believe our friend to be telling the truth 90% of the time. For similar reasons, if the friend says "red," we will attach a probability of 93.1% to this claim. Simply put, our initial belief that there are more red marbles than blue ones pulls our confidence in a "blue" announcement downward, and our confidence in a "red" announcement upward, from the 90% level.

Now, if we believe that there are an *equal* number of red and blue marbles in the bag, then, as it turns out, we will attach 90% probability to "blue" announcements and to "red" announcements as well. But *this is a special case.* In general, the probabilities we attach to each of our friend's statements will be different from the frequency with which we think he is telling the truth. Furthermore: if we have *no idea* about the mix of marbles in the bag, then we will be *unable* to set probabilities for our friend's statements, because we will be unable to run the calculation for how often his "blue" statements are true and his "red" statements are true. In other words, *we cannot justify simply setting our probability equal, by default, to the test's confidence level.*

This story has two morals. (1) The probability of a statement is one thing, and the success rate of a procedure that tries to come up with true statements is another. (2) Our prior beliefs about the conditions of an experiment are an unavoidable element in our interpretation of any sample data.

Let us apply these lessons to the business of finding confidence intervals for population means. When we create a 0.90 confidence interval, we will in general *not* be 90% sure that the interval contains the mean. It could *happen* to turn out that we were 90% sure, but this will depend on what ideas we had about the population mean going in. Suppose we were fairly sure, to start with, that the population mean lay somewhere between 10 and 20, and suppose we then took a sample that led to the construction of a 0.90 confidence interval which ran from 31 to 46. We would *not* conclude, with 90% certainty, that the mean lay between 31 and 46. Instead, we would have a probability lower than that, because previously we thought the mean was outside that range. At the same time, we would be much more ready to believe that the mean lay between 31 and 46 than we were before, because, after all, a procedure with a 90% success rate produced that prediction. Our exact probability for the "between 31 and 46" statement would depend on our entire initial probability distribution for values of the population mean—something we would have a hard time coming up with, if the question were put to us. Thus, under normal circumstances, our exact level of certainty about the confidence interval could not be calculated.

So the general point made in the text holds, even if we think of a population mean as a variable. The procedure for finding a confidence interval of confidence level *c* does not, in general, produce a statement (about the value of a population mean) that has a probability *c* of being true.

Confidence Intervals for *p* (Section 8.3)

The result obtained in the Teaching Tip for Chapter 5 has implications for the confidence interval for *p*: the most conservative interval estimate of *p*, the widest possible confidence interval in a given situation, is obtained when $E = z_c\sqrt{pq/n}$ is calculated using $p = 1/2$.

CHAPTER 9 HYPOTHESIS TESTING

What a Hypothesis Test Tells Us (Sections 9.1–9.3)

The procedure for hypothesis testing with significance levels may at first confuse some students, especially since the levels are chosen somewhat arbitrarily. Why, the students may wonder, don't we just calculate the likelihood that the null hypothesis is true? Or is that really what we're doing, when we find the *P* value?

Once again we run the risk of confusion over the role of probability in our statistical conclusions. The *P* value is *not* the same thing as the probability, in light of the data, of the null hypothesis. Instead, the *P* value is the probability that the data would turn out the way it did, assuming the null hypothesis to be true. Just as with confidence intervals, here we have to be careful not to think we are finding the probability of a given statement when in fact we are doing something else.

To illustrate: consider two coins in a sack, one fair and one two-headed. One of these coins is pulled out at random and flipped. It comes up heads. Let us take, as our null hypothesis, the statement "The flipped coin was the fair one." The probability of the outcome, given the null hypothesis, is 1/2, because a fair coin will come up heads half the time. This probability is in fact the *P* value of the outcome. On the other hand, the probability that the null hypothesis is true, given the evidence, is 1/3, since out of all the heads outcomes one will see in many such trials, 1/3 are from the fair coin.

