

Teaching Tips for Business Statistics: A First Course, Sixth Edition

Educational Philosophy

In our many years of teaching business statistics, we have continually searched for ways to improve the teaching of business statistics courses. We actively participate in Decision Sciences Institute (DSI), American Statistical Association (ASA), and Making Statistics More Effective in Schools and Business (MSMESB) conferences. We use the Guidelines for Assessment and Learning (GAISE) reports as well as our reflections on teaching business statistics to a diverse group of students at several large universities. These experiences have shaped our vision that has come to include these key principles:

1. Students need to be shown the relevance of statistics.

- Students need a frame of reference when learning statistics, especially when statistics is not their major. That frame of reference for business students should be the functional areas of business—accounting, finance, information systems, management, and marketing. Each statistical topic needs to be presented in an applied context related to at least one of these functional areas.
- The focus in teaching each topic should be on its application in business, the interpretation of results, the presentation of assumptions, the evaluation of the assumptions, and the discussion of what should be done if the assumptions are violated.

2. Students need to be familiar with the software used in the business world.

- Integrating spreadsheet and/or statistical software into all aspects of an introductory statistics course allows the course to focus on interpretation of results instead of computations.
- Introductory business statistics courses should recognize that in business, spreadsheet software is typically available on a decision-maker's desktop and statistical software is often also available.

3. Students need to be given sufficient guidance on using software.

- Textbooks should provide enough instructions so that students can effectively use the software integrated with the study of statistics, without having the software instruction dominate the course.

4. Students need ample practice in order to understand how statistics is used in business.

- Both classroom examples and homework exercises should involve actual or realistic data as much as possible.
- Students should work with data sets, both small and large, and be encouraged to look beyond the statistical analysis of data to the interpretation of results in a managerial context.

Chapter 1

One way to begin the course is by discussing why a businessperson needs to know about statistics. This will provide a reason for why the student has been required to take this course. The first day is extremely important to driving out preexisting fear that some students have concerning the course. You should talk about the fact that this course is different from a mathematics course in that you will not be deriving any proofs, but instead will be focusing on how statistics can be used in the different functional areas of business. You can talk a little about the DCOVA problem solving approach that will be used starting in Chapter 2 as a way of solving problems that will occur on the job when students graduate.

This can be followed by a brief explanation of the basic vocabulary of statistics including the distinction between descriptive and inferential statistics. Then you need to discuss the types of variables. Be sure to discuss the different types carefully since the ability to distinguish between categorical and numerical data will be crucial later in the course. Go over examples of each type of variable and have students provide examples of each type.

At this time, if you are going to use Microsoft[®] Excel, ask the class whether they are familiar with Microsoft[®] Excel. Depending on the degree to which the instructor wishes to integrate Excel into the course, point out that, since most managers will have integrated products such as Microsoft[®] Office on their desktops, it would be useful to understand how Excel can be used to analyze data. If you are going to use Minitab, you can talk about statistical packages and the advantages of using them instead of or in addition to using Excel in certain circumstances.

The chapter also introduces a continuing case related to the *Managing Ashland MultiComm Services* that appears at the end of many chapters. The Digital cases are introduced in this chapter also. In these cases, students visit Web sites related to companies and issues raised in the Using Statistics scenarios that start each chapter. The goal of the Digital cases is for students to develop skills needed to identify misuses of statistical information. As would be the situation with many real world cases, in Digital cases students often need to sift through claims and assorted information in order to discover the data most relevant to a case task. They will then have to examine whether the conclusions and claims are supported by the data. (Instructional tips for using the *Managing Ashland MultiComm Services* and Digital cases and solutions to the *Managing*

Ashland MultiComm Services and Digital cases are included in this *Instructor's Solutions Manual*.).

If you are using Excel, make sure that students read the Excel Guide at the end of the chapter. The first section explains the different types of Excel instructions. The *In-Depth Excel* instructions provide step-by-step instructions and live worksheets that automatically update when data changes. These instructions can also be used with Mac Excel and OpenOffice.org Calc 3. The *PHStat2 add-in* instructions provide instructions for using the PHStat2 add-in. *Analysis ToolPak* instructions provide instructions for using the Analysis ToolPak, the Excel add-in packages with many versions of Excel.

The remainder of the Excel Guide in this chapter goes over Excel basics including entering data, opening and saving workbooks, creating, printing, and copying worksheets, cell references, and entering formulas. Those students with limited knowledge of Excel should read Appendix B and F and all students should read Appendices C, D, and G.

If you are using Minitab, make sure that students read the Minitab Guide at the end of the chapter. Topics covered include entering data, opening, creating, and saving worksheets and projects, and printing projects and/or worksheets. Those students with limited knowledge of computers should read Appendix B.

Chapter 2

This book uses a systematic approach for meeting a business objective or solving a business problem. This approach goes across all the topics in the book and most importantly can be used as a framework in real world situations when students graduate. The approach has the acronym **DCOVA**, which stands for **Define**, **Collect**, **Organize**, **Visualize**, and **Analyze**.

- **Define** the business objective or problem to be solved and then define the variables to be studied.
- **Collect** the data from appropriate sources
- **Organize** the data
- **Visualize** the data by developing charts
- **Analyze** the data by using statistical methods to reach conclusions.

