INSTRUCTOR'S SOLUTIONS MANUAL

GAIL ILLICH
McLennan Community College

PAUL ILLICH
Southeast Community College

STATISTICS FOR MANAGERS
USING MICROSOFT® EXCEL®
NINTH EDITION

David Levine
Baruch College, City University of New York

David Stephan
Two Bridges Instructional Technology

Kathryn Szabat
La Salle University

Pearson

Reproduced by Pearson from electronic files supplied by the author.

ISBN-13: 978-0-13-597019-5
ISBN-10: 0-13-597019-9

Table of Contents

Preface

The first part of the ***Instructor's Solutions Manual*** contains our educational philosophy and teaching tips for each chapter of the text. Solutions to End-of-Section Problems and Chapter Review Problems in each chapter follow. Instructional tips and solutions for the digital cases are included next. Answers to the *Brynne Packaging* Case, the *CardioGood Fitness* Case, the *Choice Is Yours/More Descriptive Choices Follow-up Case*, the *Clear Mountain State Student Surveys Case,* the *Craybill Instrumentation Company* Case, the *Managing Ashland MultiComm Services* Case, the *Mountain States Potato Company Case* and the *Sure Value Convenience Stores* Case are included last.

The purpose of this ***Instructor's Solutions Manual*** is to facilitate grading of assignments or exams by instructors and/or teaching assistants. Screen shots using output from PHStat are integrated throughout. Most of the problems are solved using PHStat. To present the steps involved in solving a problem, some intermediate numerical results are presented accurate to only a reasonable number of significant digits. Hence, instructors are reminded that the final results presented in this manual that are obtained using PHStat can sometimes be different from those obtained with a hand calculator computed using the intermediate values due to rounding.

Teaching Tips for
Statistics for Managers using Microsoft® Excel 9th Ed.

Our Starting Point

Over a generation ago, advances in "data processing" led to new business opportunities as first centralized and then desktop computing proliferated. The Information Age was born. Computer science became much more than just an adjunct to a mathematics curriculum, and whole new fields of studies, such as computer information systems, emerged.

More recently, further advances in information technologies have combined with data analysis techniques to create new opportunities in what is more data *science* than data *processing* or *computer* science. The world of business statistics has grown larger, bumping into other disciplines. And, in a reprise of something that occurred a generation ago, new fields of study, this time with names such as informatics, data analytics, and decision science, have emerged.

This time of change makes what is taught in business statistics and how it is taught all the more critical. These new fields of study all share statistics as a foundation for further learning. We are accustomed to thinking about change, as seeking ways to continuously improve the teaching of business statistics have always guided our efforts. 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 ASA's Guidelines for Assessment and Instruction (GAISE) reports and combine them with our experiences teaching business statistics to a diverse student body at several large universities.

What to teach and how to teach it are particularly significant questions to ask during a time of change. As an author team, we bring a unique collection of experiences that we believe helps us find the proper perspective in balancing the old and the new. Our lead author, David M. Levine, was the first educator, along with Mark L. Berenson, to create a business statistics textbook that discussed using statistical software and incorporated "computer output" as illustrations—just the first of many teaching and curricular innovations in his many years of teaching business statistics. Our second author, David F. Stephan, developed courses and teaching methods in computer information systems and digital media during the information revolution, creating, and then teaching in, one of the first personal computer *classrooms* in a large school of business along the way. Early in his career, he introduced spreadsheet applications to a business statistics faculty audience that included David Levine, an introduction that would eventually led to the first edition of this textbook. Our newest co-author, Kathryn A. Szabat, has provided statistical advice to various business and non-business communities. Her background in

statistics and operations research and her experiences interacting with professionals in practice have guided her, as departmental chair, in developing a new, interdisciplinary academic department, Business Systems and Analytics, in response to the technology- and data-driven changes in business today.

All three of us benefit from our many years teaching undergraduate business subjects and the diversity of interests and efforts of our past co-authors, Mark Berenson and Timothy Krehbiel. Two of us (Stephan and Szabat) also benefit from formal training and background in educational methods and instructional design.

Educational Philosophy

As in prior editions of *Statistics for Managers Using Microsoft Excel*, we are guided by these key learning principles:

1. **Help students see the relevance of statistics to their own careers by providing examples drawn from the functional areas in which they may be specializing.** 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, such as accounting, finance, information systems, management, and marketing. Each statistics 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 evaluation of the assumptions, and the discussion of what should be done if the assumptions are violated.
2. **Emphasize interpretation of statistical results over mathematical computation.** Introductory business statistics courses should recognize the growing need to *interpret* statistical results that computerized processes create. This makes the interpretation of results more important than knowing how to execute the tedious hand calculations required to produce them.
3. **Give students ample practice in understanding how to apply statistics to 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.
4. **Familiarize students with how to use statistical software to assist business decision-making.** Introductory business statistics courses should recognize that programs with statistical functions are commonly found on a business decision maker's desktop computer. Integrating statistical software into all aspects of an introductory statistics course allows the course to focus on interpretation of results instead of computations (see point 2).
5. **Provide clear instructions to students for using statistical applications.** Books should explain clearly how to use programs such as Microsoft Excel with the study of statistics, without having those instructions dominate the book or distract from the learning of statistical concepts.

4 Teaching Tips

First Things First

In a time of change, you can never know exactly what knowledge and background students bring into an introductory business statistics classroom. Add that to the need to curb the fear factor about learning statistics that so many students begin with, and there's a lot to cover even before you teach your first statistical concept.

We created "First Things First" to meet this challenge. This unit sets the context for explaining what statistics is (not what students may think!) while ensuring that all students share an understanding of the forces that make learning business statistics critically important today. Especially designed for instructors teaching with course management tools, including those teaching hybrid or online courses, "First Things First" has been developed to be posted online or otherwise distributed before the first class section begins and is available from the download page for this book that is discussed in Appendix Section C.1.

We would argue that the most important class is the first class. First impressions are critically important. You have the opportunity to set the tone to create a new impression that the course will be important to their business education. Make the following points:

- This course is not a math course.
- State that you will be learning analytical skills for making business decisions.
- Explain that the focus will be on how statistics can be used in the functional areas of business.

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 **D**efine, **C**ollect, **O**rganize, **V**isualize, and **A**nalyze.

- **D**efine the business objective or problem to be solved and then define the variables to be studied.
- **C**ollect the data from appropriate sources
- **O**rganize the data
- **V**isualize the data by developing charts
- **A**nalyze the data by using statistical methods to reach conclusions.

To this, you can add **C** for Communicate which is critically important

You can begin by emphasizing the importance of defining your objective or problem. Then, discuss the importance of operational definitions of variables to be considered and define variable, data, and statistics.

Just as computers are used not just in the computer course, students need to know that statistics is used not just in the statistics course. This leads you to a discussion of business analytics in which data is

used to make decisions. Make the point that analytics should be part of the competitive strategy of every organization especially since “big data” meaning data collected in huge volumes at very fast rates. needs to be analyzed.

Inform the students that there is an Excel Guide at the end of each chapter. Strongly encourage or require students to read the Excel Guide at the end of this chapter so that they will be ready to use Excel with this book.

Chapter 1

You need to continue the discussion of the Define task by establishing 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. Then, if you wish, you can cover the different measurement scales.

Then move on to the C of the DCOVA approach, collecting data. Mention the different sources of data and make sure to cover the fact that data often needs to be cleaned of errors.

Then, you could spend some time discussing sampling, 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 *Consider This* essay discusses the important issue of the use of Web-based surveys.

The chapter also introduces two continuing cases related to *Managing Ashland MultiComm Services* and *CardioGood Fit*ness that appear 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*.).

Make sure that students read the Excel Guide at the end of each chapter. Ways of Working With Excel on pages 7 and 8 explains the different type of Excel instructions. The *Workbook* instructions provide step-by-step instructions and live worksheets that automatically update when data changes. 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 package that is included with many versions of Excel.

Chapter 2

This chapter moves on to the organizing and visualizing steps of the DCOVA framework. If you are going to collect sample data to use in Chapters 2 and 3, you can illustrate sampling by conducting a survey of students in your class. 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 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 worksheet and develop tables to help you prepare charts and analyze your data. Begin your discussion for categorical data with the example on p. 34 concerning how people pay for purchases and other transactions. 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. 38. 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 Excel to construct many different types of charts. Return to the purchase payment data previously discussed and illustrate how a bar chart, pie, and doughnut chart can be constructed using Excel. Mention the advantages and disadvantages of each chart. A good example is to show the data on incomplete ATM transactions on p. 49 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, the retirement funds data, 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.

8 Teaching Tips

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. If you cover the time series plot, you might also want to mention sparklines that are discussed in Section 2.6.

Also, if possible, you may want to discuss how multidimensional tables allow you to drill down to individual cells of the table. You can follow this with further discussion of PivotTables and Excel slicers that enable you to see panels for each variable being studied.

If the opportunity is available, we believe that it is worth the time to cover Section 2.7 on Challenges in Organizing and Visualizing Variables. 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 2.26 – 2.28. Ask students what is "bad" about these figures. Follow up with a homework assignment involving Problems 2.69 – 2.73 (*USA Today* is a great source).

You will find that the chapter review problems provide large data sets with numerous variables. Report writing exercises 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 CardioGood Fitness case focuses on developing a customer profile for a market research team. The Choice Is Yours Follow-up expands on the chapter discussion of the mutual funds data. The Clear Mountain State Student Survey provides data collected from a sample of undergraduate students.

The Excel Guide for this and the remaining chapters are organized according to the sections of the chapter. It is quite extensive since it covers both organizing and visualizing many different graphs. The Excel Guide includes instructions for Workbook, PHStat2, and the Analysis ToolPak.

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, use 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. If time permits, you can also discuss the geometric mean which is heavily used in finance.

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. 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. You may also want to briefly mention the property of kurtosis which is the relative concentration of values in the center of the distribution as compared to the tails. This statistic is provided by Excel through an Excel function or the Analysis Toolpak.

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

Now that these measures 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 students have taken, and the quantile scores that they have received (97th percentile, 48th percentile, 12th percentile, etc.). 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. Mention 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 discuss 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 118 that indicates the shape of the boxplot for four different distributions. Then, use PHStat2 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. 127 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.

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 CardioGood Fitness case focuses on developing a customer profile for a market research team. More Descriptive Choices Follow-up expands on the discussion of the mutual funds data. The Clear Mountain State Student Survey provides data collected from a sample of undergraduate students.

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

Chapter 4

The chapter on probability represents a bridge between the descriptive statistics already covered and the topics of statistical inference, regression, time series, and business analytics 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 the Excel Workbook instructions 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 (an online topic) can be covered. The *Consider This* concerning email SPAM is a wonderful way of helping students 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. The Digital case in this chapter extends the evaluation of the EndRun Financial Services to consider claims made about various probabilities. The CardioGood Fitness, More Descriptive Choices Follow-up, and Clear Mountain State Student Survey each involve developing contingency tables to be able to compute and interpret conditional and marginal 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, Poisson, or hypergeometric 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 0.0009. 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, an Excel function, or the binomial tables (See the Online **Binomial.pdf** tables).

Once the binomial distribution has been covered, if time permits, other discrete probability distributions can be presented. If you 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, an Excel function or the Poisson tables (See the Online **Poisson.pdf** tables).

If you so desire, you can discuss the covariance of a probability distribution (online Section 5.4), which is of particular importance to students majoring in finance. It is referred to in various finance courses including those on portfolio management and corporate finance. Use the example in the text to illustrate the covariance. If desired, continue with coverage of portfolio expected return and portfolio risk. Note that the PHStat2 Covariance and Portfolio Management menu selection allows you to readily compute the pertinent statistics. It also allows you to demonstrate changes in either the probabilities or the returns and their effect on the results. If you are using Workbook, you can start with the Portfolio.xls workbook and show how various Excel functions can be used to compute the desired statistics.

The hypergeometric distribution (online Section 5.5) can be developed for the situation in which one is sampling without replacement. Once again, use PHStat2 or an Excel function.

The *Managing Ashland MultiComm Services* case for this chapter relates to the binomial distribution. The Digital case involves the expected value and standard deviation of a probability distribution and applications of the covariance in finance.

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. 193. 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. 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 difficulty 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, or various Excel functions to compute normal probabilities. You can also use the Visual Explorations in Statistics Normal distribution procedure on p. 199. 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 *Consider 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 charts.

If time permits, you may want to cover the uniform distribution and refer to the table of random numbers as an example of this distribution. If you plan to cover the exponential distribution (which is an online topic), it is useful to discuss applications of this distribution in queuing (waiting line) theory. In

addition, be sure to point out that Equation (6.10) provides the probability of an arrival in less than or equal to a given amount of time. Be sure to mention that you can use PHStat2 or an Excel function to compute exponential probabilities.

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. The CardioGood Fitness, More Descriptive Choices Follow-up, and Clear Mountain State Student Survey each involve developing normal probability plots.

You can use either Excel functions or the PHStat add-in to compute normal and exponential probabilities and to construct normal probability plots.

Chapter 7

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 compute probabilities concerning the mean, not concerning individual values. It is helpful to display Figure 7.4 on p. 224 to show how the Central Limit Theorem applies to different 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 also involves the sampling distribution of the mean.

You might want to have students experiment with using the Visual Explorations add-in workbook to explore sampling distributions. You can also use either Excel functions, the PHStat add-in, or the Analysis ToolPak to develop sampling distribution simulations.

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 Ricknel Home Centers 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. If time permits, you may wish to discuss the effect of the finite population (this is an Online Topic) on the width of the confidence interval and the sample size needed. Point out that the correction factor should always be used when dealing with a finite population, but will have only a small effect when the sample size is a small proportion of the population size.

Due to the existence of a large number of accounting majors in many business schools, we have included an online section on applications of estimation in auditing. Two applications are included, the estimation of the total, and difference estimation. In estimating the total, point out that estimating the total is similar to estimating the mean, except that you are multiplying both the mean and the width of the confidence interval by the population size. When discussing difference estimation, be sure that the

students realize that all differences of zero must be accounted for in computing the mean difference and the standard deviation of the difference when using Equations (8.8) and (8.9).

Since the formulas for the confidence interval estimates and sample sizes discussed in this chapter are straightforward, using PHStat2 or Workbook 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. This chapter marks the first appearance of the Sure Value Convenience Store case which places the student in the role of someone working in the corporate office of a nationwide convenience store franchise. This case will appear in the next three chapters, Chapters 9 – 12, and also in Chapter 15.The CardioGood Fitness, More Descriptive Choices Follow-up, and Clear Mountain State Student Survey each involve developing confidence interval estimates.

You can use either Excel functions or the PHStat add-in to construct confidence intervals for means and proportions and to determine the sample size for means and proportions.

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 wrong) 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 (more extensive coverage of the power of a test is included in Section 9.6 which is an Online Topic).

Set up an example of a sampling distribution such as Figure 9.1 on p. 273, 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_o must go." and the rules for rejecting the null hypothesis and indicate that the *p*-value approach is a natural approach when using Excel, since the *p*-value can be determined by using PHStat, Excel functions, or the Analysis Toolpak.

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.

The *Managing Ashland MultiComm Services* case, Digital case, and the Sure Value Convenience Store case each involves the use of the one-sample test of hypothesis for the mean.

You can use either Excel functions or the PHStat add-in to carry out the hypothesis tests for means and proportions.

Chapter 10

This chapter discusses tests of hypothesis for the differences between two groups. The chapter begins with *t* tests for the difference between the means, then covers the *Z* test for the difference between two proportions, and concludes with the *F* test for the ratio of two variances.

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 Workbook approach, PHStat2, or the Analysis Toolpak 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 *Consider 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, if you have time in your course, 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 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 Workbook approach, PHStat2, or the Data Analysis tool 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 will often be a difference between males and females, you can then ask the class how to go about determining whether the results are statistically significant.

The *F*-test for the difference between two 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 placed 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 Workbook approach, PHStat2, or the Analysis Toolpak can be used to determine the test statistic and *p*-value.
The online section on effect size is particularly appropriate when you have big data with very large sample sizes.

Be aware that the *Managing Ashland MultiComm Services* case since it contains both independent sample and matched sample aspects, involves all the sections of the chapter except the test for the difference between two proportions. The Digital case is based on two independent samples. Thus, only the sections on the *t* test for independent samples and the *F* test for the difference between two variances are involved. The Sure Value Convenience Store case now involves a decision between two prices for coffee. The CardioGood Fitness, More Descriptive Choices Follow-up, and Clear Mountain State Student Survey each involve the determination of differences between two groups on both numerical and categorical variables.

You can use either Excel functions, the PHStat add-in, or the Analysis ToolPak to carry out the hypothesis tests for the differences between means and variances and for the paired *t* test. You can also use Excel functions or the PHStat add-in to carry out the hypothesis test for the differences between two proportions.

Chapter 11

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 Workbook, PHStat2, or the Analysis Toolpak. It is also useful to show how to obtain the critical *F* value by either referring to Table E.5 or the Excel 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. Workbook or PHStat2 can be used to compute the results for this test.

Once the Analysis of Variance has been covered, if time permits, 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.7. Be sure that students compare each paired difference between the means to the critical range. Note that you can use Workbook or PHStat2 to compute Tukey-Kramer multiple comparisons.

The factorial design model in Section 11.2 provides coverage of the two-way analysis of variance with equal number of observations for each combination of factor *A* and factor *B*. The approach taken in the text is primarily conceptual since, due to the complexity of the computations, the Analysis ToolPak, or PHStat2 should be used to perform the computations. You should develop the concept of partitioning the total sum of squares (*SST*) into factor *A* variation (*SSA*), factor *B* variation (*SSB*), interaction (*SSAB*) and random variation (*SSE*). Then move on to the development of the ANOVA table displayed in Table 11.6 on p. 367. Perhaps the most difficult concept to teach in the factorial design model is that of interaction. We believe that the display of an interaction graph such as the one shown in Figure 11.13 on

p. 371 is helpful. In addition, showing an example such as Example 11.2 on page 371 is particularly important, so that students observe the lack of parallel lines when significant interaction is present. Be sure to emphasize that the interaction effect is always tested prior to the main effects of *A* and *B*, since the interpretation of effects *A* and *B* will be affected by whether the interaction is significant.

The randomized block model which is an online topic is an extension of the paired *t* test in Chapter 10. Slowly go over the partitioning of the total sum of squares (*SST*) into Among Group variation (*SSA*), Among Block variation (*SSBL*), and Random variation (*SSE*). Discuss the ANOVA table and be sure students realize that Excel can be used to perform the computations. Finish this topic with a brief discussion of the relative efficiency of using the randomized block model and the use of the Tukey procedure for multiple comparisons. The online Section 11.4 briefly discusses the difference between the *F* tests involved when there are fixed and random effects.

The *Managing Ashland MultiComm Services* case for this chapter involves the one way ANOVA and the two-factor factorial design. The Digital case uses the One Way ANOVA. The Sure Value Convenience Store case now involves a decision among four prices for coffee. The CardioGood Fitness, More Descriptive Choices Follow-up, and Clear Mountain State Student Survey each involves using the one-way ANOVA to determine whether differences in numerical variables exist among three or more groups

In this chapter, using Workbook is more complicated than in other chapters, so you may want to focus on using the Analysis ToolPak or PHStat2.

Chapter 12

This chapter covers chi-square tests and nonparametric tests. The Using Statistics example concerning hotels relates to the first 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 will often 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. If you have time, you can develop the Marascuilo procedure to determine which groups differ.

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.

If you will be covering the Wilcoxon rank sum test, begin by noting that if the normality assumption was seriously violated, this test would be a good alternative to the *t* test for the difference between the means of two independent samples. Be sure to discuss the need to rank all the data values without regard to group. Review the fact that the statistic T_1 refers to the sum of the ranks for the group with the smaller sample size. If small samples are involved, be sure to point out that the null hypothesis is rejected if the test statistic T_1 is less than or equal to the lower critical value or greater than or equal to the upper critical value. In addition, explain when the normal approximation can be used. Point out that Workbook or PHStat2 can be used for the Wilcoxon rank sum test.

If the Kruskal-Wallis rank test is to be covered, you can explain that if the assumption of normality has been seriously violated, the Kruskal-Wallis rank test may be a better test procedure than the one-way ANOVA. Once again, be sure to discuss the need to rank all the data values without regard to

group. Go over how to find the critical values of the chi-square statistic using Table E.4. As was the case with the Wilcoxon rank sum test, Workbook or PHStat2 can be used for the Kruskal-Wallis rank test.
If you wish, you can briefly discuss the McNemar test which is an online topic. Explain that just like you used the paired-*t* test when you had related samples of numerical data, you use the McNemar test instead of the chi-square test when you have related samples of categorical data. Make sure to state that for two samples of related categorical data, the McNemar test is more powerful than the chi-square test.

You can then move on, if you wish, to the one sample test for the variance which is an online topic. Remind the students that if they are doing a two-tail test, they also need to find the lower critical value in the lower tail of the chi-square distribution.

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. The Sure Value Convenience Store case and the CardioGood Fitness cases involve using the Kruskal-Wallis test instead of the one-way ANOVA, The More Descriptive Choices Follow-up and Clear Mountain State Student Survey cases involve both contingency tables and nonparametric tests.

You can use Workbook or PHStat2.for testing differences between the proportions, tests of independence, and also for the Wilcoxon rank sum test and the Kruskal-Wallis test.

Chapter 13

Regression analysis is probably the most widely used and misused statistical method in business and economics. 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 chain of clothing stores, 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 Statistics Simple Linear Regression procedure on p. 436. This procedure produces a scatter plot 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 13.4 on p. 431. 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 Workbook, the Analysis ToolPak or PHStat2 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 have been violated for the model that has been fit. This leads into a discussion of residual analysis. Emphasize that Excel 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 13.9 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 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.

The *Managing Ashland MultiComm Services* case, the Digital case, and the Brynne case each involves a simple linear regression analysis of a set of data.

To perform simple linear regression, you can use Workbook, the Analysis ToolPak, or PHStat2.

Chapter 14

If time is available in the course, you can now move on to multiple regression. You should point out that Microsoft Excel 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 in simple regression -- prediction, residual analysis, coefficient of determination, and standard error of the estimate. If possible, the coefficient of partial determination is important to cover in order to be able to evaluate the contribution of each *X* variable to the model. Remind the students that to compute the coefficient of partial determination, they will need the total sum of squares, the regression sum of squares of the model that includes both variables, and the regression sum of squares for each independent variable given that the other independent variable is already included in the model.

If sufficient time is available, you can move on to 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.

Logistic regression is a topic that has become more important with the growth of business analytics since often there is the need to predict a categorical dependent variable. Explain that unlike Least Squares regression, you are predicting the odds ratio and the probability of an event of interest not a numerical value. You may need to briefly mention natural logarithms and refer students to Appendix A. Make the point that Excel will perform the complex computations involved in logistic regression and all the student will need to do is interpret the results provided.

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

To perform multiple regression, you can use Workbook, the Analysis ToolPak, or PHStat2. To perform logistic regression, you can use Workbook or PHStat2.

Chapter 15

The amount of coverage that can be given to multiple regression in a one-semester course is often limited or not even possible. However, in a two-semester course, additional topics can be covered. Collinearity should at least be mentioned when multiple regression is covered, since it represents one of the problems that can occur with multiple regression models. In terms of the coverage of the quadratic regression model, note that it can be considered as a multiple regression model in which the second independent variable is the square of the first independent variable.

If you are teaching a two-semester course or a course that focuses more on regression, you may be able to cover various topics and also include an introduction to transformations and the capstone topic in regression, model building. This text focuses on the more modern and inclusive approach called best subsets regression that allows the examination of all possible regression models. Excel with PHStat2 includes model building using this approach and provides various statistics for each model including the C_p statistic. If you are using the example presented in Section 15.4, be sure to show the results of all the models. Carefully discuss the steps involved in model building presented in Exhibit 15.1 and the Figure 15.15 roadmap on page 542 for model building.

The Mountain States Potato Company case and the Craybill Instrumentation case provide rich data sets for model building. The Sure Value Convenience Stores case provides an opportunity to fit a quadratic regression model. The Digital case here expands the Digital case presented in Chapter 14 to consider additional variables.

To perform quadratic regression, you can define the quadratic terms using Excel and then use Workbook, the Analysis ToolPak, or PHStat2. To build multiple regression models, you need to use PHStat.

Chapter 16

A good way to begin the discussion of time-series models is to indicate how these models are different from the regression models considered in the previous chapters. In particular, you should focus on the fact that three types of models will be considered, (1) classical models that use least-squares regression in which the independent variable is the time period, (2) moving average and exponential smoothing methods in which no trend is assumed to be present, and (3) autoregressive models in which the independent variable(s) represent values of the dependent variable that have been lagged by one or more time periods.

You may wish to begin by discussing moving average and exponential trend methods. Emphasize the fact that these models are appropriate for smoothing a series when the nature of the trend is unclear or no trend is thought to exist. Point out the fact that the moving average method is not used to forecast into the future and the exponential smoothing method is used to forecast only one period into the future. Be sure to indicate that there is a certain amount of subjectivity involved in any forecast in exponential smoothing since the choice of a weight is somewhat arbitrary. Be sure that students are aware that Excel functions and the Analysis ToolPak can be used to compute moving averages and exponential smoothing results.

You can then move on to the Least Squares trend models and consider three models -- the linear trend model, the curvilinear or quadratic trend model, and the exponential trend model. Several points should be made before beginning the discussion. First, to make the interpretation simpler, the first year of the time series is coded with an X value of zero. Second, remind students that the computations can be done using Workbook, PHStat2, or with the Analysis ToolPak. Third, be sure to indicate that we use the Principle of Parsimony in choosing a model. This principle states that if a simpler model is as good as a more complex one, the simpler model should be chosen. If the exponential trend model is to be covered, remind the students that since the model is linear in the logarithms, antilogarithms of the regression coefficients must be taken in order to express the model in the original units of measurement and to express the predicted Log Y values in the original units for calculating the magnitude of the residuals for model comparison statistics. Point out also that if 1 is subtracted from the antilogarithm of the slope, the rate of growth predicted by the model will be obtained. Reiterate that the exponential model is most appropriate in situations in which the time series is changing at an increasing rate so that the percentage difference from period to period is constant.

An additional approach to forecasting involves autoregressive modeling. Go over the fact that in an autoregressive model, the independent variable is a lagged dependent variable from a previous time period. A first-order autoregressive model has its independent variable as the dependent variable from the

previous time period, while a second-order model has an additional independent variable from a time period that is two periods prior to the one being considered. You might also mention the fact that these autoregressive models are simpler versions of the widely used autoregressive integrated moving average (ARIMA) models.

Now that numerous models have been considered for forecasting purposes, you can turn to the critical issue of choosing the most appropriate model. Emphasize the fact that there are two considerations, the pattern of the residuals and the amount of error in the forecast. Point out the importance of choosing a model that does not have a pattern in the residuals. Also mention that the mean absolute deviation approach is widely used, but that there are other alternative measures that could be considered.

Discussion in the next section focuses on quarterly or monthly data. The approach used in the text involves regression in which dummy variables are used to represent the months or quarters. Use Excel to obtain the results of this complex dummy variable model and slowly go over the interpretation of the intercept, the regression coefficient that refers to time, and the coefficients of the dummy variables. Be sure to note that for monthly data, each dummy variable relates to the multiplier for that month relative to December (for quarterly data each quarter is relative to the fourth quarter).

Index numbers is an Online topic. Begin with the simple price index and then point out that indexes for a group of commodities are common in business. Mention the Consumer Price Index as an example of an aggregate price index. Point out the difference between an unweighted aggregate price index and weighted price indexes that consider the consumption quantities of each commodity.

The *Managing Ashland MultiComm Services* case and the Digital case involve forecasting future sales for monthly data. The Digital case involves a comparison of models for two different sets of data.

To compute moving averages, you use Excel or the Analysis ToolPak. To compute exponentially smoothed values, you use Excel or the Analysis ToolPak. To compute linear, quadratic, exponential trend, autoregressive, and monthly/quarterly models, you use Excel functions followed by Workbook, the Analysis ToolPak, or PHStat2.

Chapter 17

Sections 17.1 and 17.2 represent the culmination of the book. All too often students who complete an introductory business statistics course or courses are faced with a situation in subsequent business courses or new business situations of trying to figure out what statistical methods are appropriate. Whereas, when they were learning methods in a specific chapter, they could assume that their solution lay with the methods covered in the chapter, now things are more open ended.

This chapter provides a roadmap for helping students deal with this situation. The chapter breaks the task down according to whether you are dealing with numerical variables or categorical variables. Then, a series of questions are asked and answers provided for each one of these circumstances.

A good strategy may be to make students aware of Chapter 17 as you proceed through the semester especially when you reach different hypothesis testing procedures. After you complete a chapter (for example, Chapter 10), you can refer to the questions in Chapter 17 so that students will have a better chance of seeing the big picture.

Sections 173, 17.4, and 17.5 on Business analytics are greatly expanded in this edition. If you are teaching a one semester course it is possible that you could integrate some of the descriptive analytics material in Section 17.4 into the discussion of tables and charts. However, a full discussion of the material in Sections 17.3 – 17.5 would be more appropriate in a two semester course or a separate course on business analytics.

You may wish to begin with a discussion of big data and the need for techniques for dealing with big data. Then you can provide examples of the various graphs used in Section 17.1 such as dashboards, gauges, and tree maps.

Section 17. introduces classification and regression trees. Classification trees are used when the dependent variable is categorical. Regression trees are used when the dependent variable is numerical. With classification trees, the dependent variable is broken down according to values of the independent variables. A good approach might be to use the same example that you may have used in logistic regression and then compare the results. With regression trees, you may want to use the same example that you may have used in Chapter 14 or 15 and then compare the results.

Chapter 18 (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 18.2 and 18.4 and variables charts in Section 18.5.

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 18.2. Be sure that students are aware that Excel or PHStat2 can be used to construct the *p* chart. If time permits, you may wish to also cover the *c* chart. If you choose to do so, be sure to focus on the fact that the variable involved represents the number of nonconformities per unit (an area of opportunity).
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, or PHStat2 can be used to construct both *R* and $\overline{X}$ charts.

If time allows, you may wish to discuss the topic of process capability. This topic reinforces any previous coverage of the normal distribution. Be sure to go over the distinction between control limits and specification limits and the differences between the various capability statistics.

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 18.3). This experiment allows the student to see the distinction between the two types of variation. The amount of time spent on Sections 18.8 and 18.9 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 18.9) 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. You can also mention Lean Six Sigma.

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

Chapter 19 (Online Chapter)

This chapter expands on the development of the expected value and standard deviation of a probability distribution and Bayes' theorem to develop additional concepts in decision making. In this chapter, all topics refer to the Using Statistics example of the mutual fund and the marketing of organic salad dressings first discussed in Example 19.1. Begin the chapter with the payoff table and the notion of alternative courses of action (some prefer using decision trees). Reiterate that payoffs are often available or can be determined from the profit or cost structure of a problem as shown in problems 19.3 – 19.5. When teaching opportunity loss, be sure to emphasize that you are finding the optimal action and the opportunity loss for each event (row of our payoff table).

The coverage of criteria for decision making covers several criteria including maximax, maximin, expected monetary value, expected opportunity loss, and the return-to-risk ratio. Be sure to remind students that, the expected monetary value and the return to risk ratio may lead to different optimal actions. Note that PHStat2 includes the Decision Making menu selection which provides computations for the various criteria for a given payoff table and event probabilities (or you can use Workbook). It also allows you to demonstrate changes in either the probabilities or the returns and their effect on the results. If time permits, Bayes' theorem can be used to revise probabilities based on sample information and the utility concept can be introduced.

Chapter 1

1.1 (a) The type of beverage sold yields categorical or "qualitative" responses rather than numerical responses. Each beverage type represents a separate category.
(b) The type of beverage category expresses no order or ranking.

1.2 Business size represents a categorical variable because each size represents a particular category. Because of the different sizes, order is implied, but this variable includes no information about the quantity of differences among the three sizes.

1.3 (a) The time it takes to download a video from the Internet is a continuous numerical or "quantitative" variable because time can have any value from 0 to any reasonable unit of time
(b) The time it takes to download a video from the Internet is a ratio-scaled variable because it is an ordered scale that includes a true zero point.

1.4 (a) The number of cellphones is a numerical variable that is discrete because the outcome is a count.
(b) Monthly data usage is a numerical variable that is continuous because any value within a range of values can occur.
(c) Number of text messages exchanged per month is a numerical variable that is discrete because the outcome is a count.
(d) Voice usage per month is a numerical variable that is continuous because any value within a range of values can occur.
(e) Whether a cellphone is used for streaming video is a categorical variable because the answer can be only yes or no.

1.5 (a) numerical, ratio
(b) numerical, ratio
(c) categorical, nominal
(d) categorical, nominal

1.6 (a) Categorical, nominal scale
(b) Numerical, continuous, ratio scale
(c) Categorical, nominal scale
(d) Numerical, discrete, ratio scale
(e) Categorical, nominal scale

1.7 (a) numerical, ratio scale, continuous
(b) categorical, nominal scale
(c) categorical, nominal scale
(d) numerical, ratio scale, discrete

1.8 (a) numerical, continuous, ratio scale
(b) numerical, discrete, ratio scale
(c) numerical, continuous, ratio scale
(d) categorical, nominal scale

1.9 (a) Income may be considered discrete if we "count" our money. It may be considered continuous if we "measure" our money; we are only limited by the way a country's monetary system treats its currency.
(b) The first format would provide more information because it includes a ratio scale value while the second measure would only include a range of values for each choice category.

1.10 The underlying variable, ability of the students, may be continuous, but the measuring device, the test, does not have enough precision to distinguish between the two students.

1.11 (a) The population is "all working women from the metropolitan area." A systematic or random sample could be taken of women from the metropolitan area. The director might wish to collect both numerical and categorical data.
(b) Three categorical questions might be occupation, marital status, type of clothing. Numerical questions might be age, average monthly hours shopping for clothing, income.

1.12 (a) Data distributed by an organization or individual.
(b) The American Community Survey is based on a sample.

1.13 The answer depends on the specific story.

1.14 The answer depends on the specific story.

1.15 The transportation engineers and planners should use primary data collected through an observational study of the driving characteristics of drivers over the course of a month.

1.16 The information presented there is based mainly on a mixture of data distributed by an organization and data collected by ongoing business activities.

1.17 (a) 001
(b) 040
(c) 902

1.18 Sample without replacement: Read from left to right in 3-digit sequences and continue unfinished sequences from end of row to beginning of next row.
Row 05: 338 505 855 551 438 855 077 186 579 488 767 833 170
Rows 05–06: 897
Row 06: 340 033 648 847 204 334 639 193 639 411 095 924
Rows 06–07: 707
Row 07: 054 329 776 100 871 007 255 980 646 886 823 920 461
Row 08: 893 829 380 900 796 959 453 410 181 277 660 908 887
Rows 08–09: 237
Row 09: 818 721 426 714 050 785 223 801 670 353 362 449
Rows 09–10: 406
Note: All sequences above 902 and duplicates are discarded.

1.19 (a) Row 29: 12 47 83 76 22 99 65 93 10 65 83 61 36 98 89 58 86 92 71
Note: All sequences above 93 and all repeating sequences are discarded.
(b) Row 29: 12 47 83 76 22 99 65 93 10 65 83 61 36 98 89 58 86
Note: All sequences above 93 are discarded. Elements 65 and 83 are repeated.

1.20 A simple random sample would be less practical for personal interviews because of travel costs (unless interviewees are paid to attend a central interviewing location).

1.21 This is a probability sample because the selection is based on chance. It is not a simple random sample because A is more likely to be selected than B or C.

1.22 Here all members of the population are equally likely to be selected and the sample selection mechanism is based on chance. But not every sample of size 2 has the same chance of being selected. For example the sample "B and C" is impossible.

1.23 (a) Since a complete roster of full-time students exists, a simple random sample of 200 students could be taken. If student satisfaction with the quality of campus life randomly fluctuates across the student body, a systematic 1-in-20 sample could also be taken from the population frame. If student satisfaction with the quality of life may differ by gender and by experience/class level, a stratified sample using eight strata, female freshmen through female seniors and male freshmen through male seniors, could be selected. If student satisfaction with the quality of life is thought to fluctuate as much within clusters as between them, a cluster sample could be taken.

(b) A simple random sample is one of the simplest to select. The population frame is the registrar's file of 4,000 student names.

(c) A systematic sample is easier to select by hand from the registrar's records than a simple random sample, since an initial person at random is selected and then every 20th person thereafter would be sampled. The systematic sample would have the additional benefit that the alphabetic distribution of sampled students' names would be more comparable to the alphabetic distribution of student names in the campus population.

(d) If rosters by gender and class designations are readily available, a stratified sample should be taken. Since student satisfaction with the quality of life may indeed differ by gender and class level, the use of a stratified sampling design will not only ensure all strata are represented in the sample, it will also generate a more representative sample and produce estimates of the population parameter that have greater precision.

(e) If all 4,000 full-time students reside in one of 10 on-campus residence halls which fully integrate students by gender and by class, a cluster sample should be taken. A cluster could be defined as an entire residence hall, and the students of a single randomly selected residence hall could be sampled. Since each dormitory has 400 students, a systematic sample of 200 students can then be selected from the chosen cluster of 400 students. Alternately, a cluster could be defined as a floor of one of the 10 dormitories. Suppose there are four floors in each dormitory with 100 students on each floor. Two floors could be randomly sampled to produce the required 200 student sample. Selection of an entire dormitory may make distribution and collection of the survey easier to accomplish. In contrast, if there is some variable other than gender or class that differs across dormitories, sampling by floor may produce a more representative sample.

1.24 (a)
Row 16: 2323 6737 5131 8888 1718 0654 6832 4647 6510 4877
Row 17: 4579 4269 2615 1308 2455 7830 5550 5852 5514 7182
Row 18: 0989 3205 0514 2256 8514 4642 7567 8896 2977 8822
Row 19: 5438 2745 9891 4991 4523 6847 9276 8646 1628 3554
Row 20: 9475 0899 2337 0892 0048 8033 6945 9826 9403 6858
Row 21: 7029 7341 3553 1403 3340 4205 0823 4144 1048 2949
Row 22: 8515 7479 5432 9792 6575 5760 0408 8112 2507 3742
Row 23: 1110 0023 4012 8607 4697 9664 4894 3928 7072 5815
Row 24: 3687 1507 7530 5925 7143 1738 1688 5625 8533 5041
Row 25: 2391 3483 5763 3081 6090 5169 0546
Note: All sequences above 5000 are discarded. There were no repeating sequences.

(b)
089 189 289 389 489 589 689 789 889 989
1089 1189 1289 1389 1489 1589 1689 1789 1889 1989
2089 2189 2289 2389 2489 2589 2689 2789 2889 2989
3089 3189 3289 3389 3489 3589 3689 3789 3889 3989
4089 4189 4289 4389 4489 4589 4689 4789 4889 4989

(c) With the single exception of invoice #0989, the invoices selected in the simple random sample are not the same as those selected in the systematic sample. It would be highly unlikely that a random process would select the same units as a systematic process.

1.25 (a) A stratified sample should be taken so that each of the three strata will be proportionately represented.

(b) The number of observations in each of the three strata out of the total of 100 should reflect the proportion of the three categories in the customer database. For example, 500/1000 = 50% so 50% of 100 = 50 customers should be selected from the potential customers; similarly, 300/1000 = 30% so 30 customers should be selected from those who have purchased once, and 200/1000 = 20% so 20 customers from the repeat buyers.

(c) It is not simple random sampling because, unlike the simple random sampling, it ensures proportionate representation across the entire population.

1.26 (a) For the third value, Apple is spelled incorrectly. The twelfth value should be Blackberry not Blueberry. The fifteenth value, APPLE, may lead to an irregularity. The eighteenth value should be Samsung not Samsun.

(b) This list contains 19 names, where one would expect to find 20 names.

1.27 Only the second value, 2.7MB, contains units and the eighth value, 1,079, might be confused with 1, 79.

1.28 (a) The times for each of the hotels would be arranged in separate columns.

(b) The hotel names would be in one column and the times would be in a second column.

1.29 A recoded variable PriceLevel could be defined, assigning the value Budget for hotels with budget-priced rooms, Moderate for hotels with moderate-priced rooms, and Deluxe for hotels with deluxe-priced rooms.

1.30 Before accepting the results of a survey of college students, you might want to know, for example:
Who funded the survey? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained?

1.30 cont. Were survey questions field-tested? What questions were asked? Were they clear, accurate, unbiased, valid? What operational definition of "vast majority" was used? What was the response rate? What was the sample size?

1.31 (a) Possible coverage error: Only employees in a specific division of the company were sampled.
(b) Possible nonresponse error: No attempt is made to contact nonrespondents to urge them to complete the evaluation of job satisfaction.
(c) Possible sampling error: The sample statistics obtained from the sample will not be equal to the parameters of interest in the population.
(d) Possible measurement error: Ambiguous wording in questions asked on the questionnaire.

1.32 Coverage error could result if bank executives are systematically excluded from the population thereby not allowing them to be part of any sample that is used to generate the results. This could lead to selection bias. Non-response error that results in non-response bias could result if not all bank executives who were selected for inclusion in the sample are contacted even after multiple attempts to do so. In this case, data that was desired and necessary for inclusion in the sample would then not be present. Sampling error reflects the variability in outcomes when taking different samples. Sampling error is unavoidable. One can obtain an impression of the size of the sampling error by creating interval estimates. Measurement error could arise if the bank executives self-report results or if the methods of reporting are not standardized, i.e. if questions are not asked in the same manner from respondent to respondent, or if those conducting the survey do not do so in a consistent manner.

1.33 Before accepting the results of the survey, you might want to know, for example:
Who funded the survey? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What questions were asked? Were they clear, accurate, unbiased, valid? What was the response rate? What was the margin of error? What was the sample size? What frame was used?

1.34 Before accepting the results of the survey, you might want to know, for example:
Who funded the survey? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What questions were asked? Were they clear, accurate, unbiased, valid? What was the response rate? What was the margin of error? What was the sample size? What frame was used?

1.35 A population contains all the items of interest whereas a sample contains only a portion of the items in the population.

1.36 A statistic is a summary measure describing a sample whereas a parameter is a summary measure describing an entire population.

1.37 Categorical random variables yield categorical responses such as yes or no answers. Numerical random variables yield numerical responses such as your height in inches.

1.38 Discrete random variables produce numerical responses that arise from a counting process. Continuous random variables produce numerical responses that arise from a measuring process.

1.39 Both nominal scaled and ordinal scaled variables are categorical variables but no ranking is implied in nominal scaled variable such as male or female while ranking is implied in ordinal scaled variable such as a student's grade of A, B, C, D and F.

1.40 Both interval scaled and ratio scaled variables are numerical variables in which the difference between measurements is meaningful but an interval scaled variable does not involve a true zero such as standardized exam scores while a ratio scaled variable involves a true zero such as height.

1.41 Items or individuals in a probability sampling are selected based on known probabilities while items or individuals in a nonprobability samplings are selected without knowing their probabilities of selection.

1.42 Missing values are values that were not collected for a variable. Outliers are values that seem excessively different from most of the other values

1.43 In unstacked arrangements, separate numerical variables are created for each group in the data. For example, you might create a variable for the weights of men and a second variable for the weights of women. In stacked arrangements, a single numerical variable is paired with a categorical variable that represents the categories. For example, all weights would be in one variable, with a categorical variable indicating male or female.

1.44 Coverage error is error generated due to an improperly or inappropriately framed population which can result in a sample that may not be representative of the population that one wishes to study. Non-response error is error generated due to members of a chosen sample not being contacted even after repeated attempts so that information that should be provided is missing.

1.45 Sampling error results from the variability of outcomes of different samples. This sample to sample variation is inevitably connected to the sampling process. Measurement error is error that results from either self-reported data or data that is collected in an inconsistent manner by those who are responsible for collecting and summarizing the desired information.

1.46 Microsoft Excel:
This product features a spreadsheet-based interface that allows users to organize, calculate, and organize data. Excel also contains many statistical functions to assist in the description of a dataset. Excel can be used to develop worksheets and workbooks to calculate a variety of statistics including introductory and advanced statistics. Excel also includes interactive tools to create graphs, charts, and pivot tables. Excel can be used to summarize data to better understand a population of interest, compare across groups, predict outcomes, and to develop forecasting models. These capabilities represent those that are generally relevant to the current course.

Excel also includes many other statistical capabilities that can be further explored on the Microsoft Office Excel official website.

1.47 (a) The population of interest include banking executives representing institutions of various sizes and U.S. geographic locations.
(b) The collected sample includes 163 banking executives from institutions of various sizes and U.S. geographic locations.

1.47 cont.
(c) A parameter of interest is the percentage of the population of banking executives that identify customer experience initiatives as an area where increased spending is expected.
(d) A statistic used to the estimate the parameter in (c) is the percentage of the 163 banking executives included in the sample who identify customer experience initiatives as an area where increased spending is expected. In this case, the statistic is 55%.

1.48 The answers are based on an article titled "U.S. Satisfaction Still Running at Improved Level" and written by Lydia Saad (August 15, 2018). The article is located on the following site: https://news.gallup.com/poll/240911/satisfaction-running-improved-level.aspx?g_source=link_NEWSV9&g_medium=NEWSFEED&g_campaign=item_&g_content=U.S.%2520Satisfaction%2520Still%2520Running%2520at%2520Improved%2520Level

(a) The population of interest includes all individuals aged 18 and older who live within the 50 U.S. states and the District of Columbia.
(b) The collected sample includes a random sample of 1,024 individuals aged 18 and older who live within the 50 U.S. states and the District of Columbia.
(c) A parameter of interest is the percentage of the population of individuals aged 18 and older and live within the 50 U.S. states and the District of Columbia who are satisfied with the direction of the U.S.
(d) A statistic used to the estimate the parameter in (c) is the percentage of the 1,024 individuals included in the sample. In this case, the statistic is 36%.

1.49 The answers were based on information obtained from the following site:
(a) The population of interest is U.S. CEOs
(b) The sample included 1,000 U.S. CEOs.
(c) A parameter of interest would be the percentage of CEOs among the population of interest that believe that AI will significantly change the way they will do business in the next five years.
(d) The statistic used to estimate the parameter in (c) is the percentage of CEOs among the 1,000 CEOs included in the sample who believe that AI will significantly change the way they will do business in the next five years. In this case, the statistic is 80% agree with this statement

1.50 (a) One variable collected with the American Community Survey is marital status with the following possible responses: now married, widowed, divorced, separated, and never married.
(b) The variable in (a) represents a categorical variable.
(c) Because the variable in (a) is a categorical, this question is not applicable. If one had chosen age in years from the American Community Survey as the variable, the answer to (c) would be discrete.

1.51 Answers will vary depending on the specific sample survey used. The below answers were based on the sample survey located at: bit.ly/21qjI6F

(a) An example of a categorical variable included in the survey is gender with male or female as possible answers.
(b) An example of a numerical variable included in the survey would be the number of phone calls made or received from or to ones direct supervisor in an average week.

1.52 (a) The population of interest consisted of 10,000 benefited employees of the University of Utah.
(b) The sample consisted of 3,095 employees of the University of Utah.

1.52 cont. (c) Gender, marital status, and employment category represent categorical variables. Age in years, education level in years completed, and household income represent numerical variables.

1.53 (a) Key social media platforms used represents a categorical variable. The frequency of social media usage represents a discrete numerical variable. Demographics of key social media platform users represent categorical variables.

(b) 1. Which of the following is your preferred social media platform: YouTube, Facebook, or Twitter?
2. What time of the day do you spend the most amount of time using social media: morning, afternoon, or evening?
3. Please indicate your ethnicity?
4. Which of the following do you most often use to access social media: mobile device, laptop computer, desktop computer, other device?
5. Please indicate whether you are a home owner: Yes or No?

(c) 1. For the past week, how many hours did you spend using social media?
2. Please indicate your current age in years.
3. What was your annual income this past year?
4. Currently, how many friends have you accepted on Facebook?
5. Currently, how many twitter followers do you have?

Chapter 2

2.1 (a)

Category	Frequency	Percentage
A	13	26%
B	28	56%
C	9	18%

(b) Category "B" is the majority.

2.2 (a) Table frequencies for all student responses

	Student Major Categories			
Gender	A	C	M	Totals
Male	14	9	2	25
Female	6	6	3	15
Totals	20	15	5	40

(b) Table percentages based on overall student responses

	Student Major Categories			
Gender	A	C	M	Totals
Male	35.0%	22.5%	5.0%	62.5%
Female	15.0%	15.0%	7.5%	37.5%
Totals	50.0%	37.5%	12.5%	100.0%

Table based on row percentages

	Student Major Categories			
Gender	A	C	M	Totals
Male	56.0%	36.0%	8.0%	100.0%
Female	40.0%	40.0%	20.0%	100.0%
Totals	50.0%	37.5%	12.5%	100.0%

Table based on column percentages

	Student Major Categories			
Gender	A	C	M	Totals
Male	70.0%	60.0%	40.0%	62.5%
Female	30.0%	40.0%	60.0%	37.5%
Totals	100.0%	100.0%	100.0%	100.0%

2.3 (a) You can conclude Apple, Samsung, and LG dominated the market from the third quarter of 2017 through the third quarter of 2018. Apple has the largest market share and the largest gain in market share increasing from 33% in the third quarter of 2017 to 39% in the third quarter of 2018.

(b) Apple, Samsung, and Motorola increased market share while LG and Others decreased market share.

2.4 (a)

Category	Total	Percentages
Bank Account or Service	202	9.330%
Consumer Loan	132	6.097%
Credit Card	175	8.083%
Credit Reporting	581	26.836%
Debt Collection	486	22.448%
Mortgage	442	20.416%
Student Loan	75	3.464%
Other	72	3.326%
Grand Total	2165	

(b) There are more complaints for credit reporting, debt collection, and mortgage than the other categories. These categories account for about 70% of all the complaints.

(c)

Company	Total	Percentage
Bank of America	42	3.64%
Capital One	93	8.07%
Citibank	59	5.12%
Ditech Financial	31	2.69%
Equifax	217	18.82%
Experian	177	15.35%
JPMorgan	128	11.10%
Nationstar Mortgage	39	3.38%
Navient	38	3.30%
Ocwen	41	3.56%
Synchrony	43	3.73%
Trans-Union	168	14.57%
Wells Fargo	77	6.68%
Grand Total	1153	

(d) Equifax, Trans-Union, and Experion, all of which are credit score companies, have the most complaints.

2.5 The respondents from the sample of companies included in the survey indicated that they would be most likely to use business analytics within the next five years with more than half planning to use this technology. Slightly less than half of the respondents plan to use machine language technology and only 21% of the respondents plaining to use self-learning robots. Nearly half of the respondents had no plans to use self-learning robots. It could be concluded that business analytics will likely be the most impactful technology within the next five years followed by machine language technology and self-learning robots.

2.6 The largest sources of electricity in the United States are Natural gas followed by Coal. Nuclear is the third most used source of energy. Hydropower and Wind are about the same followed by Solar, Biomass, Petroleum coke and liquids, Geothermal, and Other sources.

2.7 (a)

Technologies	Frequency	Percentage
Wearable technology	9	10.00%
Blockchain technology	9	10.00%
Artificial intelligence	17	18.89%
Iot: retail insurance	23	25.56%
Iot: commerical insurance	5	5.56%
Social media	27	30.00%
Grand Total	90	

(b) Professionals expect to be using Social media and Iot: retail insurance technologies the most over the next year followed by Artificial intelligence. Professionals do not expect to be using Wearable, Blockchain, and Iot: commercial insurance technologies much over the next year.