Now suppose that instead of containing two coins of known character, the sack contains an unknown mix—some fair coins, some two-headed coins, and possibly some two-tailed coins, as well. Then we can still calculate the *P* value of a heads outcome, because the probability of "heads" with a fair coin is still 1/2. But the probability of the coin's being fair, given that we're seeing heads, *cannot be calculated,* because we know nothing about the mix of coins in the bag. So the *P* value of the outcome is one thing, and the probability of the null hypothesis is another.

The lesson should be now be familiar: without some prior ideas about the character of an experiment, either based on a theoretical model or on previous outcomes, we cannot attach a definite probability to a statement about the experimental setup or its outcome.

This is the usual situation in hypothesis testing. We normally lack the information needed to calculate probabilities for the null hypothesis and its alternative. What we do instead is to take the null hypothesis as defining a well-understood scenario from which we *can* calculate the likelihoods of various outcomes—the probabilities, that is, of various kinds of sample results, given that the null hypothesis is true. By contrast, the alternative hypothesis includes all sorts of scenarios, in some of which (for instance) two population means are only slightly different, in others of which the two means are far apart, and so on. Unless we have identified the likelihoods of all these possibilities, relative to each other and to the null hypothesis, we will not have the background information needed to calculate the probability of the null hypothesis from sample data.

In fact, we will not have the data necessary to calculate what the text calls the power, $1-\beta$, of a hypothesis test. This is what the text means when it says that finding the power requires knowing the H_1 distribution. Because we cannot specify the H_1 distribution when we are concerned with things like diagnosing disease (instead of drawing coins from a sack and the like), we normally cannot determine the probability of the null hypothesis in light of the evidence. Instead, we have to content ourselves with quantifying the risk, α, of rejecting the hypothesis when it is true.

A Paradox About Hypothesis Tests

The way hypothesis tests work (see the AP Hints for discussion) leads to a result that at first seems surprising. It can sometimes happen that, at a given level of significance, a one-tailed test leads to a rejection of the null hypothesis while a two-tailed test would not. Apparently, one can be justified in concluding that $\mu > k$ (or $\mu < k$ as the case may be) but not justified in concluding that $\mu \neq k$ —even though the latter conclusion follows from the former! What is going on here?

This paradox dissolves when one remembers that a one-tailed test is used only when one has appropriate information. With the null hypothesis H_0: $\mu = k$, we choose the alternative hypotheses H_1: $\mu > k$ only if *we are already sure* that μ is not less than H_1: $\mu < k$. This assumption, in effect, boosts the force of any evidence that μ does not equal k—and if it is not less than or equal to k, it must be greater.

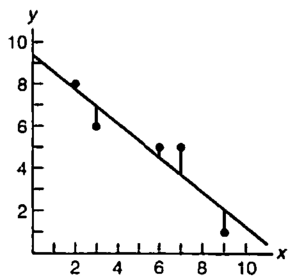
In other words, when a right-tailed test is appropriate, rejecting the null hypothesis means concluding *both* that $\mu > k$ and that $\mu \neq k$. But when there is no justification for a one-tailed test, one must use a two-tailed test and needs somewhat stronger evidence before concluding that $\mu \neq k$.

CHAPTER 10 REGRESSION AND CORRELATION

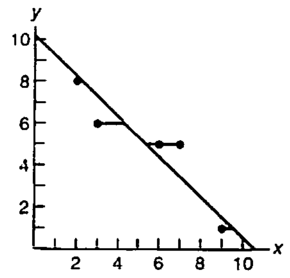
Least-Squares Criteria (Section 10.2)

With some sets of paired data, it will not be obvious which is the explanatory variable and which is the response variable. Here it may be worth mentioning that for linear regression, the choice matters. The results of a linear regression analysis will differ, depending on which variable is chosen as the explanatory variable and which one as the response variable. This is not immediately obvious. We might think that with x as the explanatory variable, we could just solve the regression equation $y = a + bx$ for x in terms of y to obtain the regression equation that we would get if we took y as the explanatory variable instead. But this would be a mistake.

The figure below shows the vertical distances from data points to the line of best fit. The line is defined so as to make the sum of the squares of these vertical distances as small as possible.