You can begin by emphasizing the importance of defining your objective or problem. Go back to the importance of operational definitions of variables discussed in Chapter 1 and the importance of knowing the type for each variable.

Continue by discussing the need for collecting data and then introduce sources of data including the World Wide Web. Next, if you are going to collect sample data to use in Chapters 2 and 3, you can illustrate the need for sampling by conducting a survey of students in your class (otherwise leave the discussion of sampling until chapter 7). Ask each student to collect his or her own personal data concerning the time it takes to get ready to go to class in the morning or the time it takes to get to school or home from school. First, ask the students to write down a definition of how they plan to measure this time. Then, collect the various answers and read them to the class. Then, a single definition could be provided (such as the time to get ready is the time measured from when you get out of bed to when you leave your dorm or home, recorded to the nearest minute). In the next class, select a random sample of students and use the data collected (depending on the sample size) in class when Chapters 2 and 3 are discussed.

Then, move on to the Organize step that involves setting up your data in an Excel (or Minitab) worksheet and developing tables to help you prepare charts and analyze your data. Begin your discussion for categorical data with the example on p. 28 on how people pay their bills. Show the summary table and then if you wish, explain that you can sometimes organize the data into a two-way table that has one variable in the row and another in the column.

Continue with organizing data (but now for numerical data) by referring to the cost of a restaurant meal on p. 32. Show the simple ordered array and how a frequency distribution,

percentage distribution, or cumulative distribution can summarize the raw data in a way that is more useful.

Now you are ready to tackle the Visualize step. A good way of starting this part of the chapter is to display the following quote.

"A picture is worth a thousand words."

Students will almost certainly be familiar with Microsoft® Word and may have already used Excel to construct charts that they have pasted into Word documents. Now you will be using either Excel or Minitab to construct many different types of charts. Return to the bill payment data previously discussed and illustrate how a bar chart and pie chart can be constructed. Mention their advantages and disadvantages. A good example is to show the data on incomplete ATM transactions on p. 43 and how the Pareto chart enables you to focus on the vital few categories. If time permits, you can discuss the side-by-side bar chart for a contingency table.

To examine charts for numerical variables you can either use the restaurant data previously mentioned or data that you have collected from your class. You may want to begin with a simple stem-and-leaf display that both organizes the data and shows a bar type chart. Then move on to the histogram and the various polygons, pointing out the advantages and disadvantages of each.

If time permits, you can discuss the scatter plot and the time-series plot for two numerical variables. Otherwise, you can wait until you get to regression analysis later in the semester. Also, if time permits, you can discuss how multidimensional tables allow you to drill down to individual cells of the table.

If the opportunity is available, we believe that it is worth the time to cover Section 2.8 on Misuses and Common Errors in Visualizing Data. This is a topic that students very much enjoy since it allows for a great deal of classroom interaction. After discussing the fundamental principles of good graphs, try to illustrate some of the improper displays shown in Figures 2.23 – 2.25. Ask students what is “bad” about these figures. Follow up with a homework assignment involving Problems 2.60 – 2.65 (*USA Today* is a great source).

You will find that the chapter review problems provide large data sets with numerous variables. Report writing exercises, the team project, and the student survey database provide the opportunity for students to integrate written and or oral presentation with the statistics they have learned.

The *Managing Ashland MultiComm Services* case enables students to examine the use of statistics in an actual business environment. The Digital case refers to the EndRun Financial Services and claims that have been made.

The Excel and Minitab Guides for this and the remaining chapters are organized according to the sections of the chapter. These guides are quite extensive since they cover both organizing and visualizing many different graphs. The Excel Guide includes instructions for In-Depth Excel, PHStat2, and the Analysis ToolPak. Pick and choose the approaches that you prefer.

Chapter 3

This chapter on descriptive numerical statistical measures represents the initial presentation of statistical symbols in the text. Students who need to review arithmetic and algebraic concepts may wish to refer to Appendix A for a quick review or to appropriate texts (see www.pearson.com) or videos (www.videoaidedinstruction.com). Once again, as with the tables and charts constructed for numerical data, it is useful to provide an interesting set of data for classroom discussion. If a sample of students was selected earlier in the semester and data concerning student time to get ready or commuting time was collected (see Chapters 1 and 2), use these data in developing the numerous descriptive summary measures in this chapter. (If they have not been developed, obtain other data for classroom illustration.)

Discussion of the chapter begins with the property of central tendency. We have found that almost all students are familiar with the arithmetic mean (which they know as the average) and most students are familiar with the median. A good way to begin is to compute the mean for your classroom example. Emphasize the effect of extreme values on the arithmetic mean and point out that the mean is like the center of a seesaw -- a balance point. Note that you will return to this concept later when you discuss the variance and the standard deviation. You might want to introduce summation notation at this point and express the arithmetic mean in formula notation as in Equation (3.1). (Alternatively, you could wait until you cover the variance and standard deviation.) A classroom example in which summation notation is reviewed is usually worthwhile. Remind the students again that Appendix A includes a review of arithmetic and algebra and summation notation [or refer them to other text sources such as those found at www.pearson.com or videos (see www.videoaidedinstruction.com)].