2.8 (a) Table of row percentages:

	GENDER		
OVERLOADED	Male	Female	Total
Yes	44.08%	55.92%	100.00%
No	53.54%	46.46%	100.00%
Total	51.64%	48.36%	100.00%

Table of column percentages:

	GENDER		
OVERLOADED	Male	Female	Total
Yes	17.07%	23.13%	20.00%
No	82.93%	76.87%	80.00%
Total	100.00%	100.00%	100.00%

Table of total percentages:

	GENDER		
OVERLOADED	Male	Female	Total
Yes	8.82%	11.18%	20.00%
No	42.83%	37.17%	80.00%
Total	51.64%	48.36%	100.00%

(b) Approximately the same percentages of males and females as a percentage of the total number of people surveyed feel overloaded with too much information. As percentages of those who do and do not feel overloaded, the genders differ mildly. However, four times as many people do not feel overloaded at work than those that do.

2.9 (a)

Column Percentage

	OUTCOME		
CATEGORY	**Successful**	**Not Successful**	**Total**
Film & Video	36.02%	36.81%	36.51%
Games	15.44%	18.24%	17.19%
Music	40.20%	24.38%	30.34%
Technology	8.34%	20.56%	15.96%
Total	100.00%	100.00%	100.00%

Row Percentages

	OUTCOME		
CATEGORY	**Successful**	**Not Successful**	**Total**
Film & Video	37.15%	62.85%	100.00%
Games	33.84%	66.16%	100.00%
Music	49.91%	50.09%	100.00%
Technology	19.69%	80.31%	100.00%
Total	37.67%	62.33%	100.00%

Total Percentages

	OUTCOME		
CATEGORY	**Successful**	**Not Successful**	**Total**
Film & Video	13.57%	22.95%	36.51%
Games	5.82%	11.37%	17.19%
Music	15.14%	15.20%	30.34%
Technology	3.14%	12.82%	15.96%
Total	37.67%	62.33%	100.00%

(b) The row percentages are most informative because they provide a percentage of successful projects within each category which allows one to compare across categories.

(c) Music kick starter projects were the most successful with approximately 50% of the projects succeeding compared to less than 20% of the Technology projects. The Film & Video and Games categories had success rates in between the Music and Technology categories, with success rates of 37% and 34% respectively.

2.10 Approximately 11.7% of the total people surveyed answered window tinting as their preferred luxury upgrade. A higher percentage of women than men prefer window tinting as a luxury upgrade.

2.11 Ordered array: 63 64 68 71 75 88 94

2.12 Ordered array: 73 78 78 78 85 88 91

2.13 (a) (166 + 100)/591 * 100 = 45.01%
(b) (124 + 77)/591 * 100 = 34.01%
(c) (59 + 65)/591 * 100 = 20.98%
(d) 45% of the incidents took fewer than 2 days and 66% of the incidents were detected in less than 8 days. 79% of the incidents were detected in less than 31 days.

2.14 $\frac{261,000 - 61,000}{6} = 33,333.33$ so choose 40,000 as interval width

(a) $60,000 – under $100,000; $100,000 – under $140,000; $140,000 – under $180,000; $180,000 – under $220,000; $220,000 – under $260,000; $260,000 – under $300,000

(b) $40,000

(c) $\frac{60,000 + 100,000}{2} = \$80,000$ similarly, the remaining class midpoints are $120,000; $160,000; $200,000; $240,000; $280,000

2.15 (a)

4.0	5.0	5.0	5.5	5.6	5.8	5.8	5.9
	6.0	6.1	6.3	6.3	6.4		
6.5	6.7	7.3	7.3	7.8	7.8	7.8	7.9
	7.9	8.0	8.4	8.8	10.2		
10.5	12.8	13.3	14.3				

(b)

Work Hours Needed	Frequency	Percentage
4 but less than 6	8	27%
6 but less than 8	14	47%
8 but less than 10	3	10%
10 but less than 12	2	7%
12 but less than 14	2	7%
14 but less than 16	1	3%
Total	30	100%

(c) 47% of the average number of work hours necessary to afford an NBA basketball game are between 6 hours and 8 hours with 27% of the costs between 4 hours and 6 hours.

2.16 (a)

Electricity Costs

Electricity Costs	Frequency	Percentage
$80 but less than $100	4	8%
$100 but less than $120	7	14%
$120 but less than $140	9	18%
$140 but less than $160	13	26%
$160 but less than $180	9	18%
$180 but less than $200	5	10%
$200 but less than $220	3	6%

2.16 (b) cont.

Electricity Costs	Frequency	Percentage	Cumulative %
$ 99	4	8.00%	8.00%
$119	7	14.00%	22.00%
$139	9	18.00%	40.00%
$159	13	26.00%	66.00%
$179	9	18.00%	84.00%
$199	5	10.00%	94.00%
$219	3	6.00%	100.00%

(c) The majority of utility charges are clustered between $120 and $180.

2.17 (a)

Commuting Time (minutes)	Frequency	Percentage
115 but less than 130	10	33%
130 but less than 145	9	30%
145 but less than 160	7	23%
160 but less than 175	3	10%
175 but less than 190	1	3%
	30	100%

(b)

Commuting Time (minutes)	Frequency	Percentage	Cumulative Percent %
115 but less than 130	10	33%	33%
130 but less than 145	9	30%	63%
145 but less than 160	7	23%	87%
160 but less than 175	3	10%	97%
175 but less than 190	1	3%	100%
	30	100%	

(c) The majority of commuters living in or near cities spend from 115 up to 145 minutes commuting each week. 97% of commuters spend from 115 up to 175 minutes commuting each week.

2.18 (a), (b)

Credit Score	Frequency	Percent (%)	Cumulative Percent (%)
560 – under 580	4	0.16	0.16
580 – under 600	24	0.93	1.09
600 – under 620	68	2.65	3.74
620 – under 640	290	11.28	15.02
640 – under 660	548	21.32	36.34
660 – under 680	560	21.79	58.13
680 – under 700	507	19.73	77.86
700 – under 720	378	14.71	92.57
720 – under 740	168	6.54	99.11
740 – under 760	22	0.86	99.96
760 – under 780	1	0.04	100.00

(c) The average credit scores are concentrated between 620 and 720.

2.19 (a), (b)

Bin	Frequency	Percentage	Cumulative %
–0.00350 but less than –0.00201	13	13.00%	13.00%
–0.00200 but less than –0.00051	26	26.00%	39.00%
–0.00050 but less than 0.00099	32	32.00%	71.00%
0.00100 but less than 0.00249	20	20.00%	91.00%
0.00250 but less than 0.00399	8	8.00%	99.00%
0.004 but less than 0.00549	1	1.00%	100.00%

(c) Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

2.20 (a), (b)

Time in Seconds	Frequency	Percent (%)
5 – under 10	8	16%
10 – under 15	15	30%
15 – under 20	18	36%
20 – under 25	6	12%
25 – under 30	3	6%

(b)

Time in Seconds	Percentage Less Than
5	0
10	16
15	46
20	82
25	94
30	100

2.20 cont. (c) The target is being met since 82% of the calls are being answered in less than 20 seconds

2.21 (a)

Call Duration (seconds)	Frequency	Percentage
60 up to 119	7	14%
120 up to 179	12	24%
180 up to 239	11	22%
240 up to 299	11	22%
300 up to 359	4	8%
360 up to 419	3	6%
420 and longer	2	4%
	50	100%

(b)

Call Duration (seconds)	Frequency	Percentage	Cumulative %
60 up to 119	7	14%	14%
120 up to 179	12	24%	38%
180 up to 239	11	22%	60%
240 up to 299	11	22%	82%
300 up to 359	4	8%	90%
360 up to 419	3	6%	96%
420 and longer	2	4%	100%
	50	100%	

(c) The call center's target of call duration less than 240 seconds is only met for 60% of the calls in this data set.

2.22 (a)

Bulb Life			Frequency Mftr A	Percentage Mftr A	Frequency Mftr B	Percentage Mftr B
46,500	but less than	47,500	3	7.5%	0	0.0%
47,500	but less than	48,500	5	12.5%	2	5.0%
48,500	but less than	49,500	20	50.0%	8	20.0%
49,500	but less than	50,500	9	22.5%	16	40.0%
50,500	but less than	51,500	3	7.5%	9	22.5%
51,500	but less than	52,500	0	0.0%	5	12.5%
			40	100.0%	40	100.0%

(b)

% less than	Percentage Less than, Mftr A	Percentage Less than, Mftr B
47,500	7.5%	0.0%
48,500	20.0%	5.0%
49,500	70.0%	25.0%
50,500	92.5%	65.0%
51,500	100.0%	87.5%
52,500	100.0%	100.0%

2.22 cont. (c) Manufacturer B produces bulbs with longer lives than Manufacturer A. The cumulative percentage for Manufacturer B shows 65% of its bulbs lasted less than 50,500 hours, contrasted with 70% of Manufacturer A's bulbs, which lasted less than 49,500 hours. None of Manufacturer A's bulbs lasted more than 51,500 hours, but 12.5% of Manufacturer B's bulbs lasted between 51,500 and 52,500 hours. At the same time, 7.5% of Manufacturer A's bulbs lasted less than 47,500 hours, whereas all of Manufacturer B's bulbs lasted at least 47,500 hours

2.23 (a)

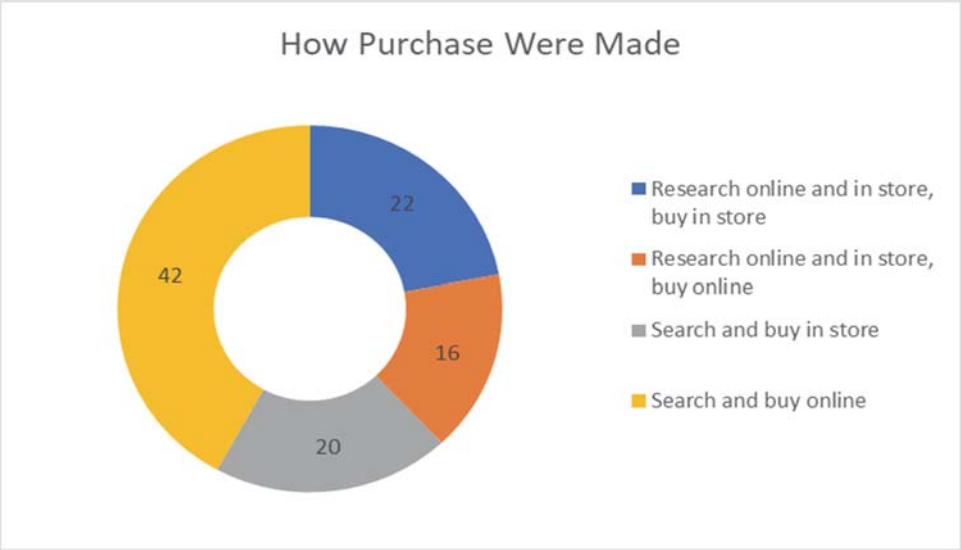
Amount of Soft Drink	Frequency	Percentage
1.850 – 1.899	1	2%
1.900 – 1.949	5	10%
1.950 – 1.999	18	36%
2.000 – 2.049	19	38%
2.050 – 2.099	6	12%
2.100 – 2.149	1	2%

Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12%
1.999	24	48%
2.049	43	86%
2.099	49	98%
2.149	50	100%

(b) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.24 (a)

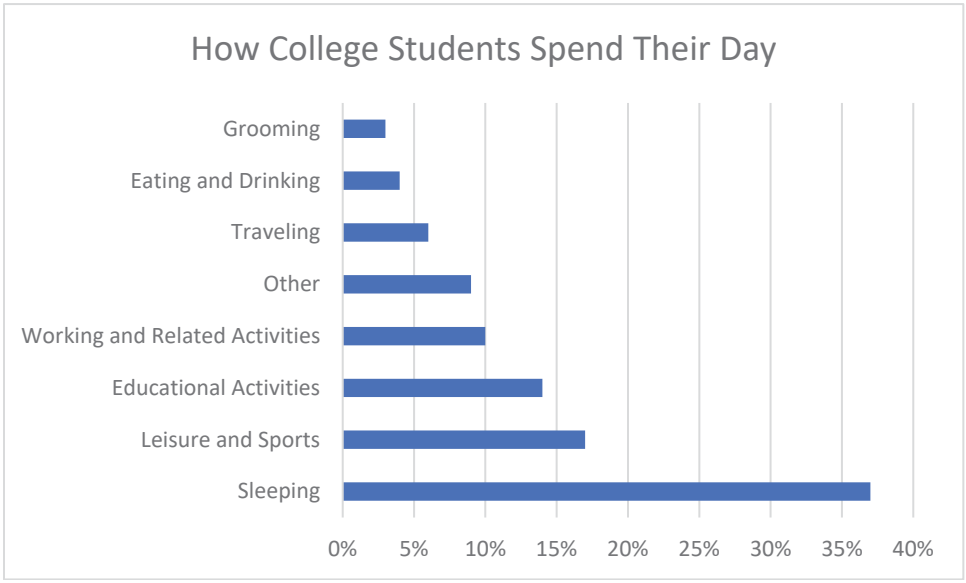
2.24 cont. (a)

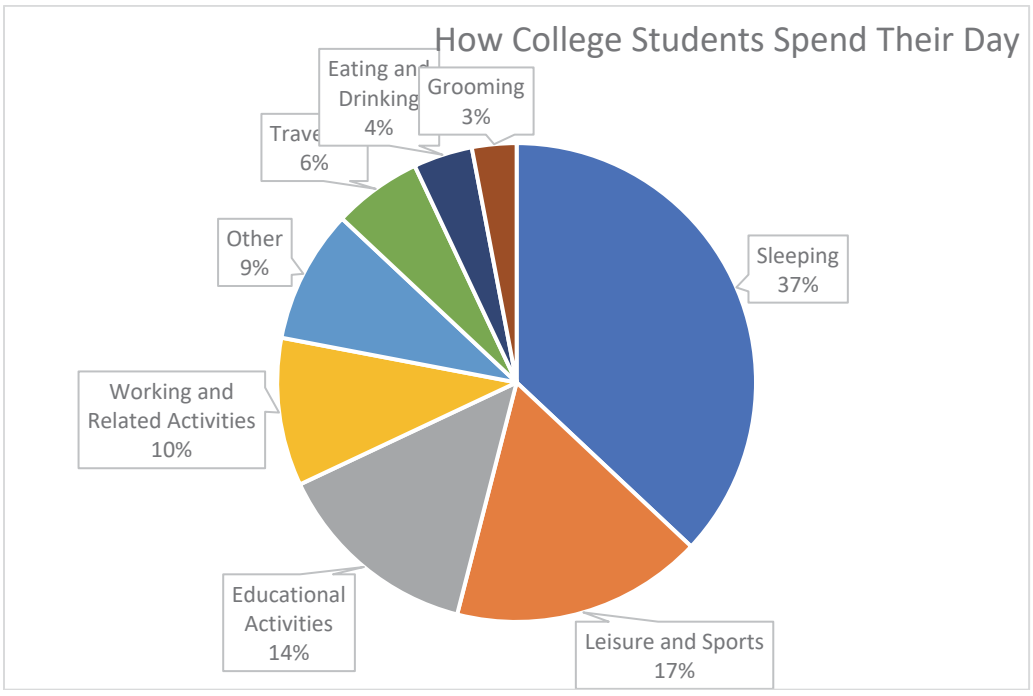


(b) The Pareto chart is best for portraying these data because it not only sorts the frequencies in descending order but also provides the cumulative line on the same chart.

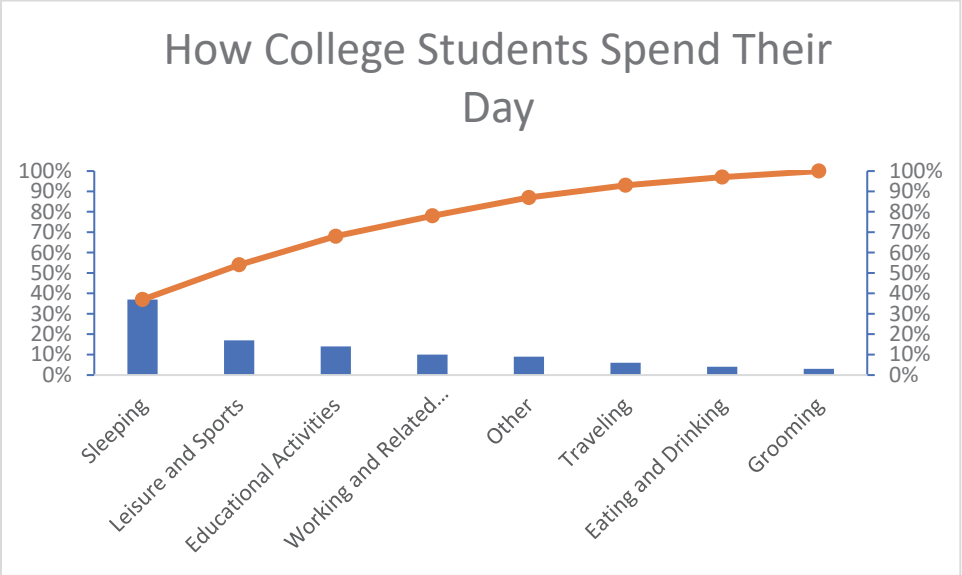
(c) You can conclude that searching and buying online was the highest category and the other three were equally likely.

2.25 (a)





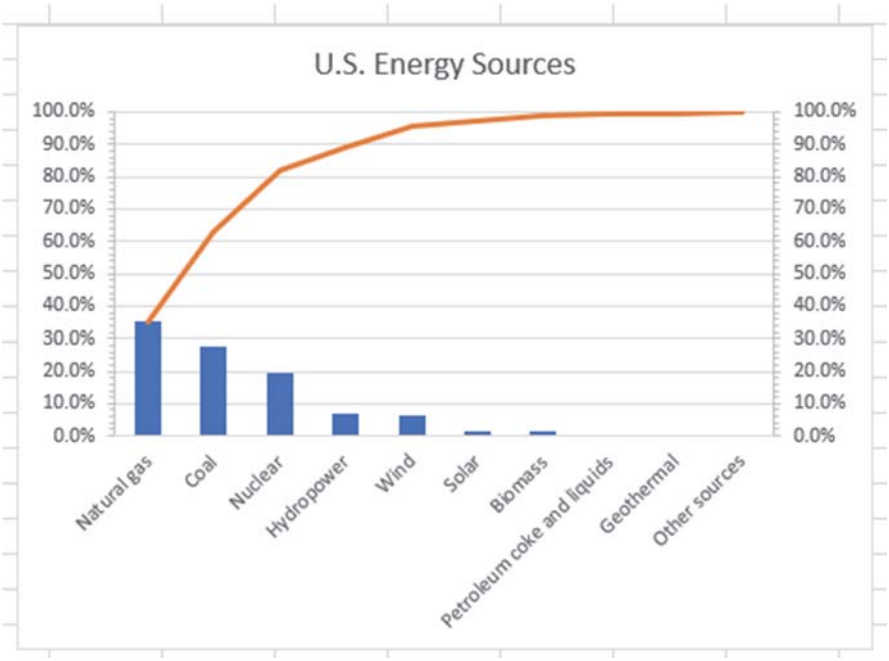
2.25 cont. (a)



(b) The Pareto diagram is better than the pie chart or the bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

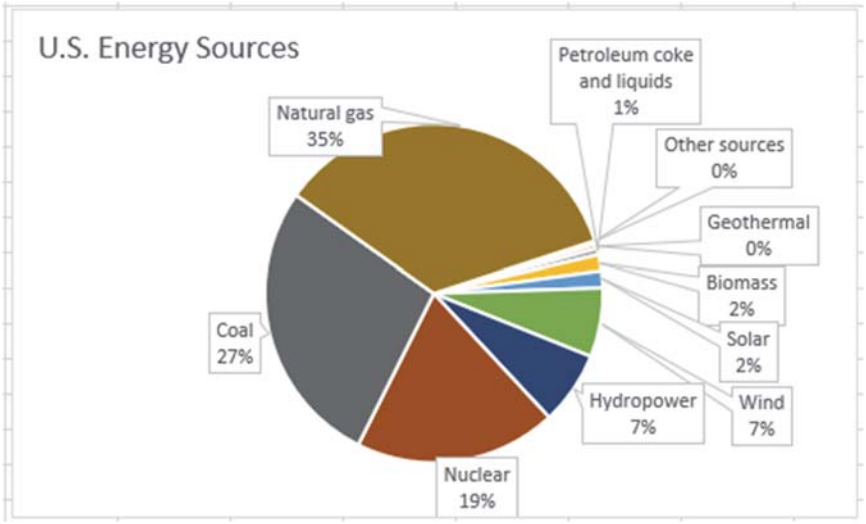
(c) From the Pareto diagram it is obvious that more than 50% of their day is spent sleeping and taking part in leisure and sports.

2.26 (a)



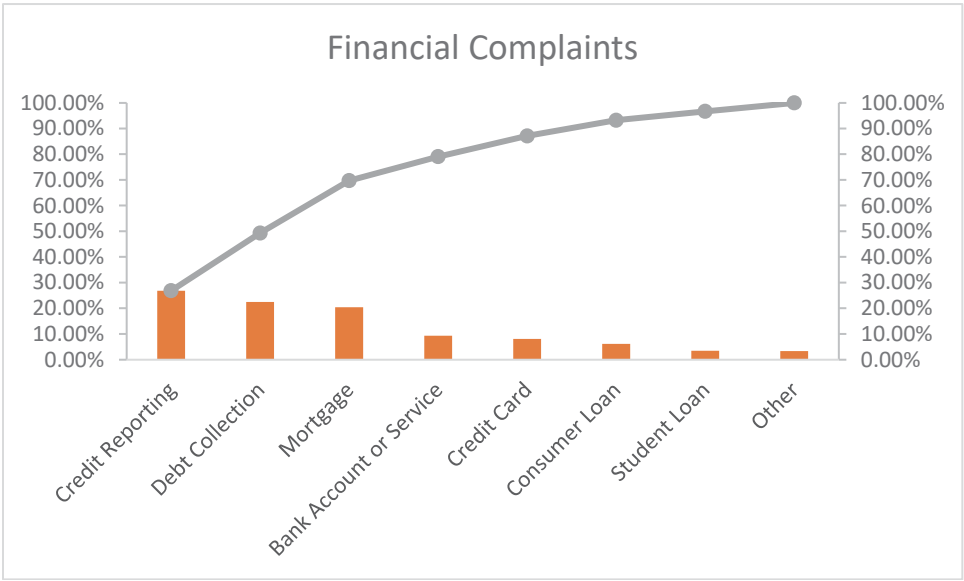
(b) 27.4% + 19.3% + 35.2% = 81.9%

2.26 cont. (c)



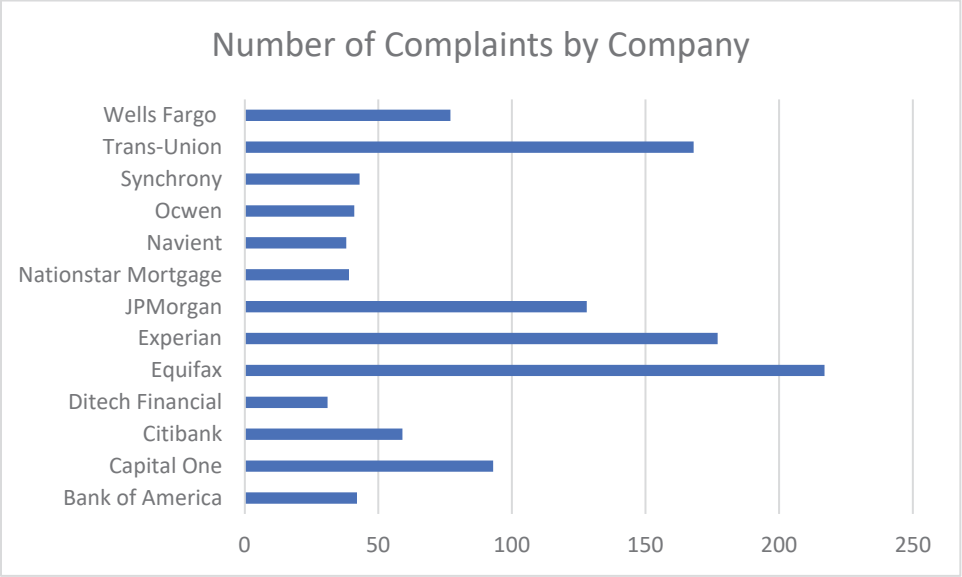
(d) The Pareto diagram is better than the pie chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

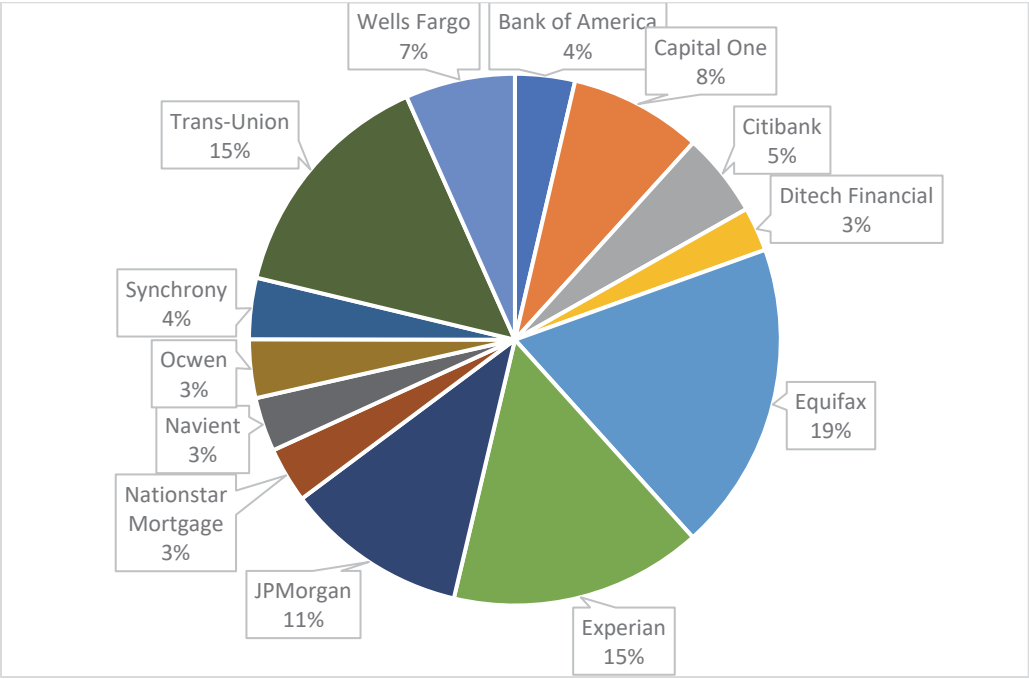
2.27 (a)



(b) The "vital few" reasons for the categories of complaints are "Credit Reporting", "Debt Collection", and "Mortgage" which account for 70% of the complaints. The remaining are the "trivial many" which make up 30% of the complaints.

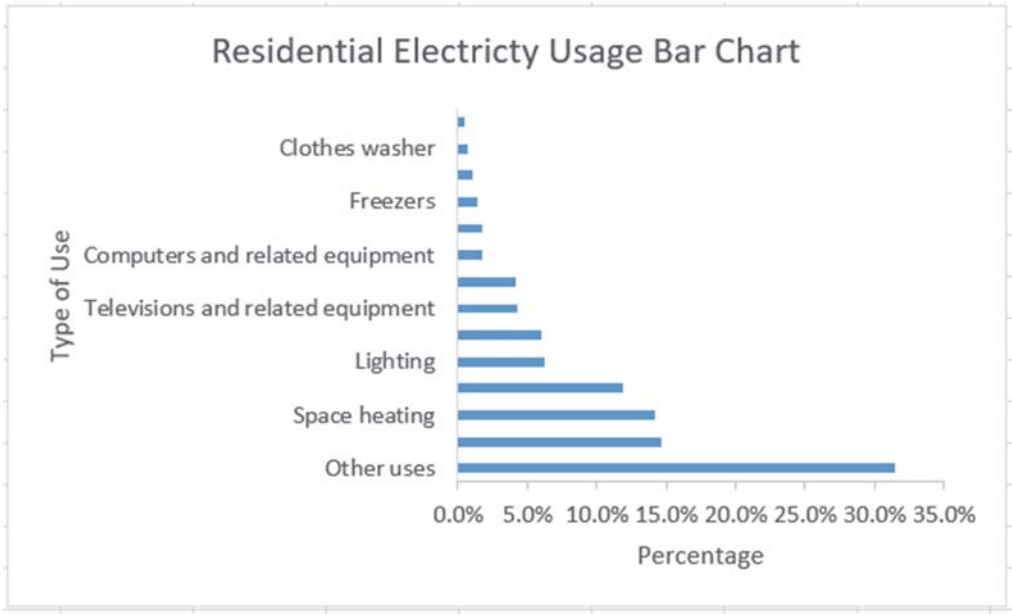
2.27 cont. (c)

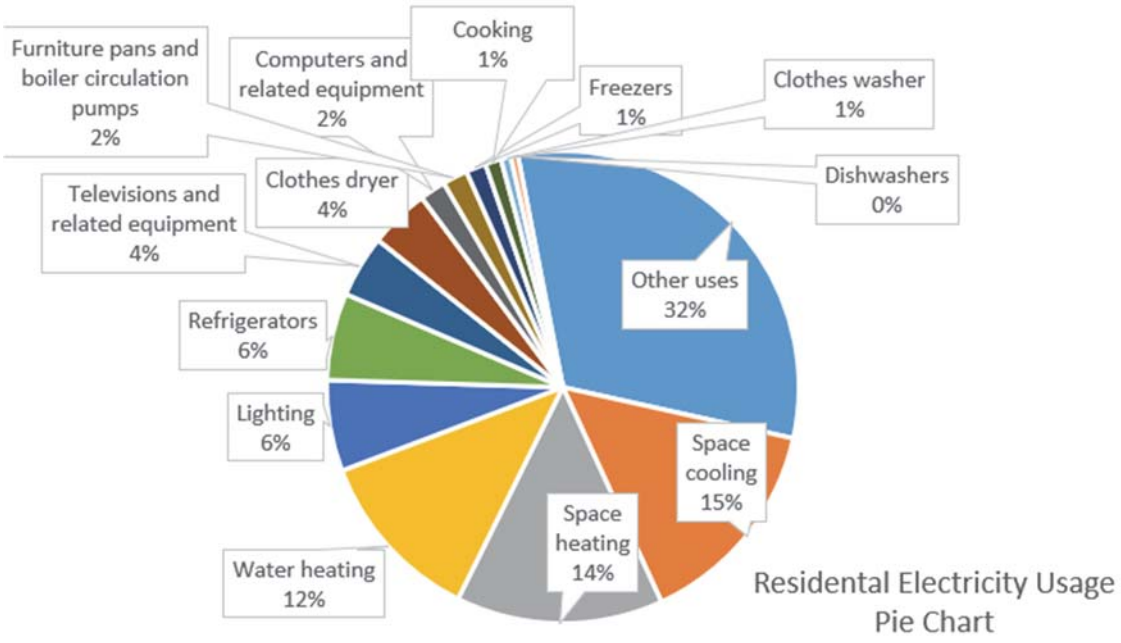




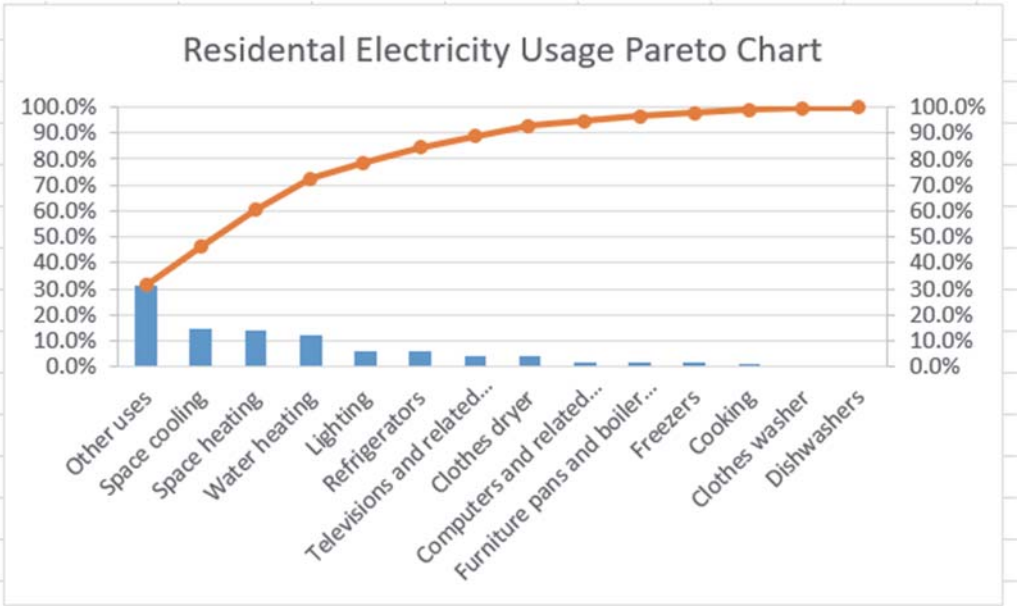
(d) The Pareto diagram is better than the pie chart and bar chart because it allows you to see which companies account for most of the complaints.

2.28 (a)





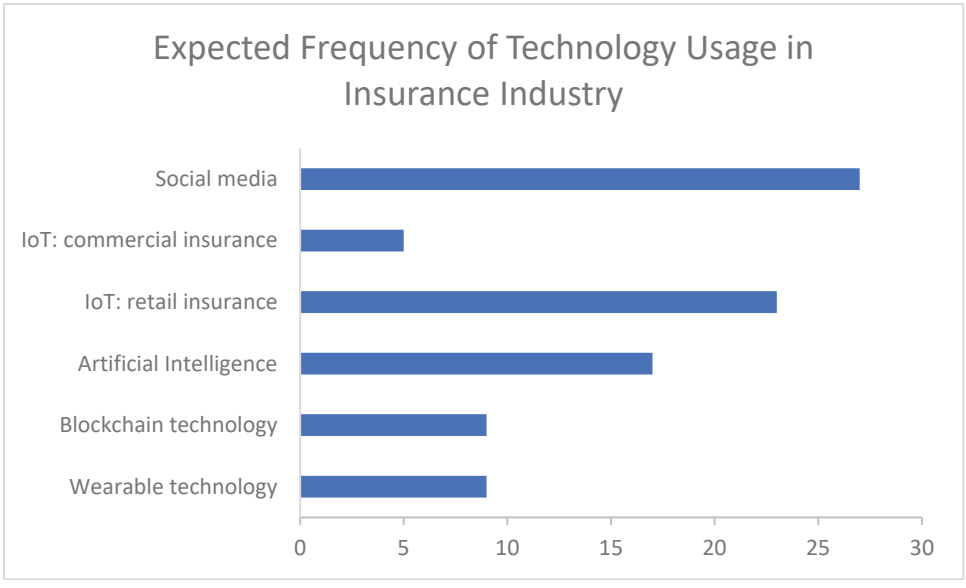
2.28 (a) cont.



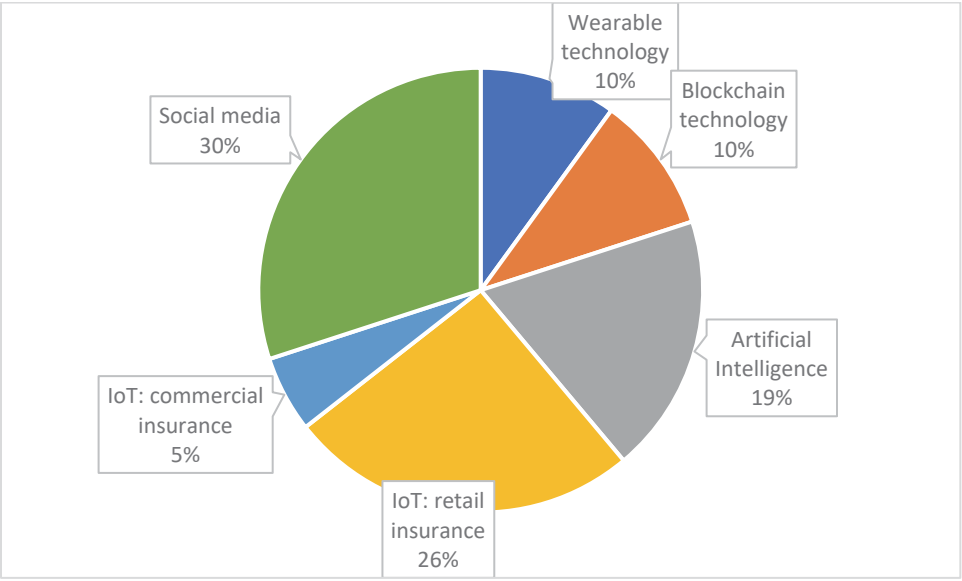
(b) The Pareto diagram is better than the pie chart and bar chart because it not only sorts the frequencies in descending order; it also provides the cumulative polygon on the same scale.

(c) Other, cooling, heating and lighting accounted for 78.5% of the residential electricity consumption in the United States.

2.29 (a)

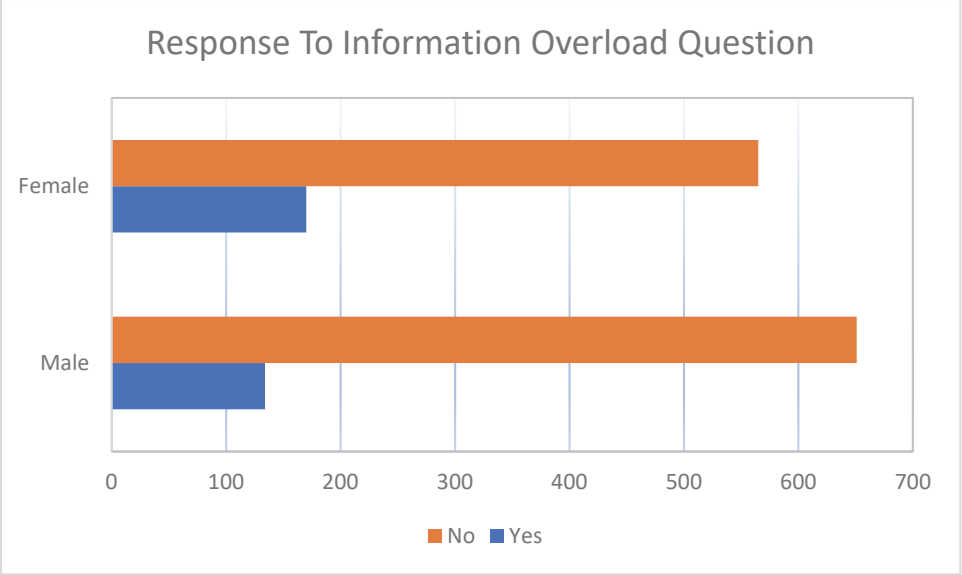


2.29 cont. (a)



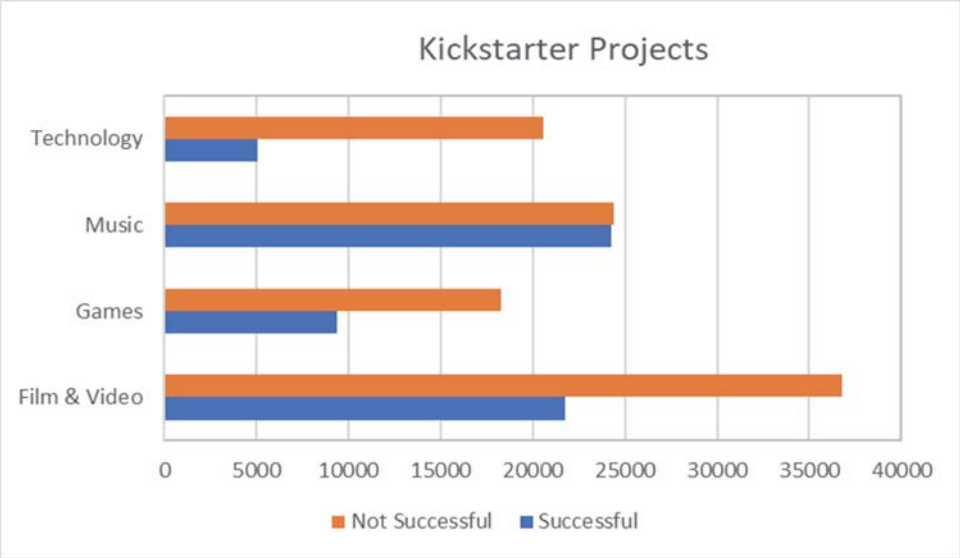
(b) Insurance professionals expect Social media, AI, and IOT retail insurance to be most used in the insurance industry in the coming year.

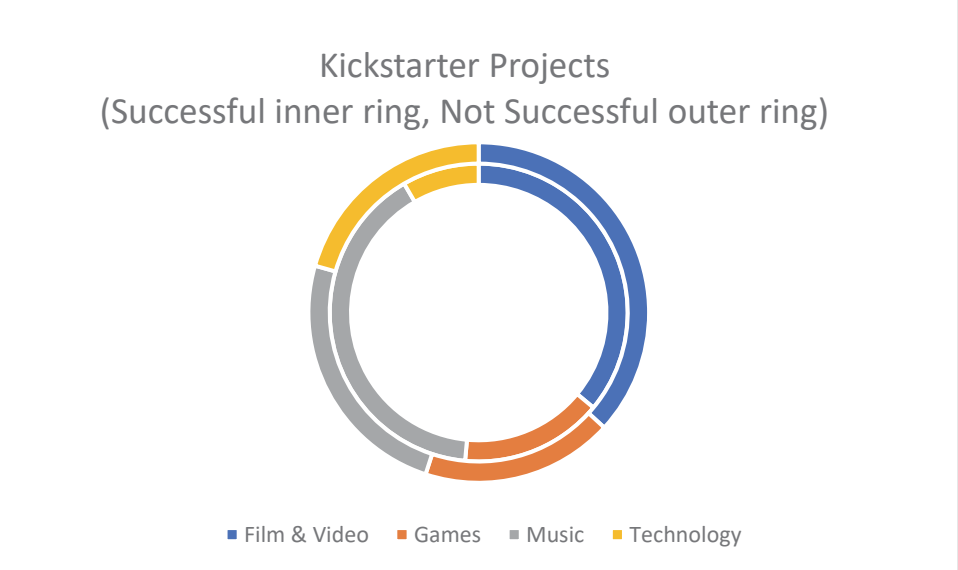
2.30 (a)



(b) Females are more likely to be overloaded with information.

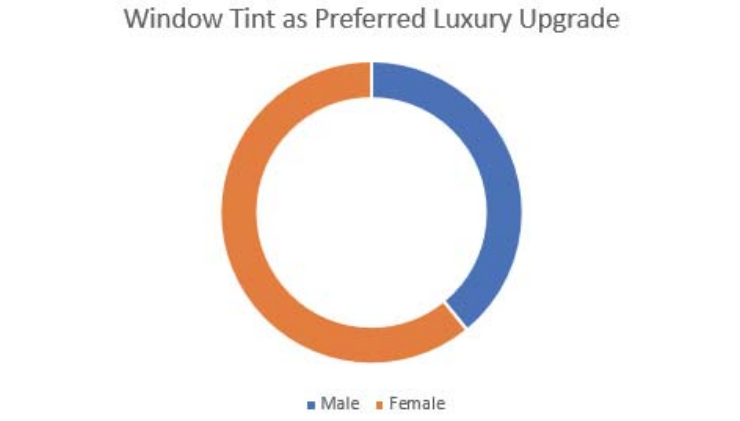
2.31 (a)

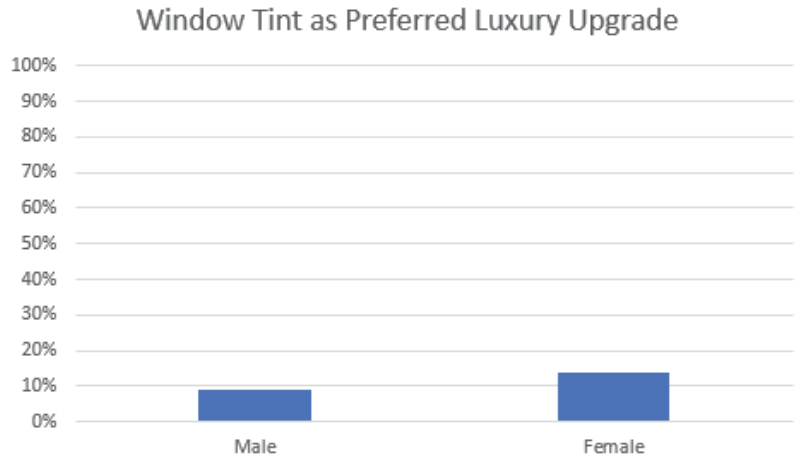




(b) Of the successful kickstarter projects, music projects make up the largest part.

2.32 (a)





(b) More females then men prefer window tint as a luxury upgrade.

2.33

Stem-and-leaf of Finance Scores	
5	34
6	9
7	4
9	38

2.34 Ordered array: 50 74 74 76 81 89 92

2.35 (a) Ordered array: 9.1 9.4 9.7 10.0 10.2 10.2 10.3 10.8 11.1 11.2
11.5 11.5 11.6 11.6 11.7 11.7 11.7 12.2 12.2 12.3
12.4 12.8 12.9 13.0 13.2

(b) The stem-and-leaf display conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display.

(c) The most likely gasoline purchase is between 11 and 11.7 gallons.

2.35 cont. (d) Yes, the third row is the most frequently occurring stem in the display and it is located in the center of the distribution.

2.36 (a)

stem	leaf
4	0
5	0056889
6	0133457
7	3388899
8	048
9	
10	25
11	
12	8
13	3
14	3

(b) The results are concentrated between 5 hours and 8 hours.

2.37 (a)

Download Speed	6.5	29.4	31.1	32.5	32.8	36.3	37.1	53.3
Upload Speed	3.7	4.0	5.8	12.9	13.0	15.6	16.9	17.5

(b) Download Speeds: Stem unit :10 Leaf rounded to nearest integer

Stem	Leaf
0	7
1	
2	9
3	13367
4	
5	3

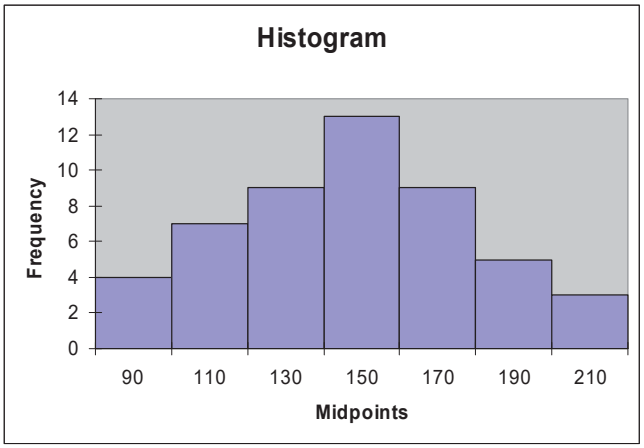
Upload Speeds: Stem unit 1

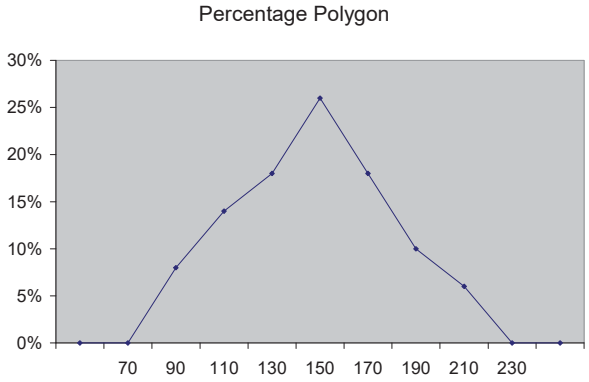
Stem	Leaf
3	7
4	0
5	8
6	
7	
8	
9	
10	
11	
12	9
13	0
14	
15	6
16	9
17	5

2.37 cont. (b) The stem-and-leaf display conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display.

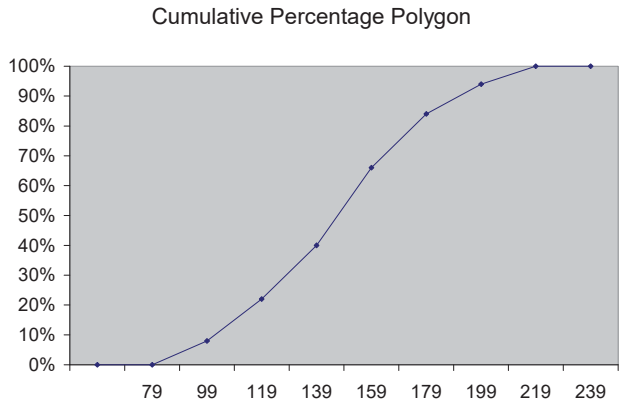
(c) Download speeds are concentrated around 30 mbps and Upload speeds are varied with a group around 3 to 5 Mbps and a group around 13 to 17 Mbps.

2.38 (a)





(b)

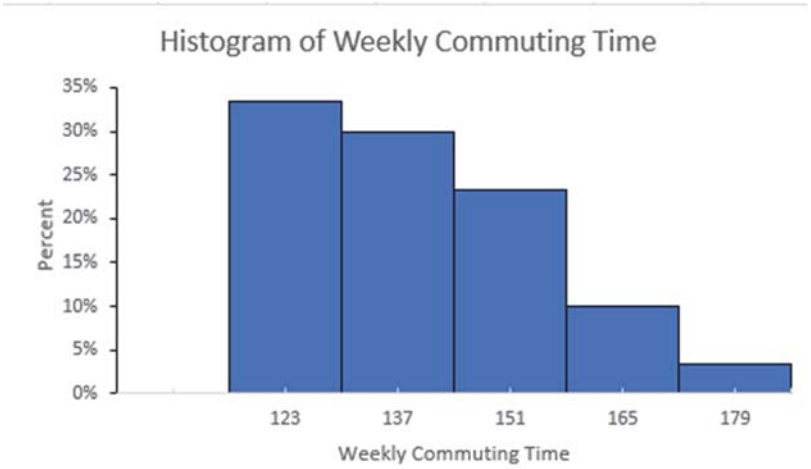


(c) The majority of utility charges are clustered between \$120 and \$180.

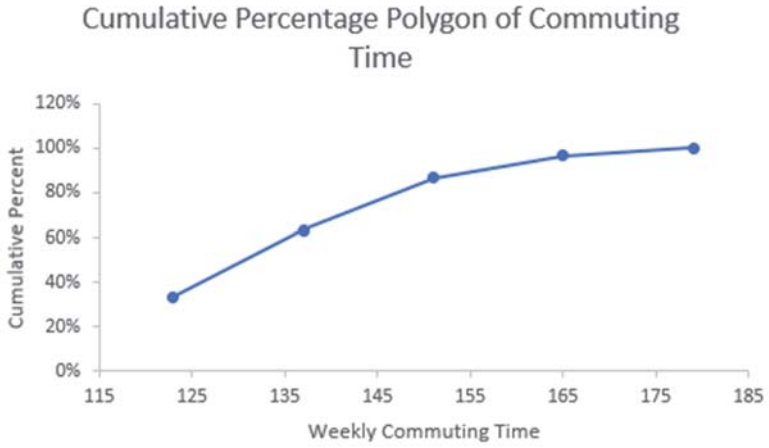
2.39 The cost of attending a baseball game is concentrated around $65 with twelve teams at that cost. Five teams have costs of $85 and one team is has the highest cost of $115.