The next figure, now, shows the *horizontal* distances from the data points to the same line. These are the distances whose sum of squares would be minimized if the explanatory and response variables switched roles. With such a switch, the graph would be flipped over, and the horizontal distances would become vertical ones. But the line that minimizes the sum of squares for vertical distances is not, in general, the same line that minimizes the sum of squares for horizontal distances.



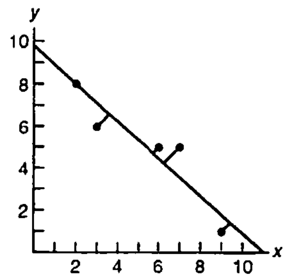
So there is more than one way, mathematically, to definite the line of best fit for a set of paired data. This raises a question: what is the *proper* way to define the line of best fit?

Let us turn this question around: under what circumstances is a best fit based on *vertical* distances the right way to go? Well, intuitively, the distance from a data point to the line of best fit represents some sort of deviation from the ideal value. We can most easily conceptualize this in terms of measurement error. If, now, we treat the error as a strictly vertical distance, then we are saying that in each data pair, the second value is possibly off but the first value is exactly correct. In other words, the least-squares method with vertical distances assumes that the first value in each data pair is measured with essentially perfect accuracy, while the second is measured only imperfectly.

An illustration shows how these assumptions can be realistic. Suppose we are measuring the explosive force generated by the ignition of varying amounts of gunpowder. The weight of the gunpowder is the explanatory variable, and the force of the resulting explosion is the response variable. It could easily happen that we were able to measure the weight of gunpowder with great exactitude—down to the thousandth-ounce—but that our means of measuring explosion force was quite crude, such as the height to which a wooden block was flung into the air by the explosion. We would then have an experiment with a good deal of error in the response variable measurement but for, practical purposes, no error in the explanatory variable measurement. This would all be perfectly in accord with the vertical-distance criterion for finding the line of best fit by the least-squares method.

But now consider a different version of the gunpowder experiment. This time we have a highly refined means of measuring explosive force (some sort of electronic device, let us say) and at the same time we have only a very crude means of measuring gunpowder mass (perhaps a rusty pan balance). In this version of the story, the error would be in the measurement of the response variable, and a horizontal least-squares criterion would be called for.

Now, the most common situation is one in which both the explanatory and the response variables contain some error. The preceding discussion suggests that the most appropriate least-squares criterion for goodness of fit for a line through the cluster of data points would be a criterion in which error was represented as a line lying at some slant, as in the figure.



To apply such a criterion, we would have to figure out to define distance in two dimensions when the x and y axis have different units of measure. We will not try to solve that puzzle here. Instead we just summarize what we have learned: there is more than one least-squares criterion for fitting a line to a set of data points. and the choice of which criterion to use implies an assumption about which variable(s) is affected by the error (or other deviation) that moves points off the line representing ideal results.

And, finally, we now see that the standard use of vertical distances in the least-squares method *implies an assumption that the error is predominantly in the response variable.* This is often a reasonable assumption to make, since the explanatory variable is frequently a *control* variable, that is, a variable under the experimenter's control and thus generally capable of being adjusted with a fair amount of precision. The response variable, by contrast, is the one that must simply be measured and cannot be fine-tuned through an experimenter's adjustment. However, it is worth noting that this is only the typical relationship, not a necessary one (as the second gunpowder scenario shows).

Finally, it is also worth nothing that both the vertical and the horizontal least-squares criteria will produce a line that passes through the point $(\bar{x}, \bar{y})$. Thus the vertical- and horizontal-least-squares lines must either coincide (which is atypical but not impossible) or intersect at $(\bar{x}, \bar{y})$. The other thing the two lines have in common is the correlation coefficient, r. It is easy to see, looking at the formula for r, that the value of r does not depend on which variable is chosen as the explanatory one and which as the response one.