The next statistic to compute is the median. Be sure to remind the students that the median as a measure of position must have all the values ranked in order from lowest to highest. Be sure to have the students compare the arithmetic mean to the median and explain that this tells us something about another property of data (skewness). Following the median, the mode can be briefly discussed. Once again, have the students compare this result to those of the arithmetic mean and median for your data set.

The completion of the discussion of central tendency leads to the second characteristic of data, variability. Mention that all measures of variation have several things in common: (1) they can never be negative, (2) they will be equal to 0 when all items are the same, (3) they will be small when there isn't much variation, and (4) they will be large when there is a great deal of variation.

The first measure of variability to consider is the simplest one, the range. Be sure to point out that the range only provides information about the extremes, not about the distribution between the extremes. Point out that the range lacks one important ingredient, the ability to take into account each data value. Bring up the idea of computing the differences around the mean, but then return to the fact that as the balance point of the seesaw, these differences add up to zero. At that point, ask the students what they can do mathematically to remove the negative sign for some of the values. Most likely, they will answer by telling you to square them (although someone may realize that the absolute value could be taken). Next, you may want to define the squared differences as a sum of squares. Now you need to have the students realize that the number of values being considered affects the magnitude of the sum of squared differences. Therefore, it makes sense to divide by the number of values and compute a measure called the variance. If a population is involved, you divide by N , the population size, but if you are using a sample, you divide by $n - 1$, to make the sample result a better estimate of the population variance. You can finish the development of variation by noting that since the variance is in squared units, you need to take the square root to compute the standard deviation.

Another measure of variation that can be discussed is the coefficient of variation. Be sure to illustrate the usefulness of this as a measure of relative variation by using an example in which two data sets have vastly different standard deviations, but also vastly different means. A good example is one that involves the volatility of stock prices. Point out that the variation of the price should be considered in the context of the magnitude of the arithmetic mean. At this point you may want to have the students use the Visual Explorations in Exploring Descriptive Statistics procedure (see p. 107). By changing values in the data provided, students can observe how the mean, median, and standard deviation are affected.

The final measure of variation is the Z score. Point out that this provides a measure of variation in standard deviation units. You can also say that you will return to Z scores in Chapter 6 when the normal distribution will be discussed.

You are now ready to move on to the third characteristic of data, shape. Be sure to clearly define and illustrate both symmetric and skewed distributions by comparing the mean and median.

Once these three characteristics have been discussed, you are ready to show how they can be computed using Excel or Minitab.

Now that these measures of central tendency and variation are understood, you can further explore data by computing the quartiles, the interquartile range, the five-number summary, and constructing a boxplot. You begin by determining the quartiles. Reference here

can be made to the standardized exams that most students have taken, and the quantile scores that they have received (97th percentile, 48th percentile, and 12th percentile). Explain that the 1st and 3rd quartiles are merely two special quantiles -- the 25th and 75th, that unlike the median (the 2nd quartile), are not at the center of the distribution. Once the quartiles have been computed, the interquartile range can be determined. It is useful to note that the interquartile range computes the variation in the center of the distribution as compared to the difference in the extremes computed by the range.

You can then mention the five-number summary of minimum value, first quartile, median, third quartile, and maximum value. Then, you construct the boxplot. Present this plot from the perspective of serving as a tool for determining the location, variability, and symmetry of a distribution by visual inspection, and as a graphical tool for comparing the distribution of several groups. It is useful to display Figure 3.6 on page 114 that indicates the shape of the boxplot for four different distributions. Then, use PHStat2 or Minitab to construct a boxplot. Note that you can construct the boxplot for a single group or for multiple groups.

If you desire, you can discuss descriptive measures for a population and introduce the empirical rule and the Chebyshev rule.

If time permits, and you have covered scatter plots in Chapter 2, you can briefly discuss the covariance and the coefficient of correlation as a measure of the strength of the association between two numerical variables. Point out that the coefficient of correlation has the advantage as compared to the covariance of being on a scale that goes from -1 to +1. Figure 3.9 on p. 124 is useful in depicting scatter plots for different coefficients of correlation.

Once again, you will find that the chapter review problems provide large data sets with numerous variables. Report writing exercises, the team project, and the survey database provide the opportunity for students to integrate written and or oral presentation with the statistics they have learned.

The *Managing Ashland MultiComm Services* case enables students to examine the use of descriptive statistics in an actual business environment. The Digital case continues the evaluation of the EndRun Financial Services discussed in the Digital case in Chapter 2.

The Excel Guide for the chapter includes instructions on using different Excel functions to compute various statistics. Alternatively, you can use the Analysis ToolPak to compute a list of many statistics. PHStat2 can be used to construct a boxplot. The Minitab Guide includes instructions for various descriptive statistics and boxplots.

Chapter 4

The chapter on probability represents a bridge between the descriptive statistics already covered and the topics of statistical inference, regression, and quality improvement to be covered in subsequent chapters. In many traditional statistics courses, often a great deal of time is spent on probability topics that are of little direct applicability in basic statistics. The approach in this text is to cover only those topics that are of direct applicability in the remainder of the text.

You need to begin with a relatively concise discussion of some probability rules. Essentially, students really just need to know that (1) no probability can be negative, (2) no probability can be more than 1, and (3) the sum of the probabilities of a set of mutually exclusive events adds to 1.0. Students often understand the subject best if it is taught intuitively with a minimum of formulas, with an example that relates to a business application shown as a two-way contingency table (see the Using Statistics example). If desired, you can use In-Depth Excel or PHStat2 to compute probabilities from the contingency table.