2.40 Property taxes on a $176K home seem concentrated between $700 and $2,200 and also between $3,200 and $3,700.

2.41 (a)

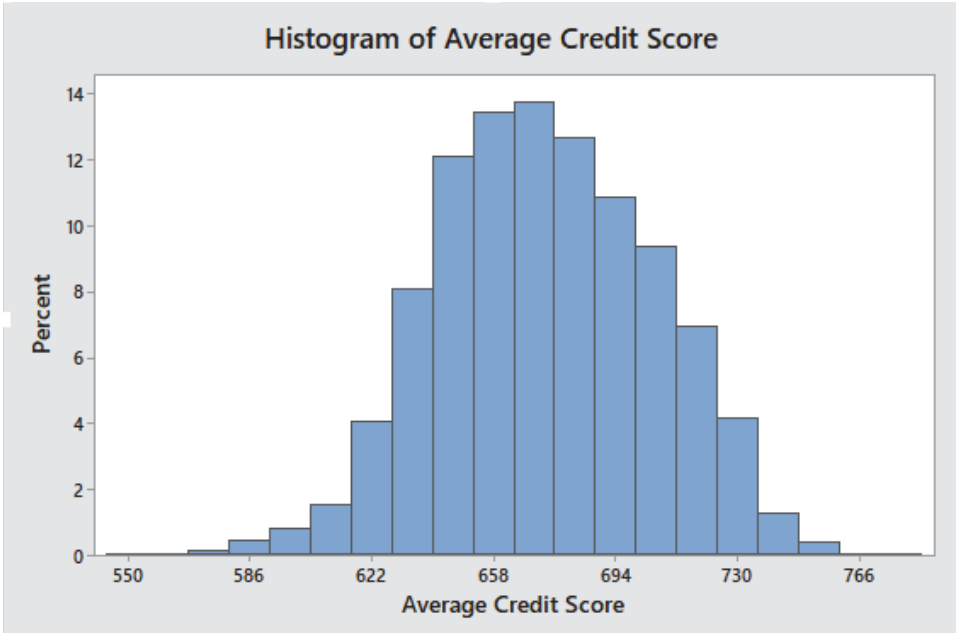


(b)

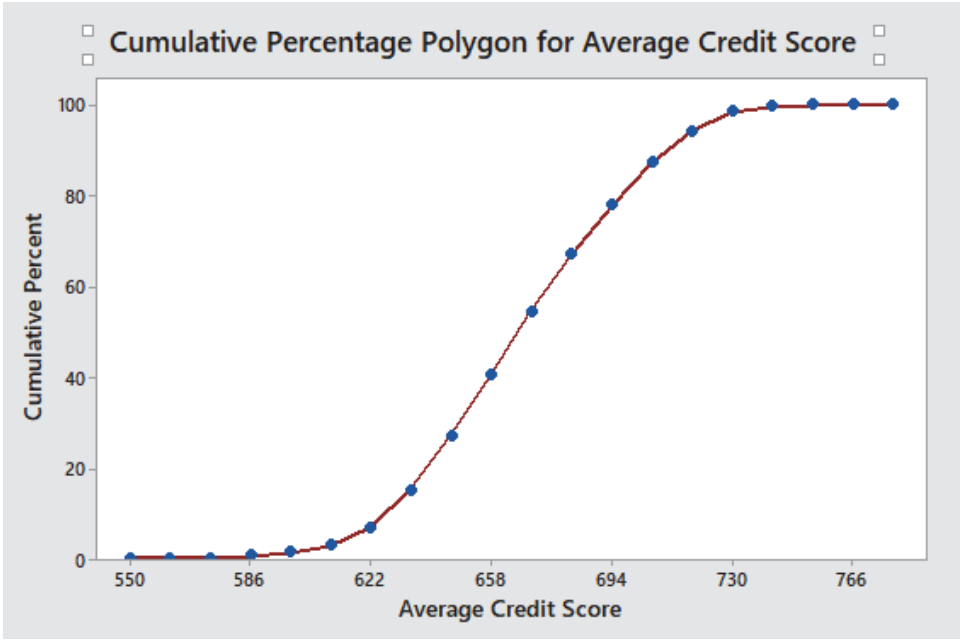


(c) The majority of commuters living in or near cities spend from 115 but less than 145 minutes commuting each week. 87% of commuters spend from 115 but less than 160 minutes commuting each week.

2.42 (a)

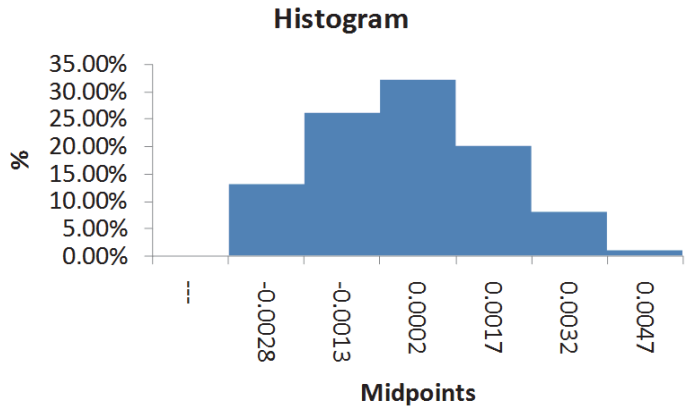


(b)



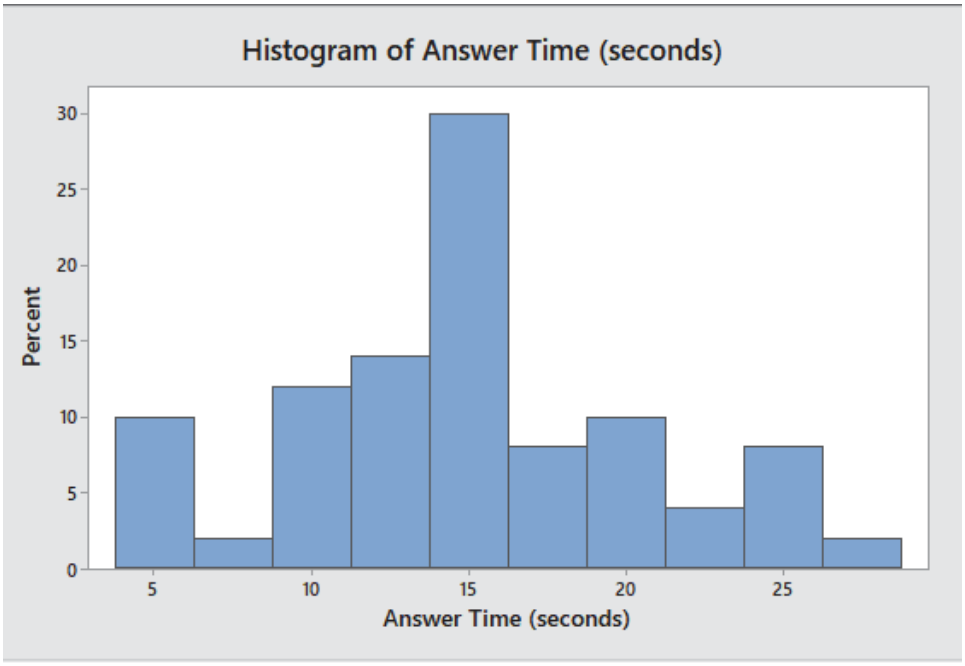
(c) The average credit scores are concentrated between 622 and 730.

2.43 (a)

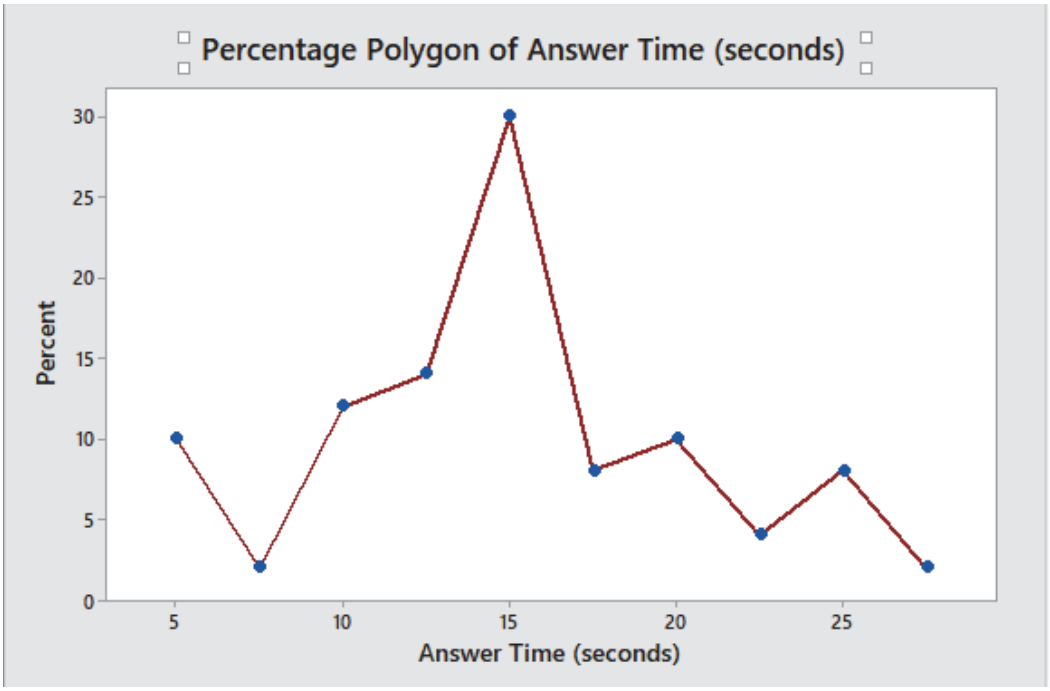


(b) Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

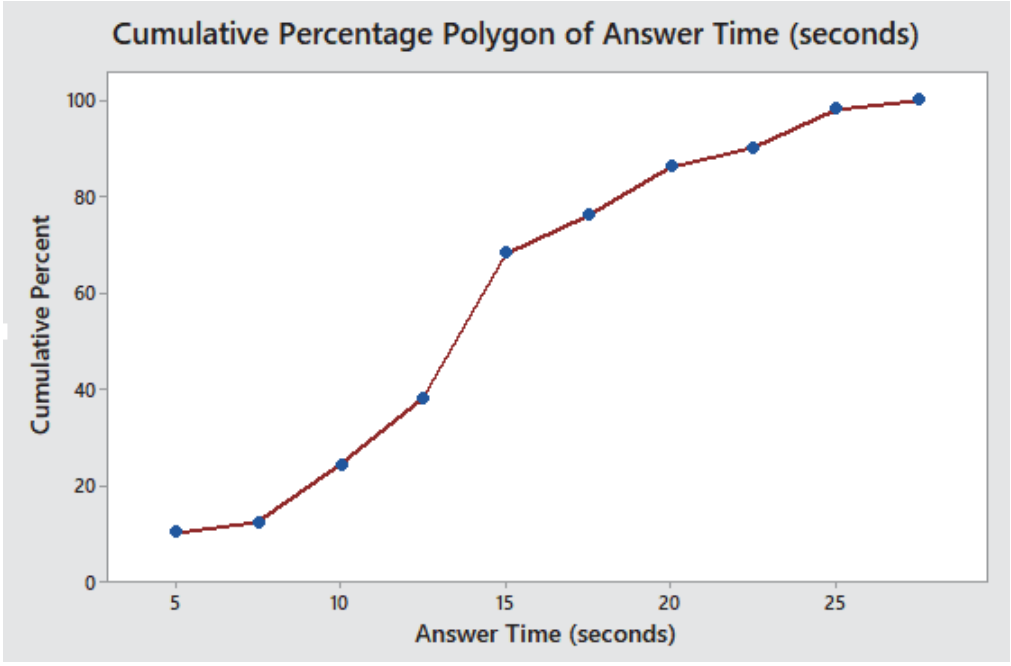
2.44 (a)



2.44 cont. (a)

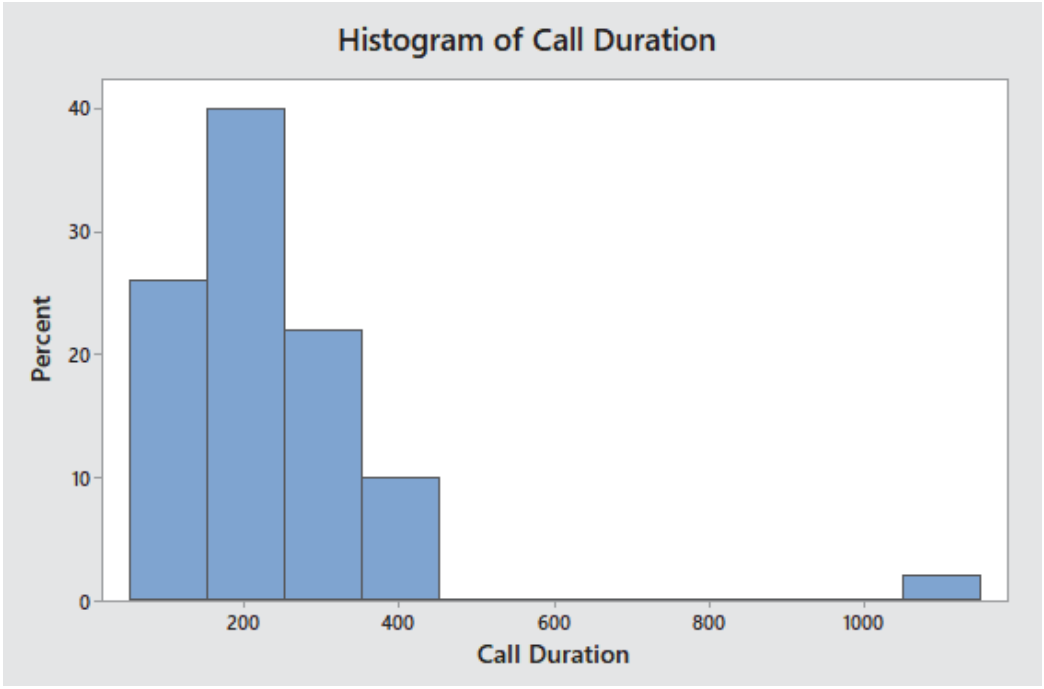


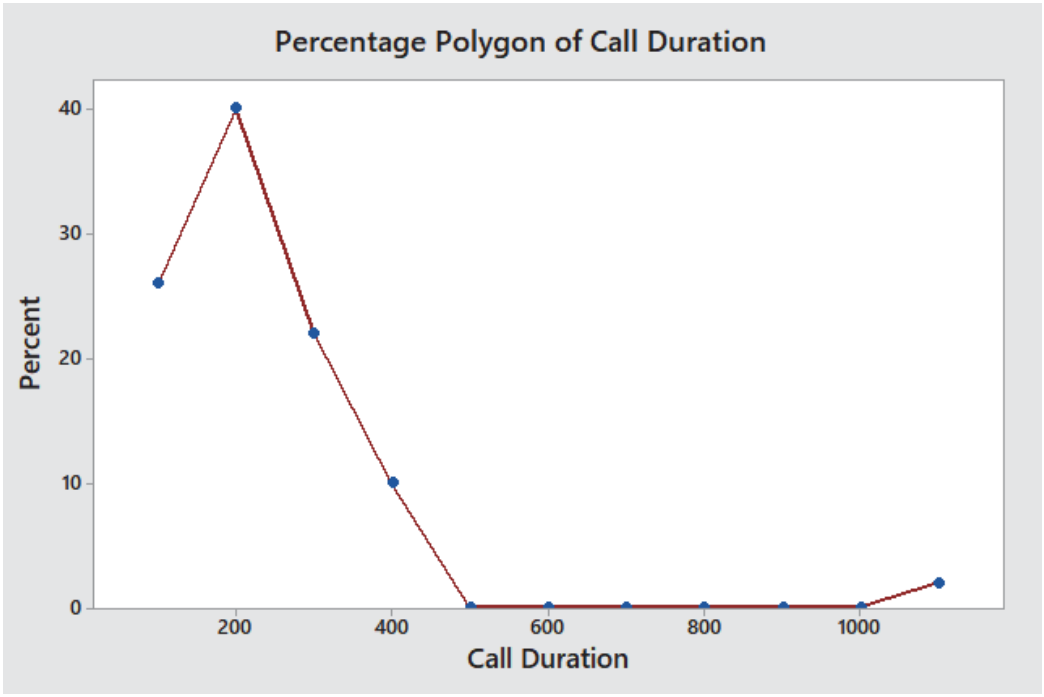
(b)



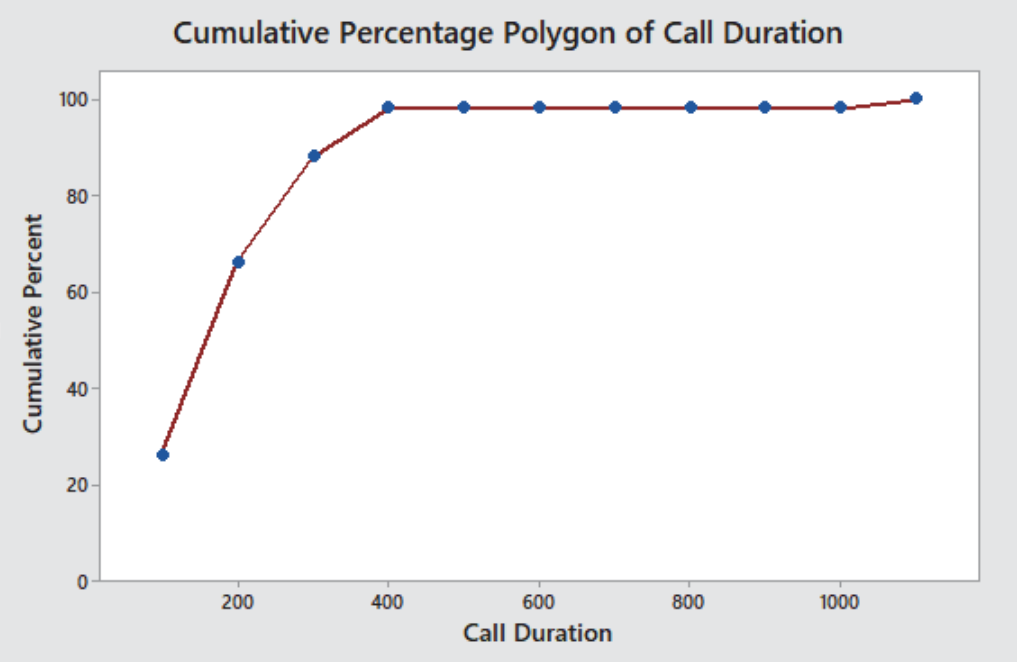
(c) The target is being met since 82% of the calls are being answered in less than 20 seconds.

2.45 (a)



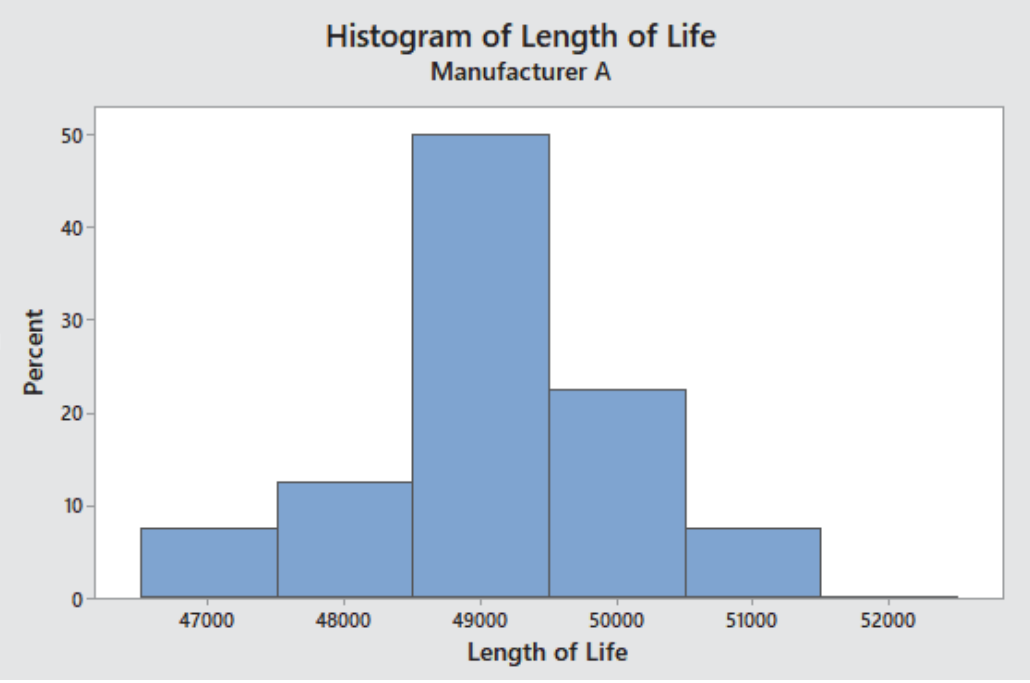


2.45 (b) cont.

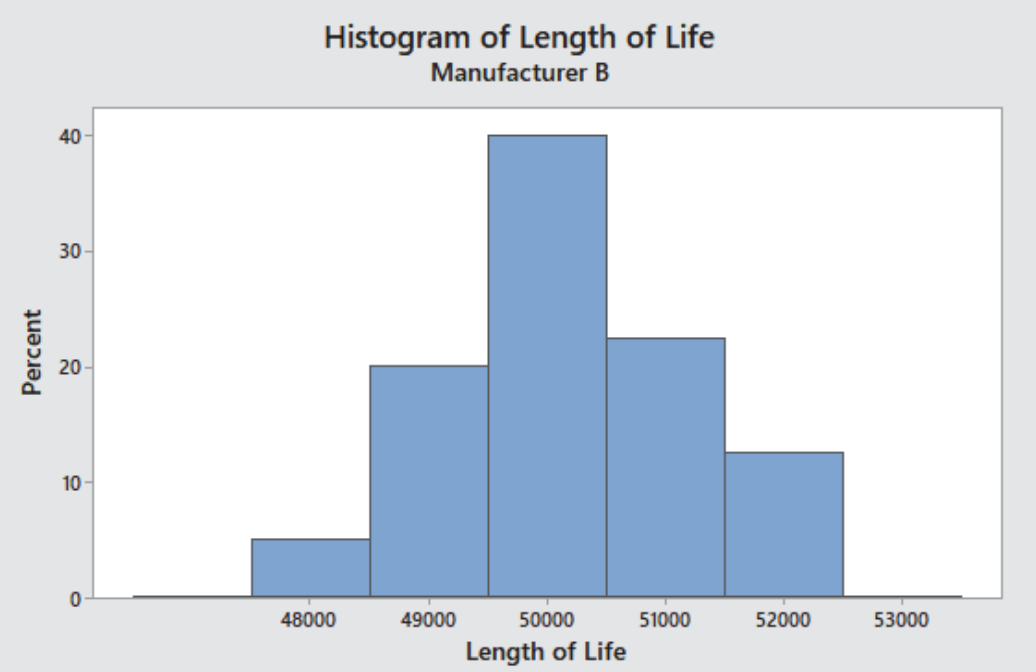


(c) The call center's target of call duration less than 240 seconds is only met for 60% of the calls in this data set.

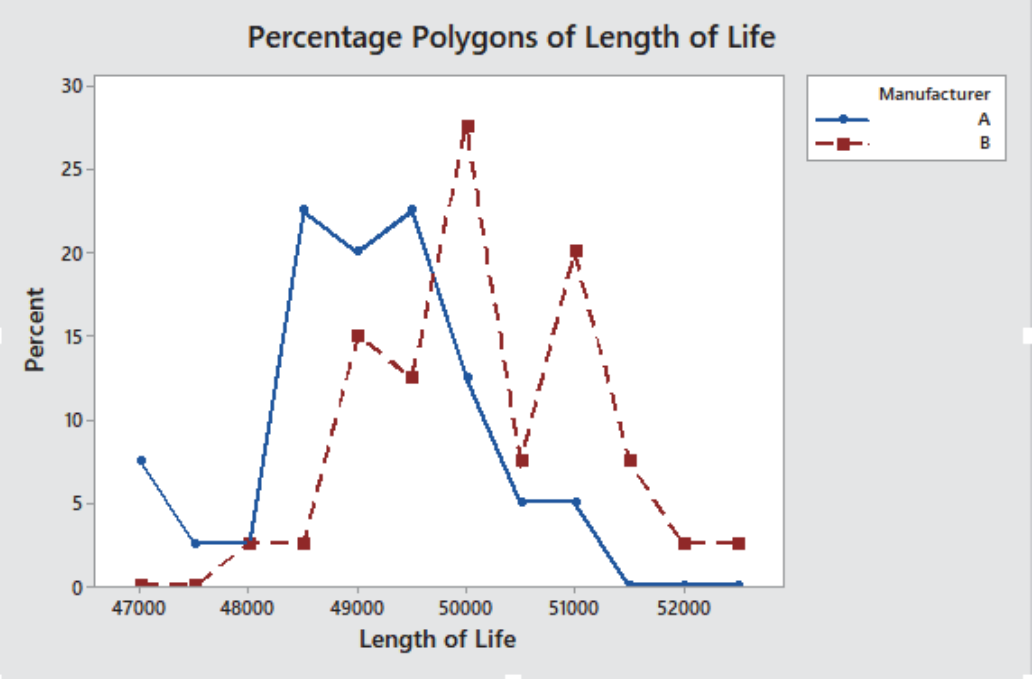
2.46 (a)



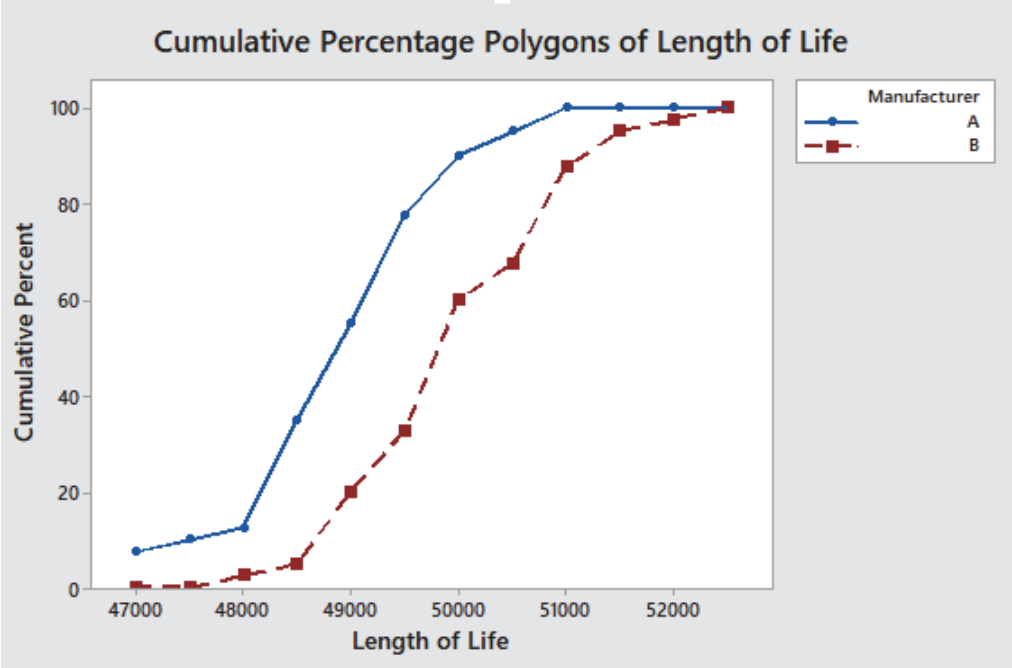
2.46 cont. (a)



(b)

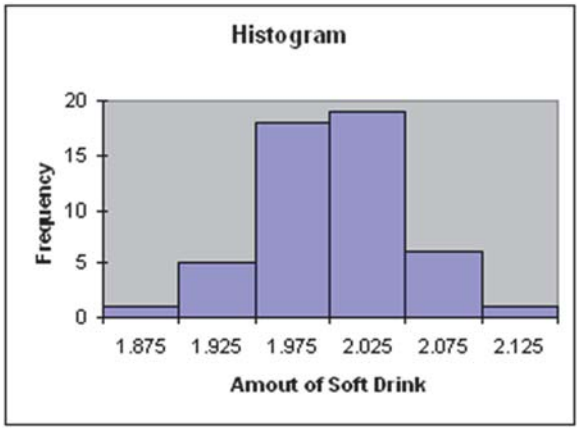


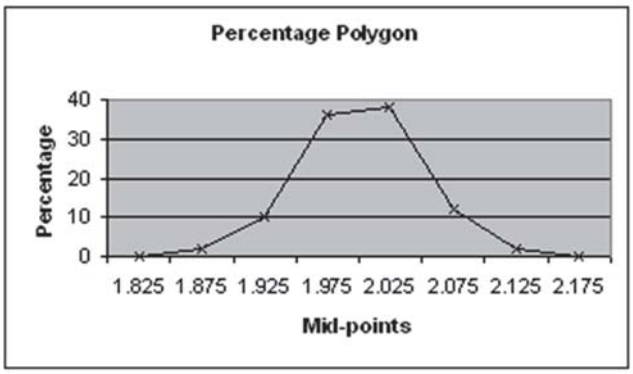
2.46 (b) cont.



(c) Manufacturer B produces bulbs with longer lives than Manufacturer A

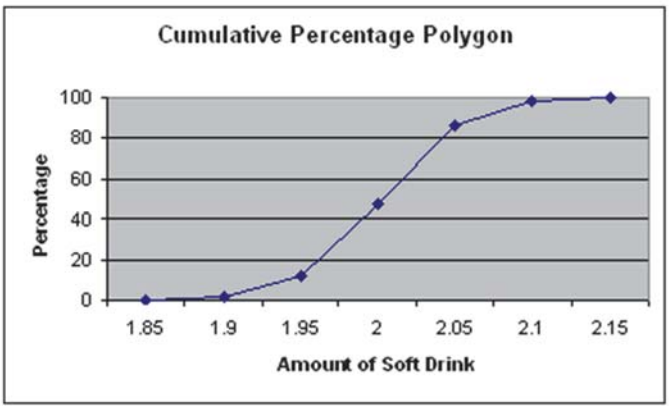
2.47 (a)





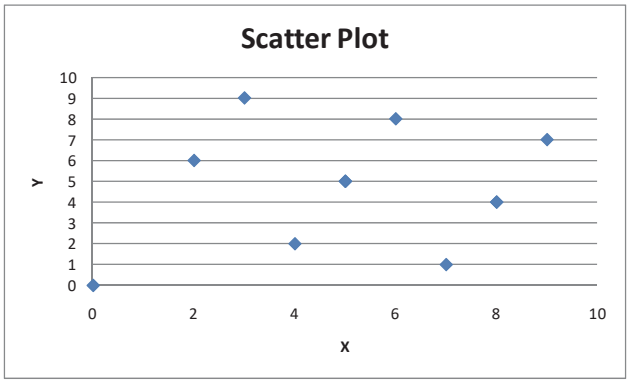
2.47 (b) cont.

Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12
1.999	24	48
2.049	43	86
2.099	49	98
2.149	50	100



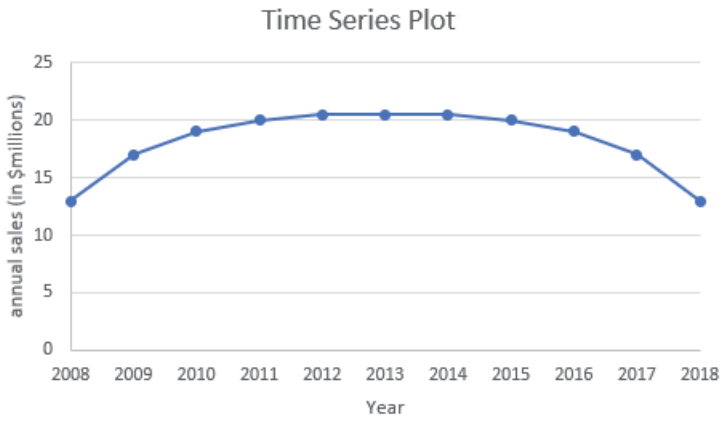
(c) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.48 (a)



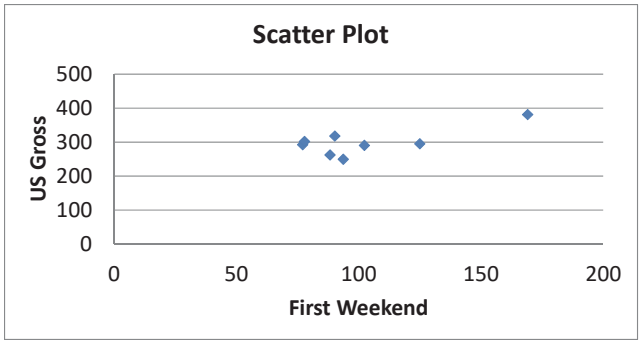
(b) There is no relationship between X and Y.

2.49 (a)

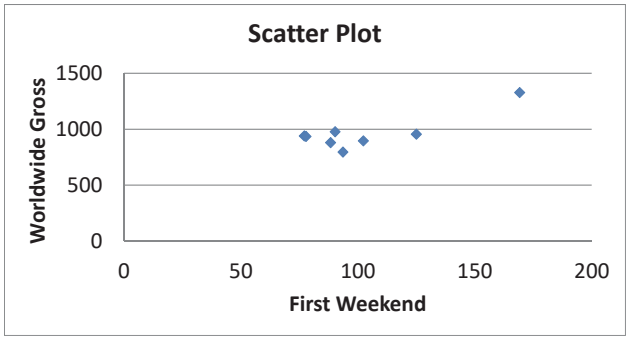


(b) Annual sales appear to be increasing in the earlier years before 2011, remain flat from 2012 to 2014 and then start to decline after 2014.

2.50 (a)

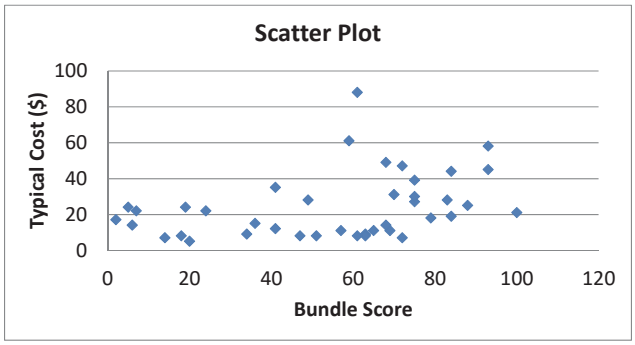


(b)



(c) There appears to be a linear relationship between the first weekend gross and either the U.S. gross or the worldwide gross of Harry Potter movies. However, this relationship is greatly affected by the results of the last movie, *Deathly Hallows*, *Part II*.

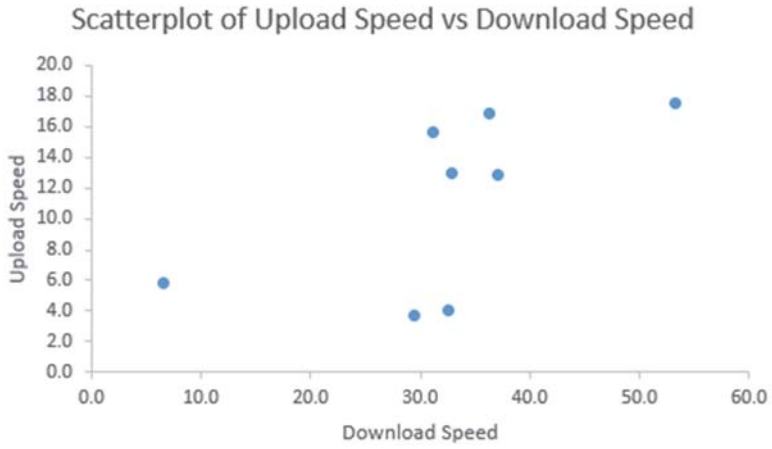
2.51 (a)



(b) There appears to be a positive relationship between Bundle score and typical cost.

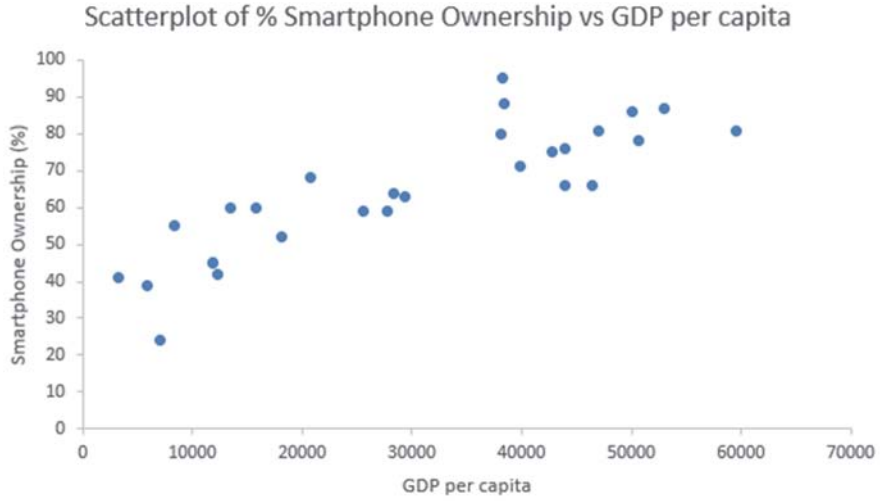
2.52 (a) There appears to be a positive relationship between the download speed and the upload speed.

(b)



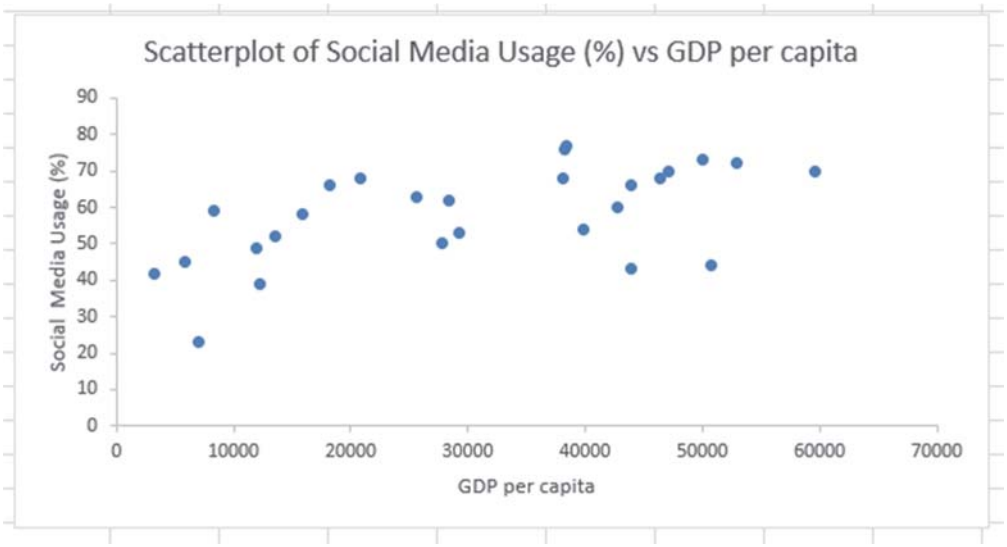
(c) Yes, this is borne out by the data

2.53 (a)



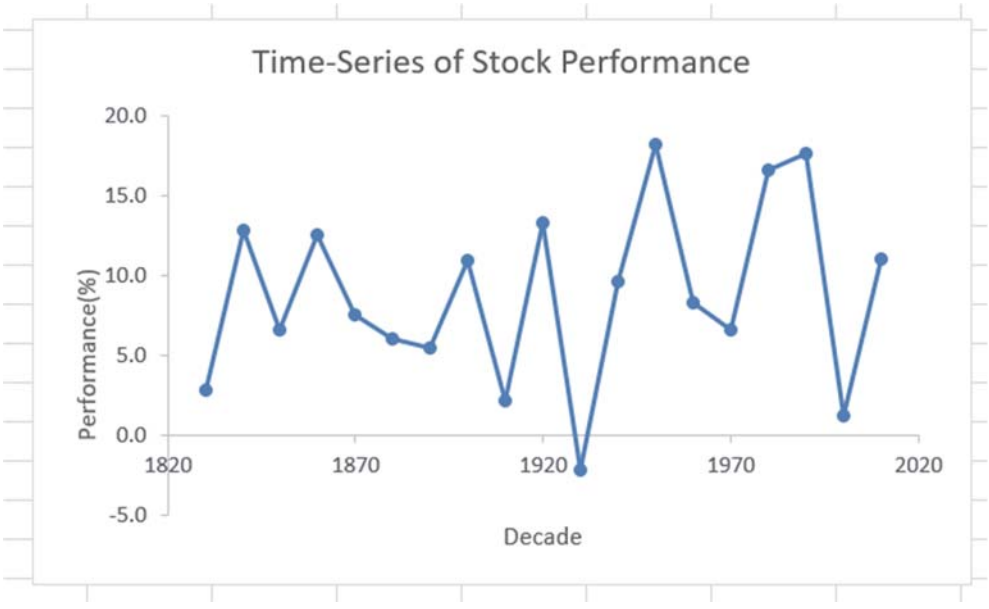
2.53 cont. (b) There is a positive relationship between GDP and percentage of smartphone ownership.

(c)



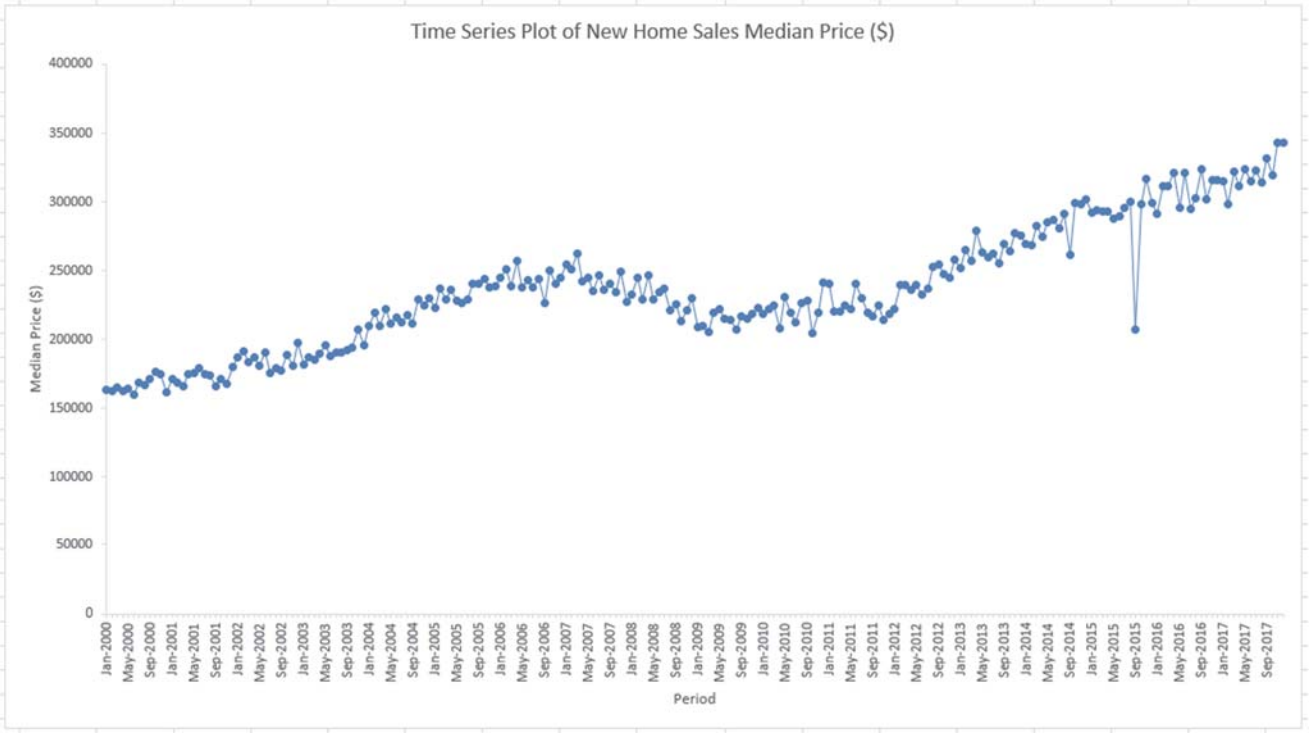
(d) There does appear to be some relationship between GDP and percentage social media usage.

2.54 (a) Excel output:



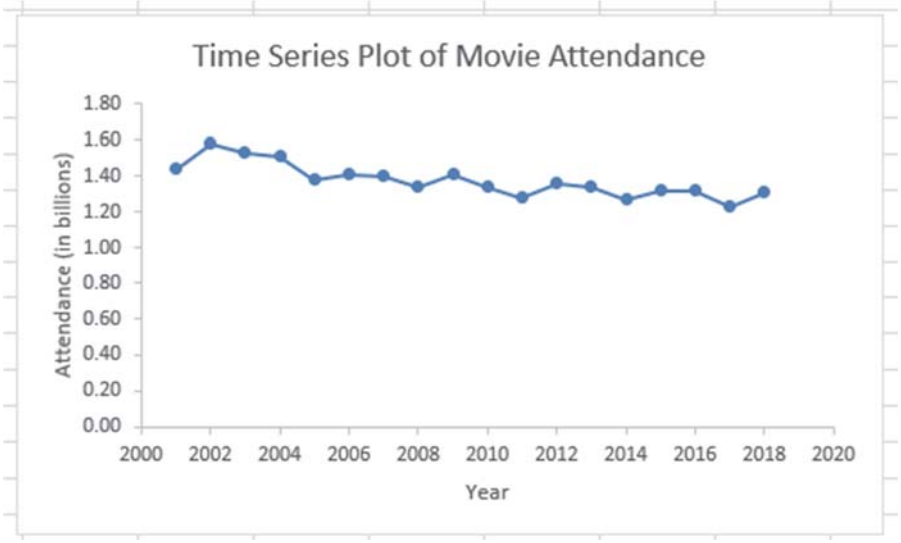
(b) There is a great deal of variation in the returns from decade to decade. Most of the returns are between 5% and 15%. The 1950s, 1980s, and 1990s had exceptionally high returns, and only the 1930s had negative returns.

2.55 (a)



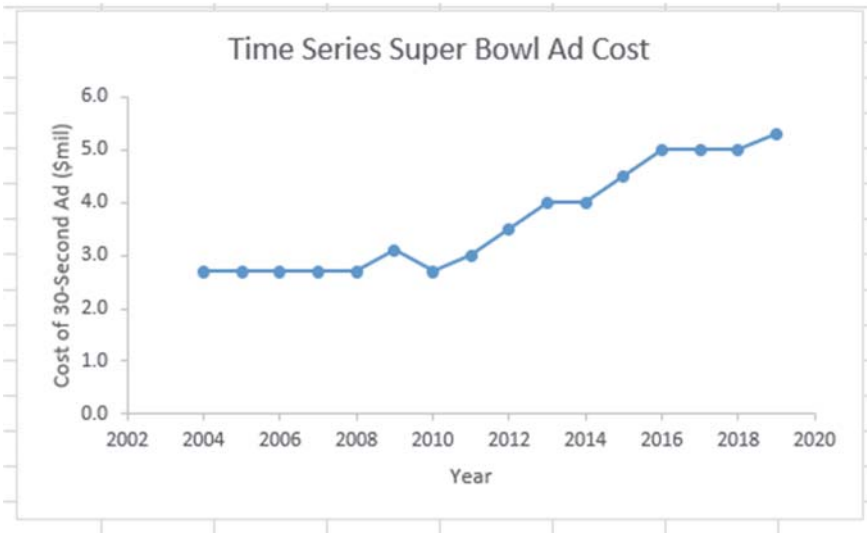
(b) There is an upward trend in home sales price until 2006. Prices decline or remain flat from 2006 – 2011. From 2011 – 2016 there is an upward trend in median price of new home sales. There is a huge drop in median prices in September 2015. This should be investigated further and may be just an error in the data file.

2.56 (a)



(b) There was a decline in movie attendance from 2002 to 2018. Movie attendance increased from 2001 to 2002 but then by 2018 decreased to a level below 2001.

2.57 (a)



(b) From 2004 to 2008 the cost of a 30-second ad was constant at 2.7 million dollars, Since 2010 the cost has increased to its highest level of 5.3 million dollars in 2019.

2.58 (a) Pivot Table in terms of %

Count of Type	Star Rating					
Type	**One**	**Two**	**Three**	**Four**	**Five**	**Grand Total**
Growth	**5.43%**	**17.12%**	**27.35%**	**11.27%**	**2.71%**	**63.88%**
Large	3.76%	7.72%	13.57%	5.43%	1.67%	32.15%
Mid-Cap	1.25%	5.43%	7.52%	3.13%	0.63%	17.96%
Small	0.42%	3.97%	6.26%	2.71%	0.42%	13.78%
Value	**2.92%**	**10.65%**	**13.99%**	**7.31%**	**1.25%**	**36.12%**
Large	2.09%	6.68%	9.19%	3.97%	1.25%	23.18%
Mid-Cap	0.63%	2.09%	2.71%	1.04%	0.00%	6.47%
Small	0.21%	1.88%	2.09%	2.30%	0.00%	6.48%
Grand Total	**8.35%**	**27.77%**	**41.34%**	**18.58%**	**3.97%**	**100.00%**

(b) The growth and value funds have similar patterns in terms of star rating and type. Both growth and value funds have more funds with a rating of three. Very few funds have ratings of five.

(c) Pivot Table in terms of Average Three-Year Return

Count of Type	Star Rating					
Type	**One**	**Two**	**Three**	**Four**	**Five**	**Grand Total**
Growth	**5.41**	**7.04**	**8.94**	**10.14**	**12.83**	**8.51**
Large	6.97	9.43	10.62	11.83	14.25	10.30
Mid-Cap	2.27	5.07	7.93	8.77	11.22	6.93
Small	0.78	5.09	6.52	8.35	9.53	6.39
Value	**4.43**	**5.49**	**7.29**	**8.34**	**10.23**	**6.84**
Large	5.23	6.05	7.58	8.85	10.23	7.29
Mid-Cap	2.79	5.77	7.32	9.26	–	6.69
Small	1.33	3.20	5.93	7.04	–	5.39
Grand Total	**5.07**	**6.45**	**8.38**	**9.43**	**12.01**	**7.91**

(d) There are 65 large cap growth funds with a rating of three. Their average three year return is 10.62.