Variables and the Issue of Cause and Effect (Sections 10.2 and 10.3)

As remarked at the end of Section 10.3, the relationship between two measured variables x and y may not, in physical terms, be one of cause and effect, respectively. It often is, of course, but it may instead happen that y is the cause and x is the effect. Note that in the example where x = cricket chirps per second and y = air temperature, y is obviously the cause and x the effect. In other situations, x and y will be two effects of a common, possibly unknown, cause. For example, x might be a patient's blood sugar level and y might be the patient's body temperature. Both of these variables could be caused by an illness, which might be quantified in terms of a count of bacterial activity. The point to remember is that although the x-causes-y scenario is typical, strictly speaking the designations "explanatory variable" and "response variable" should be understood not in terms of a causal relationship but in terms of which quantity is initially known and which one is inferred.

CHAPTER 11 CHI-SQUARE AND *F* DISTRIBUTIONS

Emphasize that both the χ^2 distribution and the F distribution are not symmetrical and have only non-negative values.

Emphasize that the applications of the χ^2 distribution include the test for independence of two factors, goodness of fit of a present distribution to a given distribution, and whether a variance (or standard deviation) has changed or varies from a known population variance (or standard deviation). The χ^2 distribution is also used to find a confidence interval for a variance (or standard deviation).

Emphasize that the applications of the F distribution includes the test of whether the variances (or equivalently, standard deviations) of two independent, normal distributions are equal. A second application of the F distribution is the one-way ANOVA test which determines whether a significant difference exists between any of several sample means of groups taken from populations that are each assumed to be normally distributed, independent of one another, and the groups come from distributions with approximately the same standard deviation. A third application of the F distribution is a two-way ANOVA test: a test of whether differences exist in the population means of varying levels of two factors where each level of each factor is assumed to be from a normal distribution and all levels of both factors are assumed to have equal variances.

CHAPTER 12 NONPARAMETRIC STATISTICS

Review the classifications of data discussed in Chapter 1: ratio, interval, ordinal, and nominal.

Emphasize that the methods of nonparametric statistics are quite general and are applied when no assumptions are known about the population distributions from which samples are drawn, such as that the distributions are normal or binomial, for example.

Emphasize that the sign test is used when comparing sample distributions from two populations that are not independent, such as when a sample is measured twice, as in a "before-and-after" study. Emphasize that the sign-test requires that the number of positive and negative signs between the samples number at least 12. Point out that since the proportion of plus signs to total number of plus and minus signs of the sampling distribution for x follows a normal distribution, the critical values for the sign test are based on z values from a normal distribution.

Emphasize that the rank-sum test for testing the difference between sample means can be used when it is not known whether the populations the samples come from are normally distributed or when assumptions about equal population variances are not satisfied. An important point to emphasize is that the rank-sum test requires that the sample size of each sample be at least 8. Emphasize that since the sampling distribution for the sum of ranks R follows a normal distribution, the critical values and sample statistics of the test are z values from a normal distribution.

Emphasize that the Spearman rank correlation is used to compare ranked data from two sources.

Emphasize that for the Spearman rank correlation coefficient r_S, $-1 \le r_S \le 1$, and discuss the meanings of $r_S = 1$, $r_S = -1$, $r_S = 0$, r_S close to 1, and r_s close to -1.

Compare the similarity of r_s to the correlation coefficent r from Chapter 10.

Hints for Distance Education Courses

Distance education uses various media, each of which can be used in one-way or interactive mode. Here is a representative list:

		One-way	Interactive
Medium:	Audio	Cassette tapes	Phone
	Audiovisual	Videotapes, CD-ROMs	Teleconferencing
	Data	Computer-resident tutorials, web tutorials	E-mail, chat rooms, discussion lists
	Print	Texts, workbooks	Mailed-in assignments, mailed-back instructor comments, fax exchanges

Sometimes the modes are given as asynchronous (students working on their schedules) versus synchronous (students and instructors working at the same time), but synchronous scheduling normally makes sense only when this enables some element of interactivity in the instruction.

Naturally the media and modes may be mixed and matched. A course might, for instance, use a one-way video feed with interactive audio, plus discussion lists.