Once these basic elements of probability have been discussed, if there is time and you desire, conditional probability and Bayes' theorem can be covered. The *Think About This* concerning email SPAM is a wonderful way of helping student realize the application of probability to everyday life. Be aware that in a one-semester course where time is particularly limited, these topics may be of marginal importance. In addition, you may wish to spend a bit of time going over counting rules, especially if the binomial distribution will be covered in Chapter 5.

The Digital case in this chapter extends the evaluation of the EndRun Financial Services to consider claims made about various probabilities.

Chapter 5

Now that the basic principles of probability have been discussed, the probability distribution is developed and the expected value and variance (and standard deviation) are computed and interpreted. Given that a probability distribution has been defined, you can now discuss some specific distributions. Although every introductory course undoubtedly covers the normal distribution to be discussed in Chapter 6, the decision about whether to cover the binomial or Poisson distributions is matter of personal choice and depends on whether the course is part of a two-course sequence.

If the binomial distribution is covered, an interesting way of developing the binomial formula is to follow the Using Statistics example that involves an accounting information system. Note, in this example, the value for p is 0.10. (It is best not to use an example with $p = 0.50$ since this represents a special case). The discussion proceeds by asking how you could get three tagged order forms in a sample of 4. Usually a response will be elicited that provides three items of interest out of four selections in a particular order such as Tagged Tagged Not Tagged Tagged. Ask the class, what would be the probability of getting Tagged on the first selection? When someone responds 0.1, ask them how they found that answer and what would be the probability of getting Tagged on the second selection. When they answer 0.1 again, you will be able to make the point that in saying 0.1 again, they are assuming that the probability of Tagged stays constant from trial to trial. When you get to the third selection and the students respond 0.9, point out that this is a second assumption of the binomial distribution -- that only two outcomes are possible -- in this case Tagged and Not Tagged, and the sum of the probabilities of Tagged and Not Tagged must add to 1.0. Now you can compute the probability of three out of four in this order by multiplying $(0.1)(0.1)(0.9)(0.1)$ to get 0.0036. Ask the class if this is the answer to the original question. Point out that this is just one way of getting three Tagged out of four selections in a specific order, and, that there are four ways to get three Tagged out of four selections This leads to the development of the binomial formula Equation (5.5). You might want to do another example at this point that calls for adding several probabilities such as three or more Tagged, less than three Tagged, etc. Complete the discussion of the binomial distribution with the computation of the mean and standard deviation of the distribution. Be sure to point out that for samples greater than five, computations can become unwieldy and the student should use PHStat2, the BINOMDIST function in Excel, Minitab, or the binomial tables (See the Online **Binomial.pdf** tables).

Once the binomial distribution has been covered, if time permits, you can cover the Poisson distribution. Point out the distinction between the binomial and Poisson distributions. Note that the Poisson is based on an area of opportunity in which you are counting occurrences within an area such as time or space. Contrast this with the binomial distribution in which each value is classified as of interest or not of interest. Point out the equations for the mean and standard deviation of the Poisson distribution and indicate that the mean is equal to the variance. Since the computation of probabilities from these discrete probability distributions can become tedious for other than small sample sizes, it is important to discuss PHStat2, the POISSON Excel function, or Minitab.

The *Managing Ashland MultiComm Services* case for this chapter relates to the binomial distribution.

Chapter 6

Now that probability and probability distributions have been discussed in Chapters 4 and 5, you are ready to introduce the normal distribution. We recommend that you begin by mentioning some reasons that the normal distribution is so important and discuss several of its properties. We would also recommend that you do not show Equation (6.1) in class as it will just intimidate some students. You might begin by focusing on the fact that any normal distribution is defined by its mean and standard deviation and display Figure 6.3 on p. 204. Then, an example can be introduced and you can explain that if you subtracted the mean from a particular value, and divided by the standard deviation, the difference between the value and the mean would be expressed as a standardized normal or Z score that was discussed in Chapter 3. Next, use Table E.2, the cumulative normal distribution, to find probabilities under the normal curve. In the text, the cumulative normal distribution is used since this table is consistent with results provided by Excel and Minitab. (The standardized normal distribution is shown in Table E.9) Make sure that all the students can find the appropriate area under the normal curve in their cumulative normal distribution tables. If anyone cannot, show them how to find the correct value. Be sure to remind the class that since the total area under the curve adds to 1.0, the word area is synonymous with the word probability. Once this has been accomplished, a good approach is to work through a series of examples with the class, having a different student explain how to find each answer. The example that will undoubtedly cause the most problems will be finding the values corresponding to known probabilities. Slowly go over the fact that in this type of example, the probability is known and the Z value needs to be determined, which is the opposite of what the student has done in previous examples. Also point out that in cases in which the unknown X value is below the mean, the negative sign must be assigned to the Z value. Once the normal distribution has been covered, you can use PHStat2, various Excel functions, or Minitab to compute normal probabilities. It is also useful to use the Visual Explorations Exploring the Normal Distribution procedure on p. 213. This will be useful if you intend to use examples that explore the effect on the probabilities obtained by changing the X value, the population mean, μ , or the standard deviation, σ . The *Think About This* essay provides a historical perspective of the application of the normal distribution.