2.59 (a) Pivot table of tallies in terms of counts:

Count of Star Rating	Column Labels					
Row Labels	Five	Four	Three	Two	One	Grand Total
Large	**14**	**45**	**109**	**69**	**28**	**265**
Low	8	24	50	25	9	116
Average	3	20	55	40	11	129
High	3	1	4	4	8	20
MidCap	**3**	**20**	**49**	**36**	**9**	**117**
Low	2	11	7	5	1	26
Average	1	9	34	15	3	62
High			8	16	5	29
Small	**2**	**24**	**40**	**28**	**3**	**97**
Low	1	3		1		5
Average		13	15	5		33
High	1	8	25	22	3	59
Grand Total	**19**	**89**	**198**	**133**	**40**	**479**

Pivot table of tallies in terms of % of grand total:

Count of Star Rating	Column Labels					
Row Labels	Five	Four	Three	Two	One	Grand Total
Large	**2.92%**	**9.39%**	**22.76%**	**14.41%**	**5.85%**	**55.32%**
Low	1.67%	5.01%	10.44%	5.22%	1.88%	24.22%
Average	0.63%	4.18%	11.48%	8.35%	2.30%	26.93%
High	0.63%	0.21%	0.84%	0.84%	1.67%	4.18%
MidCap	**0.63%**	**4.18%**	**10.23%**	**7.52%**	**1.88%**	**24.43%**
Low	0.42%	2.30%	1.46%	1.04%	0.21%	5.43%
Average	0.21%	1.88%	7.10%	3.13%	0.63%	12.94%
High	0.00%	0.00%	1.67%	3.34%	1.04%	6.05%
Small	**0.42%**	**5.01%**	**8.35%**	**5.85%**	**0.63%**	**20.25%**
Low	0.21%	0.63%	0.00%	0.21%	0.00%	1.04%
Average	0.00%	2.71%	3.13%	1.04%	0.00%	6.89%
High	0.21%	1.67%	5.22%	4.59%	0.63%	12.32%
Grand Total	**3.97%**	**18.58%**	**41.34%**	**27.77%**	**8.35%**	**100.00%**

(b) For the large-cap funds, the three-star rating category had the highest percentage of funds, followed by two-star, four-star, one-star, and five-star. Very few large-cap funds had ratings of five. This pattern was also seen with the mid-cap funds as a group. The same pattern was observed with the small-cap funds. However, the pattern was more subtle in that the differences in percentage were less in many cases.

Within the large-cap fund category, the highest percentage of funds were in the average-risk category followed by the low-risk and high-risk categories. Within the mid-cap category, the highest percentage of funds were in the average-risk category followed by the high and low risk categories. Within the small-cap category, the highest percentage of funds were in the high-risk category followed by the average and low risk categories.

2.59 cont. (c)

Average of 3YrReturn	Column Labels					
Row Labels	Five	Four	Three	Two	One	Grand Total
Large	**12.53**	**10.57**	**9.39**	**7.86**	**6.35**	**9.04**
Low	12.36	9.91	8.57	7.59	6.94	8.77
Average	10.73	11.35	10.00	7.75	7.03	9.27
High	14.80	10.92	11.36	10.68	4.76	9.07
MidCap	**11.22**	**8.89**	**7.77**	**5.27**	**2.44**	**6.87**
Low	11.74	9.02	8.48	7.13	3.90	8.52
Average	10.18	8.73	7.40	5.74	0.72	6.91
High			8.73	4.24	3.18	5.29
Small	**9.53**	**7.75**	**6.38**	**4.48**	**0.96**	**6.07**
Low	9.09	5.98		7.60		6.92
Average		7.73	6.61	2.85		6.48
High	9.96	8.44	6.24	4.71	0.96	5.76
Grand Total	**12.01**	**9.43**	**8.38**	**6.45**	**5.07**	**7.91**

(d) There are four high-risk large-cap funds with a three-star rating. Their average three-year return is 11.36.

2.60

Count of Type	Star Rating					
Type	One	Two	Three	Four	Five	Grand Total
Growth	**5.43%**	**17.12%**	**27.35%**	**11.27%**	**2.71%**	**63.88%**
Large	1.25%	2.09%	4.80%	3.55%	1.46%	13.15%
Mid-Cap	1.67%	7.72%	15.87%	6.05%	0.42%	31.73%
Small	2.51%	7.31%	6.68%	1.67%	0.84%	19.00%
Value	**2.92%**	**10.65%**	**13.99%**	**7.31%**	**1.25%**	**36.12%**
Large	0.84%	4.38%	7.10%	4.38%	0.84%	17.54%
Mid-Cap	1.25%	4.80%	5.85%	2.71%	0.42%	15.03%
Small	0.84%	1.46%	1.04%	0.21%	0.00%	3.55%
Grand Total	**8.35%**	**27.77%**	**41.34%**	**18.58%**	**3.96%**	**100.00%**

(b) Patterns of star rating conditioned on risk:
For the growth funds as a group, most are rated as three-star, followed by two-star, four-star, one-star, and five-star. The pattern of star rating is different among the various risk growth funds.

For the value funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. Among the high-risk value funds, more are two-star than three-star.

Most of the growth funds are rated as average-risk, followed by high-risk and then low-risk. The pattern is not the same among all the rating categories.

(b) Most of the value funds are rated as low-risk, followed by average-risk and then high-risk. The pattern is the same among the three-star, four-star, and five-star value funds. Among the one-star and two-star funds, there are more average risk funds than low risk funds.

2.60 (c) cont.

Count of Type	Star Rating					
Type	One	Two	Three	Four	Five	Grand Total
Growth	**5.41**	**7.04**	**8.94**	**10.14**	**12.83**	**8.51**
Large	7.53	8.60	9.89	10.29	12.64	9.87
Mid-Cap	6.17	7.99	9.28	10.43	11.96	9.06
Small	3.83	5.59	7.45	8.76	13.59	6.64
Value	**4.43**	**5.49**	**7.29**	**8.34**	**10.23**	**6.84**
Large	5.29	7.00	7.66	8.57	10.74	7.76
Mid-Cap	5.01	4.98	6.97	7.96	9.23	6.41
Small	2.71	2.63	6.53	8.39	–	4.13
Grand Total	**5.07**	**6.45**	**8.38**	**9.43**	**12.01**	**7.91**

The three-year returns for growth funds is higher than for value funds. The return is higher for funds with higher ratings than lower ratings. This pattern holds for the growth funds for each risk level. For the low risk and average risk value funds, the return is lowest for the funds with a two-star rating.

(d) There are 32 growth funds with high risk with a rating of three. These funds have an average three-year return of 7.45.

2.61 (a) Pivot table of tallies in terms of counts:

Row Labels	Five	Four	Three	Two	One	Grand Total
Growth	**13**	**54**	**131**	**82**	**26**	**306**
Large	**8**	**26**	**65**	**37**	**18**	**154**
High	3	1	3	4	6	17
Average	1	16	43	25	6	91
Low	4	9	19	8	6	46
MidCap	**3**	**15**	**36**	**26**	**6**	**86**
High			8	13	4	25
Average	1	7	24	11	2	45
Low	2	8	4	2		16
Small	**2**	**13**	**30**	**19**	**2**	**66**
High	1	7	21	18	2	49
Average		6	9	1		16
Low	1					1
Value	**6**	**35**	**67**	**51**	**14**	**173**
Large	**6**	**19**	**44**	**32**	**10**	**111**
High			1		2	3
Average	2	4	12	15	5	38
Low	4	15	31	17	3	70
MidCap		**5**	**13**	**10**	**3**	**31**
High				3	1	4
Average		2	10	4	1	17
Low		3	3	3	1	10
Small		**11**	**10**	**9**	**1**	**31**
High		1	4	4	1	10
Average		7	6	4		17
Low		3		1		4
Grand Total	**19**	**89**	**198**	**133**	**40**	**479**

2.61 cont. (a) Pivot table of tallies in terms of % of grand total:

Count of Star Rating	Column Labels					
Row Labels	Five	Four	Three	Two	One	Grand Total
Growth	**2.71%**	**11.27%**	**27.35%**	**17.12%**	**5.43%**	**63.88%**
Large	**1.67%**	**5.43%**	**13.57%**	**7.72%**	**3.76%**	**32.15%**
High	0.63%	0.21%	0.63%	0.84%	1.25%	3.55%
Average	0.21%	3.34%	8.98%	5.22%	1.25%	19.00%
Low	0.84%	1.88%	3.97%	1.67%	1.25%	9.60%
MidCap	**0.63%**	**3.13%**	**7.52%**	**5.43%**	**1.25%**	**17.95%**
High	0.00%	0.00%	1.67%	2.71%	0.84%	5.22%
Average	0.21%	1.46%	5.01%	2.30%	0.42%	9.39%
Low	0.42%	1.67%	0.84%	0.42%	0.00%	3.34%
Small	**0.42%**	**2.71%**	**6.26%**	**3.97%**	**0.42%**	**13.78%**
High	0.21%	1.46%	4.38%	3.76%	0.42%	10.23%
Average	0.00%	1.25%	1.88%	0.21%	0.00%	3.34%
Low	0.21%	0.00%	0.00%	0.00%	0.00%	0.21%
Value	**1.25%**	**7.31%**	**13.99%**	**10.65%**	**2.92%**	**36.12%**
Large	**1.25%**	**3.97%**	**9.19%**	**6.68%**	**2.09%**	**23.17%**
High	0.00%	0.00%	0.21%	0.00%	0.42%	0.63%
Average	0.42%	0.84%	2.51%	3.13%	1.04%	7.93%
Low	0.84%	3.13%	6.47%	3.55%	0.63%	14.61%
MidCap	**0.00%**	**1.04%**	**2.71%**	**2.09%**	**0.63%**	**6.47%**
High	0.00%	0.00%	0.00%	0.63%	0.21%	0.84%
Average	0.00%	0.42%	2.09%	0.84%	0.21%	3.55%
Low	0.00%	0.63%	0.63%	0.63%	0.21%	2.09%
Small	**0.00%**	**2.30%**	**2.09%**	**1.88%**	**0.21%**	**6.47%**
High	0.00%	0.21%	0.84%	0.84%	0.21%	2.09%
Average	0.00%	1.46%	1.25%	0.84%	0.00%	3.55%
Low	0.00%	0.63%	0.00%	0.21%	0.00%	0.84%
Grand Total	**3.97%**	**18.58%**	**41.34%**	**27.77%**	**8.35%**	**100.00%**

(b) For growth funds, most are rated as three-star followed by two-star, four-star, one-star and five-star. Among the growth funds, large-cap and mid-cap had the same pattern of star-rating as observed for growth funds in general. Small-cap growth funds had the same pattern with the exception of having the same the number of funds rated as one-star and five-star. The pattern of star-rating is different among the various risk levels within the large-cap, mid-cap and small-cap growth funds.

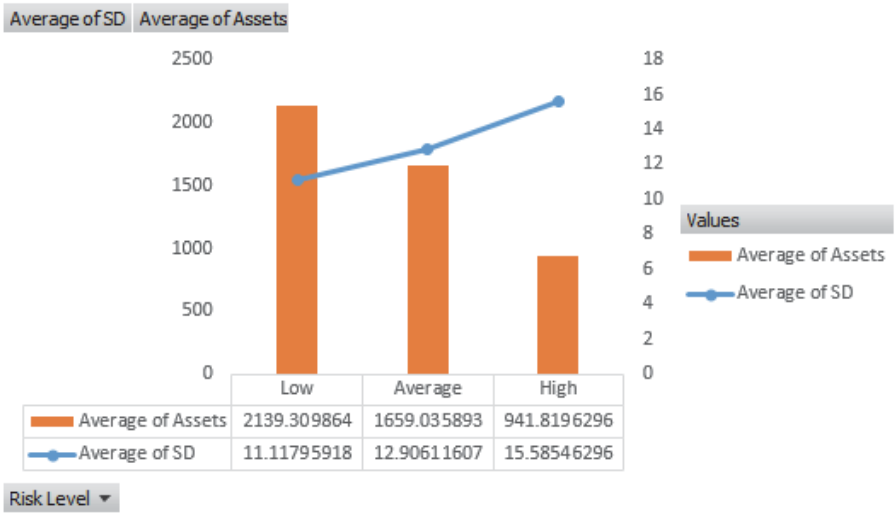
For value funds, most are rated as three-star followed by two-star, four-star, one-star, and five-star. Among the value funds, the pattern is the same for large-cap and mid-cap funds. Small-cap value funds have a different pattern. The pattern of star-rating is different among the various risk levels within the large-cap, mid-cap and small-cap funds.

(c) The tables in 2.58 through 2.60 are easier to interpret because they contain fewer fields. The table in 2.61 tallies star rating across three fields: market type, market cap, and risk level. Problems 2.58 through 2.60 tally star rating across two fields.

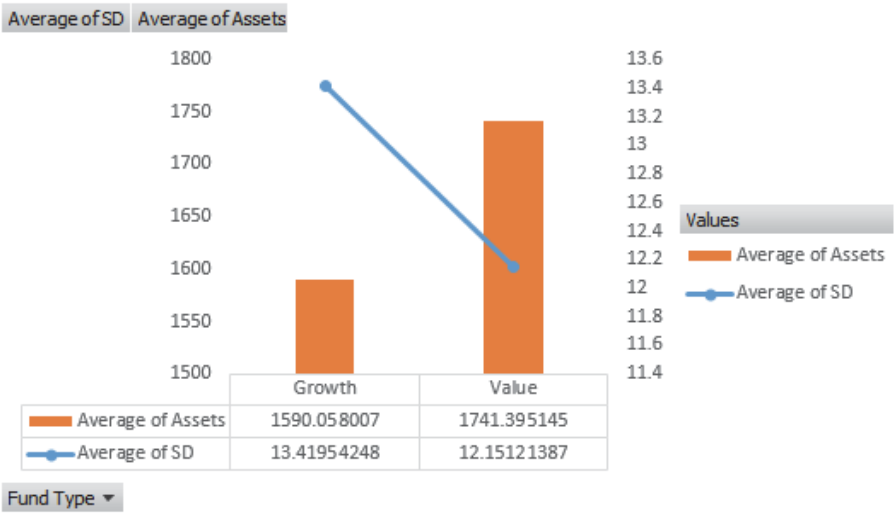
(d) Problem 2.60 reveals that most value funds are rated as low-risk followed by average-risk and high-risk. Problem 2.61 reveals that this is only the case among large-cap value funds. Most mid-cap value funds are rated as average-risk followed by low-risk and high-risk. Most small-cap value funds are rated as average-risk followed by high-risk and low-risk. Problem 2.61 also reveals that among small-cap funds rated as average-risk, most are rated as four-star, followed by three-star and two-star. Because Problem 2.61 includes four fields compared to three fields included in problems 2.58 through 2.60, additional patterns can be observed.

2.62 The fund with the highest five-year return of 15.72 is a large cap growth fund that has a four-star rating and low risk.

2.63 (a)



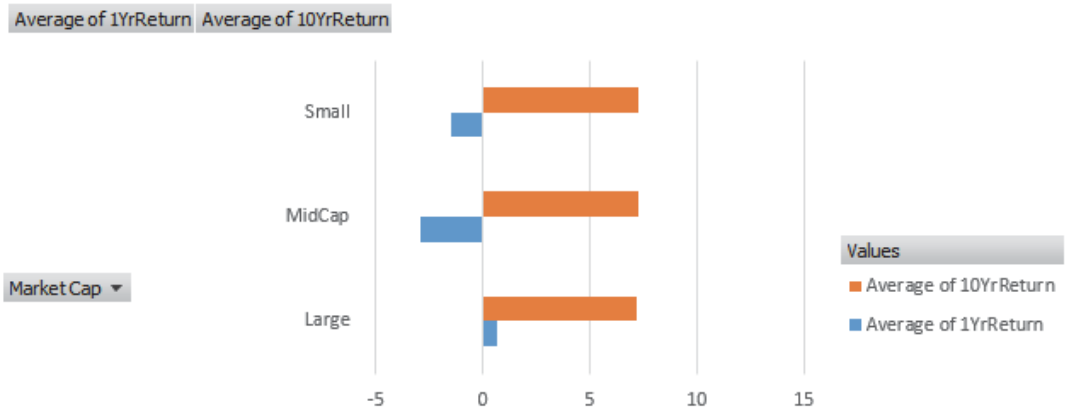
(b)



(c) The results from (a) reveal that the average of SD increases as the risk level increases while average of assets decreases as risk level increases. The results from (b) reveal that the average of SD is higher for growth funds compared to value funds. The patterns suggest that value funds are likely to be associated with less risk because the average of SD was lower among value funds and low risk funds.

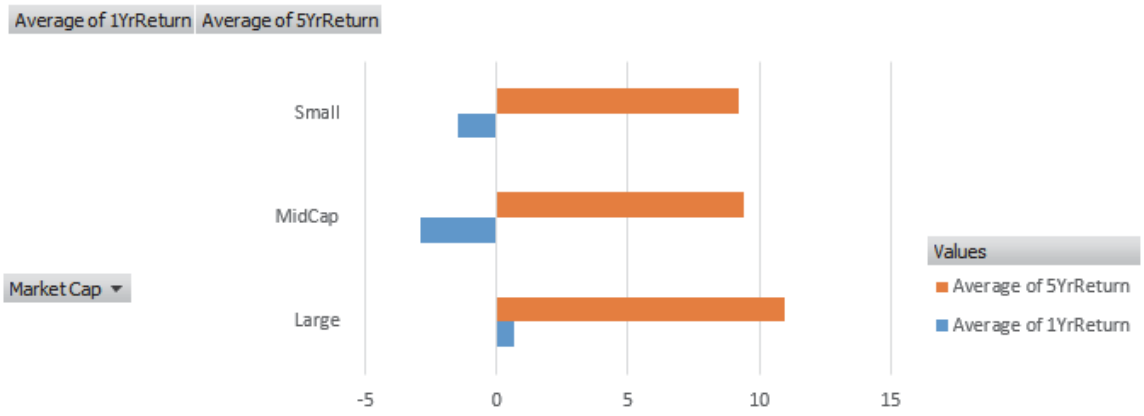
2.64 Funds 479, 471, 347, 443, and 477 have the lowest five-year return.

2.65 (a)



	Large	MidCap	Small
Average of 10YrReturn	7.16954717	7.291794872	7.260721649
Average of 1YrReturn	0.679358491	-2.876666667	-1.455876289

(b)



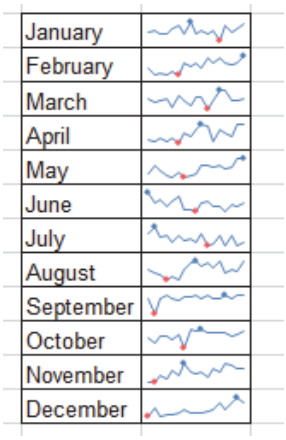
	Large	MidCap	Small
Average of 5YrReturn	10.94430189	9.409059829	9.208556701
Average of 1YrReturn	0.679358491	-2.876666667	-1.455876289

(c) For the 1-year versus 10-year return chart, the 10-year returns are much higher than the 1-year returns with similar 5-year returns near 7 percent for all three market cap categories. For the 1-year versus 5-year chart, the returns are all higher for the 5-year returns compared to the 1-year returns. The 5-year returns are higher than the 10-year returns. The large-cap 5-year return is higher than the mid-cap and small 5-year returns.

(d) Because the average 5-year returns were all higher than the 10-year returns for all market cap categories, one can conclude that the returns were lower in years 6 through 10. Without annual data, one cannot conclude that this was due to consistent lower returns across the years or the result of one or two years with lower returns.

2.66 The five funds with the lowest five-year return have (1) midcap growth, average risk, one-star rating, (2) midcap growth, high risk, two-star rating, (3) large value, average risk, two-star rating, (4) midcap growth, high risk, one-star rating, and (5) small value, average risk, two-star rating.

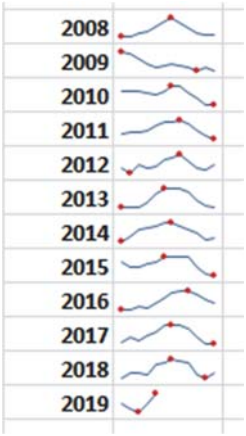
2.67 (a)



(b) The sparklines reveal that a general trend upward in home prices during the months of February, May, November, and December and they have remained steady in September after a jump from a low in 2001.

(c) In the Time-series plot one can see an upward trend in home sales price until 2006. Prices decline or remain flat from 2006 – 2011. From 2011 – 2016 there is an upward trend in median price of new home sales. With the exception of one year, the September home prices are fairly stable. This could be an error in the data.

2.68 (a)

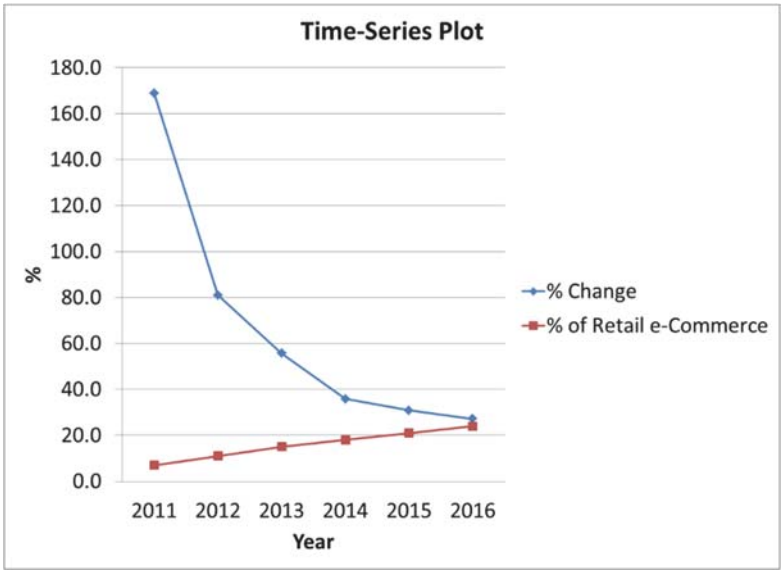


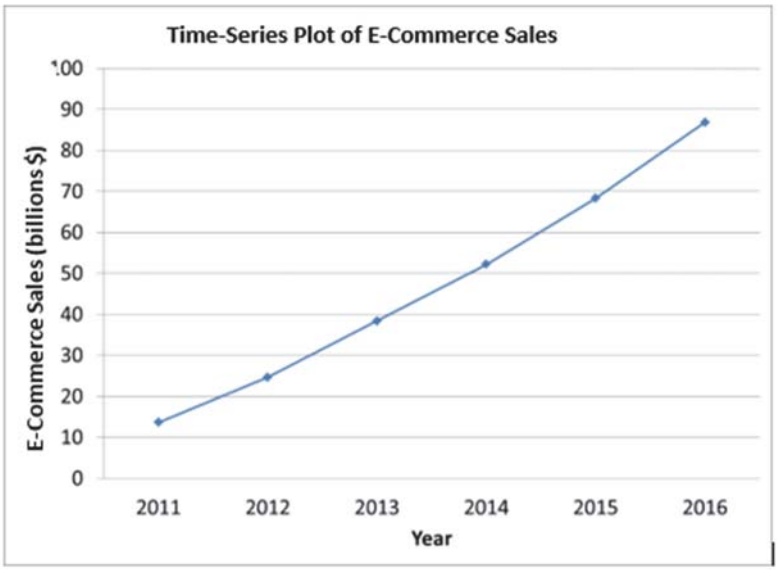
(b) There has been a decline in the price of natural gas over time. However, there is no pattern within the years. For some years, the price is higher in the beginning of the year. For other years, the price is higher in the latter part of the year. Sometimes, there is little variation within the year.

2.69 Student project answers will vary

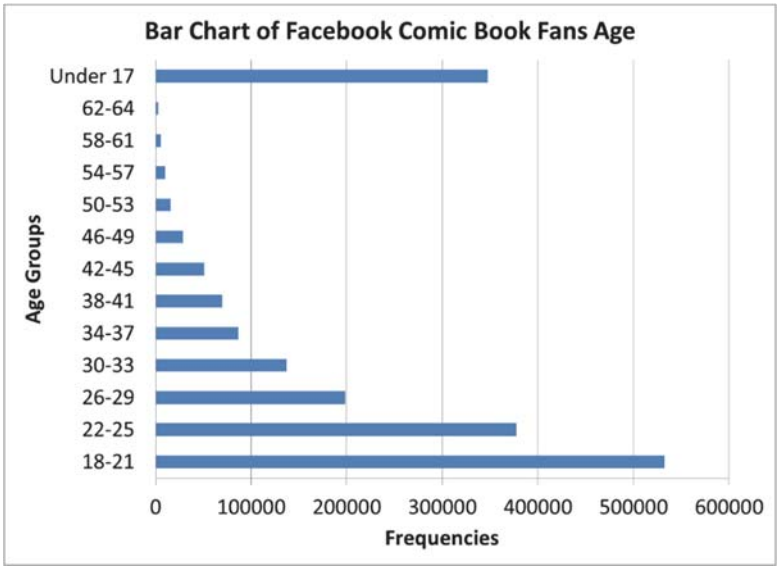
2.70 Student project answers will vary

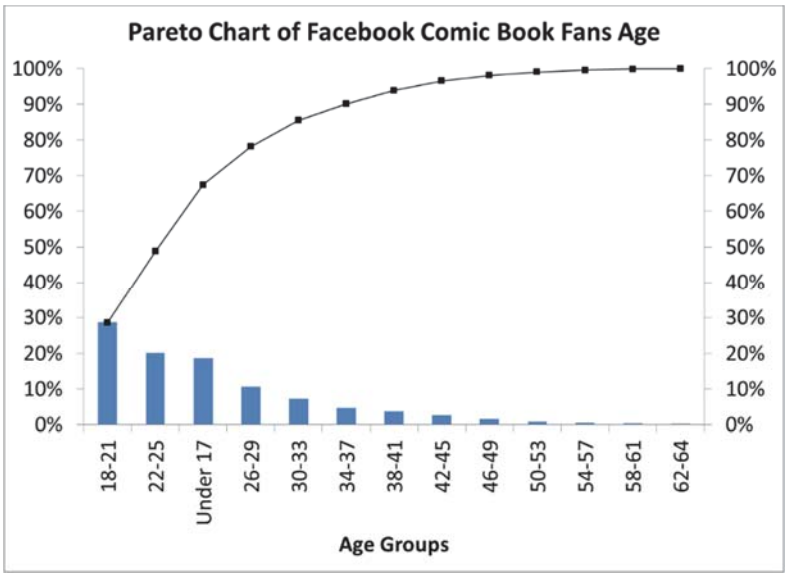
2.71 (a) There is a title.
(b) None of the axes are labeled.
(c)





2.72 (a) There is a title.
(b) The simplest possible visualization is not used.
(c)

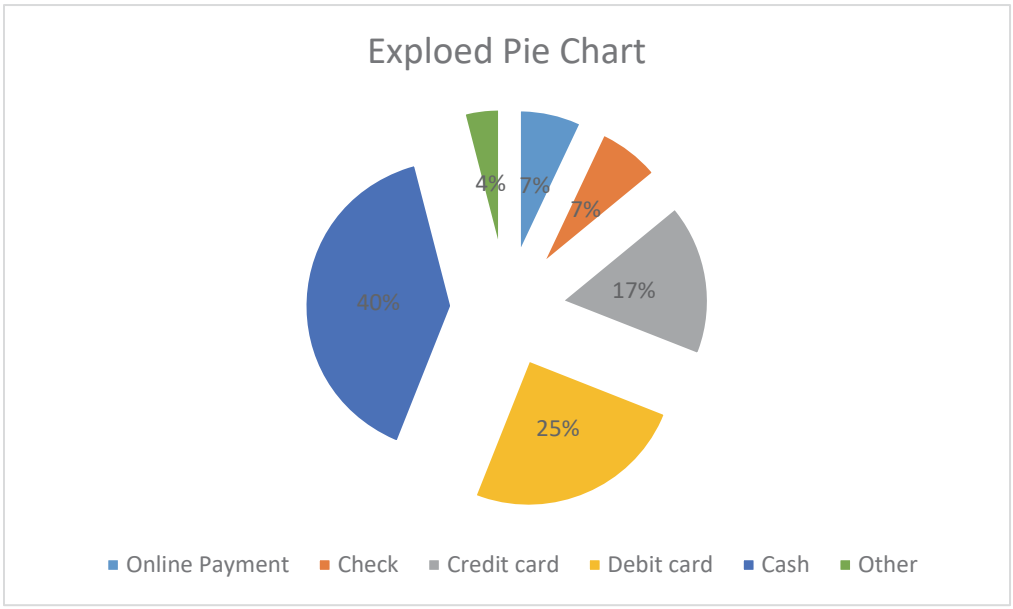




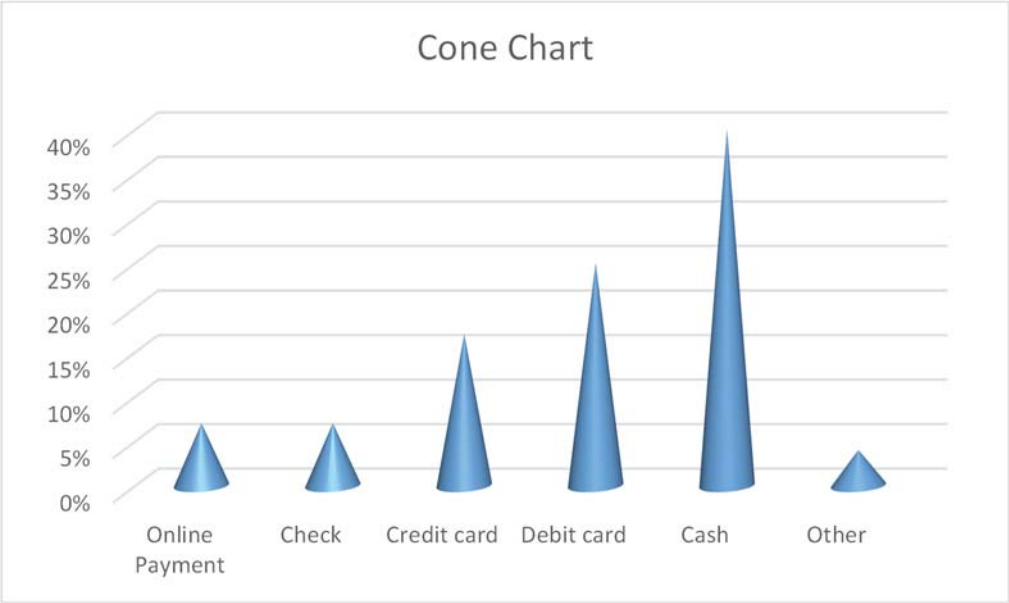
2.73 (a) None.
(b) The use of a 3D graph.
(c)

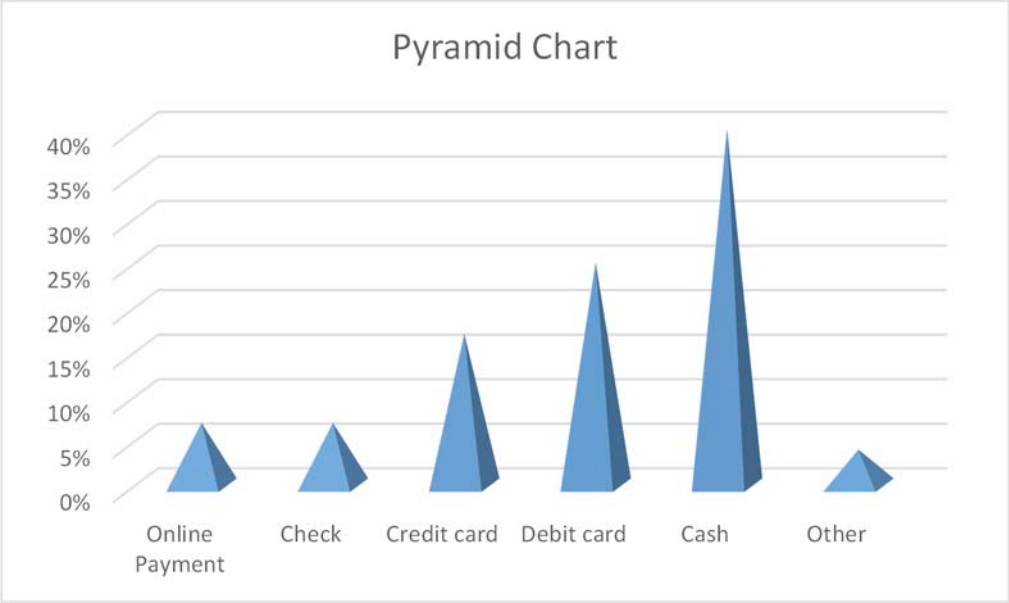
2.74 Answers will vary depending on selection of source.

2.75 (a)



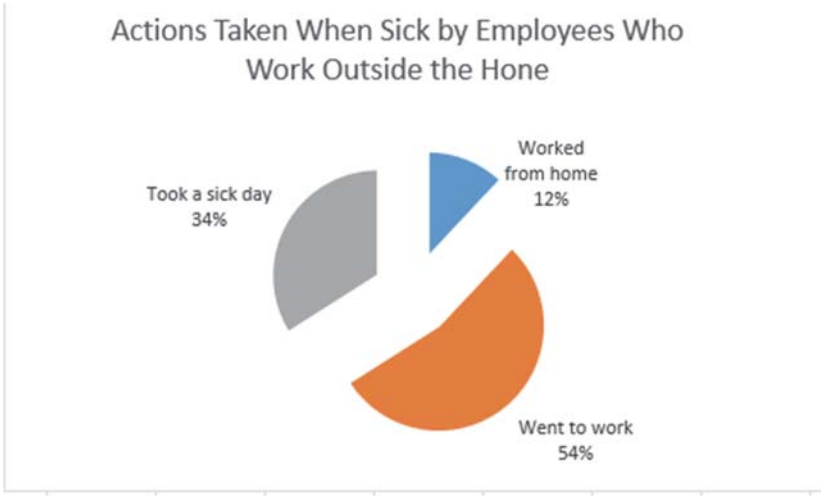
2.75 cont. (a)

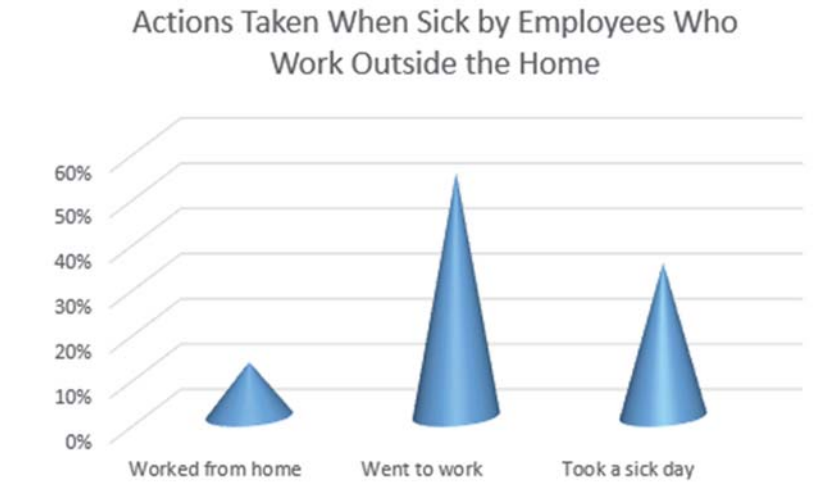


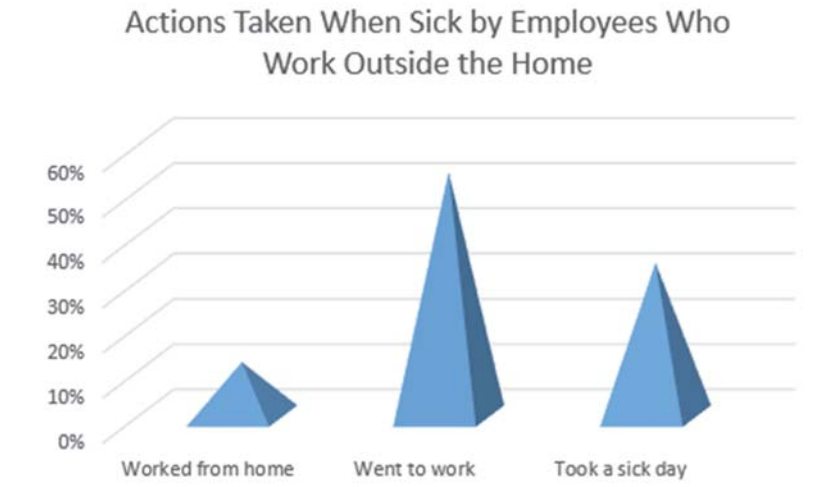


(b) The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.76 (a)







(b) The bar chart and the pie chart should be preferred over the exploded pie chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.77 A histogram uses bars to represent each class while a polygon uses a single point. The histogram should be used for only one group, while several polygons can be plotted on a single graph.

2.78 A summary table allows one to determine the frequency or percentage of occurrences in each category.

2.79 A bar chart is useful for comparing categories. A pie chart is useful when examining the portion of the whole that is in each category. A Pareto diagram is useful in focusing on the categories that make up most of the frequencies or percentages.

2.80 The bar chart for categorical data is plotted with the categories on the vertical axis and the frequencies or percentages on the horizontal axis. In addition, there is a separation between categories. The histogram is plotted with the class grouping on the horizontal axis and the frequencies or percentages on the vertical axis. This allows one to more easily determine the distribution of the data. In addition, there are no gaps between classes in the histogram.

2.81 A time-series plot is a type of scatter diagram with time on the *x*-axis.

2.82 Because the categories are arranged according to frequency or importance, it allows the user to focus attention on the categories that have the greatest frequency or importance.

2.83 Percentage breakdowns according to the total percentage, the row percentage, and/or the column percentage allow the interpretation of data in a two-way contingency table from several different perspectives.

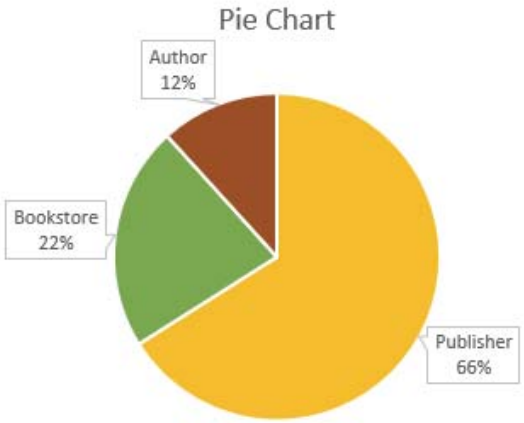
2.84 A contingency table contains information on two categorical variables whereas a multidimensional table can display information on more than two categorical variables.

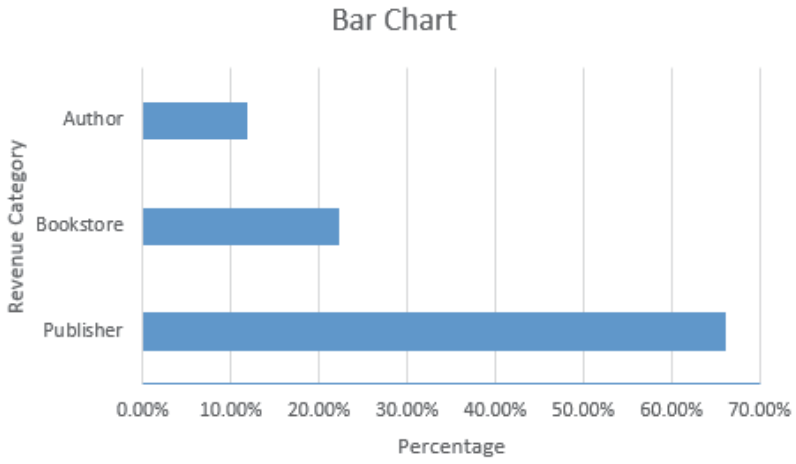
2.85 The multidimensional PivotTable can reveal additional patterns that cannot be seen in the contingency table. One can also change the statistic displayed and compute descriptive statistics which can add insight into the data.

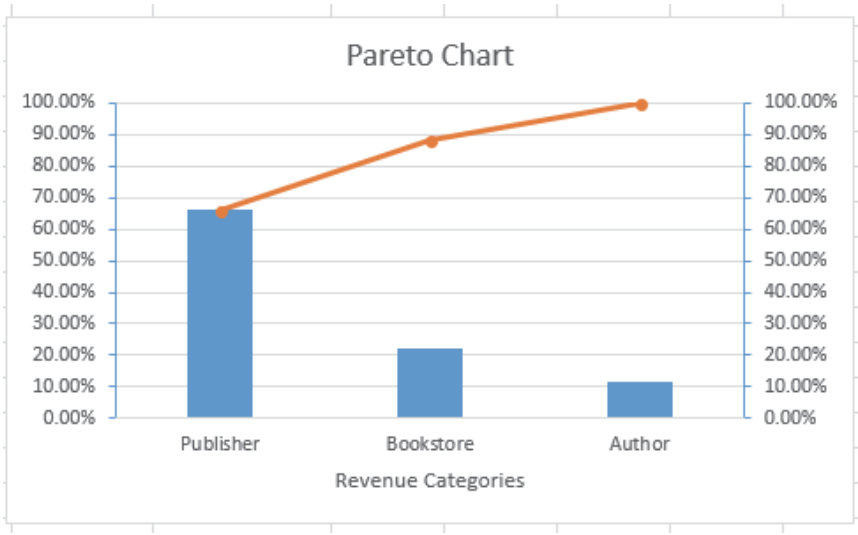
2.86 In a PivotTable in Excel, double-clicking a cell drills down and causes Excel to display the underlying data in a new worksheet to enable you to then observe the data for patterns. In Excel, a slicer is a panel of clickable buttons that appears superimposed over a worksheet to enable you to work with many variables at once in a way that avoids creating an overly complex multidimensional contingency table that would be hard to comprehend and interpret.

2.87 Sparklines are compact time-series visualizations of numerical variables. Sparklines can also be used to plot time-series data using smaller time units than a time-series plot to reveal patterns that the time-series plot may not.

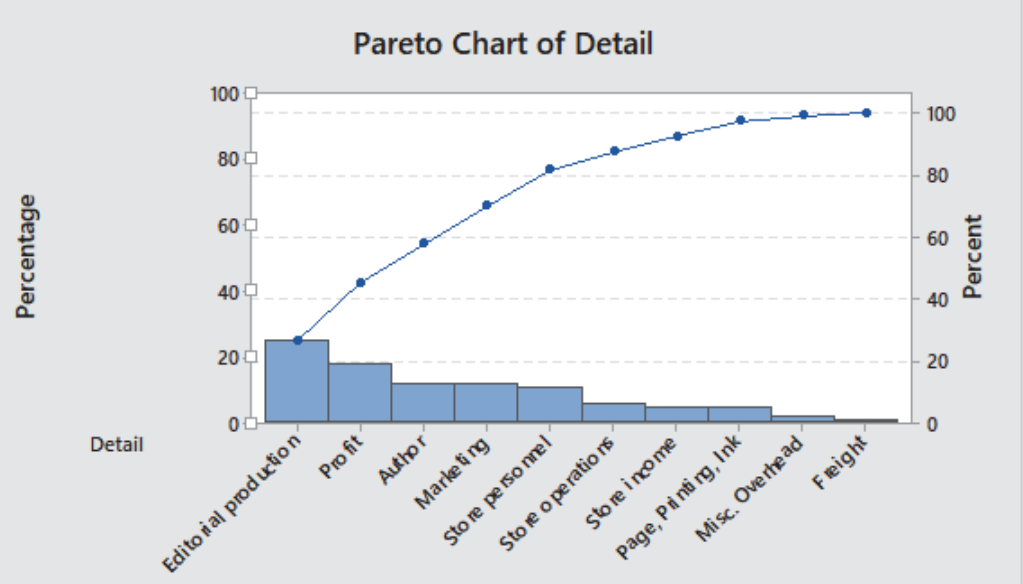
2.88 (a)





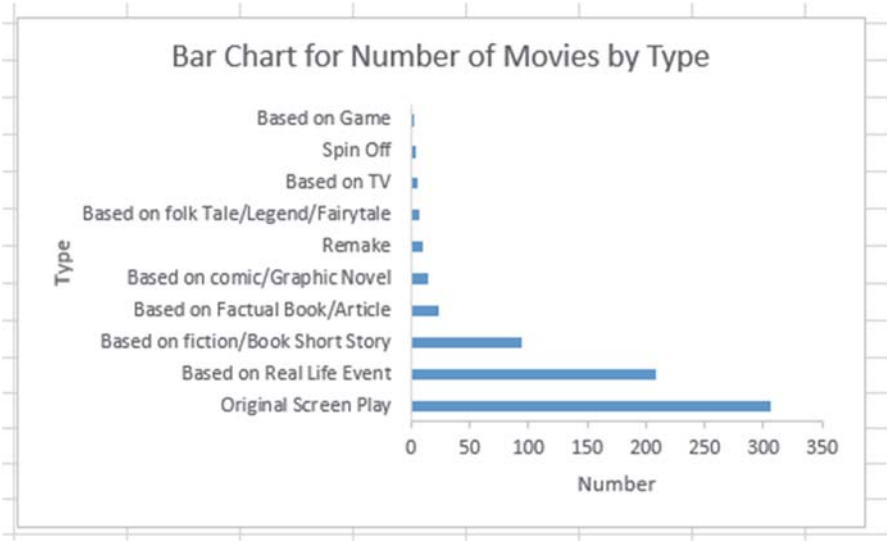


2.88 (b) cont.

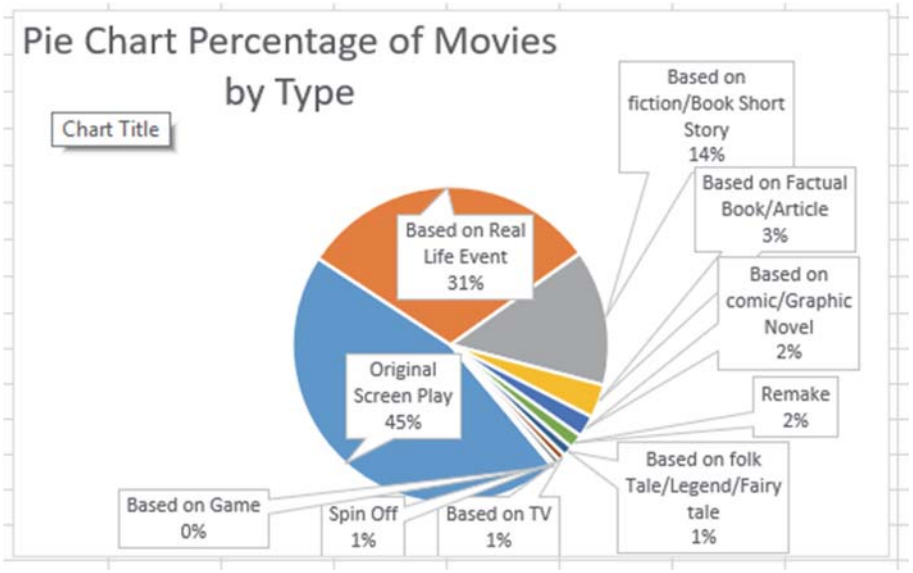


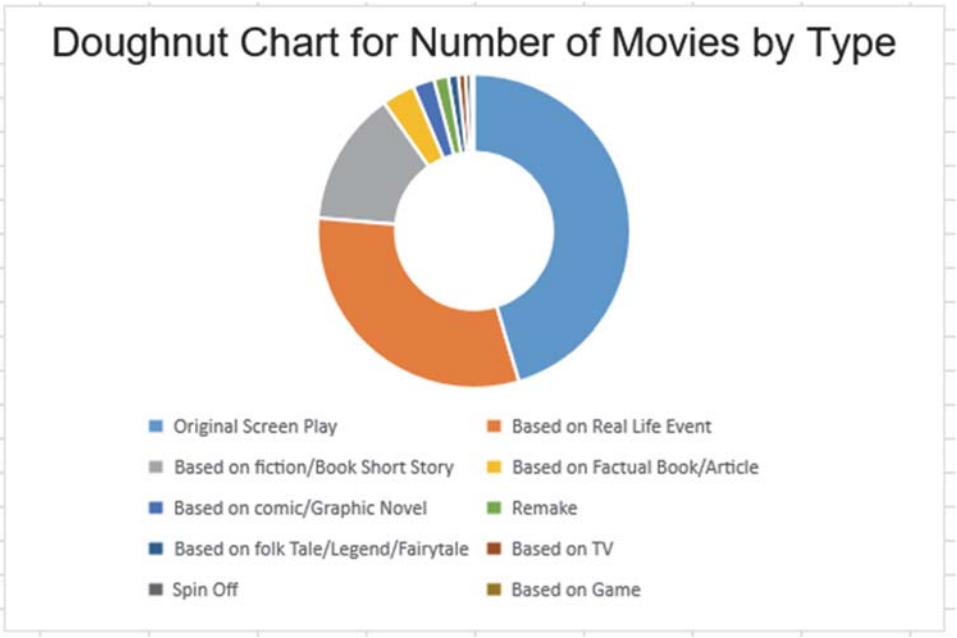
(c) The publisher gets the largest portion (66.06%) of the revenue. 24.93% is editorial production manufacturing costs. The publisher's marketing accounts for the next largest share of the revenue, at 11.6%. Author and bookstore personnel each account for around 11 to 12% of the revenue, whereas the publisher and bookstore profit and income account for more than 26% of the revenue. Yes, the bookstore gets almost twice the revenue of the authors.

2.89 (a) Number of Movies

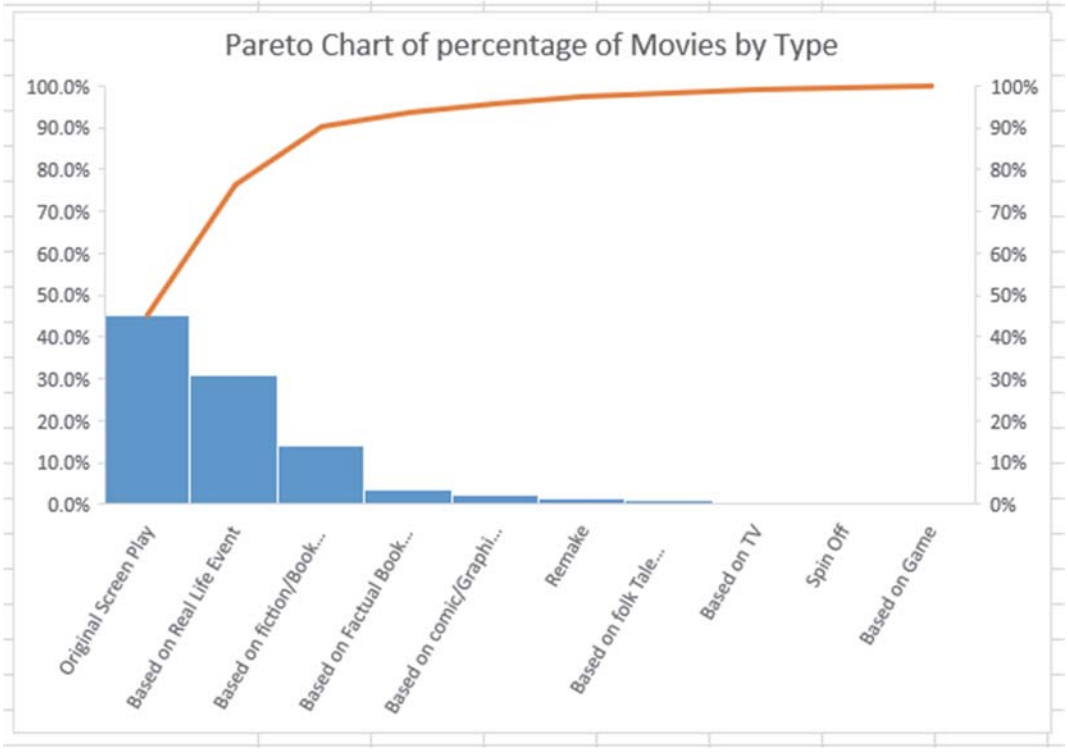


2.89 (a) cont.