THINGS TO KEEP IN MIND

Even in a very high-tech telecourse, print is a foundational part of the instruction. The textbook is *at least* as important as in a traditional course, since it is the one resource which requires no special equipment to use, and whose use is not made more difficult by the distance separating student and instructor.

Because students generally obtain all course materials at once, before instruction begins, mid-course adjustments of course content are generally not practicable. Plan the course carefully up front, so everything is in place when instruction begins.

In distance courses, students can often be assumed to have ready access to computers while working on their own. This creates the opportunity for technology-based assignments that in a traditional course might be feasible at best as optional work (for example, assignments using ComputerStat, MINITAB, or Microsoft Excel; see the corresponding guides that accompany *Understandable Statistics*). However, any time students have to spend learning how to use unfamiliar software will add to their overall workload and possibly to their frustration level. Remember this when choosing technology-based work to incorporate.

Remember that even (and perhaps especially) in distance education, students take a course because they want to interact with a human being rather than just read a book. The goal of distance instruction is to make that possible for students who cannot enroll in a traditional course. Lectures should not turn into slide shows with voice commentary, even though these may be technologically easier to transmit than, say, real-time video. Keep the human element uppermost.

All students should be self-motivated, but in real life nearly all students benefit from a little friendly supervision and encouragement. This goes double for distance education. Make an extra effort to check in with students one-on-one, ask how things are going, and remind them of things they may be forgetting or neglecting.

CHALLENGES IN DISTANCE EDUCATION

Technology malfunctions often plague distance courses. To prevent this from happening in yours:

- Don't take on too much at once. As the student sites multiply, so do the technical difficulties. Try the methodology with one or two remote sites before expanding.

- Plan all technology use well in advance and thoroughly test all equipment before the course starts.

- Have redundant and backup means for conducting class sessions. If, for instance, a two-way teleconferencing link goes down, plan for continuing the lecture by speakerphone, with students referring to predistributed printed materials as needed.

- Allow enough slack time in lectures for extra logistical tasks and occasional technical difficulties.

- If possible, do a pre-course dry run with at least some of the students, so they can get familiar with the equipment and procedures and alert you to any difficulties they run into.

- When it is feasible, have a facilitator at each student site. This person's main job is to make sure the technology at the students' end works smoothly. If the facilitator can assist with course administration and answer student questions about course material, so much the better.

In a distance course, establishing rapport with students and making them comfortable can be difficult.

- An informal lecture style, often effective in traditional classrooms, can be even more effective in a distance course. Be cheerful and use humor. (But in cross-cultural contexts, remember that what is funny to you may fall flat with your audience.)

- Remember that your voice will not reach the students with the clarity as in a traditional classroom. Speak clearly, not too fast, and avoid over-long sentences. Pause regularly.

- Early in the course, work in some concrete, real-world examples and applications to help the students relax, roll up their sleeves, and forget about the distance-learning aspect of the course.

- If the course is interactive, via teleconferencing or real-time typed communication, get students into "send" mode as soon as possible. Ask them questions. Call on individuals if you judge that this is appropriate.

- A student-site assistant with a friendly manner can also help the students quickly settle into the course.

The distance learning format will make it hard for you to gauge how well students are responding to your instruction. In a traditional course, students' incomprehension or frustration often registers in facial expression, tone of voice, muttered comments—all of which are, depending on the instructional format, either difficult or impossible to pick up on in a distance course. Have some way for students to give feedback on how well the course is going for them. Possibilities:

- Quickly written surveys ("On a scale of 1 to 5, please rate ...") every few weeks.

- Periodic "How are things going?" phone calls from you.

- A student-site assistant can act as your "eyes and ears" for this aspect of the instruction, and students may be more comfortable voicing frustrations to him or her than to you.

If students are to mail in finished work, set deadlines with allowance for due to mail delivery times, and with allowance for the possibility of lost mail. This is especially important for the end of the term, when you have a deadline for turning in grades.