If you have sufficient time in the course, the normal probability plot can be discussed. Be sure to note that all the data values need to be ranked in order from lowest to highest and that each value needs to be converted to a normal score. Again, you can either use PHStat2 to generate a normal probability plot or use Excel functions and the Chart Wizard or use Minitab.

The *Managing Ashland MultiComm Services* case for this chapter relates to the normal distribution. The Digital case involves the normal distribution and the normal probability plot.

Chapter 7

You should spend some time discussing sampling, even if it is just using the table of random numbers to select a random sample. You may want to take a bit more time and discuss the types of survey sampling methods and issues involved with survey sampling results. The *Think About This* essay discusses the important issue of the use of Web-based surveys.

The coverage of the normal distribution in Chapter 6 flows into a discussion of sampling distributions. Point out the fact that the concept of the sampling distribution of a statistic is important for statistical inference. Make sure that students realize that problems in this section will find probabilities concerning the mean, not concerning individual values. It is helpful to display Figure 7.5 on p. 242 to show how the Central Limit Theorem applies to differently shaped populations. A useful classroom or homework exercise involves using PHStat2 or Excel to form sampling distributions. This reinforces the concept of the Central Limit Theorem.

The *Managing Ashland MultiComm Services* case for this chapter relates to the sampling distribution of the mean. The Digital case involves the sampling distribution of the mean.

Chapter 8

You should begin this chapter by reviewing the concept of the sampling distribution covered in Chapter 7. It is important that the students realize that (1) an interval estimate provides a range of values for the estimate of the population parameter, (2) you can never be sure that the interval developed does include the population parameter, and (3) the proportion of intervals that include the population parameter within the interval is equal to the confidence level.

Note that the Using Statistics example for this chapter, which refers to the Saxon Home Improvement Company is actually a case study that relates to every part of the chapter. This scenario is a good candidate for use as the classroom example demonstrating an application of statistics in accounting. It also enables you to use the DCOVA approach of Define, Collect, Organize, Visualize, and Analyze in the context of statistical inference.

When introducing the t distribution for the confidence interval estimate of the population mean, be sure to point out the differences between the t and normal distributions, the assumption of normality, and the robustness of the procedure. It is useful to display Table E.3 in class to illustrate how to find the critical t value. When developing the confidence interval for the proportion, remind the students that the normal distribution may be used here as an approximation to the binomial distribution as long as the assumption of normality is valid [when $n\pi$ and $n(1 - \pi)$ are at least 5].

Having covered confidence intervals, you can move on to sample size determination by turning the initial question of estimation around, and focusing on the sample size needed for a desired confidence level and width of the interval. In discussing sample size determination for the mean, be sure to focus on the need for an estimate of the standard deviation. When discussing sample size determination for the proportion, be sure to focus on the need for an estimate of the population proportion and the fact that a value of $\pi = 0.5$ can be used in the absence of any other estimate. Since the formulas for the confidence interval estimates and sample sizes discussed in this chapter are straightforward, using PHStat2 or In-Depth Excel can remove much of the tedious nature of these computations.

The *Managing Ashland MultiComm Services* case for this chapter involves developing various confidence intervals and interpreting the results in a marketing context. The Digital case also relates to confidence interval estimation.

Chapter 9

A good way to begin the chapter is to focus on the reasons that hypothesis testing is used. We believe that it is important for students to understand the logic of hypothesis testing before they delve into the details of computing test statistics and making decisions. If you begin with the Using Statistics example concerning the filling of cereal boxes, slowly develop the rationale for the null and alternative hypotheses. Ask the students what conclusion they would reach if a sample revealed a mean of 200 grams (They will all say that something is the matter) and if a sample revealed a mean of 367.99 grams (Almost all will say that the difference between the sample result and what the mean is supposed to be is so small that it must be due to chance). Be sure to make the point that hypothesis testing allows you to take away the decision from a person's subjective judgment, and enables you to make a decision while at the same time quantifying the risks of different types of incorrect decisions. Be sure to go over the meaning of the Type I and Type II errors, and their associated probabilities α and β along with the concept of statistical power.

Set up an example of a sampling distribution such as Figure 9.1 on p. 298, and show the regions of rejection and nonrejection. Explain that the sampling distribution and the test statistic involved will change depending on the characteristic being tested. Focus on the situation where σ is unknown if you have numerical data. Emphasize that σ is virtually never known. It is also useful at this point to introduce the concept of the p -value approach as an alternative to the classical hypothesis testing approach. Define the p -value and use the phrase given in the text “If the p -value is low, H_0 must go.” and the rules for rejecting the null hypothesis and indicate that the p -value approach is a natural approach when using Excel or Minitab, since the p -value can be determined by using PHStat, Excel functions, the Data Analysis tool or Minitab.

Once the initial example of hypothesis testing has been developed, you need to focus on the differences between the tests used in various situations. The Chapter 9 summary table is useful for this since it presents a road map for determining which test is used in which circumstance. Be sure to point out that one-tail tests are used when the alternative hypothesis involved is directional (e.g., $\mu > 368$, $\pi < 0.20$). Examine the effect on the results of changing the hypothesized mean or proportion.