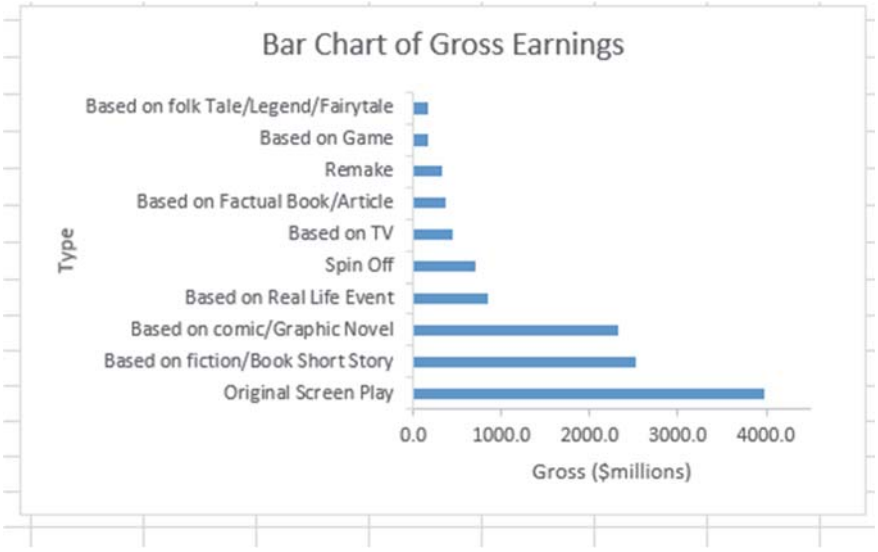




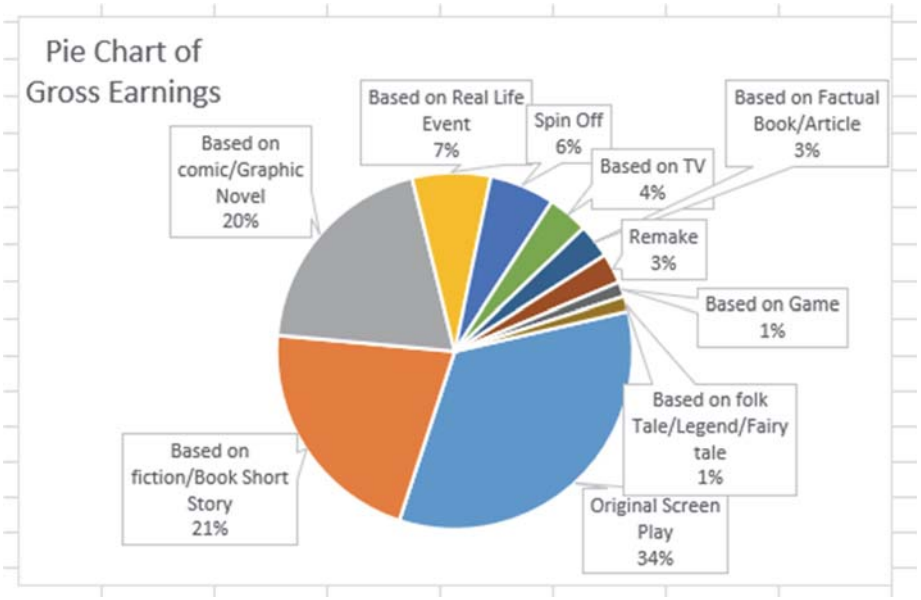
2.89 (a) cont.

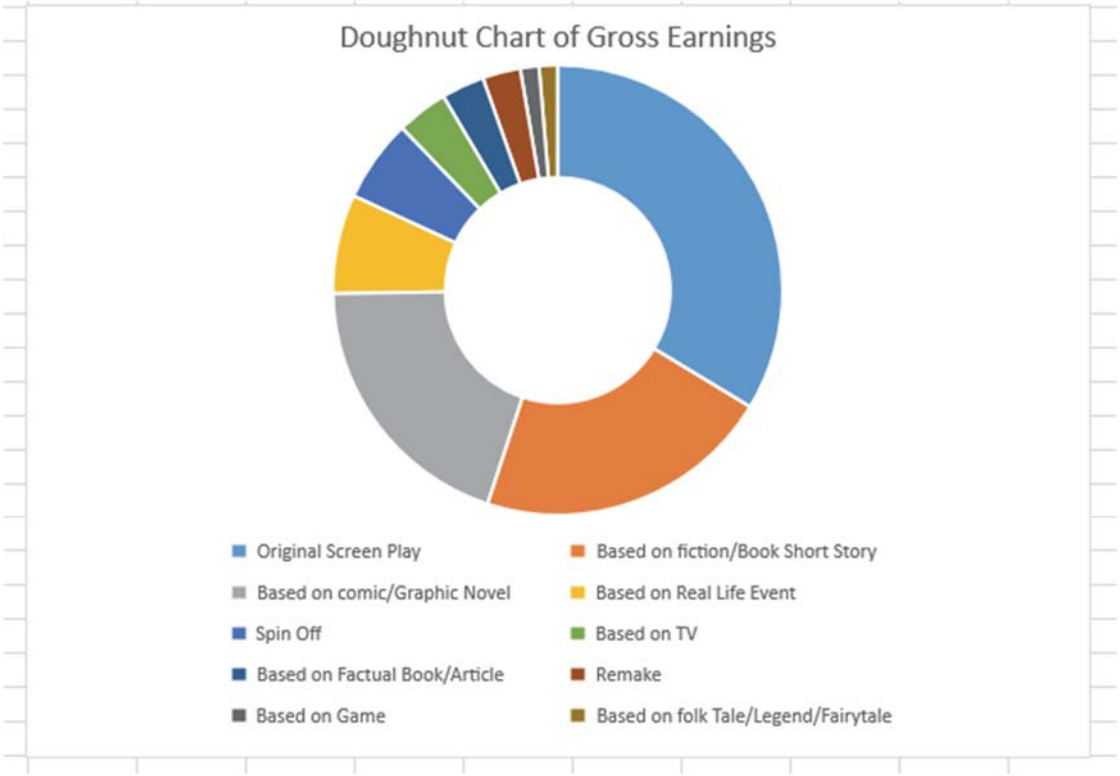


Gross

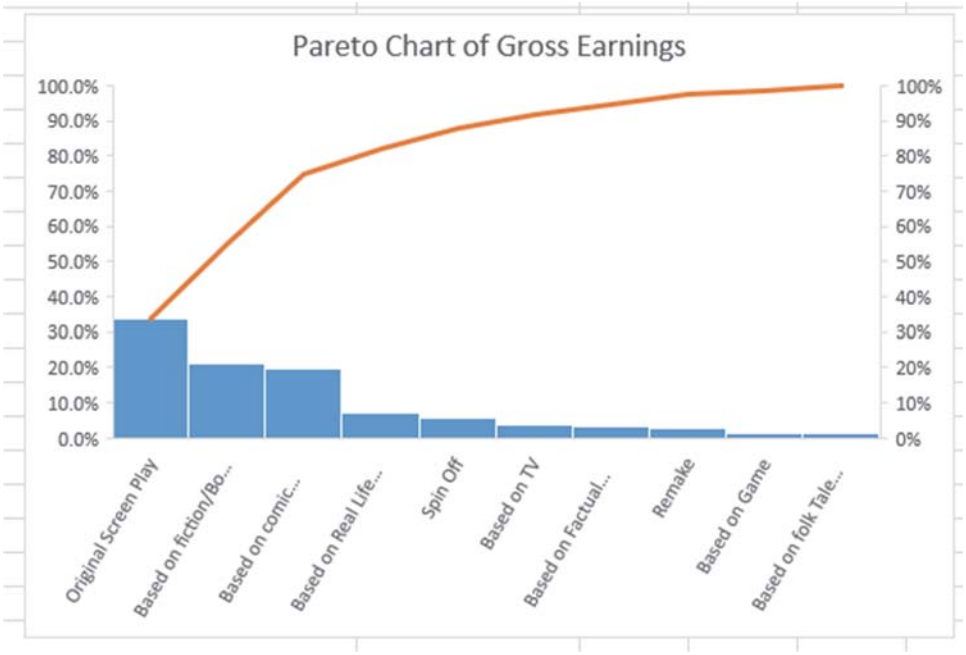


2.89 cont. (a) Gross

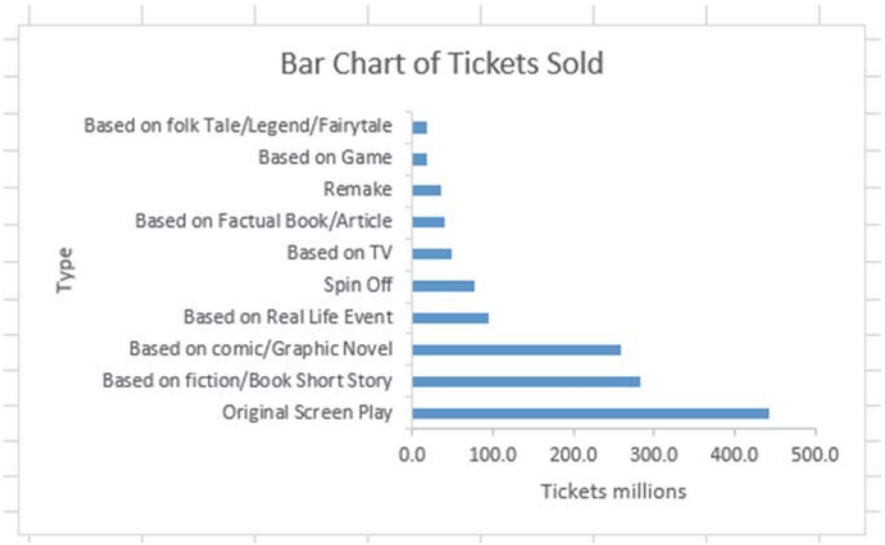




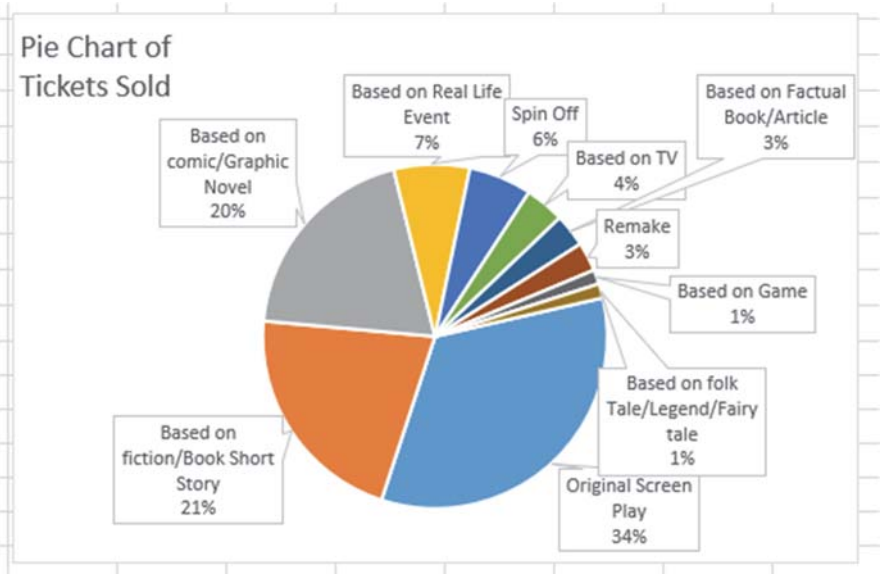
2.89 cont. (a) Gross

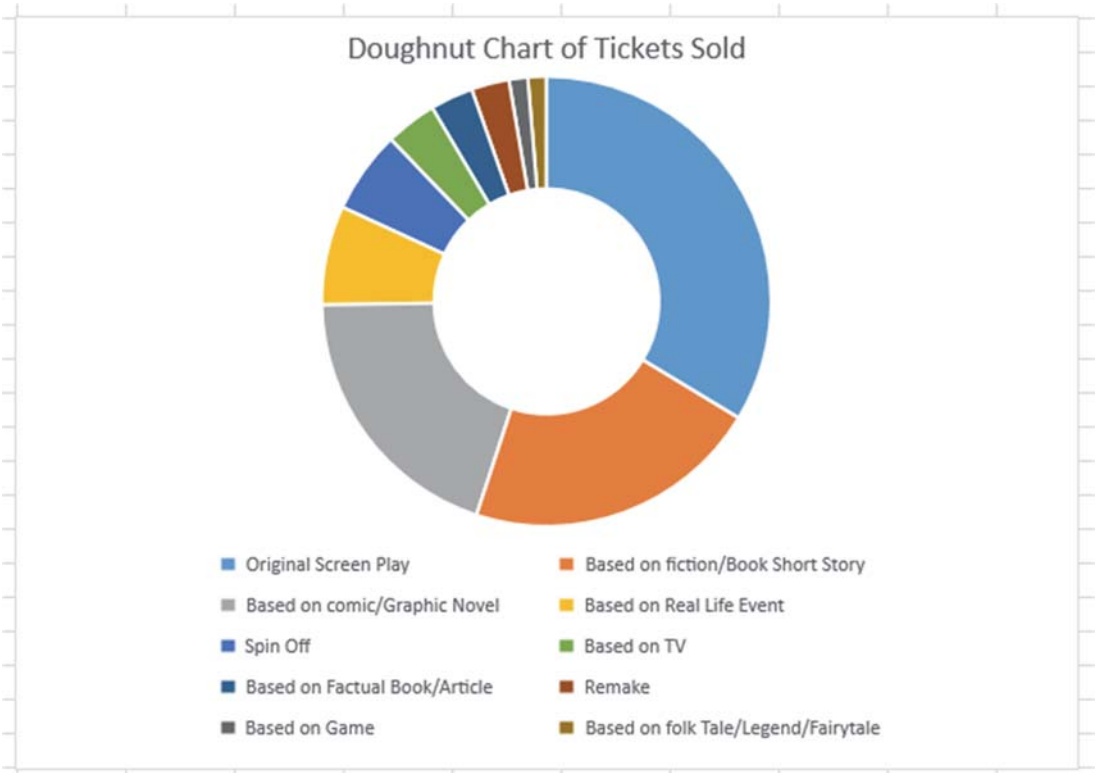


Tickets Sold

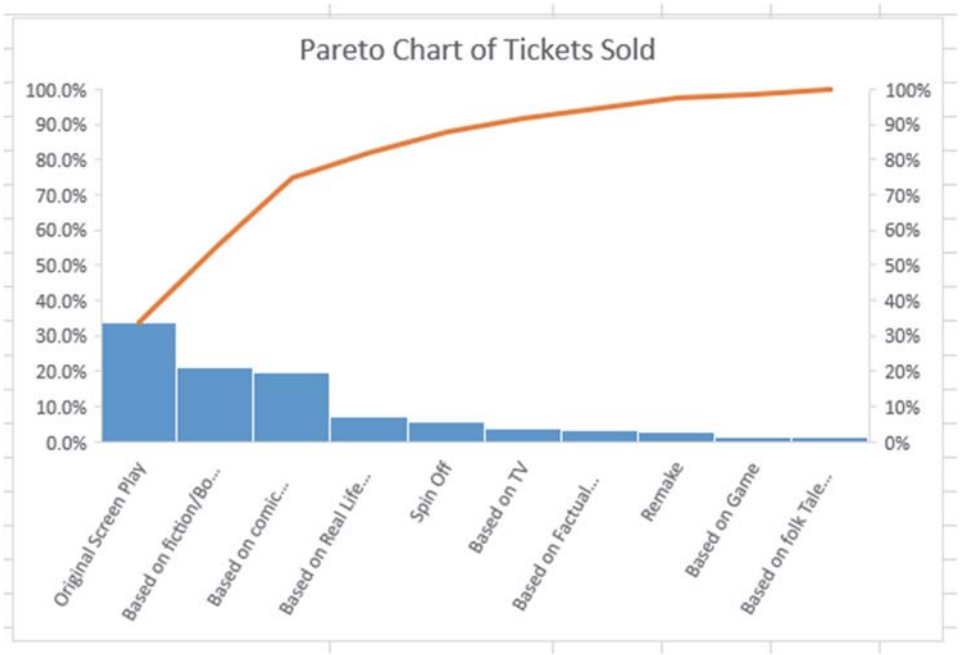


2.89 cont. (a) Tickets Sold



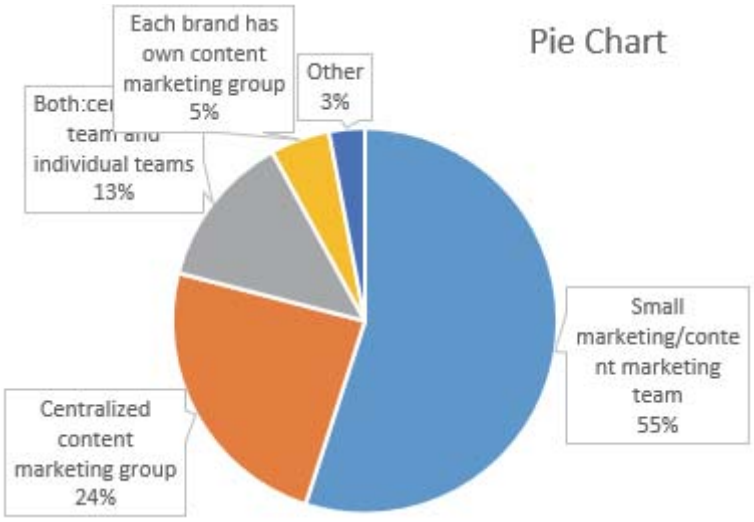


2.89 cont. (a) Tickets Sold

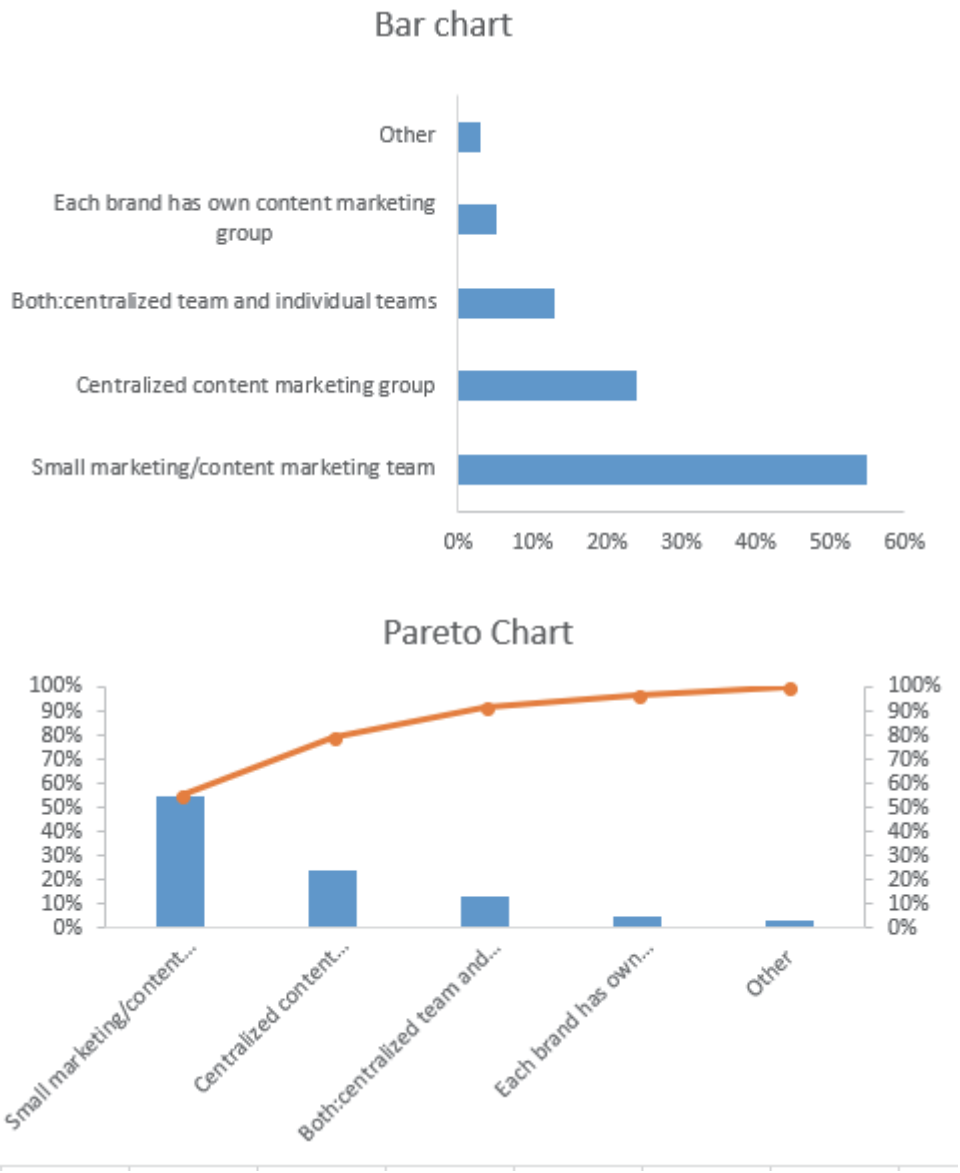


(b) Based on the Pareto chart for the number of movies, "Original screenplay", "Based on real life events" and "Based on fiction/short story" are the "vital few" and capture about 90% of the market share. According to the Pareto chart for gross (in $millions), "Original screenplay", "Based on fiction book/short story" and "Based on comic/graphic novel" are the "vital few" and capture about 74% of the market share. According to the Pareto chart for number of tickets sold (in millions), "Original screenplay", "Based on fiction book/short story" and "Based on comic/graphic novel" are the "vital few" and capture about 75% of the market share.

2.90 (a)

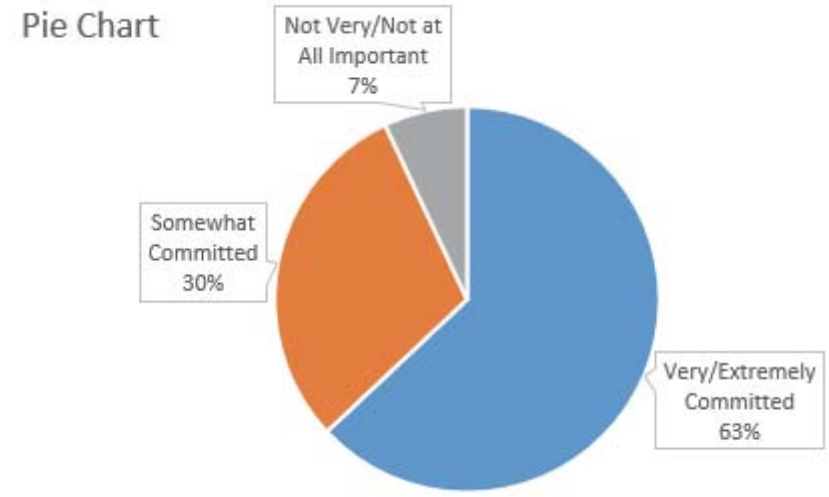


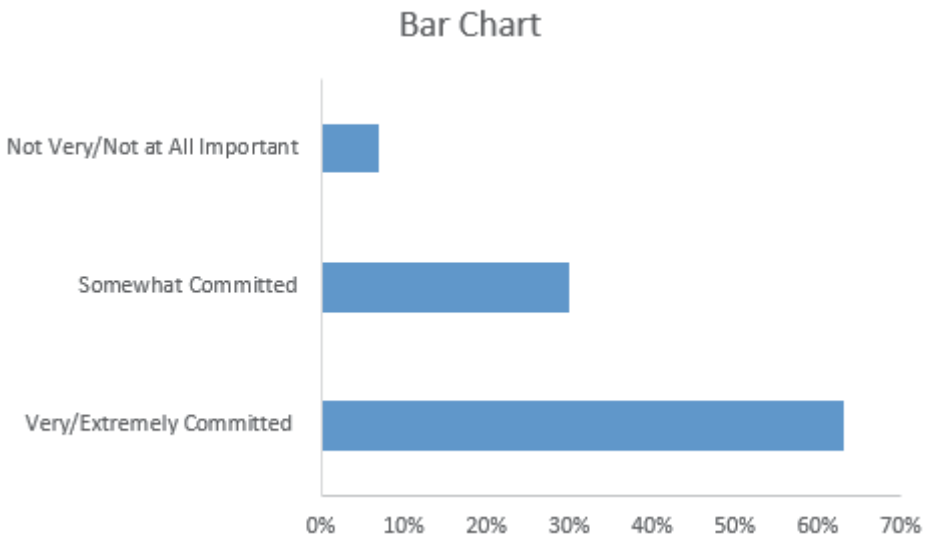
2.90 cont. (a)



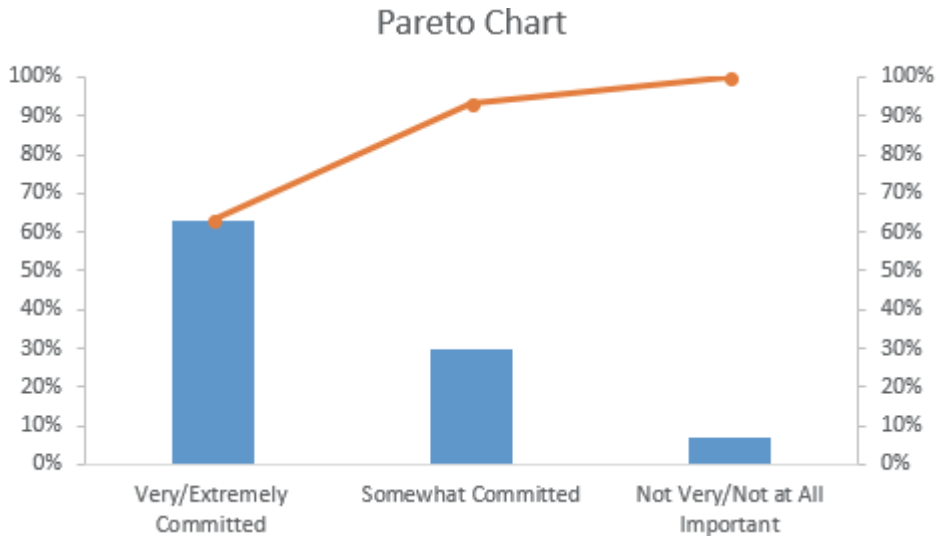
(b) The pie chart or the Pareto chart would be best. The pie chart would allow you to see each category as part of the whole, while the Pareto chart would enable you to see that Small marketing/content marketing team is the dominant category.

2.90 (c)
cont.





2.90 cont. (c)



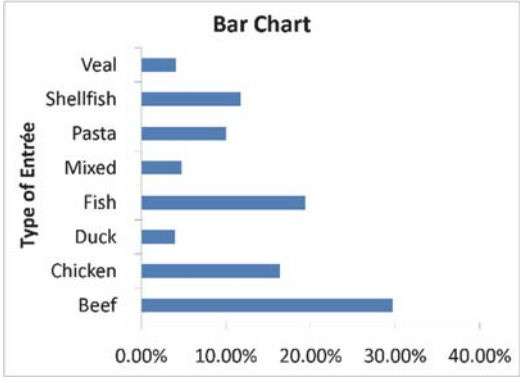
(d) The pie chart or the Pareto chart would be best. The pie chart would allow you to see each category as part of the whole while the Pareto chart would enable you to see that very committed to content marketing is the dominant category.

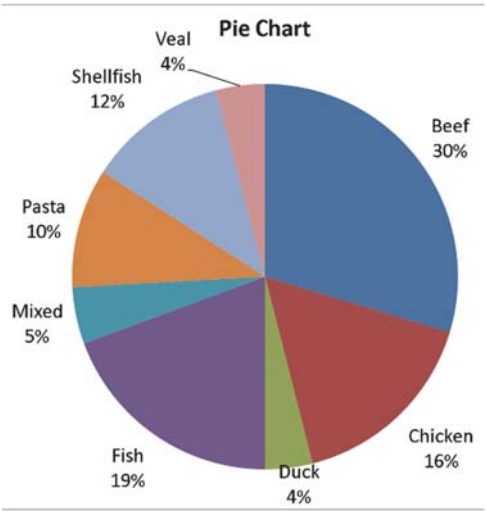
(e) Most organizations have a small marketing/content marketing team and are very committed to content marketing.

2.91 (a)

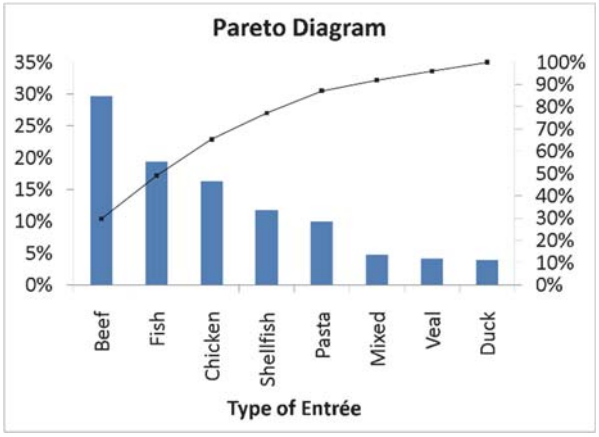
Type of Entrée	%	Number S
Beef	29.68%	187
Chicken	16.35%	103
Mixed	4.76%	30
Duck	3.97%	25
Fish	19.37%	122
Pasta	10.00%	63
Shellfish	11.75%	74
Veal	4.13%	26
Total	100.00%	630

(b)





2.91 (b) cont.



(c) The Pareto diagram has the advantage of offering the cumulative percentage view of the categories and, hence, enables the viewer to separate the "vital few" from the "trivial many".

(d) Beef and fish account for nearly 50% of all entrees ordered by weekend patrons of a continental restaurant. When chicken is included, nearly two-thirds of the entrees are accounted for.

2.92 (a)

Dessert Ordered	Gender		
	Male	Female	Total
Yes	66%	34%	100%
No	48%	52%	100%
Total	52%	48%	100%

Dessert Ordered	Gender		
	Male	Female	Total
Yes	29%	34%	100%
No	71%	52%	100%
Total	100%	48%	100%

Dessert Ordered	Gender		
	Male	Female	Total
Yes	15%	8%	23%
No	37%	40%	77%
Total	52%	48%	100%

Dessert Ordered	Gender		
	Male	Female	Total
Yes	52%	48%	100%
No	25%	75%	100%
Total	31%	69%	100%

Dessert Ordered	Gender		
	Male	Female	Total
Yes	38%	16%	23%
No	62%	84%	77%
Total	100%	100%	100%

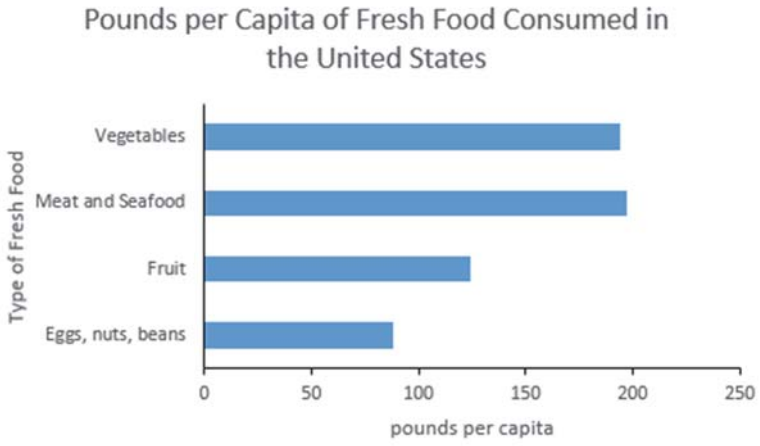
2.92 cont. (a)

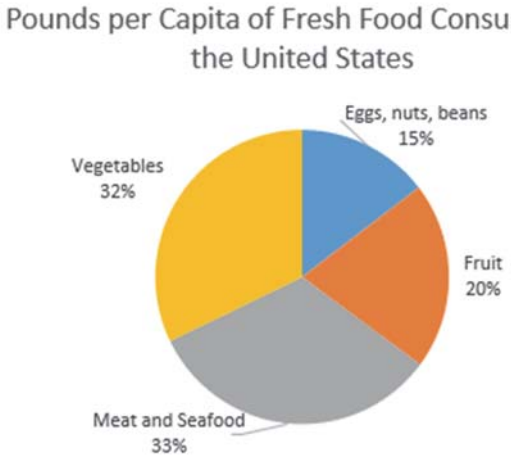
Dessert Ordered	Gender		
	Male	Female	Total
Yes	11.75%	10.79%	22.54%
No	19.52%	57.94%	77.46%
Total	31.27%	68.73%	100%

(b) If the owner is interested in finding out the percentage of males and females who order dessert or the percentage of those who order a beef entrée and a dessert among all patrons, the table of total percentages is most informative. If the owner is interested in the effect of gender on ordering of dessert or the effect of ordering a beef entrée on the ordering of dessert, the table of column percentages will be most informative. Because dessert is usually ordered after the main entrée, and the owner has no direct control over the gender of patrons, the table of row percentages is not very useful here.

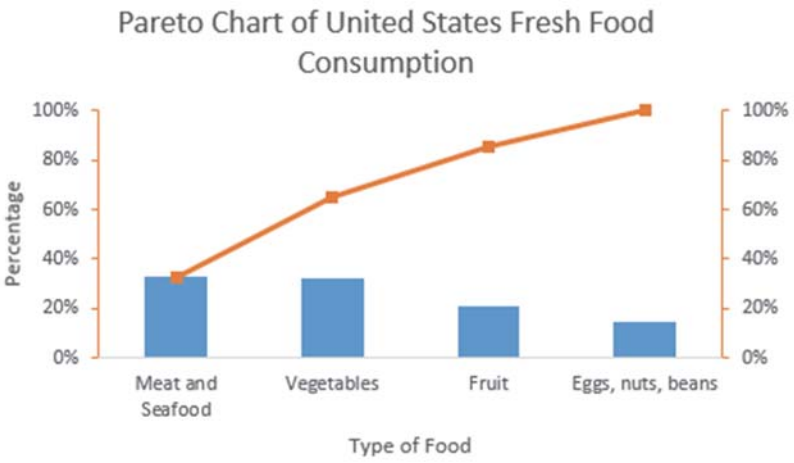
(c) 29% of the men ordered desserts, compared to 17 of the women; men are almost twice as likely to order dessert as women. Almost 38% of the patrons ordering a beef entrée ordered dessert, compared to 16% of patrons ordering all other entrées. Patrons ordering beef are more than 2.3 times as likely to order dessert as patrons ordering any other entrée.

2.93 (a) **United States Fresh Food Consumed:**

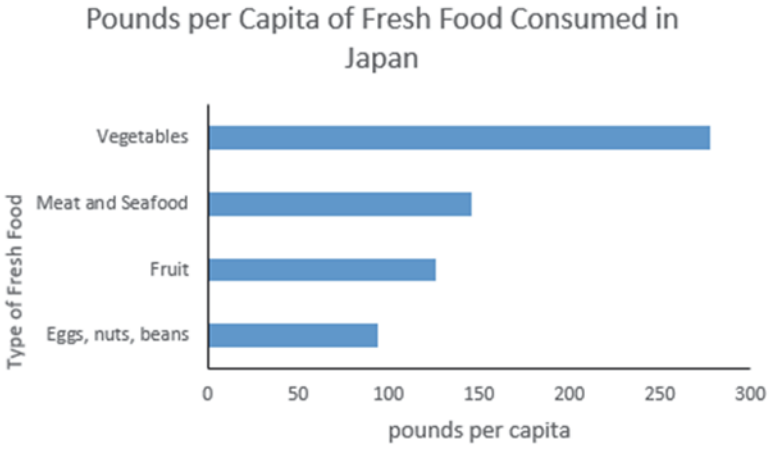


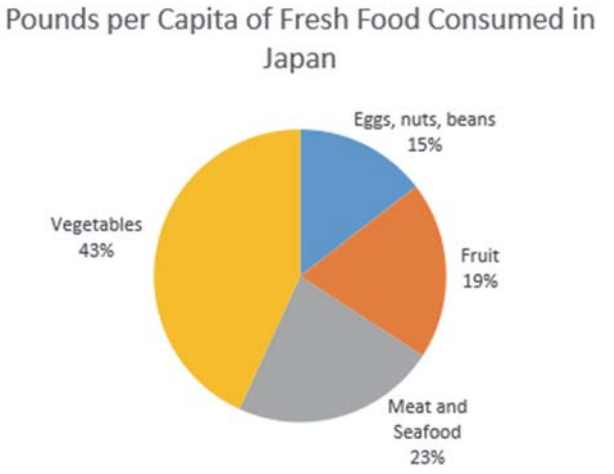


2.93 cont. (a) **United States Fresh Food Consumed:**

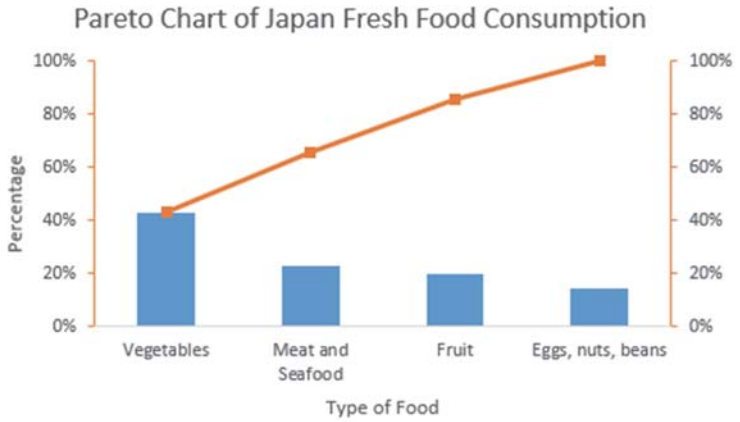


Japan Fresh Food Consumed:

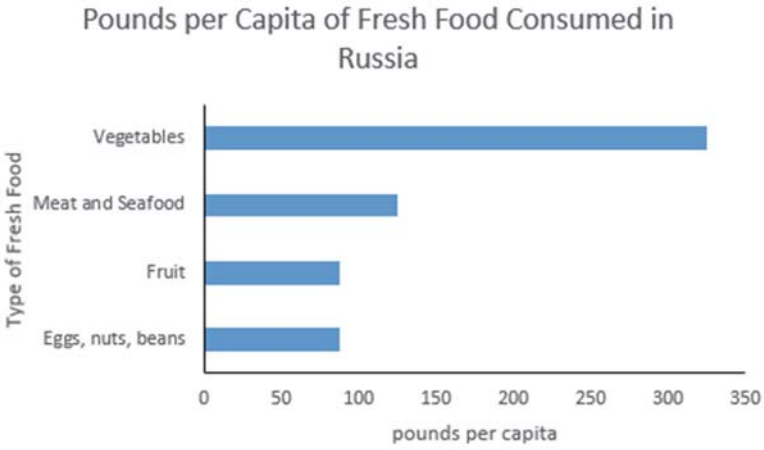


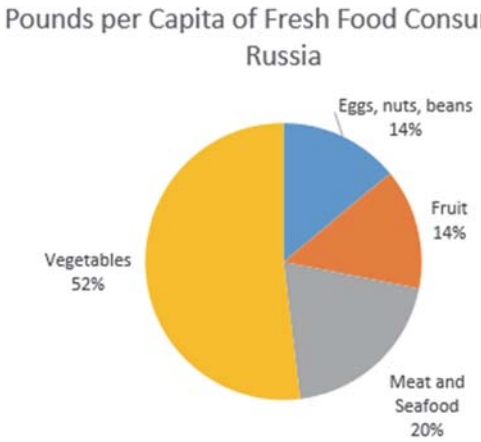


2.93 cont. (a) **Japan Fresh Food Consumed:**

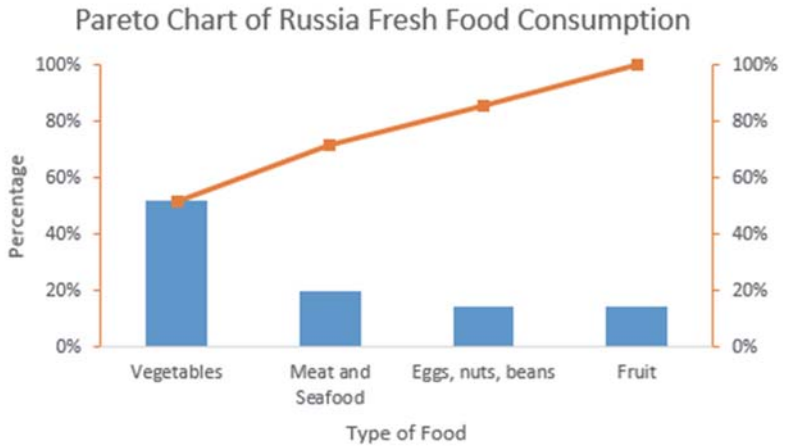


Russia Fresh Food Consumed:

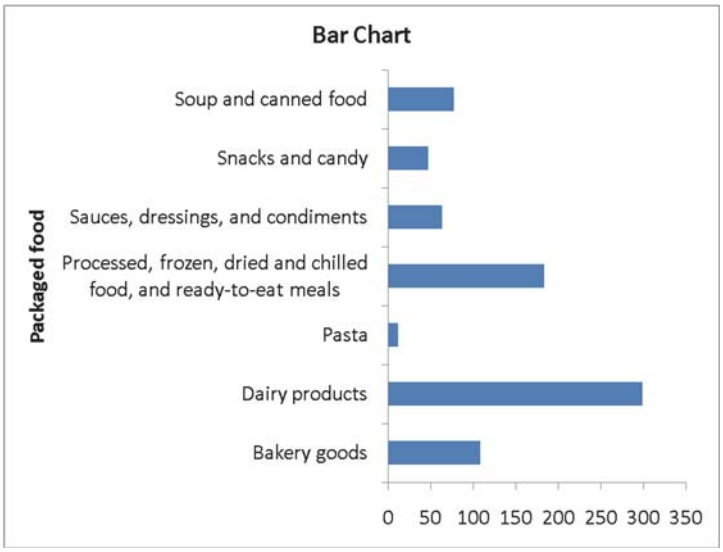


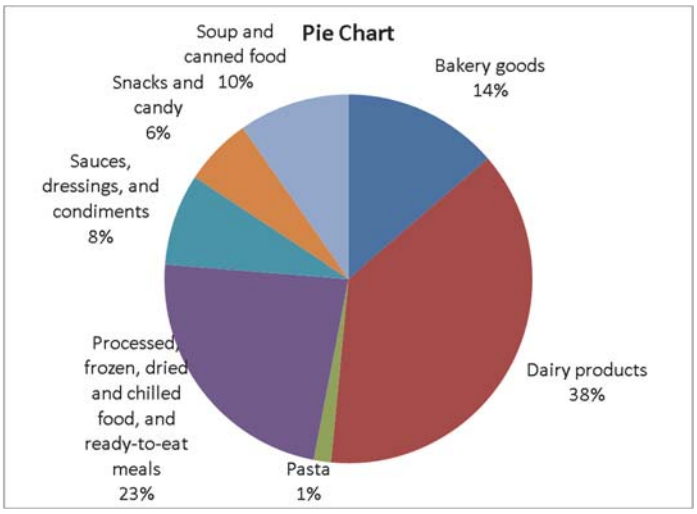


2.93 cont. (a) **Russia Fresh Food Consumed:**

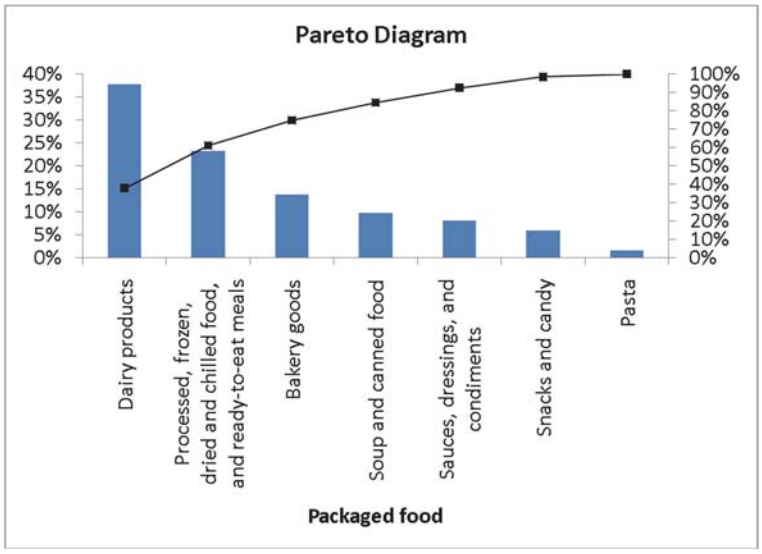


(b) **United States Packaged Food Consumed:**

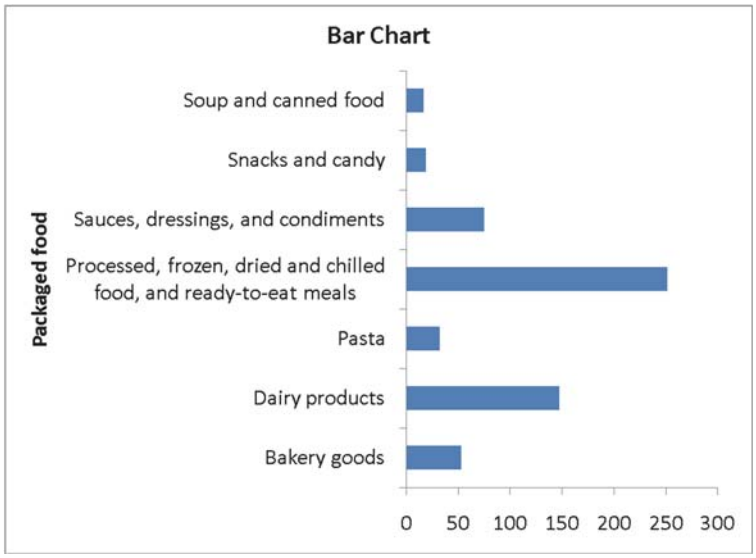




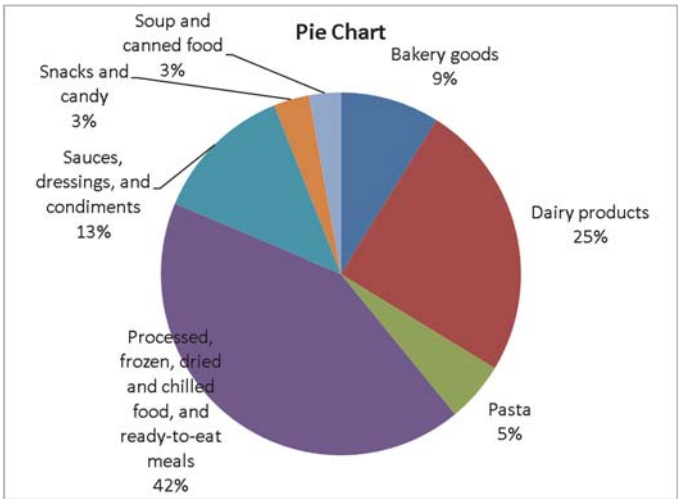
2.93 cont. (b) **United States Packaged Food Consumed:**

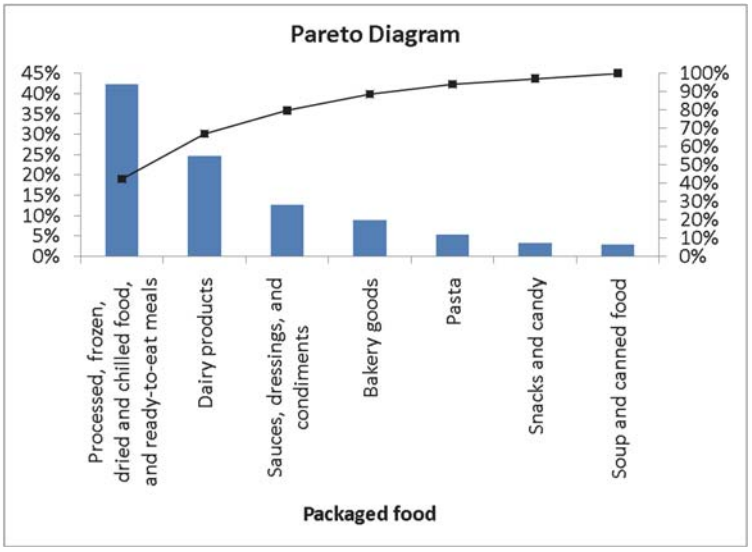


Japan Packaged Food Consumed:

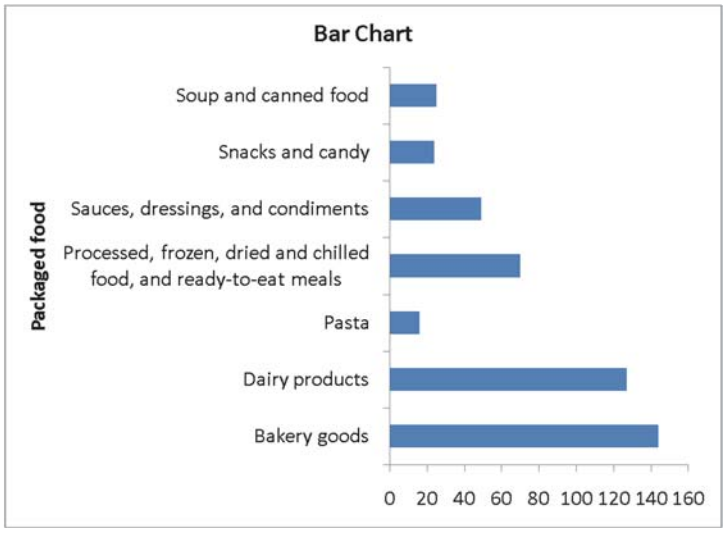


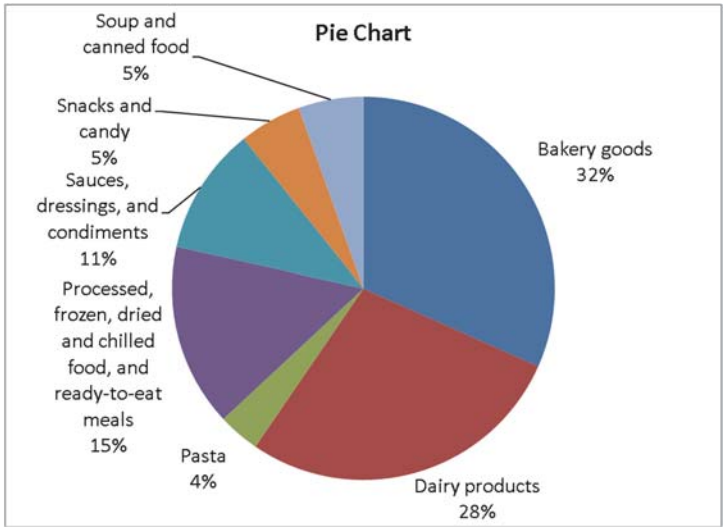
2.93 cont. (b) **Japan Packaged Food Consumed:**

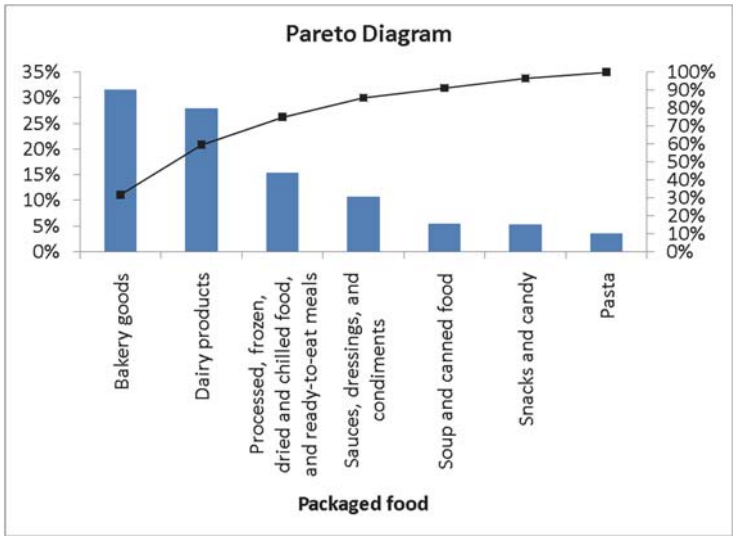




2.93 cont. (b) **Russian Packaged Food Consumed:**



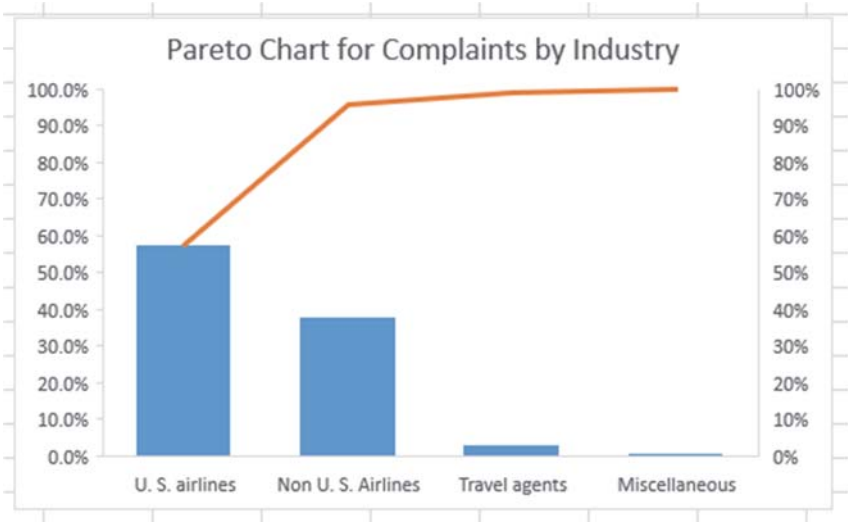




2.93 cont. (c) The fresh food consumption patterns between Japanese and Russians are quite similar with vegetables taking up the largest share followed by meats and seafood while Americans consume about the same amount of meats and seafood, and vegetables. Among the three countries, vegetables, and meats and seafood constitute more than 60% of the fresh food consumption.