Cheating is a problem in any course, but especially so in distance courses. Once again, an on-site facilitator is an asset. Another means of forestalling cheating is to have open-book exams, which takes away the advantage of sneaking a peek at the text.

Good student-instructor interaction takes conscious effort and planning in a distance course. Provide students with a variety of ways to contact you:

- E-mail is the handiest way for most students to stay in touch.
- Phone; a toll-free number is ideal. When students are most likely to be free in the evenings, set the number up for your home address and schedule evening office hours when students can count on reaching you.
- When students can make occasional in-person visits to your office, provide for that as well.

ADVANTAGES IN DISTANCE EDUCATION

Compared to traditional courses, more of the information shared in a distance course is, or can be, preserved for later review. Students can review videotaped lectures, instructor-student exchanges via e-mail can be reread, class discussions are on a reviewable discussion list, and so on.

To the extent that students are on their own, working out of texts or watching prerecorded video, course material can be modularized and customized to suit the needs of individual students: where a traditional course would necessarily be offered as a 4-unit lecture series, the counterpart distance course could be broken into four 1-unit modules, with students free to take only those modules they need. This is especially beneficial when the course is aimed at students who already have some professional experience with statistics and need to fill-in-gaps rather than comprehensive instruction.

STUDENT INTERACTION

Surprisingly, some instructors have found that students interact more with one another in a well-designed distance course than in a traditional course, even when the students are physically separated from one another. Part of the reason may be a higher level of motivation among distance learners. But another reason is the same technologies which facilitate student-instructor communication—things like e-mail and discussion lists—also facilitate student-student communication. In some cases, distance learners have actually done better than traditional learners taking the very same course. Better student interaction was thought to be the main reason.

One implication is that while group projects, involving statistical evaluations of real-world data, might seem more difficult to set up in a distance course, they are actually no harder, and the students learn just as much. The web has many real-world data sources, like the U.S. Department of Commerce (home.doc.gov) which has links to the U.S. Census Bureau (www.census.gov), the Bureau of Economic Analysis (www.bea.gov), and other agencies that compile publicly-available data.

Suggested References

THE AMERICAN STATISTICAL ASSOCIATION

Contact Information

1429 Duke Street
Alexandria, VA 22314-3415
Phone: (703) 684-1221 or toll-free: (888) 231-3473
Fax: (703) 684-2037

ASA Publications

Stats: The Magazine for Students of Statistics
CHANCE magazine
The American Statistician
AmStat News

BOOKS

Huff, Darryll and Geis, Irving (1954). *How to Lie with Statistics.* Classic text on the use and misuse of statistics.

Moore, David S. (2000) *Statistics: Concepts and Controversies,* fifth edition. Does not go deeply in to computational aspects of statistical methods. Good resource for emphasizing concepts and applications.

Tufte, Edward R. (2001). *The Visual Display of Quantitative Information,* second edition. A beautiful book, the first of three by Tufte on the use of graphic images to summarize and interpret numerical data. The books are virtual works of art in their own right.

Tanur, Judith M. (1989) *Statistics: A Guide to the Unknown,* third edition. Another excellent source of illustrations.

REFERENCES FOR DISTANCE LEARNING

Bolland, Thomas W. (1994). "Successful Customers of Statistics at a Distant Learning Site." *Proceedings of the Quality and Productivity Section, American Statistical Association,* pp. 300–304.

Lawrence, Betty and Gaines, Leonard M. (1997). "An Evaluation of the Effectiveness of an Activity-Based Approach to Statistics for Distance Learners." *Proceedings of the Section on Statistical Education, American Statistical Association,* pp. 271–272.

Wegman, Edward J. and Solka, Jeffrey L. (1999). "Implications for Distance Learning: Methodologies for Statistical Education." *Proceedings of the Section on Statistical Education, the Section on Teaching Statistics in the Health Sciences, and the Section on Statistical Consulting, American Statistical Association,* pp. 13–16.

Distance Learning: Principles for Effective Design, Delivery, and Evaluation

University of Idaho website: www.uidaho.edu/evo/distglan.html