Both the *Managing Ashland MultiComm Services* case and the Digital case involve the use of the one-sample test of hypothesis for the mean.

Chapter 10

This chapter discusses tests of hypothesis for the differences between two or more means and between two proportions. The chapter begins with t tests for the difference between the means, then covers the Z test for the difference between two proportions, the F test for the ratio of two variances, and concludes with the one way Analysis of Variance to test the difference among more than two means.

The first test of hypothesis covered is usually the test for the difference between the means of two groups for independent samples. Point out that the test statistic involves pooling of the sample variances from the two groups and assumes that the population variances are the same for the two groups. Students should be familiar with the t distribution, assuming that the confidence interval estimate for the mean has been previously covered. Point out that a stem-and-leaf display, a boxplot, or a normal probability plot can be used to evaluate the validity of the assumptions of the t test for a given set of data. This allows you to once again use the DCOVA approach of Define, Collect, Organize, Visualize, and Analyze to meet a business objective.

Once the t test has been discussed, you can use the Excel worksheets provided with the In-Depth Excel approach, PHStat2, the Data Analysis tool or Minitab to determine the test statistic and p -value. Mention that if the variances are not equal, a separate variance t test can be conducted. The *Think About This* essay is a wonderful example of how the two-sample t test was used to solve a business problem that a student had after she graduated and had taken the introductory statistics course.

At this point, having covered the test for the difference between the means of two independent groups, you can discuss a test that examines differences in the means of two paired or matched groups. The key difference is that the focus in this test is on differences between the paired values in the two groups since the data have been collected from matched pairs or repeated measurements on the same individuals or items. Once the paired t test has been discussed, the In-Depth Excel approach, PHStat2, the Data Analysis tool, or Minitab can be used to determine the test statistic and p -value.

You can continue the coverage of differences between two groups by testing for the difference between two proportions. Be sure to review the difference between numerical and categorical data emphasizing the categorical variable used here classifies each observation as of interest or not of interest. Make sure that the students realize that the test for the difference between two proportions follows the normal distribution. A good classroom example involves asking the students if they enjoy shopping for clothing and then classifying the yes and no

responses by gender. Since there often will be a difference between males and females, you can then ask the class how we might go about determining whether the results are statistically significant.

The F -test for the variances can be covered next. Be sure to carefully explain that this distribution, unlike the normal and t distributions, is not symmetric and cannot have a negative value since the statistic is the ratio of two variances. Remind the students that the larger variance is in the numerator. Be sure to mention that a boxplot of the two groups and normal probability plots can be used to determine the validity of the assumptions of the F test. This is particularly important here since this test is sensitive to non-normality in the two populations. The In-Depth Excel approach, PHStat2, the Data Analysis tool or Minitab can be used to determine the test statistic and p -value.

If the one-way ANOVA F test for the difference between c means is to be covered in your course, a good way to start is to go back to the sum of squares concept that was originally covered when the variance and standard deviation were introduced in Section 3.2. Explain that in the one-way Analysis of Variance, the sum of squared differences around the overall mean can be divided into two other sums of squares that add up to the total sum of squares. One of these measures differences among the means of the groups and thus is called sum of squares among groups (SSA), while the other measures the differences within the groups and is called the sum of squares within the groups (SSW). Be sure to remind the students that, since the variance is a sum of squares divided by degrees of freedom, a variance among the groups and a variance within the groups can be computed by dividing each sum of squares by the corresponding degrees of freedom. Make the point that the terminology used in the Analysis of Variance for variance is Mean Square, so the variances computed are called MSA , MSW , and MST . This will lead to the development of the F statistic as the ratio of two variances. A useful approach at this point when all formulas are defined, is to set up the ANOVA summary table. Try to minimize the focus on the computations by reminding students that the Analysis of Variance computations can be done using In-Depth Excel, PHStat2, the Data Analysis tool or Minitab. It is also useful to show how to obtain the critical F value by either referring to Table E.5 or the Excel or Minitab results. Be sure to mention the assumptions of the Analysis of Variance and that the boxplot and normal probability plot can be used to evaluate the validity of these assumptions for a given set of data. Levene's test can be used to test for the equality of variances. In-Depth Excel, PHStat2, or Minitab can be used to compute the results for this test.

Once the Analysis of Variance has been covered, if time permits (which it may not in a one-semester course), you will want to determine which means are different. Although many

approaches are available, this text uses the Tukey-Kramer procedure that involves the Studentized range statistic shown in Table E.6. Be sure that students compare each paired difference between the means to the critical range. Note that you can use In-Depth Excel, PHStat2, or Minitab to compute Tukey-Kramer multiple comparisons.

Be aware that the *Managing Ashland MultiComm Services* case involves all the sections of the chapter except the test for the difference between two proportions since it contains both independent sample and matched sample aspects. The Digital case is based on two or more independent samples. Thus, only the sections on the t test for independent samples, the F test for the difference between two variances, and the one way ANOVA are involved.

Chapter 11

This chapter covers chi-square tests. The Using Statistics example concerning hotels relates to all three sections of the chapter.