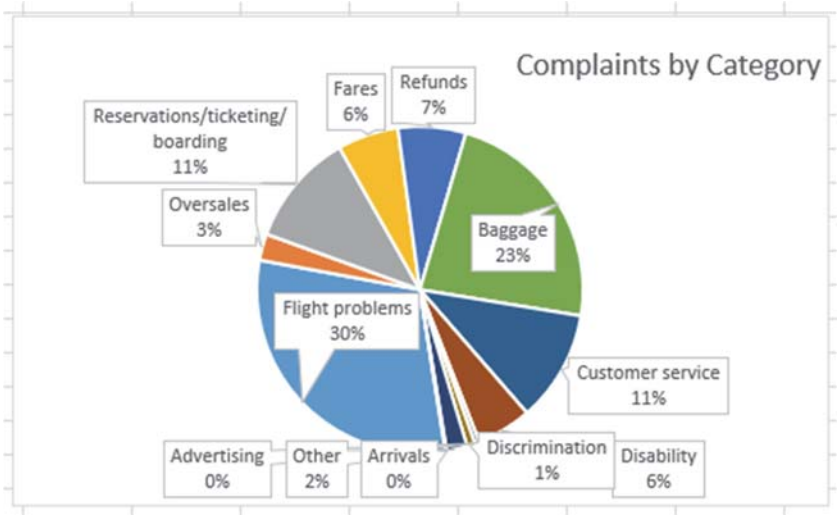
For Americans, dairy products, and processed, frozen, dried and chilled food and ready-to-eat meals make up slightly more than 60% of the packaged food consumption. For Japanese, processed, frozen, dried and chilled food, and ready-to-eat meals, and dairy products constitute more than 60% of their packaged food consumption. For the Russians, bakery goods and dairy products take up 60% of the share of their package food consumption.

2.94 (a)

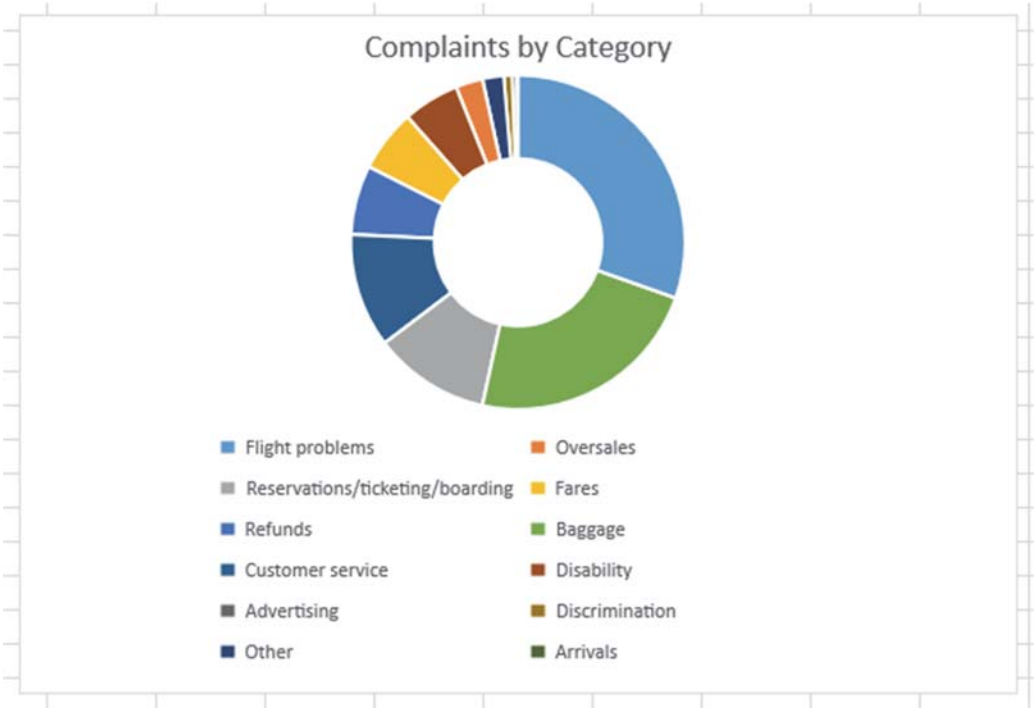


Most complaints were against U.S. airlines.

(b)



2.94 (b) cont.

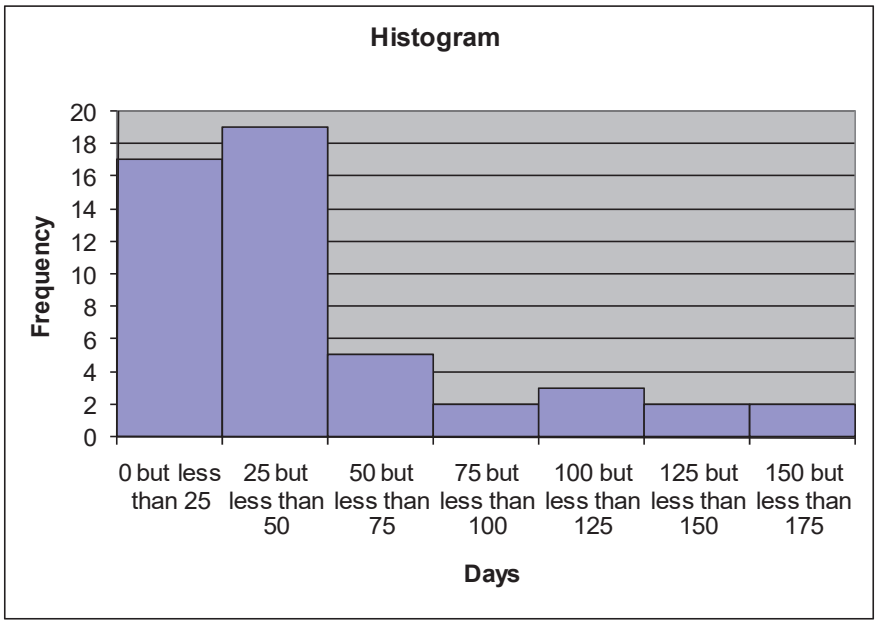


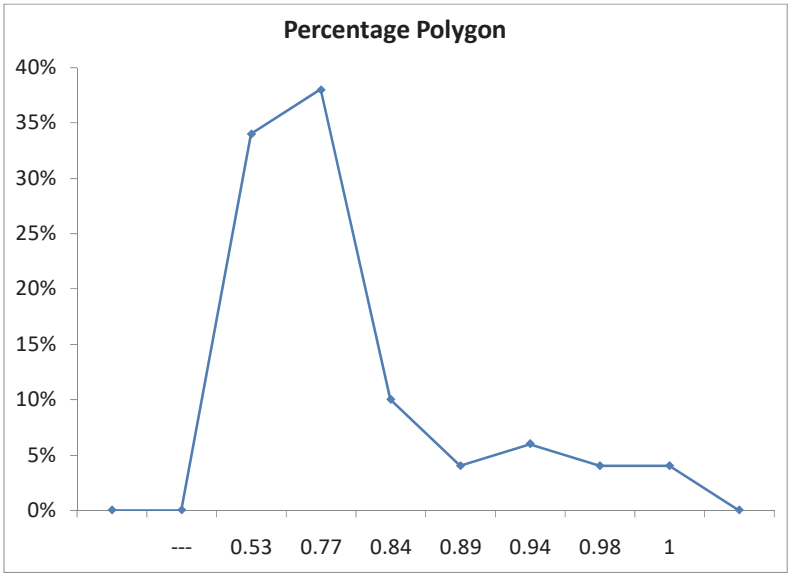
Most of the complaints were due to flight problems.

2.95 (a)

Range	*Frequency*	*Percentage*
0 but less than 25	17	34%
25 but less than 50	19	38%
50 but less than 75	5	10%
75 but less than 100	2	4%
100 but less than 125	3	6%
125 but less than 150	2	4%
150 but less than 175	2	4%

2.95 cont. (b)

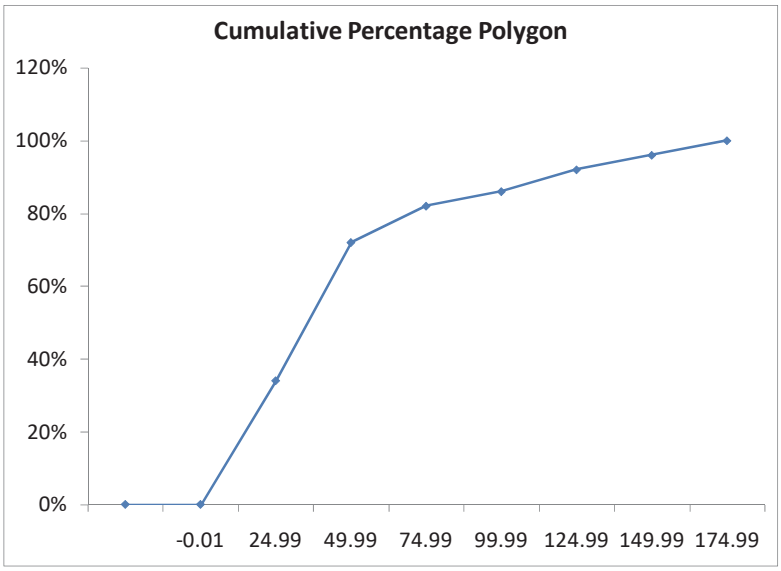




(c)

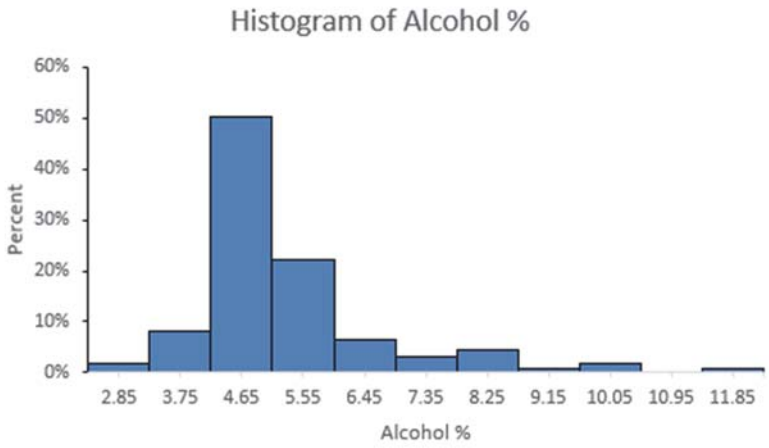
Range	*Cumulative %*
0 but less than 25	34%
25 but less than 50	72%
50 but less than 75	82%
75 but less than 100	86%
100 but less than 125	92%
125 but less than 150	96%
150 but less than 175	100%

2.95 (c) cont.

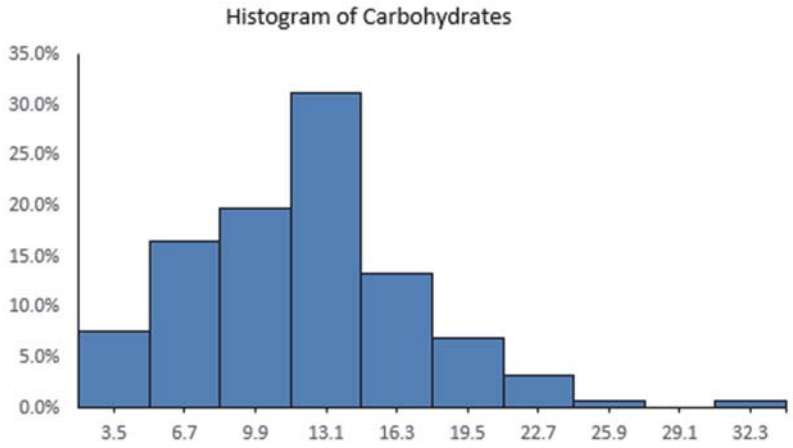


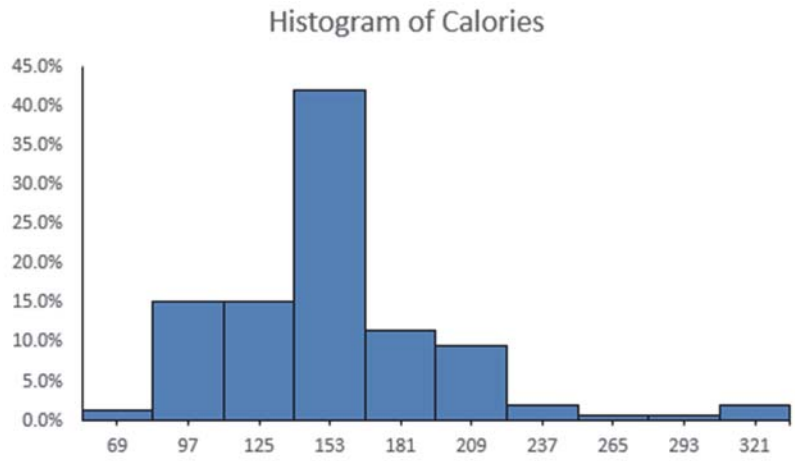
(d) You should tell the president of the company that over half of the complaints are resolved within a month, but point out that some complaints take as long as three or four months to settle.

2.96 (a)

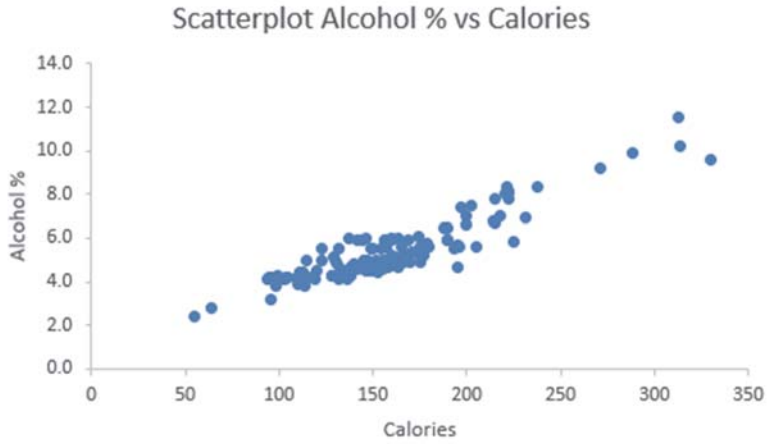


2.96 cont. (a)

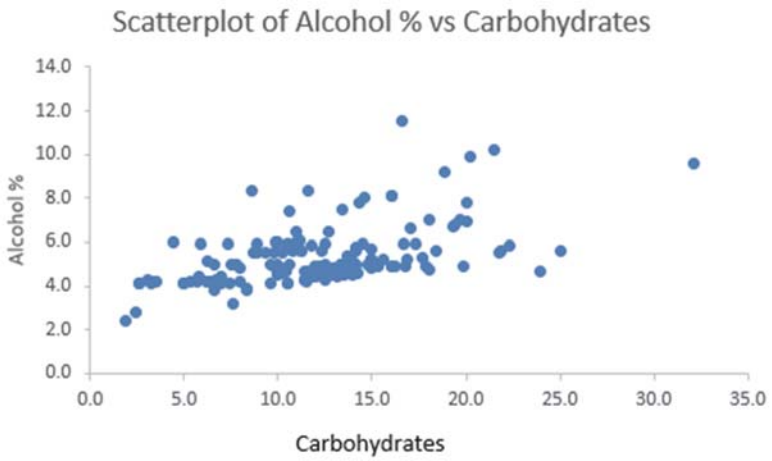


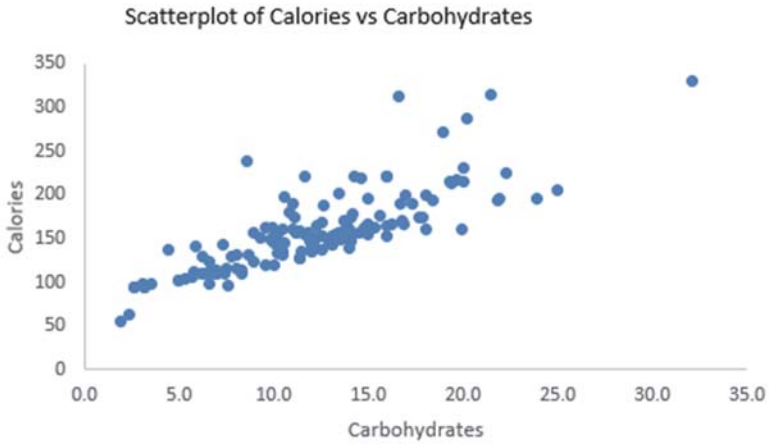


(b)



2.96 cont. (b)





(c) The alcohol percentage is concentrated between 4% and 6%, with more between 4% and 5%. The calories are concentrated between 140 and 160. The carbohydrates are concentrated between 8 and 15. There are outliers in the percentage of alcohol in both tails. There are a few beers with alcohol content as high as around 11.5%. There are a few beers with calorie content as high as around 320 and carbohydrates as high as 32. There is a positive relationship between percentage of alcohol and calories and between calories and carbohydrates, and positive relationship between percentage alcohol and carbohydrates.

2.97 (a) Ordered array of ratings of Super Bowl ads that ran before halftime

4.22	4.32	4.38	4.45	4.45	4.61	4.63	4.81	4.85	4.95
5.07	5.10	5.15	5.25	5.29	5.34	5.34	5.34	5.50	5.63
5.84	5.88	5.98	6.14	6.14	6.30	6.31	6.32	6.38	6.39
6.51									

Ordered array of ratings of Super Bowl ads that ran at or after halftime

3.63	3.97	4.50	4.55	4.59	4.84	4.96	4.96	5.04	5.11
5.14	5.29	5.52	5.59	5.64	5.80	5.84	6.11	6.18	6.37
6.41	6.43	6.95	7.05	7.07	7.34	7.69			

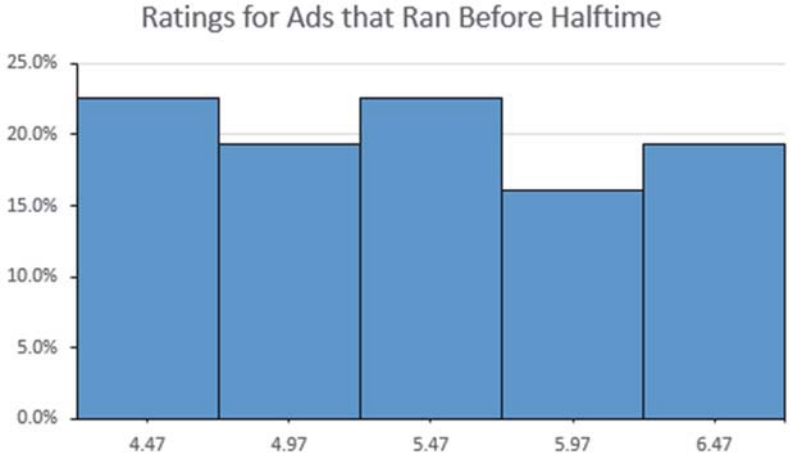
2.97 cont. (b) Stem-and-leaf display of ratings of Super Bowl ads that ran before halftime

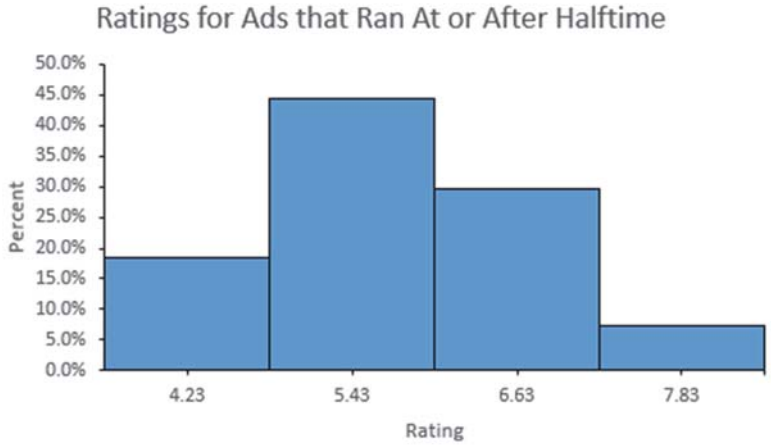
Stem	Leaf
4	234556689
5	0112333335689
6	011333445

Stem-and-leaf display of Super Bowl ads that ran at or after halftime

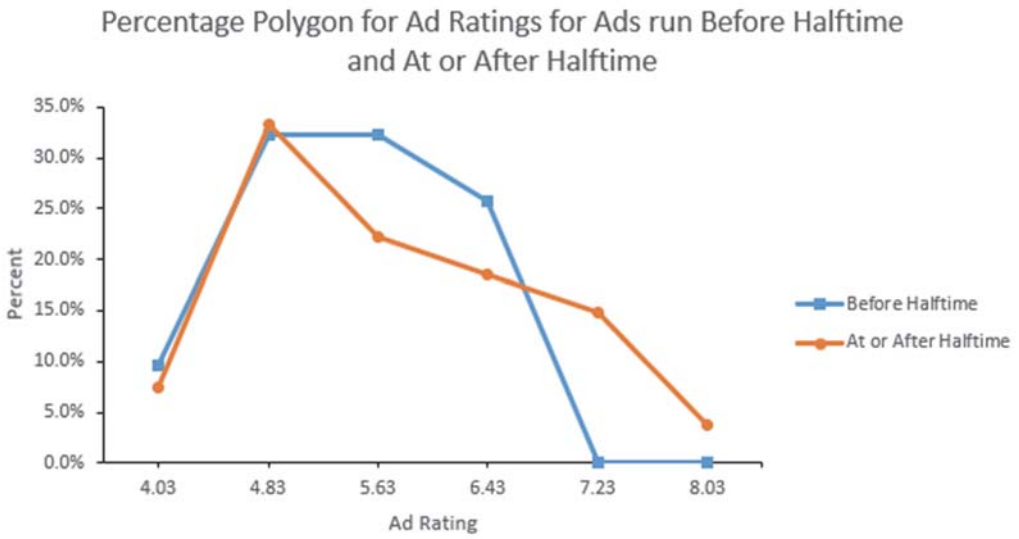
Stem	Leaf
3	6
4	05668
5	00011356688
6	12444
7	01137

(c)

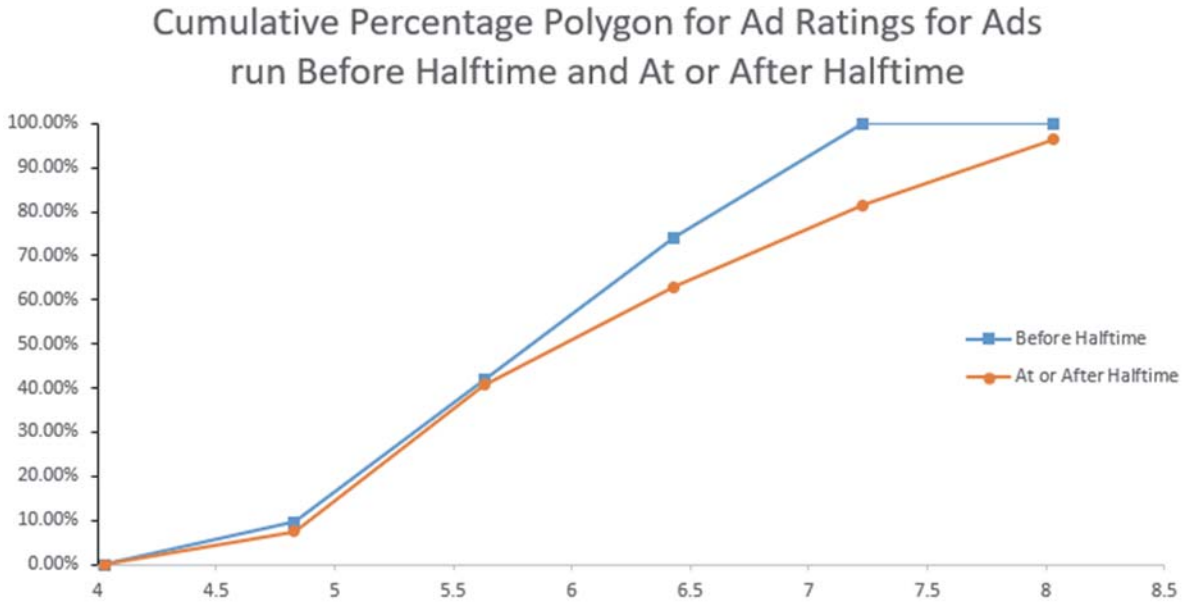




2.97 cont. (d)



(e)



(f) There is more variation in ratings for ads that ran at or after halftime then before halftime. Approximately 74% of the ads run before halftime had a rating of 5.63 or less while 63% of the ads run at or after halftime had a rating of 5.63 or lower. Five ads run at or after halftime had a rating of 7 or higher while none of the ads running before halftime had a rating that high.

2.98 (a) Stem-and -leaf of One Year CD, leaf unit = 0.1

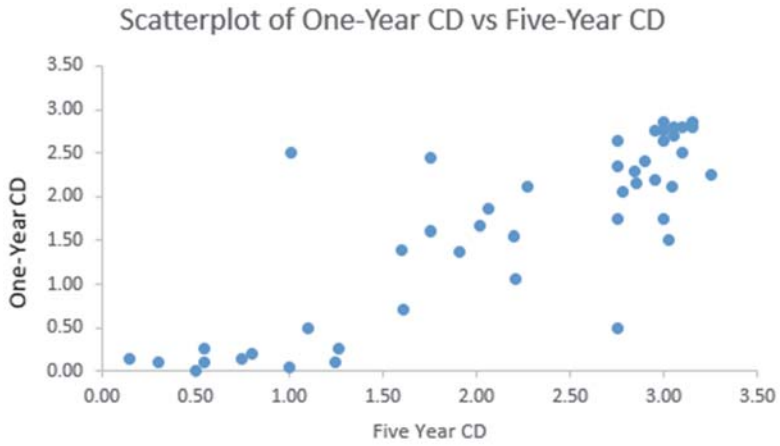
Stems	Leaf
0	01111
0	22233
0	55
0	7
0	
1	1
1	
1	445
1	667
1	889
2	111
2	2233
2	44555
2	777
2	88888899

Stem-and -leaf of five Year CD, leaf unit = 0.1

Stems	Leaf
0	
0	23
0	5
0	66
0	88
1	001
1	33
1	
1	66
1	889
2	01
2	223
2	
2	
2	88888899
3	0000000011111
3	223

(b) The yield of one-year CDs shows that most values are at least 1.0. The yield of five-year CDs shows that most values are at least 1.6.

2.98 cont. (c)

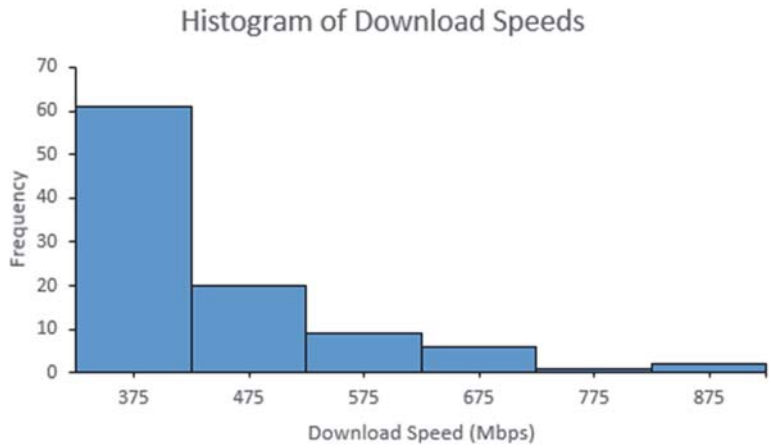


(d) There appears to be a positive relationship between the yield of the one-year CD and the five-year CD.

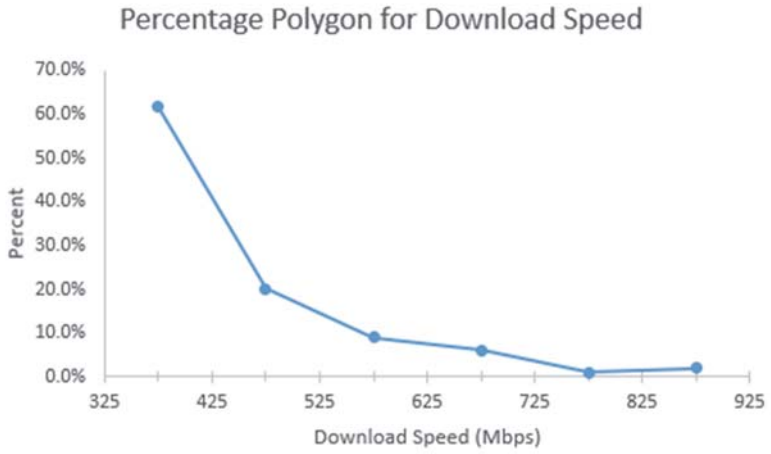
2.99 (a)

Download Speed (Mbs)	Frequency	Percentage
325 but less than 425	61	61.6%
425 but less than 525	20	20.2%
525 but less than 625	9	9.1%
625 but less than 725	6	6.1%
725 but less than 825	1	1.0%
825 but less than 925	2	2.0%
	99	100%

(b)

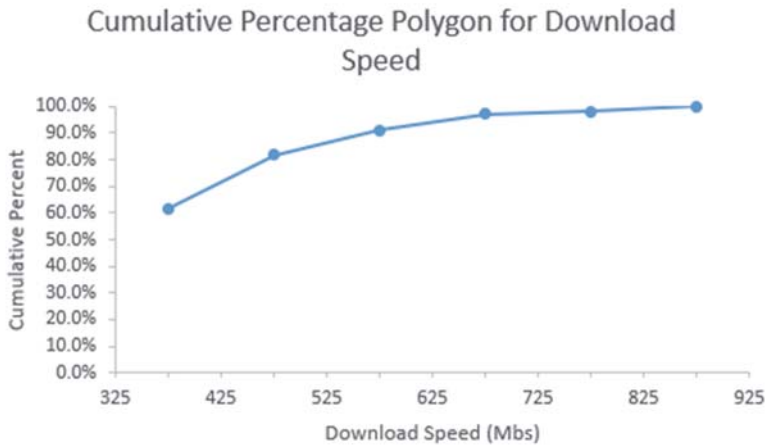


2.99 cont. (b)



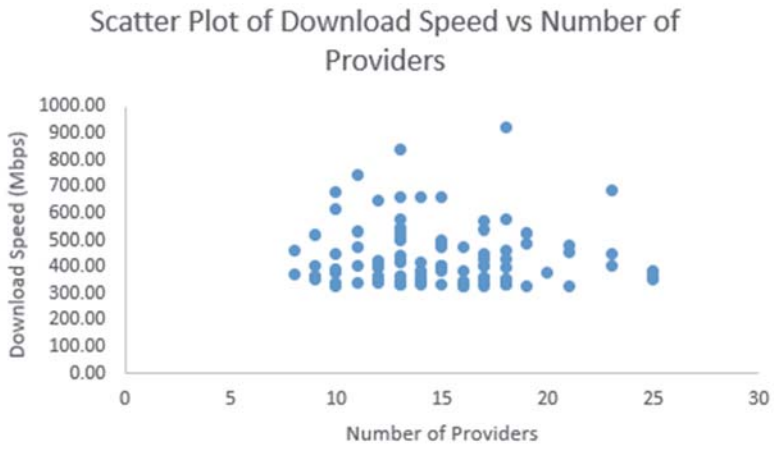
(c)

Download Speed (Mbs)	Frequency	Percentage	Cumulative Percentage
325 but less than 425	61	61.6%	61.6%
425 but less than 525	20	20.2%	81.8%
525 but less than 625	9	9.1%	90.9%
625 but less than 725	6	6.1%	97.0%
725 but less than 825	1	1.0%	98.0%
825 but less than 925	2	2.0%	100.0%
	99	100%	



(d) Approximately 80% of the cities have download speeds from 325 to 525 Mbs.

2.99 cont. (e)



(d) There does not appear to be any relationship between download speed and number of providers.

2.100 (a)

Frequencies (Boston)

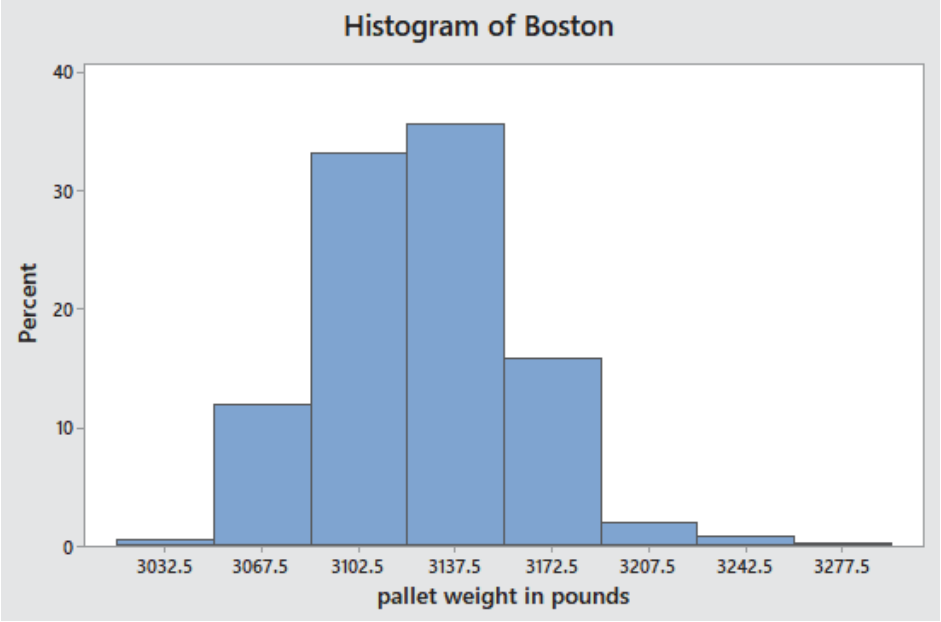
Weight (Boston)	*Frequency*	*Percentage*
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but less than 3295	1	0.27%

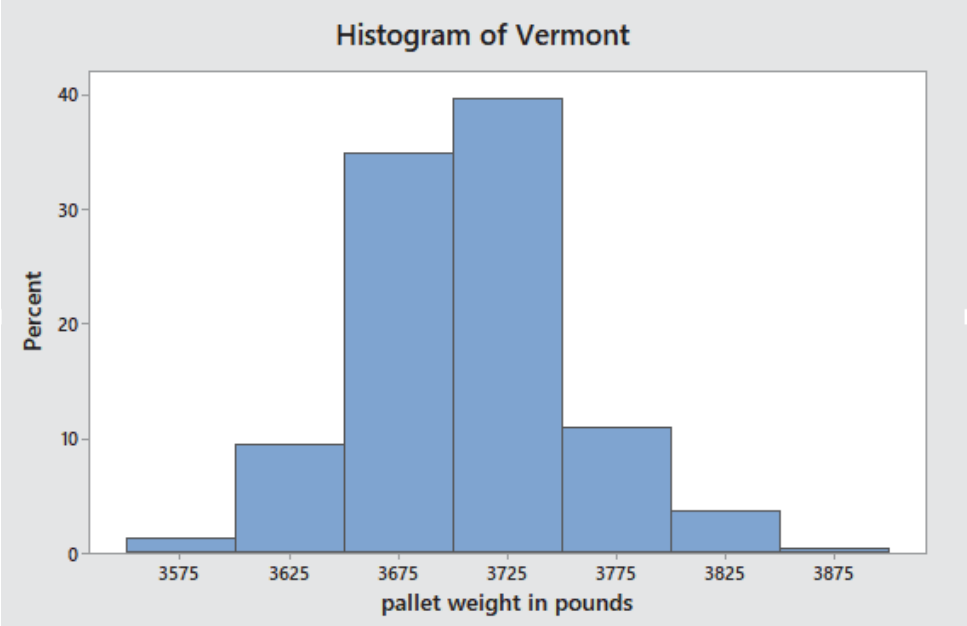
(b)

Frequencies (Vermont)

Weight (Vermont)	*Frequency*	*Percentage*
3550 but less than 3600	4	1.21%
3600 but less than 3650	31	9.39%
3650 but less than 3700	115	34.85%
3700 but less than 3750	131	39.70%
3750 but less than 3800	36	10.91%
3800 but less than 3850	12	3.64%
3850 but less than 3900	1	0.30%

2.100 (c) cont.





(d) 0.54% of the "Boston" shingles pallets are underweight while 0.27% are overweight. 1.21% of the "Vermont" shingles pallets are underweight while 3.94% are overweight.

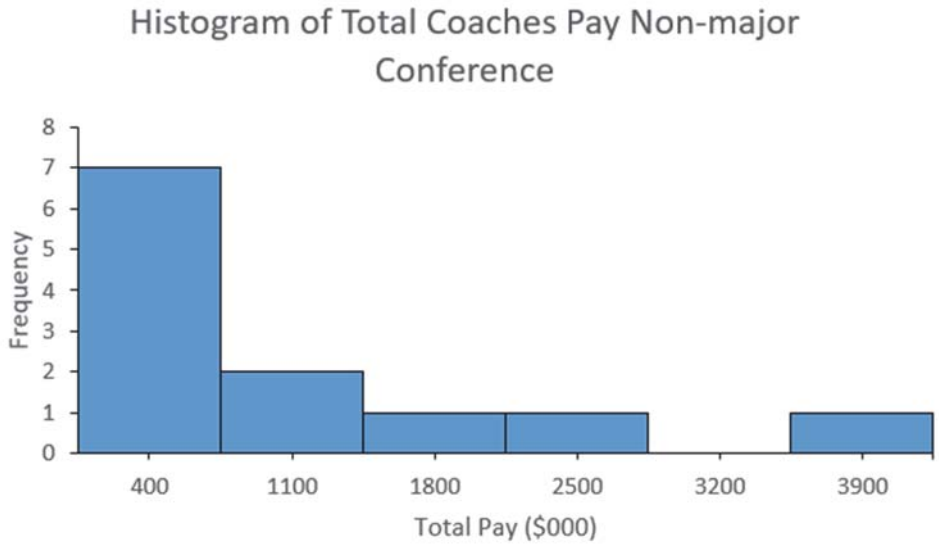
2.101 (a) Not member of major conference

Total Pay ($000)	Frequency	Percentage
50 but less than750	7	58%
750 but less than 1450	2	17%
1450 but less than 2150	1	8%
2150 but less than 2850	1	8%
2850 but less than 3550	0	0%
3550 but less than 4250	1	8%
	12	100%

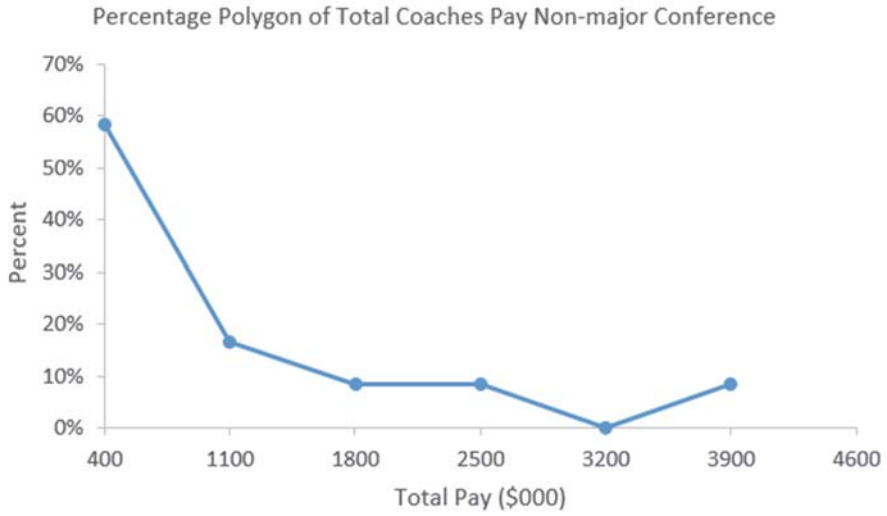
Member of major conference

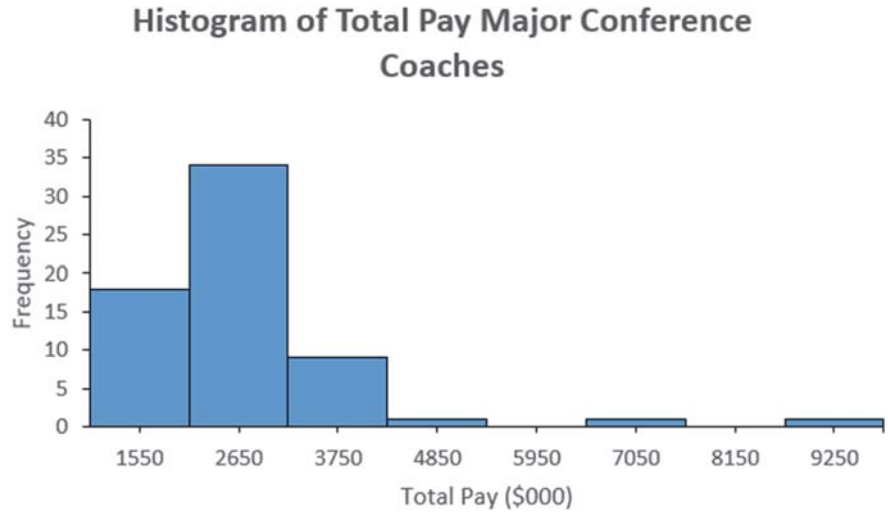
Total Pay ($000)	Frequency	Percentage
1000 but less than 2100	18	28.1%
2100 but less than 3200	34	53.1%
3200 but less than 4300	9	14.1%
4300 but less than 5400	1	1.6%
5400 but less than 6500	0	0.0%
6500 but less than 7600	1	1.6%
7600 but less than 8700	0	0.0%
8700 but less than 9800	1	1.6%
	64	100.0%

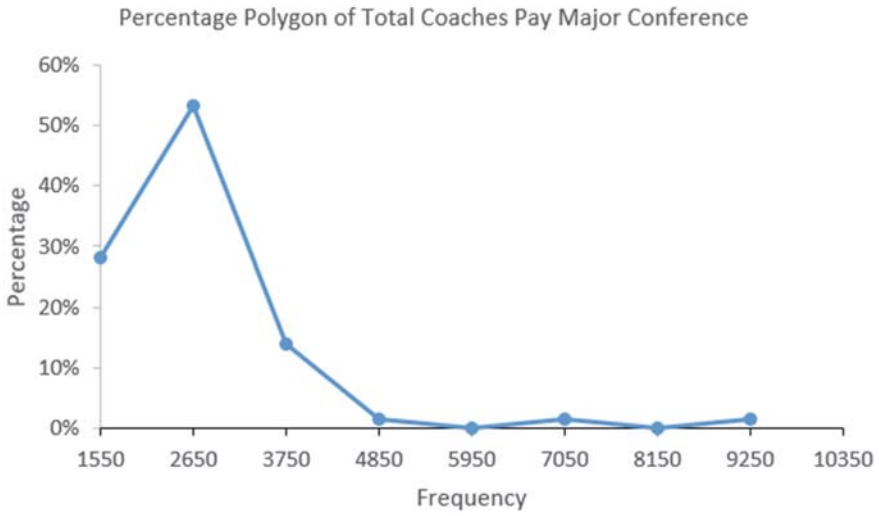
(b)



2.101 (b)
cont.

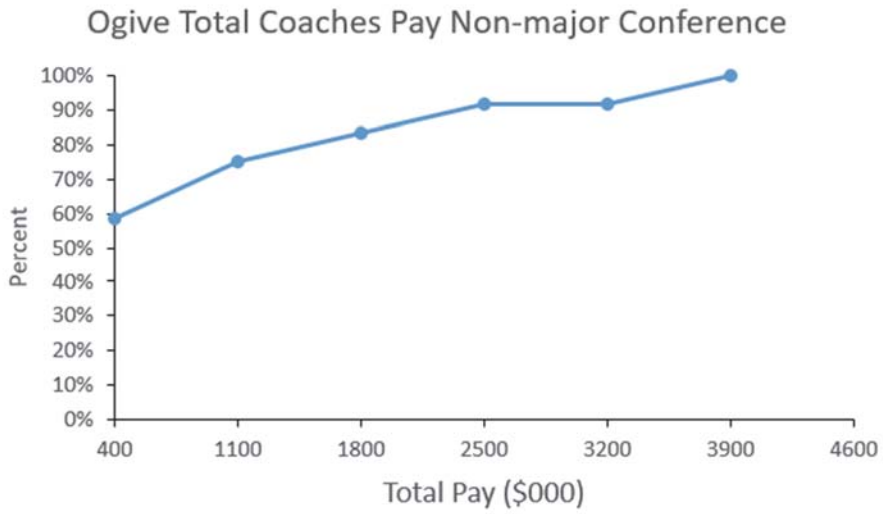






2.101 cont. (c) Not member of major conference

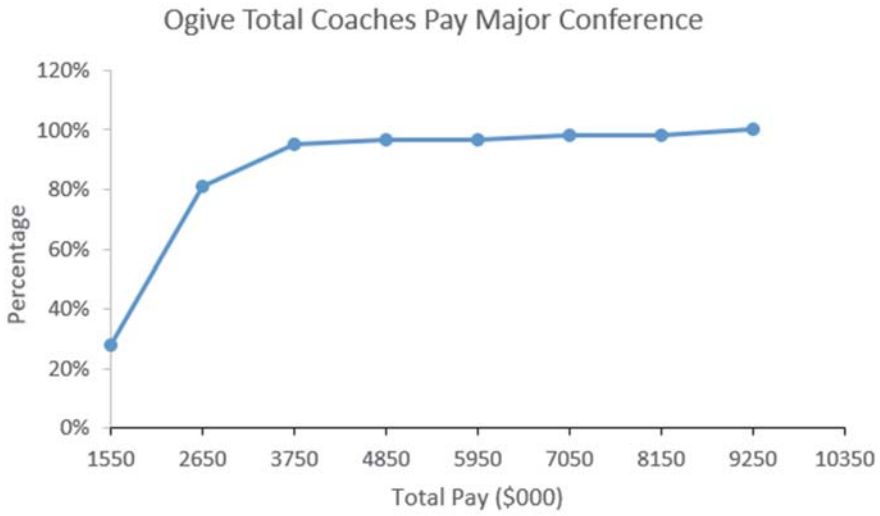
Total Pay ($000)	Frequency	Percentage	Cumulative Percentage
50 but less than750	7	58%	58%
750 but less than 1450	2	17%	75%
1450 but less than 2150	1	8%	83%
2150 but less than 2850	1	8%	92%
2850 but less than 3550	0	0%	92%
3550 but less than 4250	1	8%	100%



Member of major conference

Total Pay ($000)	Frequency	Percentage	Cumulative Percentage
1000 but less than 2100	18	28.1%	28.1%
2100 but less than 3200	34	53.1%	81.3%
3200 but less than 4300	9	14.1%	95.3%
4300 but less than 5400	1	1.6%	96.9%
5400 but less than 6500	0	0.0%	96.9%
6500 but less than 7600	1	1.6%	98.4%
7600 but less than 8700	0	0.0%	98.4%
8700 but less than 9800	1	1.6%	100.0%
	64	100.0%	

2.101 cont. (c)



(d)

Coaches Pay -- Not in a Major Conference

Stem unit: 1000

0	13444668
1	28
2	2
3	6

Coaches Pay -- Major Conference

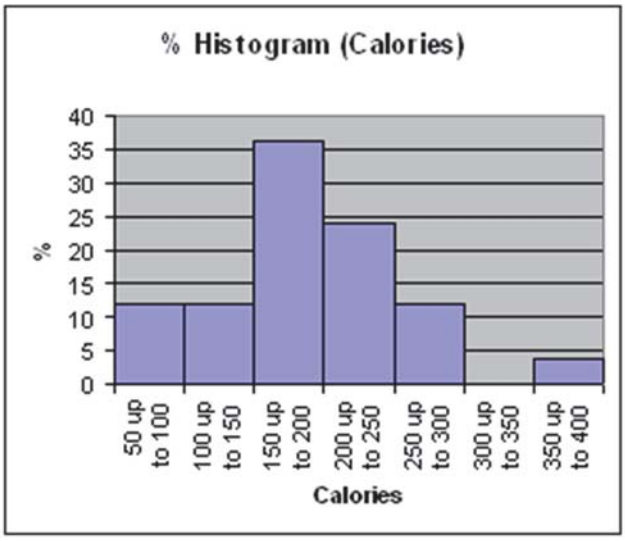
Stem unit: 1000

1	0445556688889
2	000112233444455566666777888889
3	00011222336899
4	0124
5	
6	
7	0
8	
9	3

(e) The majority of coaches in the non-major conference earn under 1 million dollars while the majority coaches in a major conference earn between 1 and 3 million dollars per year. The highest paid coach in a major conference is $9,277,000 while the highest paid coach in the non-major conference is $3,570,000.