If you covered the Z test for the difference between two proportions in Chapter 10, you can return to the example you used there and point out that the chi-square test can be used as an alternative. A good classroom example involves asking the students if they enjoy shopping for clothing (or revisiting Chapter 10's example) and then classifying the yes and no responses by gender. Since there often will be a difference between males and females, you can then ask the class how they might go about determining whether the results are statistically significant. The expected frequencies are computed by finding the mean proportion of items of interest (enjoying shopping) and items not of interest (not enjoying shopping) and multiplying by the sample sizes of males and females respectively. This leads to the computation of the test statistic. Once again as with the case of the normal, t , and F distribution, be sure to set up a picture of the chi-square distribution with its regions of rejection and non-rejection and critical values. In addition, go over the assumptions of the chi square test including the requirement for an expected frequency of at least five in each cell of the 2×2 contingency table.

Now you are ready to extend the chi-square test to more than two groups. Be sure to discuss the fact that with more than two groups, the number of degrees of freedom will change and the requirements for minimum cell expected frequencies will be somewhat less restrictive.

The discussion of the chi-square test concludes with the test of independence in the r by c table. Be sure to go over the interpretation of the null and alternative hypotheses and how they differ from the situation in which there are only two rows. You can use In-Depth Excel, PHStat2, or Minitab.

The *Managing Ashland MultiComm Services* case extends the survey discussed in Chapter 8 to analyze data from contingency tables. The Digital case also involves analyzing various contingency tables.

Chapter 12

Regression analysis is probably the most widely used and misused statistical method in business and economics. You may want to start the chapter with the *Think About This* essay on p. 471 to show students the importance of this topic in business. In an era of easily available statistical and spreadsheet applications, we believe that the best approach is one that focuses on the interpretation of regression results obtained from such applications, the assumptions of regression, how those assumptions can be evaluated, and what can be done if they are violated. Although we also feel that is useful for students to work out at least one example with the aid of a hand calculator, we believe that the focus on hand calculations should be minimized.

A good way to begin the discussion of regression analysis is to focus on developing a model that can provide a better prediction of a variable of interest. The Using Statistics example, which forecasts sales for a clothing store, is useful for this purpose. You can extend the DCOVA approach discussed earlier by defining the business objective, discussing data collection, and data organization before moving on to the visualization and analysis in this chapter. Be sure to clearly define the dependent variable and the independent variable at this point.

Once the two types of variables have been defined, the example should be introduced. Explain the goal of the analysis and how regression can be useful. Follow this with a scatter plot of the two variables. Before developing the Least Squares method, review the straight-line formula and note that different notation is used in statistics for the intercept and the slope than in mathematics. At this point, you need to develop the concept of how the straight line that best fits the data can be found. One approach involves plotting several lines on a scatter plot and asking the students how they can determine which line fits the data better than any other. This usually leads to a criterion that minimizes the differences between the actual Y value and the value that would be predicted by the regression line. Remind the class that when you computed the mean in Chapter 3, you found out that the sum of the differences around the mean was equal to zero. Tell the class that the regression line in two dimensions is similar to the mean in one dimension, and that the differences between the actual Y value and the value that would be predicted by the regression line will sum to zero. Students at this point, having covered the variance, will usually tell you just to square the differences. At this juncture, you might want to substitute the regression equation for the predicted value, and tell the students that since you are minimizing a quantity, derivatives are used. We discourage you from doing the actual proof, but mentioning derivatives may help some students realize that the calculus they may have learned in mathematics courses is actually used to develop the theory behind the statistical method. The

least-squares concepts discussed can be reinforced by using the Visual Explorations in Exploring Simple Linear Regression Coefficients procedure on p. 440. This procedure produces a scatter plot of the site selection data of Table 12.1 with an unfitted line of regression and a floating control panel of controls with which to adjust the line. The spinner buttons can be used to change the values of the slope and Y intercept to change the line of regression. As these values are changed, the difference from the minimum SSE changes.

The solution obtained from the Least Squares method allows you to find the slope and Y intercept. In this text, since the emphasis is on the interpretation of computer output, focus is now on finding the regression coefficients on the output shown in Figure 12.4 on p. 436. Once this has been done, carefully review the meaning of these regression coefficients in the problem involved. The coefficients can now be used to predict the Y value for a given X value. Be sure to discuss the problems that occur if you try to extrapolate beyond the range of the X variable. Now you can show how to use either the In-Depth Excel, the Analysis ToolPak, PHStat2, or Minitab to obtain the regression output.

Tell the students that now you need to determine the usefulness of the regression model by subdividing the total variation in Y into two component parts, explained variation or regression sum of squares (SSR) and unexplained variation or error sum of squares (SSE). Once the sum of squares has been determined and the coefficient of determination r^2 computed, be sure to focus on the interpretation. Having computed the error sum of squares (SSE), the standard error of the estimate can be computed. Make the analogy that the standard error of the estimate has the same relationship to the regression line that the standard deviation had to the arithmetic mean.

The completion of this initial model development phase allows you to begin focusing on the validity of the model fitted. First, go over the assumptions and emphasize the fact that unless the assumptions are evaluated, a correct regression analysis has not been carried out. Reiterate the point that this is one of the things that people are most likely to do incorrectly when they carry out a regression analysis.

Once the assumptions have been discussed, you are ready to begin evaluating whether they are true for the model that has been fit. This leads into a discussion of residual analysis. Emphasize that Excel or Minitab can be used to determine the residuals and that in determining whether there is a pattern in the residuals, you look for gross patterns that are obvious on the plot, *not* minor patterns that are not obvious. Be sure to note that the residual plot can also be used to evaluate the assumption of equal variance along with whether there is a pattern in the residuals over time if the data have been collected in sequential order. Point out that finding no pattern (i.e., a random pattern) means that the model fit is an appropriate one. However, it does not mean that

other alternative models involving additional variables should not be considered. Mention also, that a normal probability plot of the residuals can be helpful in determining the validity of the normality assumption. If time permits, the discussion of the Anscombe data in Section 12.9 serves as a strong reinforcement of the importance of residual analysis.

If time is available, you may wish to discuss the Durbin-Watson statistic for autocorrelation. Be sure to discuss how to find the critical values from the table of the D statistic and the fact that sometimes the results will be inconclusive.

Once the model fit has been found to be appropriate, inferences in regression can be made. First cover the t or F test for the slope by referring to the Excel or Minitab results. Here, the p -value approach is usually beneficial. Then, if time permits, you can discuss the confidence interval estimate for the mean and the prediction interval for the individual value.

Both the *Managing Ashland MultiComm Services* case and the Digital case involve a simple linear regression analysis of a set of data.

To perform simple linear regression, you can use In-Depth Excel, the Analysis ToolPak, PHStat2, or Minitab.

Chapter 13

If time is available in the course, you can now move on to multiple regression. You should point out that Microsoft Excel or Minitab needs to be used to perform the computations in multiple regression. Once you have the results, you need to focus on the interpretation of the regression coefficients and how the interpretation differs between simple linear regression and multiple regression. Mention the aspects of multiple regression that are similar in interpretation to those observed in simple regression -- prediction, residual analysis, coefficient of determination, and standard error of the estimate. If sufficient time is available, it is probably most efficiently used in covering the dummy variable model. With dummy variables, be sure to mention that the categories must be coded as 0 and 1. In addition, indicate the importance of determining whether there is an interaction between the dummy variable and the other independent variables. Further discussion can include interaction terms in regression models.

Both the *Managing Ashland MultiComm Services* case and the Digital case involve developing a multiple regression model that includes dummy variables.

Chapter 14 (Online)

In order to fully understand the role of statistics in quality management, the themes of quality management and Six Sigma need to be mentioned. Although students may wonder why this is either being discussed in a statistics class (or why they are reading non-statistical material), they usually enjoy learning about this subject since it provides a rationale for how the statistics course relates to management.

You may want to begin the discussion of control charts by demonstrating the Red Bead experiment. Tell the students that two broad categories of control charts will be considered, attribute charts in Sections 14.2 and variables charts in Section 14.4.

Once this introduction has been completed, an overview of the theory of control charts can be undertaken. Begin by referring to the normal distribution and mention Shewhart's concern about committing errors in determining special causes. Tell the students that setting the limits at three standard deviation units away from the mean is done to insure that there is only a small chance that a stable process will have special cause signals that appear and cannot be explained. Continue the discussion by noting that the integer value 3 made computations simpler in an era prior to the availability of calculators and computers, and that experience has shown that this serves the purpose of keeping false alarms to a minimum.

Once these topics have been discussed, you are ready to begin covering specific control charts. The choice of where to start is an individual one. The simplest approach is to begin with the p chart and refer to the Red Bead experiment and then use other examples such as those shown in Section 14.2. Be sure that students are aware that Excel, PHStat2, or Minitab can be used to construct the p chart

The discussion of variables charts should begin with a review of the distinction between attribute and variables charts. Briefly discuss the decisions that need to be made when sample sizes are to be determined and subgroups are to be formed. Be sure to emphasize the fact that variables charts are usually done in pairs, one for the variability and the other for the mean. Emphasize the notion that if the variability chart is out of control, you will be unable to meaningfully interpret the chart for the mean. Again, note that Excel, PHStat2, or Minitab can be used to construct both R and \bar{X} charts.

The themes of quality management and the inclusion of a discussion of the work of Deming and Shewhart allow you to distinguish between common causes of variation and special causes of variation. Perhaps the best way to reinforce this is by conducting the Red Bead experiment (see Section 14.3). This experiment allows the student to see the distinction between

the two types of variation. The amount of time spent on Sections 14.5 and 14.6 is a matter of instructor discretion. Some may wish to just list the fourteen points and have students read the section, while others will want to cover the points in detail. Regardless of which approach is taken, in order to emphasize the importance of statistics, the Shewhart-Deming PDSA cycle needs to be mentioned since the study stage typically involves the use of statistical methods. In addition, points 6 (institute training on the job) and 13 (encourage education and self-improvement for everyone) underscore the importance of everyone within an organization being familiar with the basic statistical methods required to manage a process. Students find the experiment of counting F s (see Figure 14.8) particularly intriguing since they can't believe that they have messed up such a seemingly easy set of directions.

The importance of statistics can be reinforced by briefly covering the Six Sigma[®], an approach that is being used by many large corporations. Go over the DMAIC model and compare it to Deming's 14 points.

The *Harnswell Company Sewing Machine Company* case contains several phases and uses R and \bar{X} charts. The *Managing Ashland MultiComm Services* case also has several phases and uses the p chart and R and \bar{X} charts.