2.102 (a)

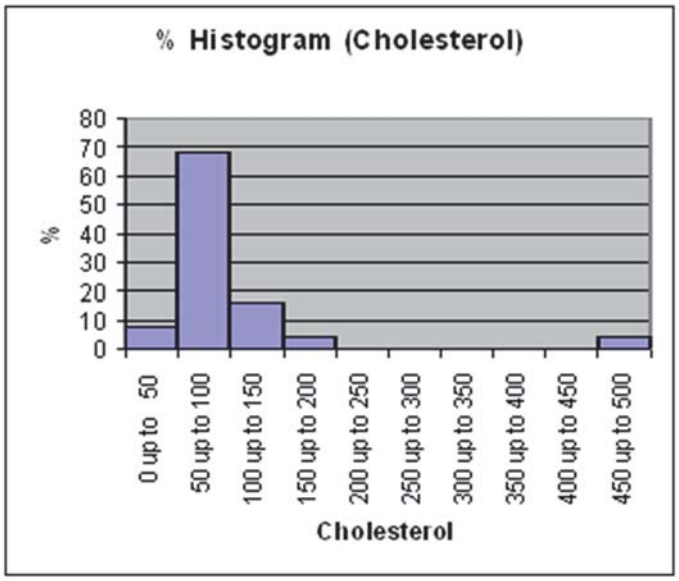
Calories	*Frequency*	*Percentage*	*Percentage Less Than*
50 up to 100	3	12%	12%
100 up to 150	3	12	24
150 up to 200	9	36	60
200 up to 250	6	24	84
250 up to 300	3	12	96
300 up to 350	0	0	96
350 up to 400	1	4	100



(b)

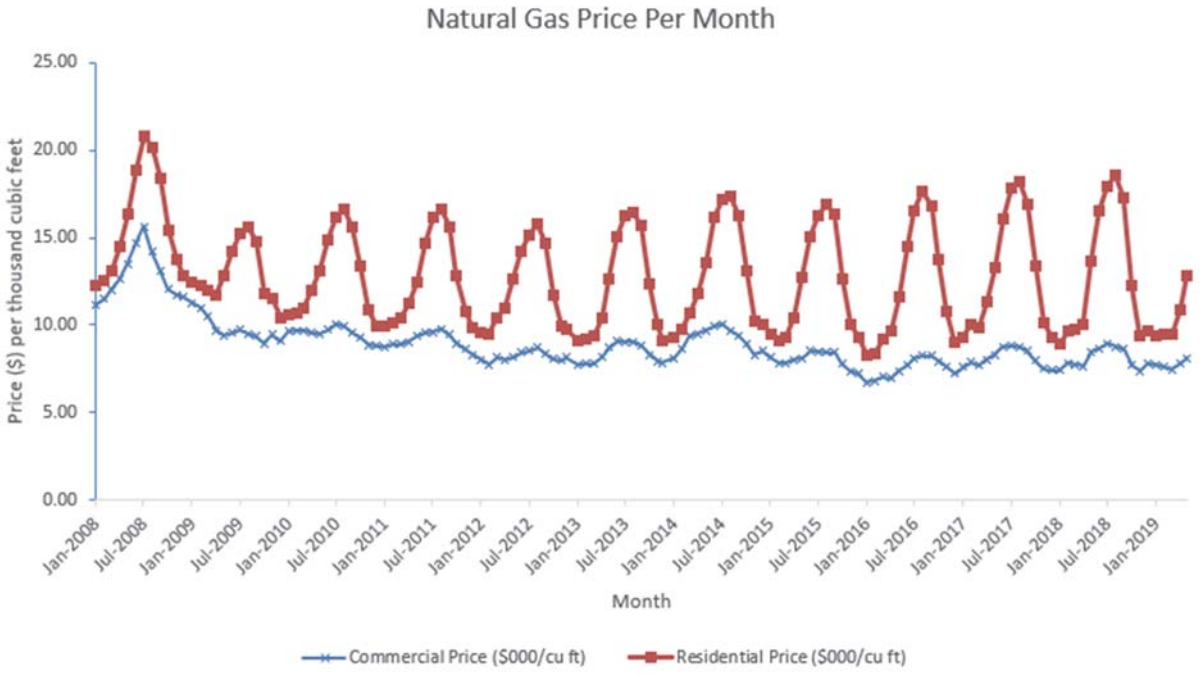
Cholesterol	*Frequency*	*Percentage*	*Percentage Less Than*
0 up to 50	2	8	8%
50 up to 100	17	68	76
100 up to 150	4	16	92
150 up to 200	1	4	96
200 up to 250	0	0	96
250 up to 300	0	0	96
300 up to 350	0	0	96
350 up to 400	0	0	96
400 up to 450	0	0	96
450 up to 500	1	4	100

2.102 (b)
cont.

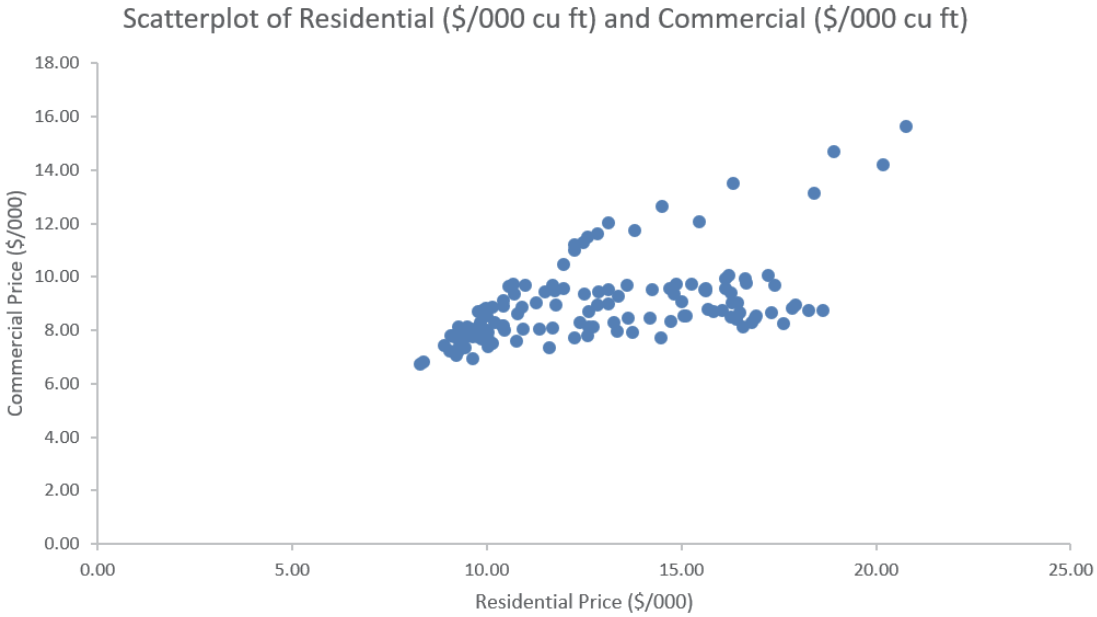


(c) The sampled fresh red meats, poultry, and fish vary from 98 to 397 calories per serving, with the highest concentration between 150 to 200 calories. One protein source, spareribs, with 397 calories, is more than 100 calories above the next highest caloric food. The protein content of the sampled foods varies from 16 to 33 grams, with 68% of the data values falling between 24 and 32 grams. Spareribs and fried liver are both very different from other foods sampled—the former on calories and the latter on cholesterol content.

2.103 (a)

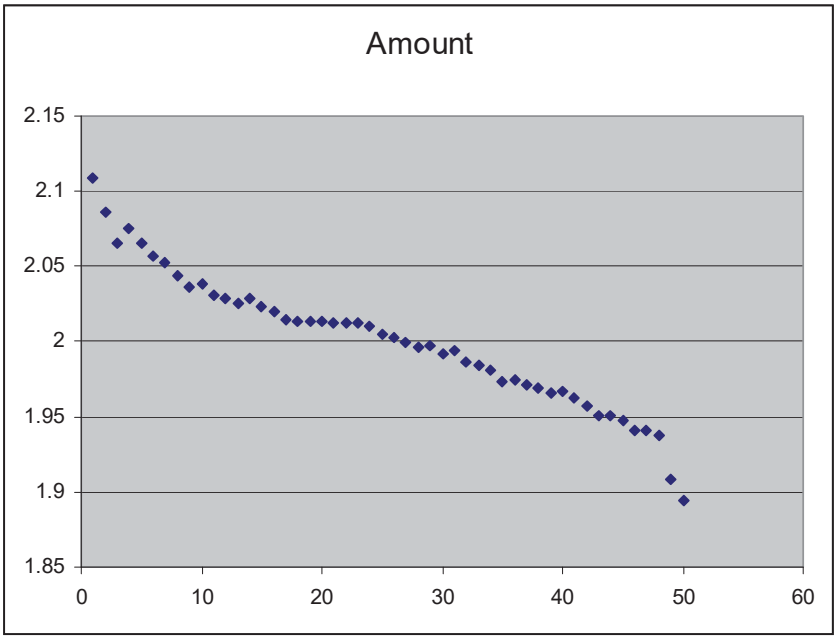


2.103 cont. (b) The commercial average price was highest in the summer of 2008 and has since declined. The residential average price of gasoline in the United States is higher in the summer in general.



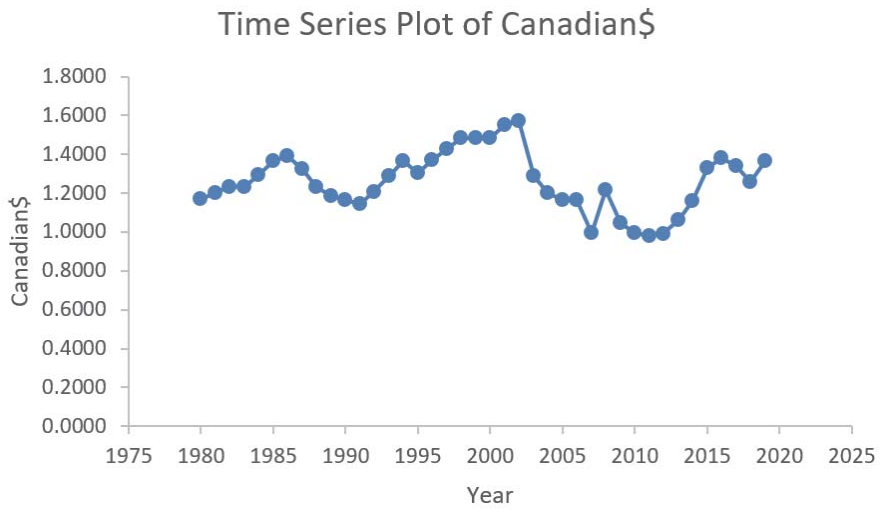
(d) There appears to be a slight positive relationship between the commercial price and residential price.

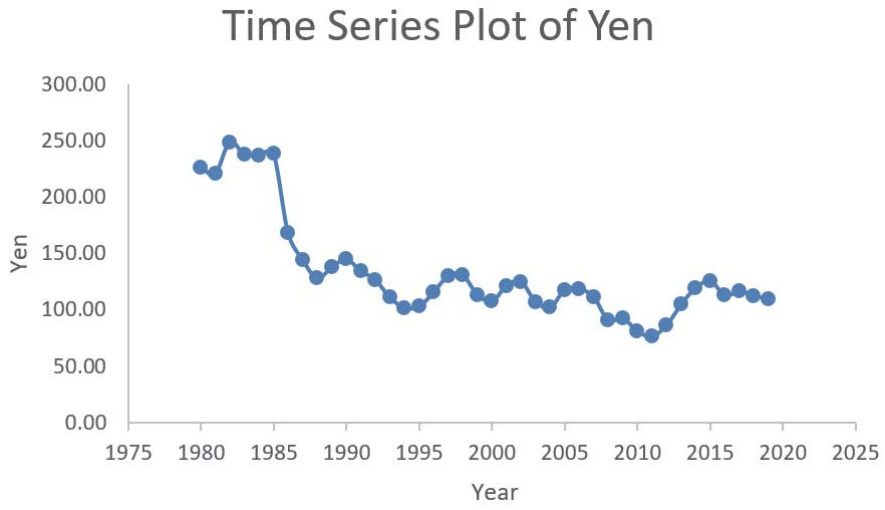
2.104 (a)

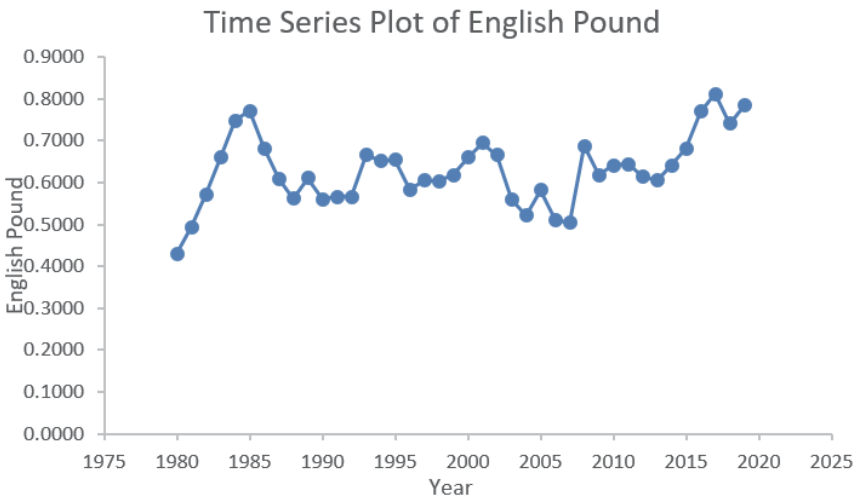


(b) There is a downward trend in the amount filled.
(c) The amount filled in the next bottle will most likely be below 1.894 liter.
(d) The scatter plot of the amount of soft drink filled against time reveals the trend of the data, whereas a histogram only provides information on the distribution of the data.

2.105 (a)



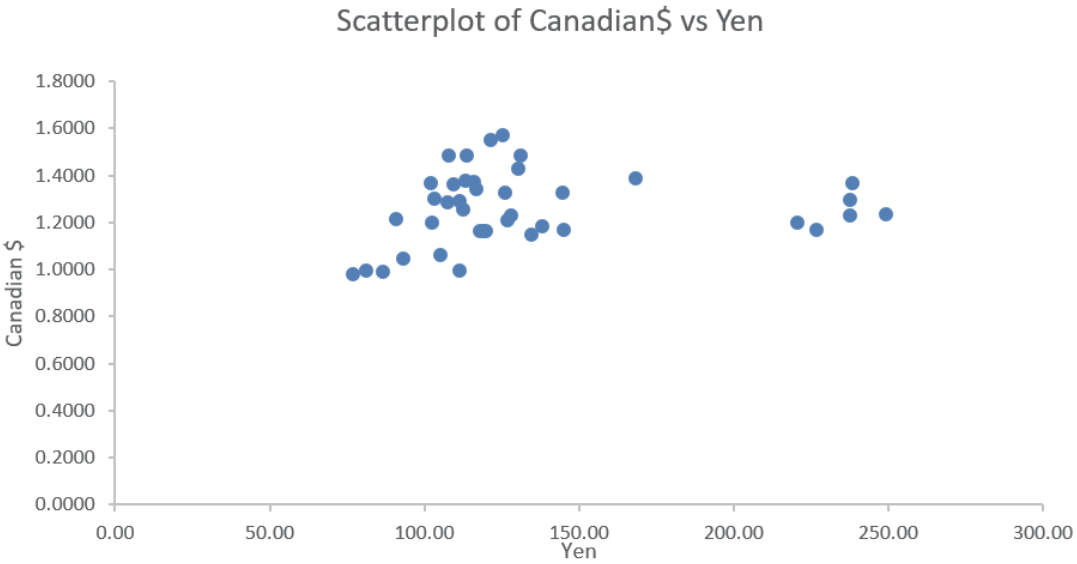


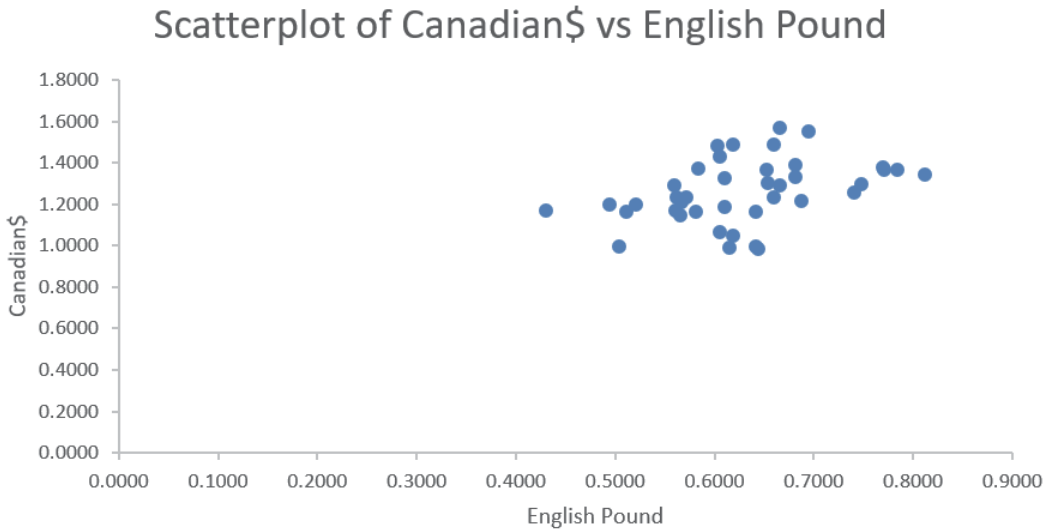


2.105 cont. (b) The Japanese yen had depreciated against the U.S. dollar since 1985 while the Canadian dollar appreciated gradually from 1980 to 1987 and from 1991 to 2002 and then started to depreciate until 2011. The English pound to U.S. dollar's exchange rate has been quite stable since 1983.

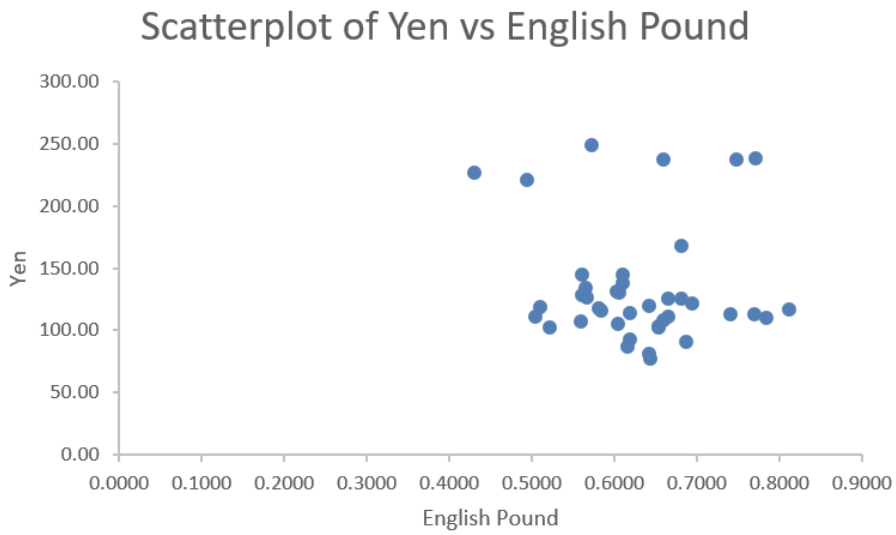
(c) The U.S. dollar has appreciated against the Japanese yen since 1980 and appreciated against the Canadian dollar since 2002 in general while the exchange rate against the English bound has been stable in general.

(d)





2.105 (d) cont.

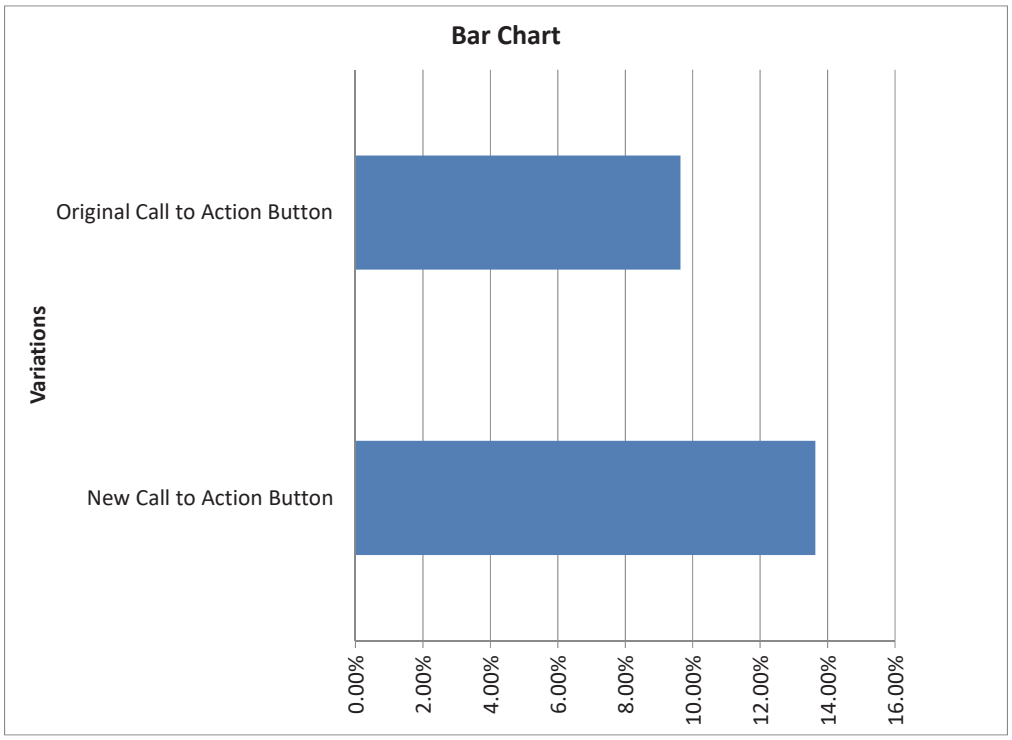


(e) There is not any obvious relationship between the Canadian dollar and Japanese yen in terms of the U.S. dollar nor any relationship between the Japanese yen and English pound. There is a slightly positive relationship between the Canadian dollar and English pound.

2.106 (a)

Variations	Percentage of Download
Original Call to Action Button	9.64%
New Call to Action Button	13.64%

(b)

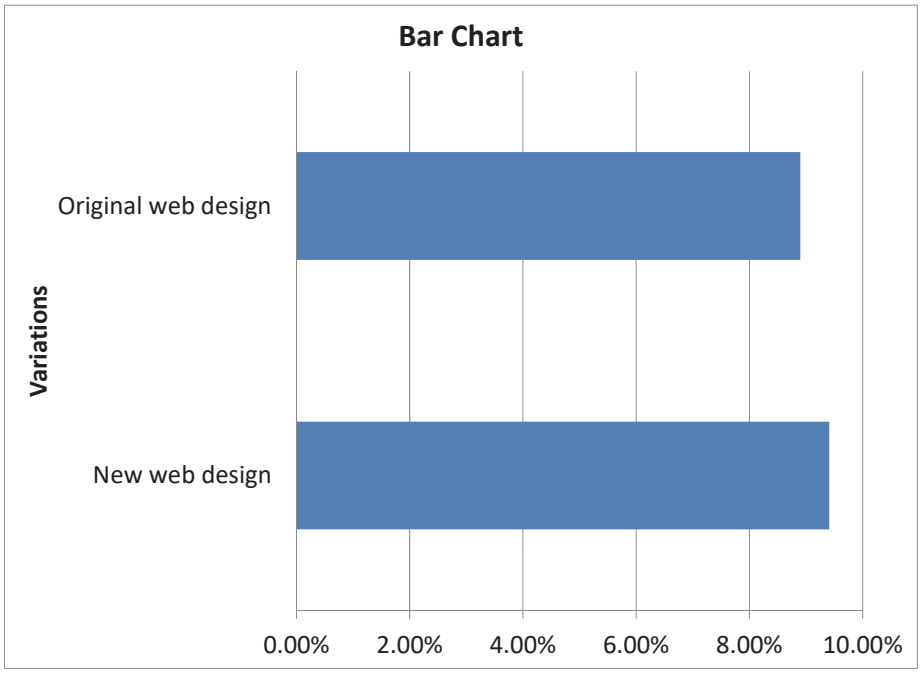


2.106 cont. (c) The New Call to Action Button has a higher percentage of downloads at 13.64% when compared to the Original Call to Action Button with a 9.64% of downloads.

(d)

Variations	Percentage of Download
Original web design	8.90%
New web design	9.41%

(e)



(f) The New web design has only a slightly higher percentage of downloads at 9.41% when compared to the Original web design with an 8.90% of downloads.

(g) The New web design is only slightly more successful than the Original web design while the New Call to Action Button is much more successful than the Original Call to Action Button with about 41% higher percentage of downloads.

(h)

Call to Action Button	Web Design	Percentage of Download
Old	Old	8.30%
New	Old	13.70%
Old	New	9.50%
New	New	17.00%

(i) The combination of the New Call to Action Button and the New web design results in slightly more than twice as high a percentage of downloads than the combination of the Old Call to Action Button and Old web design.

(j) The New web design is only slightly more successful than the Original web design while the New Call to Action Button is much more successful than the Original Call to Action Button with about 41% higher percentage of downloads. However, the combination of the New Call to Action Button and New web design results in more than twice as high a percentage of downloads than the combination of the Old Call to Action Button and Old web design.

2.107 Class project – answers will vary depending on student responses.

2.108 Class project – answers will vary depending on student responses.

2.109 A descriptive analysis of the weight of the pallets of the Boston shingles revealed that the average weight was 3124.2 pounds with a standard deviation of 34.7. The average weight of 3124.2 pounds was 74.2 pounds above the expected minimum weight of 3,050 pounds. An analysis of the Vermont shingles revealed that the average weight was 3704.0 pounds with a standard deviation of 46.7. The average weight of 3704.0 pounds was 104 pounds above the expected minimum weight of 3,600 pounds. The below table includes a number of descriptive statistics for the two shingle types.

Descriptive Statistics: Boston, Vermont

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Skewness
Boston	368	0	3124.2	1.81	34.7	3044.0	3098.0	3122.0	3146.0	3266.0	0.53
Vermont	330	38	3704.0	2.57	46.7	3566.0	3670.0	3704.0	3732.0	3856.0	0.29

A frequency distribution of the Boston shingles revealed that 0.54% of the pallets were underweight and 0.27% were overweight. A frequency distribution of the Vermont shingles revealed that 1.21% of the shingles were underweight and 3.94% were overweight. The complete results are provided in the below frequency distributions.

Frequencies (Boston)

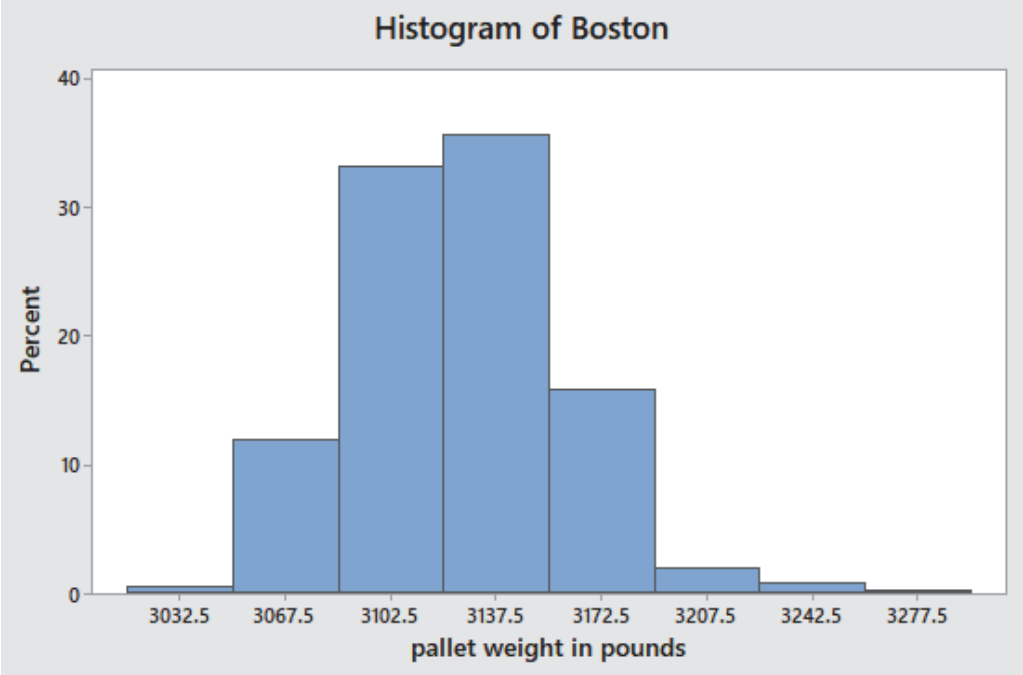
Weight (Boston)	*Frequency*	*Percentage*
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but less than 3295	1	0.27%

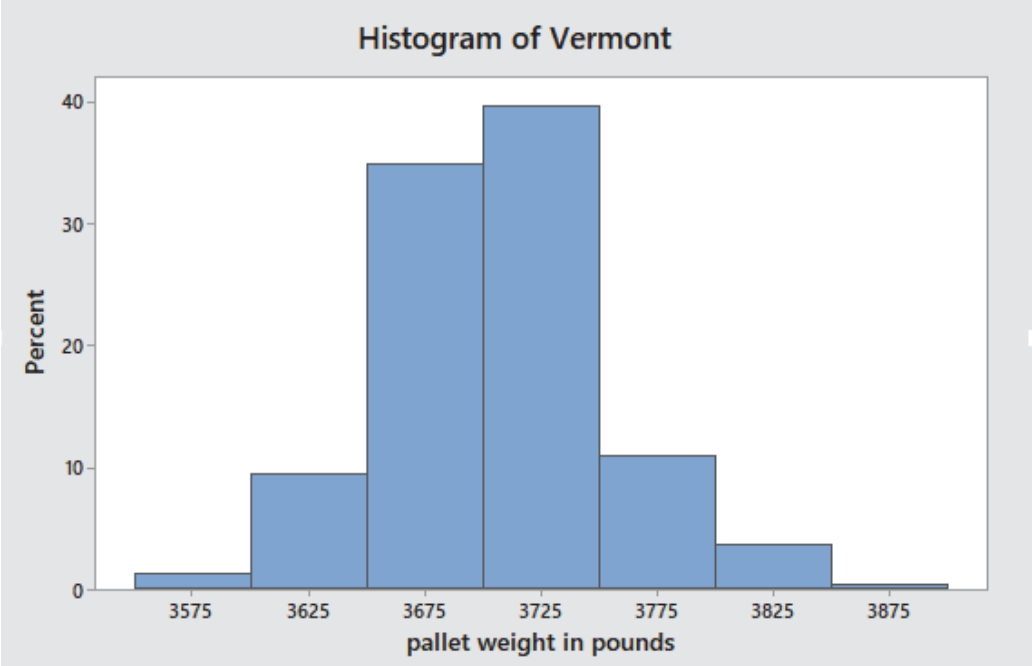
Frequencies (Vermont)

Weight (Vermont)	*Frequency*	*Percentage*
3550 but less than 3600	4	1.21%
3600 but less than 3650	31	9.39%
3650 but less than 3700	115	34.85%
3700 but less than 3750	131	39.70%
3750 but less than 3800	36	10.91%
3800 but less than 3850	12	3.64%
3850 but less than 3900	1	0.30%

Histogram graphs of the Boston shingles and the Vermont shingles, shown below, revealed that the weights of the pallets appeared to be consistent with a normal distribution. In both cases, there was slight right skewness with the Boston shingles having slightly more right skewness than the Vermont shingles.

2.109
cont.





The results of the above analyses reveal that both shingle types generally met pallet weight expectations with less than 1% of the Boston shingles weighing outside of the expected parameters and just over 5% of the Vermont shingles weighing outside of the expected parameters. The results suggest that the manufacturer should consider implementation of parameter compliance strategies for the Vermont shingles.

Chapter 3

3.1 (a) Excel output:

X	
Mean	6
Median	7
Mode	#N/A
Standard Deviation	2.915476
Sample Variance	8.5
Range	7
Minimum	2
Maximum	9
Sum	30
Count	5
First Quartile	3
Third Quartile	8.5
Interquartile Range	5.5
Coefficient of Variation	48.5913%

Mean = 6 Median = 7 There is no mode.

(b) Range = 7 Variance = 8.5

Standard deviation = 2.9 Coefficient of variation = (2.915/6) • 100% = 48.6%

(c) Z scores: 0.343, –0.686, 1.029, 0.686, –1.372

None of the Z scores is larger than 3.0 or smaller than –3.0. There is no outlier.

(d) Since the mean is less than the median, the distribution is left-skewed.

3.2 (a) Excel output:

X	
Mean	7
Median	7
Mode	7
Standard Deviation	3.286335
Sample Variance	10.8
Range	9
Minimum	3
Maximum	12
Sum	42
Count	6
First Quartile	4
Third Quartile	9
Interquartile Range	5
Coefficient of Variation	46.9476%

Mean = 7 Median = 7 Mode = 7

(b) Range = 9 Variance = 10.8

Standard deviation = 3.286

Coefficient of variation = (3.286/7) • 100% = 46.948%

3.2 cont. (c) Z scores: 0, –0.913, 0.609, 0, –1.217, 1.522
None of the Z scores is larger than 3.0 or smaller than –3.0. There is no outlier.

(d) Since the mean equals the median, the distribution is symmetrical.

3.3 (a) Excel output:

X	
Mean	6
Median	7
Mode	7
Standard Deviation	4
Sample Variance	16
Kurtosis	–0.34688
Minimum	0
Maximum	12
Sum	42
Count	7
First Quartile	3
Third Quartile	9
Interquartile Range	6
Coefficient of Variation	66.6667%

Mean = 6 Median = 7 Mode = 7

(b) Range = 12 Variance = 16 Standard deviation = 4
Coefficient of variation = (4/6) • 100% = 66.67%

(c) Z scores: 1.5, 0.25, –0.5, 0.75, –1.5, 0.25, –0.75. There is no outlier.

(d) Since the mean is less than the median, the distribution is left-skewed.

3.4 Excel output:

X	
Mean	2
Median	7
Mode	7
Standard Deviation	7.874007874
Sample Variance	62
Range	17
Minimum	–8
Maximum	9
Sum	10
Count	5
First Quartile	–6.5
Third Quartile	8
Interquartile Range	14.5
Coefficient of Variation	393.7004%

(a) Mean = 2 Median = 7 Mode = 7

(b) Range = 17 Variance = 62
Standard deviation = 7.874 Coefficient of variation = (7.874/2) • 100% = 393.7%

3.4 cont. (c) Z scores: 0.635, –0.889, –1.270, 0.635, 0.889. No outliers.

(d) Since the mean is less than the median, the distribution is left-skewed.

3.5 $\bar{R}_G = \left[(1+0.1)(1+0.3)\right]^{1/2} - 1 = 19.58\%$

3.6 $\bar{R}_G = \left[(1+0.2)(1-0.3)\right]^{1/2} - 1 = -8.348\%$

3.7 Half of the Wired readers have an income of no more than $99,874 while half of the Wired.com readers have an income of no more than $80,394.

3.8 (a)

	Grade *X*	Grade *Y*
Mean	575	575.4
Median	575	575
Standard deviation	6.4	2.1

(b) If quality is measured by central tendency, Grade *X* tires provide slightly better quality because *X*'s mean and median are both equal to the expected value, 575 mm. If, however, quality is measured by consistency, Grade *Y* provides better quality because, even though *Y*'s mean is only slightly larger than the mean for Grade *X*, *Y*'s standard deviation is much smaller. The range in values for Grade *Y* is 5 mm compared to the range in values for Grade *X*, which is 16 mm.

(c) Excel output:

Grade X	
Mean	575
Median	575
Mode	#N/A
Standard Deviation	6.403124
Sample Variance	41
Range	16
Minimum	568
Maximum	584
Sum	2875
Count	5

Grade Y	
Mean	577.4
Median	575
Mode	#N/A
Standard Deviation	6.107373
Sample Variance	37.3
Range	15
Minimum	573
Maximum	588
Sum	2887
Count	5

	Grade *X*	Grade *Y*, Altered
Mean	575	577.4
Median	575	575
Standard deviation	6.4	6.1

When the fifth *Y* tire measures 588 mm rather than 578 mm, *Y*'s mean inner diameter becomes 577.4 mm, which is larger than *X*'s mean inner diameter, and *Y*'s standard deviation increases from 2.1 mm to 6.1 mm. In this case, *X*'s tires are providing better quality in terms of the mean inner diameter, with only slightly more variation among the tires than *Y*'s.

3.9 (a) Half of the new houses were sold at a price no higher than $326,200.

(b) On average, the sales price of houses was $384,600.

(c) The sales price of new houses in 2018 is right skewed

3.10 (a), (b)

Download Speed		*Upload Speed*	
Mean	32.375	Mean	11.175
Standard Error	4.542999	Standard Error	2.048148118
Median	32.65	Median	12.95
Mode	#N/A	Mode	#N/A
Standard Deviation	12.84954	Standard Deviation	5.793037693
Sample Variance	165.1107	Sample Variance	33.55928571
Kurtosis	3.158892	Kurtosis	-1.951578447
Skewness	-0.72039	Skewness	-0.39562811
Range	46.8	Range	13.8
Minimum	6.5	Minimum	3.7
Maximum	53.3	Maximum	17.5
Sum	259	Sum	89.4
Count	8	Count	8

(c) The mean is about the same as the median for the download speed. The upload speed is indicating a left or negative skewed distribution (the skewness statistic is also negative). The kurtosis statistic is positive for the upload speed, indicating a distribution that is more peaked than a normal (bell-shaped) distribution. The kurtosis statistic is negative for download speed, indicating a distribution that is less peaked than a normal distribution.

(d) The mean download speed is much higher than the mean upload speed. The median download speed indicates that half the carriers have a download speed of at least 32.65 Mbps as compared to a median upload speed of 12.95 Mbps that indicates that half the carriers have an upload speed of at least 12.95 Mbps. There is much more variation in the download speed than the upload speed because the standard deviation is 12.8495 as compared to 5.7930.

3.11 (a), (b)

First Half		*Halftime or After*	
Mean	5.382903226	Mean	5.650741
Standard Error	0.12846337	Standard Error	0.200359
Median	5.34	Median	5.59
Mode	5.34	Mode	4.96
Standard Deviation	0.715253771	Standard Deviation	1.041094
Sample Variance	0.511587957	Sample Variance	1.083876
Kurtosis	-1.25668136	Kurtosis	-0.59102
Skewness	0.035752937	Skewness	0.140612
Range	2.29	Range	4.06
Minimum	4.22	Minimum	3.63
Maximum	6.51	Maximum	7.69
Sum	166.87	Sum	152.57
Count	31	Count	27
Coefficient of variation	13.28750938	Coefficient of variatio	18.42402

3.11 cont. (a), (b)

First Half	Zscore	Halftime or After	Zscore
6.51	1.5758	7.69	1.958766
6.39	1.408027	7.34	1.622581
6.38	1.394046	7.07	1.363239
6.31	1.296179	7.05	1.344028
6.30	1.282198	6.95	1.247975
6.32	1.31016	6.43	0.748501
6.14	1.058501	6.41	0.72929
6.14	1.058501	6.37	0.690869
5.98	0.834804	6.18	0.508368
5.88	0.694994	6.11	0.441131
5.84	0.639069	5.84	0.181789
5.63	0.345467	5.80	0.143368
5.25	-0.18581	5.64	-0.01032
5.50	0.163714	5.59	-0.05834
5.34	-0.05998	5.52	-0.12558
5.34	-0.05998	5.29	-0.3465
5.34	-0.05998	5.14	-0.49058
5.29	-0.12989	5.11	-0.5194
5.15	-0.32562	5.04	-0.58663
5.10	-0.39553	4.96	-0.66348
5.07	-0.43747	4.96	-0.66348
4.95	-0.60524	4.84	-0.77874
4.85	-0.74505	4.59	-1.01887
4.81	-0.80098	4.55	-1.05729
4.63	-1.05264	4.50	-1.10532
4.61	-1.0806	3.97	-1.6144
4.45	-1.3043	3.63	-1.94098
4.45	-1.3043		
4.38	-1.40216		
4.32	-1.48605		
4.22	-1.62586		

(c) The mean and median are approximately equal for the First Half ratings indicating the data is symmetric. The mean is more than the median for the ratings on ads running at halftime or after indicating a right or positively skewed distribution.

(d) The mean and median are both greater for the ads running at or after halftime compared to the ads running in the first half. There is more variation in the ratings of the ads running at or after halftime then those running in the first half.

3.12 (a), (b)

	Household Hours
Mean	7.5667
Median	7
Minimum	4
Maximum	14.3
Range	10.3
Variance	6.1444
Standard deviation	2.4788
Coefficient of variation	32.76%
Skewness	1.3521
Kurtosis	1.5387
Sample size	30

(c) (d) The mean work hours needed is greater than the median, indicating a right or positive skewed distribution (the skewness statistic is also positive). The kurtosis statistic is positive, indicating a distribution that is more peaked than a normal (bell-shaped) distribution.

3.13 (a), (b)

Number of Partners	
Mean	19.51111
Standard Error	1.148834
Median	18
Mode	14
Standard Deviation	7.706615
Sample Variance	59.39192
Kurtosis	0.57365
Skewness	0.804088
Range	36
Minimum	5
Maximum	41
Sum	878
Count	45
Coefficient of variation	39.4986

3.13 (a), (b) cont.

Number of Partners	Zscore
37	2.269335
41	2.788369
26	0.841989
14	–0.71511
22	0.322955
29	1.231265
36	2.139576
11	–1.10439
16	–0.4556
29	1.231265
30	1.361024
20	0.063438
20	0.063438
20	0.063438
26	0.841989
21	0.193196
17	–0.32584
21	0.193196
14	–0.71511
28	1.101507
24	0.582472
14	–0.71511
15	–0.58536
19	–0.06632
14	–0.71511
11	–1.10439
18	–0.19608
9	–1.36391
10	–1.23415
13	–0.84487
14	–0.71511
24	0.582472
25	0.712231
13	–0.84487
5	–1.88294
13	–0.84487
20	0.063438
15	–0.58536
17	–0.32584
16	–0.4556
26	0.841989
18	–0.19608
20	0.063438
16	–0.4556
11	–1.10439

There are no Z-Scores greater than 3 or less than –3, which indicates there are no outliers.

3.13 (c) Because the mean is larger than the median, the data is skewed to the right.
cont. (d) The mean number of partners in rising accounting firms is 19.5 and half of the rising accounting firms have 18 or more partners. The average scatter around the mean is 7.71 and the lowest number of partners is 5 (Krost) and the greatest number of partners is 41 (Brady, Martz & Associates).

3.14 (a), (b)

	A	B	C
1	**Mobil Commerce Penetration %**		
2			
3		**Mobile Commerce Penetration (%)**	
4	**Mean**	45.08333333	
5	**Median**	45	
6	**Mode**	46	
7	**Minimum**	30	
8	**Maximum**	60	
9	**Range**	30	
10	**Variance**	58.5145	
11	**Standard Deviation**	7.6495	
12	**Coeff. of Variation**	16.97%	
13	**Skewness**	-0.0469	
14	**Kurtosis**	-0.2779	
15	**Count**	24	
16	**Standard Error**	1.5614	
17			

	A	B	C
1	**Country**	**Mobile Commerce Penetration (%)**	Z-Scores
2	Australia	46	0.119834
3	Austria	44	-0.14162
4	Belgium	38	-0.92599
5	Brazil	43	-0.27235
6	Canada	33	-1.57963
7	Denmark	51	0.773473
8	Finland	49	0.512018
9	France	39	-0.79526
10	Germany	50	0.642745
11	Italy	41	-0.53381
12	Japan	55	1.296385
13	Netherlands	49	0.512018
14	New Zealand	44	-0.14162
15	Norway	57	1.557841
16	Poland	33	-1.57963
17	Russia	30	-1.97181
18	South Korea	47	0.250562
19	Spain	48	0.38129
20	Sweden	60	1.950024
21	Switzerland	43	-0.27235
22	Taiwan	42	-0.40308
23	Turkey	46	0.119834
24	United Kingdom	55	1.296385
25	United States	39	-0.79526

3.14 cont. (a), (b) Because there are no Z values below –3.0 or above 3.0, there are no outliers.

(c) The mean is approximately the same the median, so Mobile Commerce Penetration is symmetrical.

(d) The mean Mobile Commerce Penetration is 45.0833% and half the countries have values greater than or equal to 45%. The average scatter around the mean is 7.6495%. The lowest value is 30% (Russia), and the highest value is 60% (Sweden).

3.15 (a)

One-Year CD		*Five-Year CD*	
Mean	1.659783	Mean	2.172391
Standard Error	0.148902	Standard Error	0.142587
Median	1.96	Median	2.75
Mode	0.1	Mode	3
Standard Deviation	1.009901	Standard Deviation	0.967074
Sample Variance	1.0199	Sample Variance	0.935232
Kurtosis	-1.29325	Kurtosis	-0.98189
Skewness	-0.48514	Skewness	-0.68224
Range	2.85	Range	3.1
Minimum	0.01	Minimum	0.15
Maximum	2.86	Maximum	3.25
Sum	76.35	Sum	99.93
Count	46	Count	46
Coefficient of variation	60.84538	Coefficient of variation	44.51656

(b) Relative to the mean one-year CDs have much more variation thnn five-year CDs. The standard deviation, variance, and range are all greater for one-year CDs compared to five-year CDs.

3.16 (a),(b)

Time	
Mean	232.78
Standard Error	22.441676
Median	228
Mode	243
Standard Deviation	158.68661
Sample Variance	25181.44
Kurtosis	21.834832
Skewness	3.923888
Range	1076
Minimum	65
Maximum	1141
Sum	11639
Count	50

3.16 cont. (c) The mean time is 232.78 seconds, and half the calls last greater than or equal to 228 seconds, so call duration is slightly right-skewed. The average scatter around the mean is 158.6866 seconds. The shortest call lasted 65 seconds, and the longest call lasted 1141 seconds.

3.17 Excel output:

Waiting Time	
Mean	4.286667
Median	4.5
Mode	#N/A
Standard Deviation	1.637985
Sample Variance	2.682995
Range	6.08
Minimum	0.38
Maximum	6.46
Sum	64.3
Count	15
First Quartile	3.2
Third Quartile	5.55
Interquartile Range	2.35
Coefficient of Variation	38.2112%

(a) Mean = 4.287 Median = 4.5

(b) Variance = 2.683 Standard deviation = 1.638 Range = 6.08

Coefficient of variation = 38.21%

Z scores: –0.05, 0.77, –0.77, 0.51, 0.30, –1.19, –0.46, –0.66, 0.13, 1.11, –2.39, 0.51, 1.33, 1.16, –0.30

There are no outliers.

(c) Since the mean is less than the median, the distribution is left-skewed.

(d) The mean and median are both under 5 minutes and the distribution is left-skewed, meaning that there are more unusually low observations than there are high observations. But six of the 15 bank customers sampled (or 40%) had wait times in excess of 5 minutes. So, although the customer is more likely to be served in less than 5 minutes, the manager may have been overconfident in responding that the customer would "almost certainly" not wait longer than 5 minutes for service.

3.18 (a) Mean = 7.11 Median = 6.68

(b) Variance = 4.336 Standard Deviation = 2.082 Range = 6.67

Coefficient of variation = 29.27%

Waiting Time	*Z Score*
9.66	1.222431
5.90	–0.58336
8.02	0.434799
5.79	–0.63619
8.73	0.775786
3.82	–1.58231
8.01	0.429996
8.35	0.593286

Waiting Time	*Z Score*
10.49	1.62105
6.68	–0.20875
5.64	–0.70823
4.08	–1.45744
6.17	–0.45369
9.91	1.342497
5.47	–0.78987

Since there are no Z values below –3.0 or above 3.0, there are no outliers.

(c) Because the mean is greater than the median, the distribution is right-skewed.

3.18 cont. (d) The mean and median are both greater than five minutes. The distribution is right-skewed, meaning that there are some unusually high values. Further, 13 of the 15 bank customers sampled (or 86.7%) had waiting times greater than five minutes. So the customer is likely to experience a waiting time in excess of five minutes. The manager overstated the bank's service record in responding that the customer would "almost certainly" not wait longer than five minutes for service.

3.19 (a) $\bar{R}_G = \left[(1+0.0009)(1-0.569)\right]^{1/2} - 1 = -34.32\%$

(b) 1000(1 – 0.3432) = $656.80

(c) The rate of return was much better than that of GE, which declined in price.

3.20 (a) $\bar{R}_G = \left[(1+0.758)(1-0.257)\right]^{1/2} - 1 = 14.29\%$

(b) 1000(1+0.1429) = $1142.90

(c) The rate of return was much better than that of GE, which declined in price.

3.21 (a)

	A	B	C	D	E	F	G	H
1	Year	DJIA	S&P 500	NASDAQ				
2	2015	-2.23	-0.73	8.43				
3	2016	13.41	9.53	7.50				
4	2017	25.08	19.42	31.52				
5	2018	-5.63	-6.24	-5.98				
6	Geometric Mean	7.0%	5.0%	9.6%				
7	Excel fomula for DJIA	=((1+B2/100)*(1+B3/100)*(1+B4/100)*(1+B5/100))^(1/4)-1						
8								
9								

(b) NASDAQ had the best rate of return followed by DJIA. All of the indices had positive returns.

(c) The return on the metals was much worse than the rate of the stock indices.

3.22 (a)

	A	B	C	D	E	F
1	Year	Platinum	Gold	Silver		
2	2016	8.50	8.50	15.70		
3	2017	-0.44	10.10	2.30		
4	2018	-14.80	-2.10	-9.50		
5	Geometric Mean	-2.7%	5.4%	2.3%		
6	Excel formula for Platinum	=((1+B2/100)*(1+B3/100)*(1+B4/100))^(1/3)-1				
7						
8						

(b) Gold had the best rate of return followed by silver. Only platinum had a negative rate of return.

(c) The return on the metals was much worse than the rate of the stock indices.

3.23 (a)

Mean 3YrReturn	Risk Level			
Fund Type	Low	Average	High	Grand Total
Growth	**9.87**	**9.06**	**6.64**	**8.51**
Large	10.22	10.43	9.79	10.30
MidCap	8.93	6.86	5.78	6.93
Small	9.09	7.43	5.99	6.39
Value	**7.76**	**6.41**	**4.13**	**6.84**
Large	7.82	6.49	5.02	7.29
MidCap	7.87	7.05	2.22	6.69
Small	6.38	5.60	4.63	5.39
Grand Total	**8.66**	**8.21**	**6.25**	**7.91**

(b)

StdDev of 3YrReturn	Risk Level			
Fund Type	Low	Average	High	Grand Total
Growth	**2.42**	**2.86**	**3.39**	**3.19**
Large	2.51	2.14	4.31	2.56
MidCap	1.97	2.72	3.00	2.86
Small	---	2.09	2.56	2.52
Value	**1.72**	**2.27**	**2.64**	**2.33**
Large	1.63	2.02	2.72	1.93
MidCap	1.93	1.64	2.35	2.51
Small	2.64	3.08	2.60	2.85
Grand Total	**2.30**	**2.95**	**3.40**	**3.02**

(c) The three-year return is higher for low risk funds than average risk or high risk funds for both the growth and the value funds. However, this pattern changes when Market Cap categories are considered. For example, the three-year return percentage for growth funds with average risk is much higher for large cap funds than for midcap or small market cap funds. Also for value funds with average risk, the three-year return for midcap funds is higher than the return for large funds. The standard deviation for high risk funds is higher than average risk or low risk low risk funds for both the growth and value funds. The standard deviation for high risk large cap growth funds is much larger than any other category.

3.24 (a)

Mean of 3YrReturn%	Rating					
Type	One	Two	Three	Four	Five	Grand Total
Growth	**5.41**	**7.04**	**8.94**	**10.14**	**12.83**	**8.51**
Large	6.97	9.43	10.62	11.83	14.25	10.30
Mid-Cap	2.27	5.07	7.93	8.77	11.22	6.93
Small	0.78	5.09	6.52	8.35	9.53	6.39
Value	**4.43**	**5.49**	**7.29**	**8.34**	**10.23**	**6.84**
Large	5.23	6.05	7.58	8.85	10.23	7.29
Mid-Cap	2.79	5.77	7.32	9.26	–	6.69
Small	1.33	3.20	5.93	7.04	–	5.39

3.24 (b) cont.

StdDev of 3Yr Return%	Rating					
Type	One	Two	Three	Four	Five	Grand Total
Growth	**3.72**	**2.85**	**2.71**	**2.23**	**2.12**	**3.19**
Large	2.86	1.34	2.23	1.43	0.89	2.56
Mid-Cap	3.49	2.04	2.08	1.03	1.02	2.86
Small	0.84	2.40	2.08	2.11	0.62	2.52
Value	**2.07**	**2.40**	**1.20**	**2.09**	**1.32**	**2.33**
Large	1.81	1.68	0.98	1.63	1.32	1.93
Mid-Cap	1.00	2.90	1.13	0.99	–	2.51
Small	–	2.88	1.36	2.62	–	2.35
Grand Total	**3.24**	**2.78**	**2.44**	**2.34**	**2.24**	**3.02**

(c) The mean three-year return of small-cap funds is much lower than mid-cap and large funds. Five-star funds for all market cap categories show the highest mean three-year returns. The mean three-year returns for all combinations of type and market cap rises as the star rating rises, consistent to the mean three-year returns for all growth and value funds. The standard deviations of the three-year return for large-cap and mid-cap value funds vary greatly among star rating categories.

3.25 (a)

Mean of 3YrReturn	Star Rating					
Market Cap	One	Two	Three	Four	Five	Grand Total
Large	**6.35**	**7.86**	**9.39**	**10.57**	**12.53**	**9.04**
Low	6.94	7.59	8.57	9.91	12.36	8.77
Average	7.03	7.75	10.00	11.35	10.73	9.27
High	4.76	10.68	11.36	10.92	14.80	9.07
MidCap	**2.44**	**5.27**	**7.77**	**8.89**	**11.22**	**6.87**
Low	3.90	7.13	8.48	9.02	11.74	8.52
Average	0.72	5.74	7.40	8.73	10.18	6.91
High	3.18	4.24	8.73			5.29
Small	**0.96**	**4.48**	**6.38**	**7.75**	**9.53**	**6.07**
Low		7.60		5.98	9.09	6.92
Average		2.85	6.61	7.73		6.48
High	0.96	4.71	6.24	8.44	9.96	5.76
Grand Total	**5.07**	**6.45**	**8.38**	**9.43**	**12.01**	**7.91**

(b)

StdDev of 3YrReturn	Star Rating					
Market Cap	One	Two	Three	Four	Five	Grand Total
Large	**2.64**	**2.26**	**2.36**	**2.11**	**2.31**	**2.75**
Low	2.82	1.54	1.85	2.20	1.94	2.34
Average	2.03	2.53	2.51	1.81	2.91	2.77
High	2.78	1.65	2.85	----	0.64	4.41
MidCap	**2.82**	**2.29**	**1.89**	**1.02**	**1.02**	**2.76**
Low	----	1.28	1.96	1.19	0.67	1.99
Average	3.57	1.85	1.79	0.80	----	2.46
High	2.46	2.46	1.94			3.14
Small	**0.68**	**2.66**	**1.66**	**2.40**	**0.62**	**2.66**
Low		----		3.08	----	2.59
Average		3.28	1.31	2.76		2.77
High	0.68	2.43	1.85	1.13	----	2.59
Grand Total	**3.24**	**2.78**	**2.44**	**2.34**	**2.24**	**3.02**

3.25 cont. (c) The mean three-year return for large-cap funds is much higher than mid-cap or small-cap funds. In all risk categories except large-cap average risk, five-star funds have the highest mean three-year return. Large-cap, five-star, high-risk funds have the highest mean three -year return and the lowest standard deviation. The highest standard deviation is found in mid-cap, average-risk, one-star funds.

3.26 (a)

Mean of 3Yr Return%	Rating					
Type	One	Two	Three	Four	Five	Grand Total
Growth	**5.41**	**7.04**	**8.94**	**10.14**	**12.83**	**8.51**
Low	7.53	8.60	9.89	10.29	12.64	9.87
Average	6.17	7.99	9.28	10.43	11.96	9.06
High	3.83	5.59	7.45	8.76	13.59	6.64
Value	**4.43**	**5.49**	**7.29**	**8.34**	**10.23**	**6.84**
Low	5.29	7.00	7.66	8.57	10.74	7.76
Average	5.01	4.98	6.97	7.96	9.23	6.41
High	2.71	2.63	6.53	8.39	–	4.13
Grand Total	**5.07**	**6.45**	**8.38**	**9.43**	**12.01**	**7.91**

(b)

StdDev of 3Yr Return%	Rating					
	One	Two	Three	Four	Five	Grand Total
Growth	3.72	2.85	2.71	2.23	2.12	3.19
Low	3.27	1.57	2.02	2.05	2.04	2.42
Average	4.37	2.43	2.67	2.42	2.51	2.86
High	2.98	2.92	2.73	1.43	2.47	3.39
Value	**2.07**	**2.40**	**1.20**	**2.09**	**1.32**	**2.33**
Low	1.46	1.12	1.00	2.15	0.85	1.72
Average	2.11	2.43	1.25	2.09	1.87	2.27
High	–	2.88	1.36	2.62	–	2.35
Grand Total	**3.24**	**2.78**	**2.44**	**2.34**	**2.24**	**3.02**

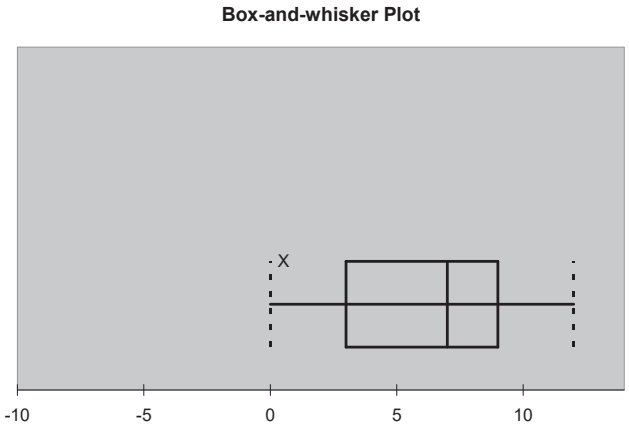
(c) The mean three-year return of high-risk funds is much lower than the other risk categories except for five-star funds. In all risk categories, five-star funds have the highest mean three-year return. The mean three-year returns for high-risk growth and value funds for one-, two-, and three-star rating funds are lower than the means for the other risk categories.

The standard deviations of the three-year return for low-risk funds show the most consistency across star rating categories and the standard deviations of the three-year return for low-risk funds are the lowest across categories. They also vary greatly among star rating categories.

3.27 (a) $Q_1 = 3$, $Q_3 = 9$, interquartile range = 6

(b) Five-number summary: 0 3 7 9 12

3.27 (c) cont.



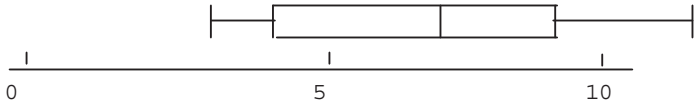
The distribution is left-skewed.

(d) Answers are the same.

3.28 (a) $Q_1 = 4$, $Q_3 = 9$, interquartile range = 5

(b) Five-number summary: 3 4 7 9 12

(c)



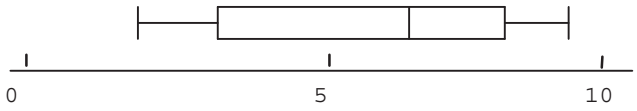
The distances between the median and the extremes are close, 4 and 5, but the differences in the tails are different (1 on the left and 3 on the right), so this distribution is slightly right-skewed.

(d) In 3.2 (d), because the mean and median are equal, the distribution is symmetric. The box part of the graph is symmetric, but the tails show right-skewness.

3.29 (a) $Q_1 = 3$, $Q_3 = 8.5$, interquartile range = 5.5

(b) Five-number summary: 2 3 7 8.5 9

(c)



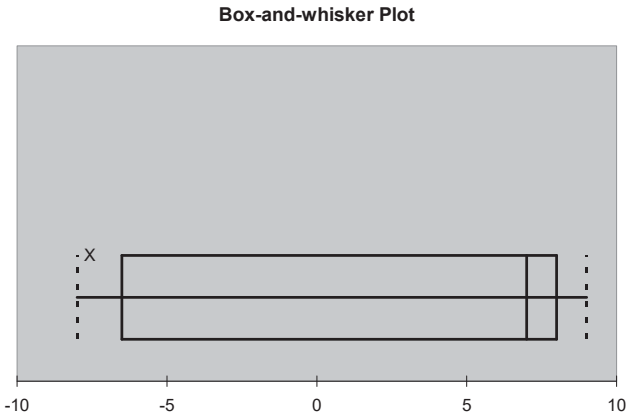
The distribution is left-skewed.

(d) Answers are the same.

3.30 (a) $Q_1 = -6.5$, $Q_3 = 8$, interquartile range = 14.5

(b) Five-number summary: – 8 – 6.5 7 8 9

3.30 (c)
cont.



The distribution is left-skewed.

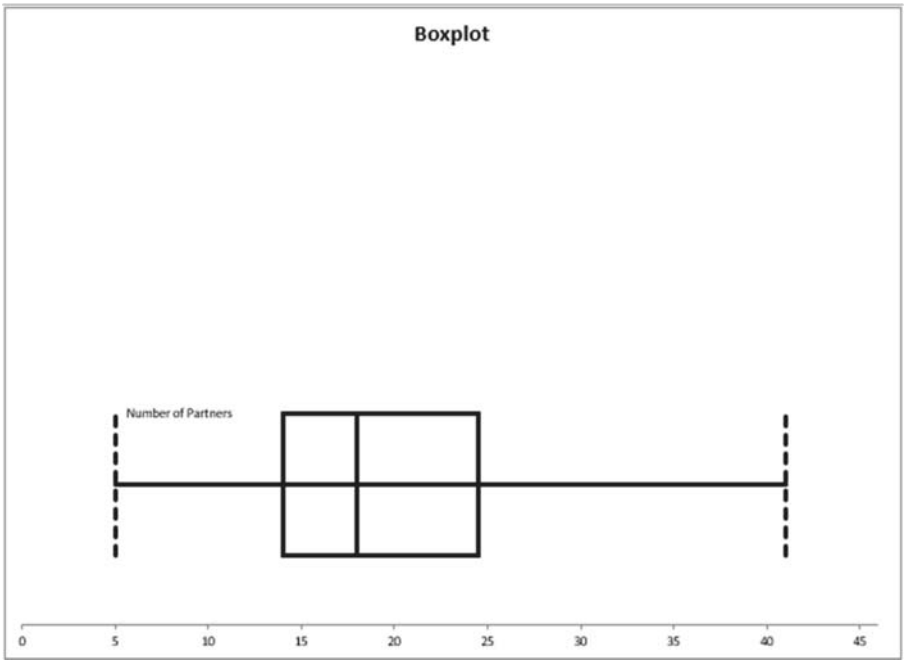
(d) This is consistent with the answer in 3.4 (d).

3.31 (a) $Q_1 = 14$, $Q_3 = 24.5$, interquartile range = 10.5

(b)

Five-Number Summary	
Minimum	5
First Quartile	14
Median	18
Third Quartile	24.5
Maximum	41

(c)



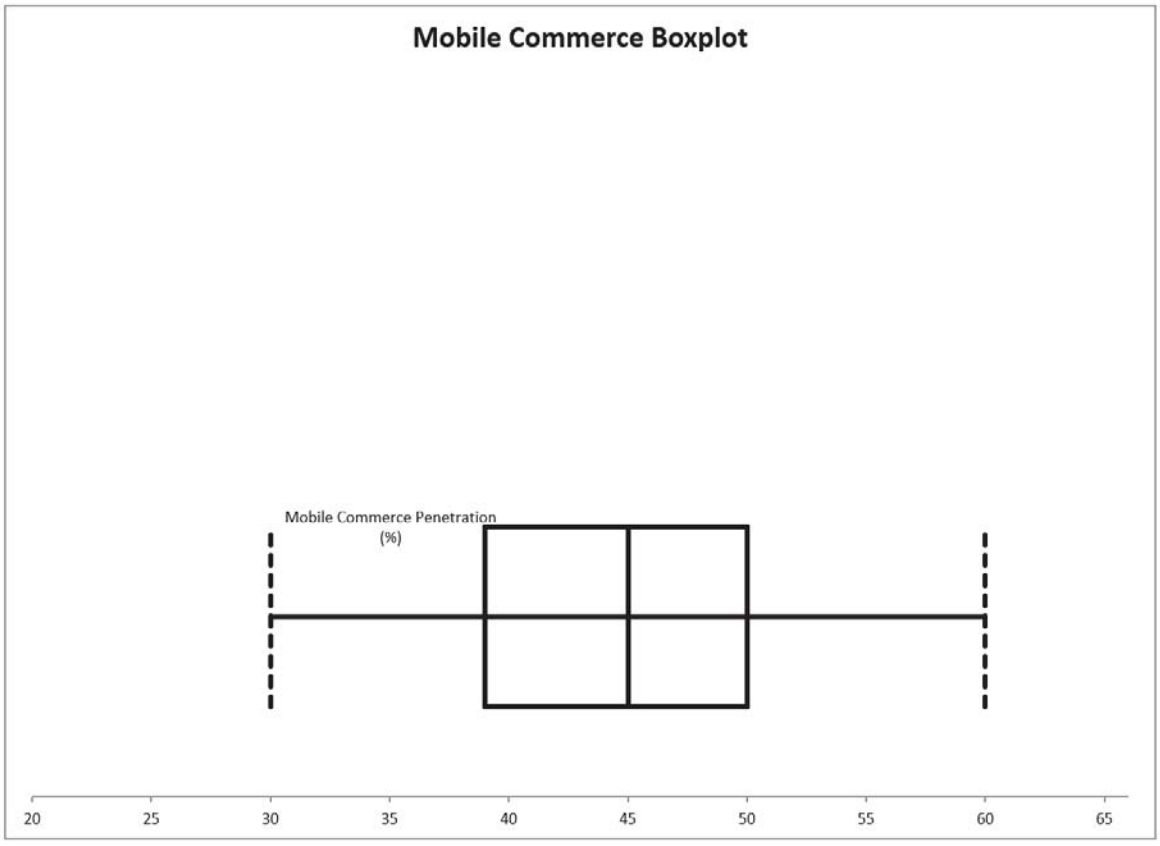
The distribution is right-skewed

3.32 (a) $Q_1 = 39$, $Q_3 = 50$, interquartile range = 11

(b)

3	Five-Number Summary	
4	Minimum	30
5	First Quartile	39
6	Median	45
7	Third Quartile	50
8	Maximum	60
9		

(c)



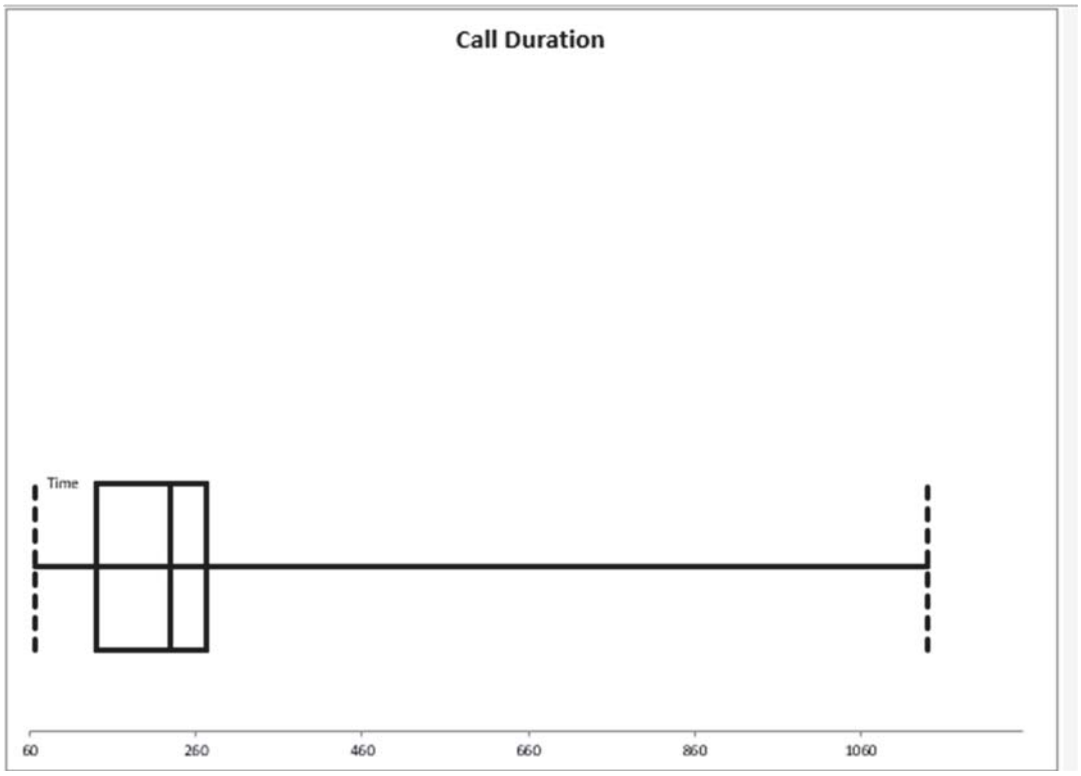
The distribution is symmetric

3.33 (a) $Q_1 = 139$, $Q_3 = 273$, interquartile range = 134

(b)

Five-Number Summary	
Minimum	65
First Quartile	139
Median	228
Third Quartile	273
Maximum	1141

3.33 (c)
cont.

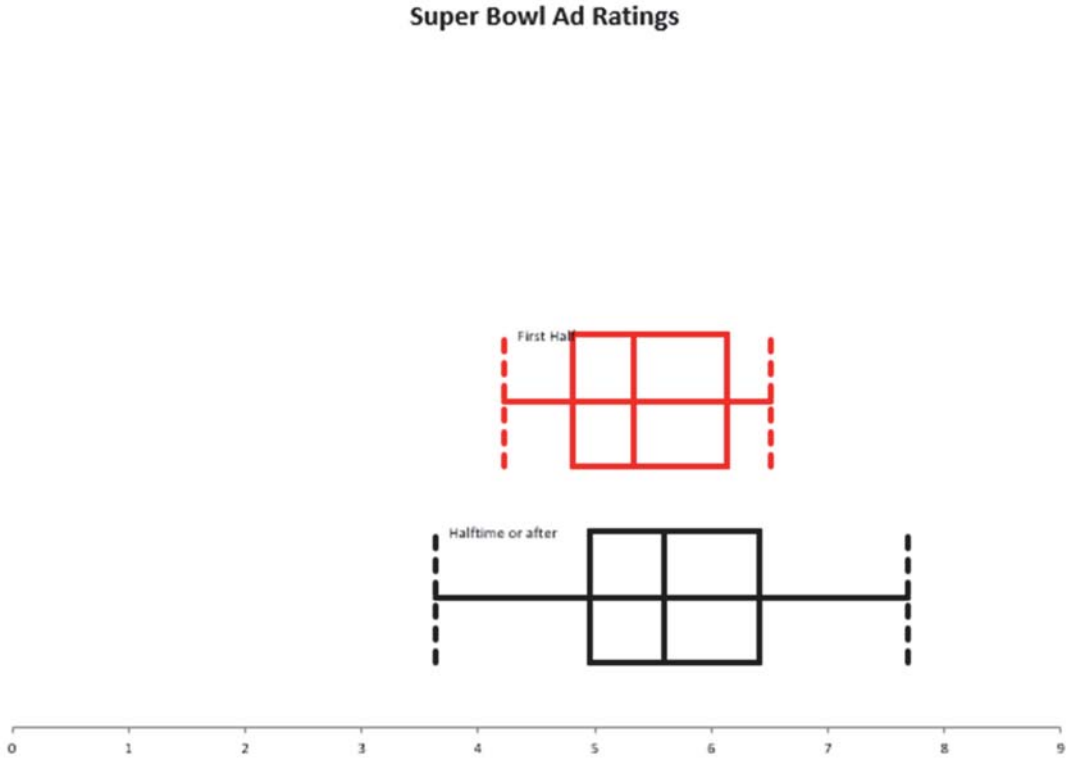


The distribution is right-skewed

3.34 (a),(b)

	Five-Number Summary		
		Halftime or After	First Half
	Minimum	3.63	4.22
	First Quartile	4.96	4.81
	Median	5.59	5.34
	Third Quartile	6.41	6.14
	Maximum	7.69	6.51
	Interquartile Range	1.45	1.33

3.34 (c)
cont.

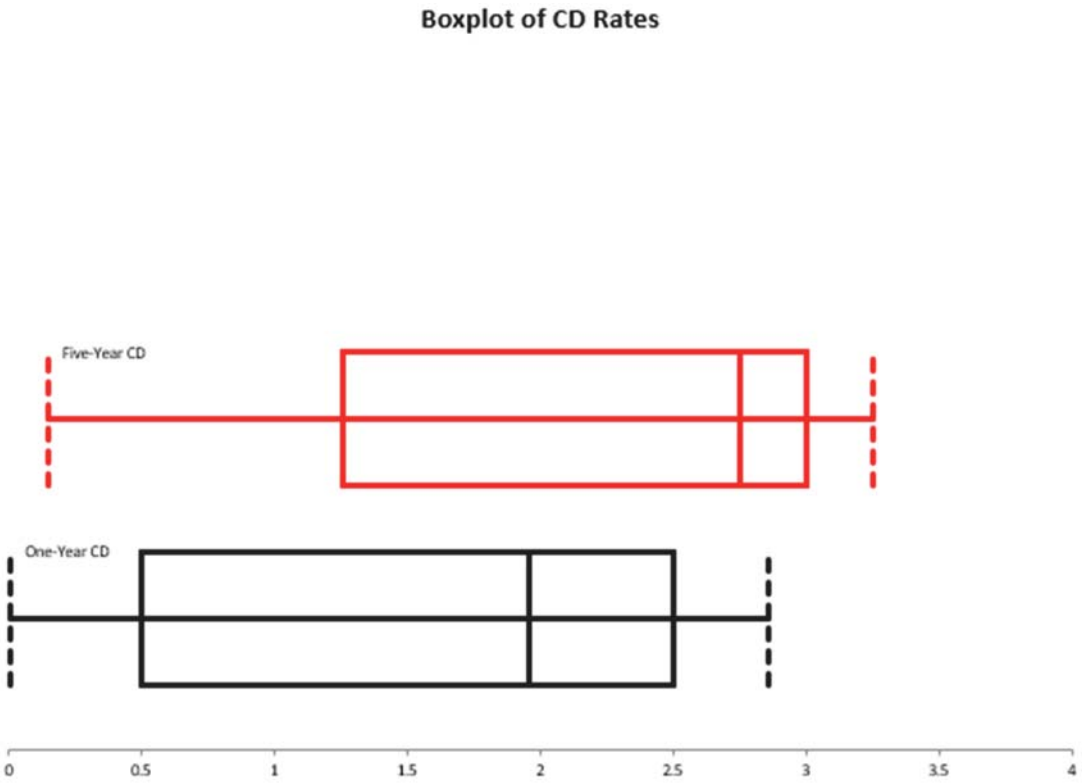


The boxplot plot for halftime or after is approximately symmetrical and the boxplot for the first or second quarter is also approximately symmetric

3.35 (a), (b)

Five-Number Summary		
	One-Year CD	Five-Year CD
Minimum	0.01	0.15
First Quartile	0.5	1.26
Median	1.96	2.75
Third Quartile	2.5	3
Maximum	2.86	3.25
Interquartile Range	2	1.74

3.35 (c)
cont.

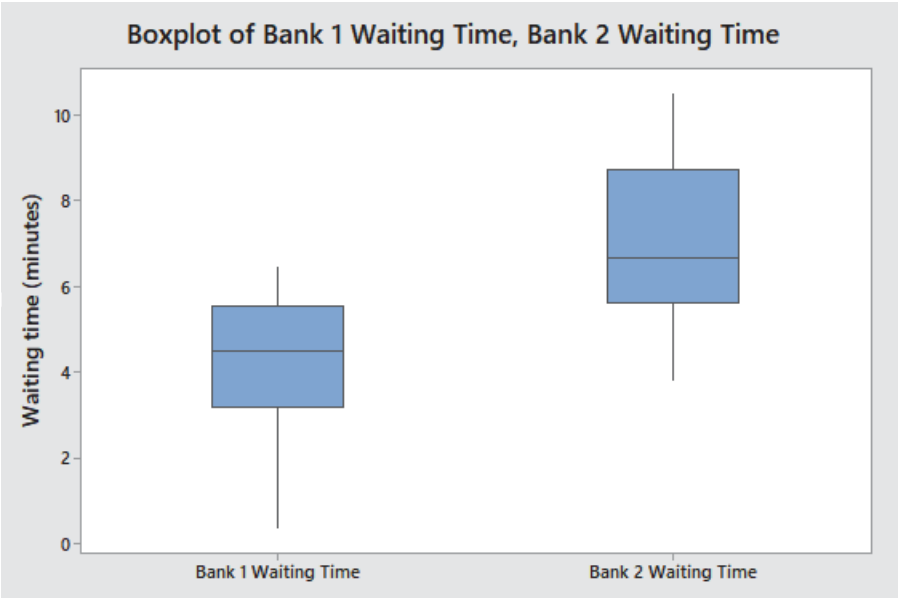


Both of the distributions are left-skewed.

3.36 (a)

Variable	Minimum	Q1	Median	Q3	Maximum
Bank 1 Waiting Time	0.380	3.200	4.500	5.550	6.460
Bank 2 Waiting Time	3.820	5.640	6.680	8.730	10.490

(b)



3.36 cont. (b) The waiting time for bank 1 in the commercial district of a city is skewed to the left. The waiting time for bank 2 in the residential area is skewed right.

(c) The central tendency of the waiting times for the bank branch located in the commercial district (bank 1) of a city is lower than that of the branch located in the residential area (bank 2). There are a few longer than normal waiting times for the branch located in the residential area whereas there are a few exceptionally short waiting times for the branch located in the commercial area.

3.37 (a) Population Mean = 6

(b) $\sigma^2 = 9.4$ $\sigma = 3.1$

3.38 (a) Population Mean = 6

(b) $\sigma^2 = 2.8$ $\sigma = 1.67$

3.39 (a)

mean	276.4902
variance	78039.01
standard deviation	279.3546

Variability in a Distribution

Data	
Mean	276.4902
Standard Deviation	279.3546

Results			% of Values Found in Intervals Around the Mean	
	Range		Chebyshev Rule (any distribution)	Empirical Rule (normal distribution)
±1 standard deviations	-2.86	555.84	At least 0%	Approximately 68%
±2 standard deviations	-282.22	835.20	At least 75%	Approximately 95%
±3 standard deviations	-561.57	1114.55	At least 88.89%	Approximately 99.7%

(b) Within one standard deviation of the mean is (–2.86,555.84) , counting the data reveals 45 states or 45/51*100 = 88.24% of the states are within this range.

Within two standad deviations of the mean is (–282.22,835.2), counting the data reveals 48 states or 48/51*100 = 94.12% of the states are within this range.

Within three standad deviations of the mean is (–561.57,1114.55), counting the data reveals 49 states or 49/51*100 = 96.08% of the states are within this range.

(c) This is slightly different from 68%, 95% and 99.7% of the empirical rule.

3.40 (a) 68%

(b) 95%

(c) at least 0 75% 88.89%

(d) $\mu - 4\sigma$ to $\mu + 4\sigma$ or –2.8 to 19.2

3.41 (a)

	Cigarette Tax
Mean	1.688117647
Standard Deviation	1.0389

(b) On average, the cigarette tax is $1.69. The typical distance between the cigarette tax in each of the 50 states and the District of Columbia and the population mean cigarette tax is $1.03.

3.42 (a)

	Average Residential Price
Mean	13.48
Standard Deviation	3.4515

(b) Within one standard deviation of the mean is (10.03,16.93) , counting the data reveals 40 states or 40/51*100 = 78.43% of the states are within this range.

Within two standad deviations of the mean is (6.57,20.38), counting the data reveals 49 states or 49/51*100 = 96.08% of the states are within this range.

Within three standad deviations of the mean is (3.12,23.83), counting the data reveals 50 states or 50/51*100 = 98.04% of the states are within this range.

(c) This is slightly different from 68%, 95% and 99.7% of the empirical rule.

. Excel Output:

	A	B	C	D	E	F
1	Variability in a Distribution					
2						
3	Data					
4	Mean	13.47706				
5	Standard Deviation	3.451484				
6						
7	Results				% of Values Found in Intervals Around the Mean	
8		Range			Chebyshev Rule (any distribution)	Empirical Rule (normal distribution)
9	±1 standard deviations	10.03	16.93		At least 0%	Approximately 68%
10	±2 standard deviations	6.57	20.38		At least 75%	Approximately 95%
11	±3 standard deviations	3.12	23.83		At least 88.89%	Approximately 99.7%
12						

3.43 (a)

	Market Cap ($bil)
mean	185.8
Standard Deviation	133.708697

(b) On average, the market capitalization for this population of 30 companies is $185.8 billion. The typical distance between the market capitalization and the mean market capitalization for this population of 30 companies is $133.7 billion.

3.44 (a) $cov(X,Y) = 65.2909$

(b) $S_X^2 = 21.7636,\ S_Y^2 = 195.8727$

$$r = \frac{\text{cov}(X,Y)}{\sqrt{S_X^2}\sqrt{S_Y^2}} = \frac{65.2909}{\sqrt{21.7636}\sqrt{195.8727}} = +1.0$$

(c) There is a perfect positive linear relationship between X and Y; all the points lie exactly on a straight line with a positive slope.

3.45 (a) The study suggests that the perceived usefulness of smartphones in an educational setting and the number of times students used their smartphone to send or read email for class purpose are positively correlated.

(b) There could be a cause and effect relationship between perceived usefulness of smartphones and the number of times students used their smartphone to send or read email for class purposes. The more a student uses their smartphone for class the more they may feel it is useful in an educational setting.

3.46 (a) $cov(X,Y) = 133.3333$

(b) $S_X^2 = 2200,\ S_Y^2 = 11.4762$

$$r = \frac{\text{cov}(X,Y)}{S_X S_Y} = 0.8391$$

(c) The correlation coefficient is more valuable for expressing the relationship between calories and sugar because it does not depend on the units used to measure calories and sugar.

(d) There is a strong positive linear relationship between calories and sugar.

3.47 (a)

	First Weekend	***US Gross***	***Worldwide Gross***
First Weekend	947.4799		
US Gross	890.3014	1576.679	
Worldwide Gross	4001.782	6045.573	24934.48

(b)

	First Weekend	***US Gross***	***Worldwide Gross***
First Weekend	1		
US Gross	0.728417	1	
Worldwide Gross	0.823319	0.964197	1

(c) The correlation coefficient is more valuable for expressing the relationship because it does not depend on the units used.

(d) There is a strong positive linear relationship between U.S. gross and worldwide gross, first weekend gross and worldwide gross and first weekend gross and U.S. gross.

3.48 Excel Output:

	A	B	C	D	E	F
1	**Carrier**	**Download Speed**	**Upload Speed**			
2	Verizon	53.3	17.5			
3	T-Mobile	36.3	16.9			
4	AT&T	37.1	12.9			
5	Metro PCS	32.8	13.0			
6	Sprint	32.5	4.0			
7	Boost	29.4	3.7			
8	Straight Talk	31.1	15.6			
9	Cricket	6.5	5.8			
10						
11	Covariance	45.5036	=COVARIANCE.S(B2:B9,C2:C9)			
12	r	0.6113	=CORREL(B2:B9,C2:C9)			
13						
14						

(a) cov(X,Y) = 45.5036
(b) Correlation = r = 0.6113
(c) The is a positive linear relationship between download speed and upload speed.

3.49 Excel Output
(a), (b)

	A	B	C	D
1	**Country**	**GDP per capita**	**Smartphone Ownership (%)**	
2	South Korea	38260	95	
3	Israel	38413	88	
4	Netherlands	52941	87	
5	Sweden	50070	86	
6	Australia	47047	81	
7	U.S.	59532	81	
8	Spain	38091	80	
9	Germany	50716	78	
10	UK	43877	76	
11	France	42779	75	
12	Italy	39817	71	
13	Argentina	20787	68	
14	Canada	46378	66	
15	Japan	43876	66	
16	Hungary	28375	64	
17	Poland	29291	63	
18	Greece	27809	59	
19	Russia	25533	59	
20	Brazil	15848	60	
21	South Africa	13498	60	
22	Philippines	8343	55	
23	Mexico	18149	52	
24	Tunisia	11911	45	
25	Indonesia	12284	42	
26	Kenya	3286	41	
27	Nigeria	5861	39	
28	India	7056	24	
29				
30	Covarince	245722.076923	=COVARIANCE.S(B2:B28,C2:C28)	
31	Correlation	0.847866	=CORREL(B2:B28,C2:C28)	
32				

	A	B	C	D
1	**Country**	**GDP per capita**	**Social Media Usage (%)**	
2	South Korea	38260	76	
3	Israel	38413	77	
4	Netherlands	52941	72	
5	Sweden	50070	73	
6	Australia	47047	70	
7	U.S.	59532	70	
8	Spain	38091	68	
9	Germany	50716	44	
10	UK	43877	66	
11	France	42779	60	
12	Italy	39817	54	
13	Argentina	20787	68	
14	Canada	46378	68	
15	Japan	43876	43	
16	Hungary	28375	62	
17	Poland	29291	53	
18	Greece	27809	50	
19	Russia	25533	63	
20	Brazil	15848	58	
21	South Africa	13498	52	
22	Philippines	8343	59	
23	Mexico	18149	66	
24	Tunisia	11911	49	
25	Indonesia	12284	39	
26	Kenya	3286	42	
27	Nigeria	5861	45	
28	India	7056	23	
29				
30	Covarince	125323.307692	=COVARIANCE.S(B2:B28,C2:C28)	
31	Correlation	0.567411	=CORREL(B2:B28,C2:C28)	
32				

3.49 cont. Excel Output (a), (b)

Data Set	*Smartphone Ownership*	*Social Media Usage*
Covariance	245722.076923	125323.307692
Coefficient of Correlation	0.847866	0.567411

(c) There is a strong positive linear relationship between the percentage of adults polled who own a smart phone and GDP. There is a positive linear relationship between social media usage and GDP.

3.50 We should look for ways to describe the typical value, the variation, and the distribution of the data within a range.

3.51 Central tendency or location refers to the fact that most sets of data show a distinct tendency to group or cluster about a certain central point.

3.52 The arithmetic mean is a simple average of all the values, but is subject to the effect of extreme values. The median is the middle ranked value, but varies more from sample to sample than the arithmetic mean, although it is less susceptible to extreme values. The mode is the most common value, but is extremely variable from sample to sample.

3.53 The first quartile is the value below which 25% of the total ranked observations will fall, the median is the value that divides the total ranked observations into two equal halves and the third quartile is the observation above which 25% of the total ranked observations will fall.

3.54 Variation is the amount of dispersion**,** or "spread," in the data.

3.55 The *Z* score measures how many standard deviations an observation in a data set is away from the mean.

3.56 The range is a simple measure, but only measures the difference between the extremes. The interquartile range measures the range of the center fifty percent of the data. The standard deviation measures variation around the mean while the variance measures the squared variation around the mean, and these are the only measures that take into account each observation. The coefficient of variation measures the variation around the mean relative to the mean. The range, standard deviation, variance and coefficient of variation are all sensitive to outliers while the interquartile range is not.

3.57 The empirical rule relates the mean and standard deviation to the percentage of values that will fall within a certain number of standard deviations of the mean.

3.58 Chebyshev's theorem applies to any type of distribution while the empirical rule applies only to data sets that are approximately bell-shaped. The empirical rule is more accurate than the Chebyshev rule in approximating the concentration of data around the mean.

3.59 Shape is the manner in which the data are distributed. The shape of a data set can be symmetrical or asymmetrical (skewed).

3.60 Skewness measures the extent to which the data values are not symmetrical around the mean. Kurtosis measures the extent to which values that are very different from the mean affect the shape of the distribution of a set of data.

3.61 For symmetrical distributions, the boxplot is also symmetrical with the median splitting the box in half and whiskers of equal length. For left skewed distributions the boxplot's left whisker will be longer and the median will be located in the right half of the box. For rights skewed distributions the boxplot's the right whisker will be longer and the median will be located in the left half of the box.

3.62 The covariance measures the strength of the linear relationship between two numerical variables while the coefficient of correlation measures the relative strength of the linear relationship. The value of the covariance depends very much on the units used to measure the two numerical variables while the value of the coefficient of correlation is totally free from the units used.

3.63 The arithmetic mean is the most common measure of central tendency and is calculated by dividing the sum of the values in the data set by the number of data values in the set. The geometric mean is used to measure the rate of change of a variable over time. It is calculated by taking the nth root of the product of the n data values, where n is the number of data values.

3.64 The geometric mean is used to measure the rate of change of a variable over time. It is calculated by taking the nth root of the product of the n data values, where n is the number of data values. The geometric rate of return measures the mean percentage return of an investment per time period.

3.65 Excel Output:

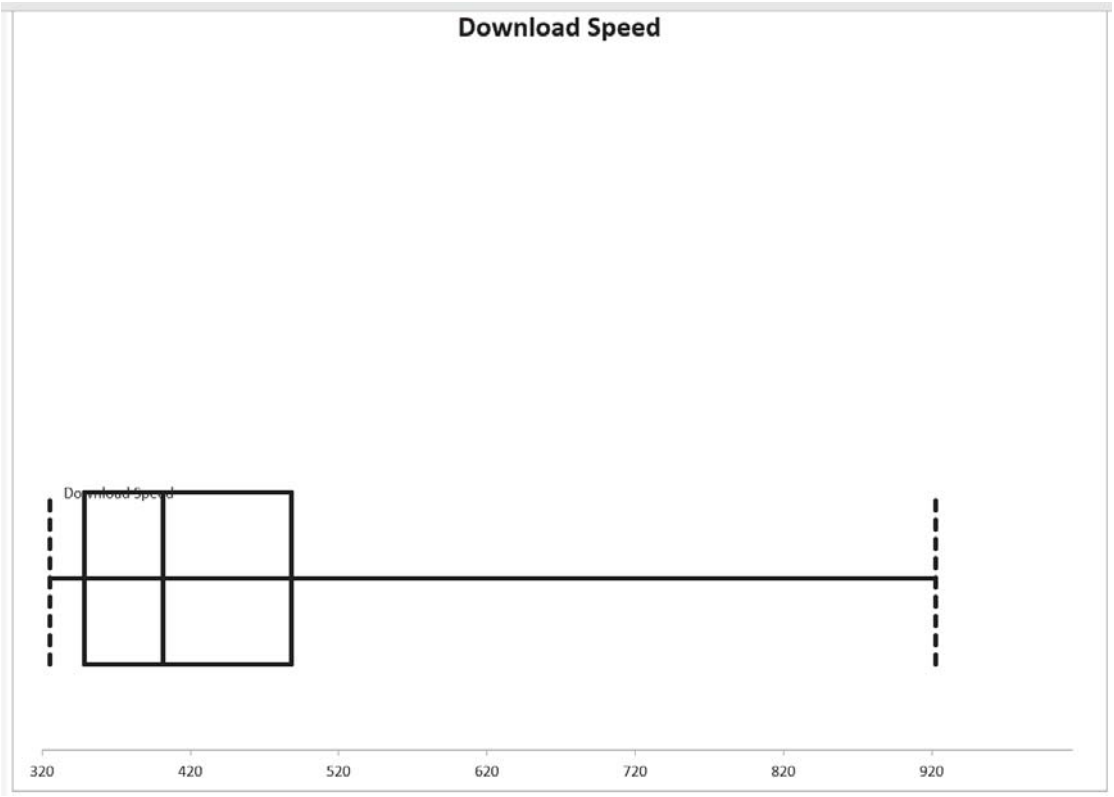
	A	B
1	Download Speed	
2		
3	Five-Number Summary	
4	Minimum	325.34
5	First Quartile	348.75
6	Median	401.38
7	Third Quartile	488.29
8	Maximum	922.39
9	IQR	139.54
10	Range	597.05
11		
12		

1	Download Speed	
2		
3		Download Speed
4	Mean	437.10
5	Median	401.38
6	Mode	#N/A
7	Minimum	325.34
8	Maximum	922.39
9	Range	597.05
10	Variance	13659.8629
11	Standard Deviation	116.8754
12	Coeff. of Variation	26.74%
13	Skewness	1.7382
14	Kurtosis	3.4361
15	Count	99
16	Standard Error	11.7464
17		

(a) Download Speed: mean=437.1, median= 401.38, first quartile=348.75, third quartile=488.29

(b) Download Speed: range=597.05, interquartile range= 139.54, variance=13,659.86, standard deviation=116.8754, coefficient of variation=26.74%

3.65 cont. (c)



(d) The mean download speed is 437.10 Mbps with 50% of the cities having a download speed less than 401.38 Mbps and 50% of the cities having download speed between 348.75 Mbps and 488.29 Mbps.

3.66 Minitab Output:

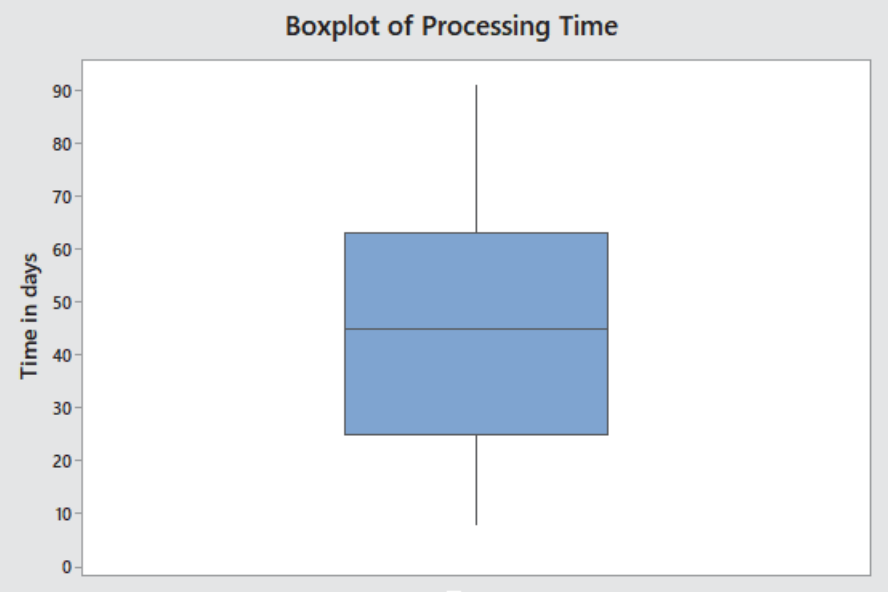
Statistics

Variable	Mean	StDev	Variance	CoefVar	Q1	Median	Q3	Range	IQR
Time	45.22	23.15	535.79	51.19	25.00	45.00	63.00	83.00	38.00

(a) Mean = 45.22 Median = 45 first quartile = 25 third quartile = 63

(b) Range = 83 Interquartile range = 38 Variance = 535.79
Standard Deviation = 23.15 Coefficient of variation = 51.19%

3.66 (c)
cont.



The distribution is approximately symmetric.

(d) The mean approval process takes 45.22 days with 50% of the policies being approved in less than 45 days. 50% of the applications are approved between 25 and 63 days. About 25% of the applicants are approved in no more than 25 days.

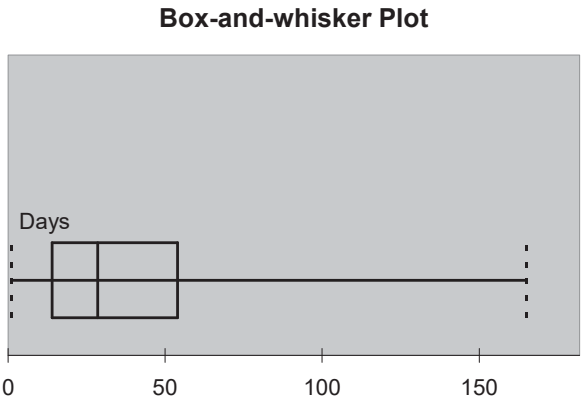
3.67 Excel output:

Days	
Mean	43.04
Median	28.5
Mode	5
Standard Deviation	41.92606
Sample Variance	1757.794
Range	164
Minimum	1
Maximum	165
First Quartile	14
Third Quartile	54
Interquartile Range	40
CV	97.41%

(a) Mean = 43.04 Median = 28.5 Q1 = 14 Q3 = 54

(b) Range = 164 Interquartile range = 40 Variance = 1,757.79
Standard deviation = 41.926 Coefficient of variation = 97.41%

3.67 cont. (c) Box-and-whisker plot for Days to Resolve Complaints



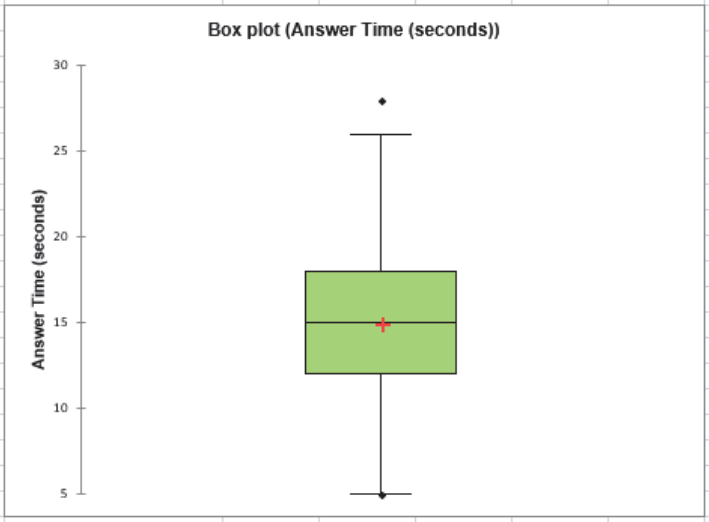
The distribution is right-skewed.

(d) Half of all customer complaints that year were resolved in less than a month (median = 28.5 days), 75% of them within 54 days. There were five complaints that were particularly difficult to settle which brought the overall mean up to 43 days. No complaint took longer than 165 days to resolve.

3.68 (a) Excel output:

Statistic	r Time (seconds)
Nbr. of observations	50
Minimum	5.000
Maximum	28.000
1st Quartile	12.000
Median	15.000
3rd Quartile	18.000
Mean	14.980
Standard deviation (n-1)	5.557

Box plots:



3.68 cont. (a) Minitab Output:

Variable	Mean	StDev	Median	Range
Answer Time (seconds)	14.980	5.557	15.000	23.000

(b) Using the formulas in the text with n= 50, Q1 = (50+1)/4 ranked value = 12.75 ranked value so choose 13th ranked value which is 12. Q3 = 3(50+1)/4 ranked value = 38.25 ranked value so choose 38th ranked value which is 18
Therefore 5 number summary is min, Q1, median, Q3, max = 5, 12, 15, 18, 28

* Note Minitab uses a slightly different formula to calculate the quartiles

Variable	Minimum	Q1	Median	Q3	Maximum
Answer Time (seconds)	5.000	11.750	15.000	18.250	28.000

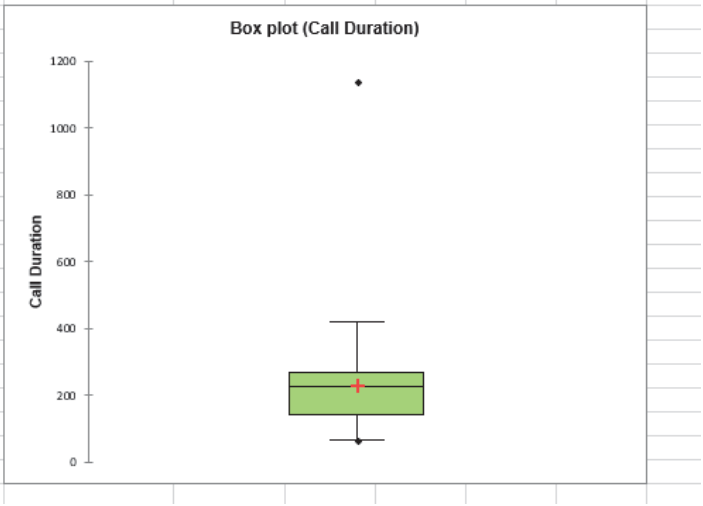
(c) The distribution is symmetric.
(d) The service level is met because 75% of the class are answered in less than 18 seconds.

3.69 (a) Excel output:

Descriptive statistics (Quantitative data):

Statistic	Call Duration
Nbr. of observations	50
Minimum	65.000
Maximum	1141.000
Range	1076.000
1st Quartile	142.000
Median	228.000
3rd Quartile	270.750
Mean	232.780
Standard deviation	158.687

Box plots:



Minitab Output:

Variable	Mean	StDev	Median	Range
Call Duration	232.8	158.7	228.0	1076.0

3.69 cont. (a) The mean call duration is 232.8 seconds. The middle ranked call duration is 228 seconds. The data is symmetric but with one high outlier. The range, calculated as the difference between the smallest and largest call duration, is 1076 seconds. The typical distance between the call duration and the mean call duration for this sample of 50 customers is called the standard deviation which is 158.7 seconds.

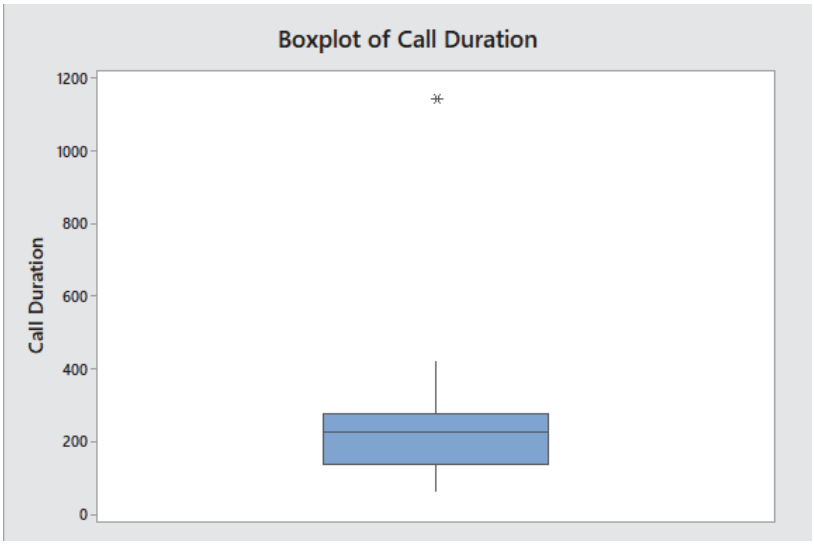
(b) Using the formulas in the text with n= 50, Q1 = (50+1)/4 ranked value = 12.75 ranked value so choose 13th ranked value which is 139. Q3 = 3(50+1)/4 ranked value = 38.25 ranked value so choose 38th ranked value which is 273
Therefore 5 number summary is min, Q1, median, Q3, max = 65, 139, 228, 273, 1141

* Note Minitab uses a slightly different formula to calculate the quartiles

Minitab output:

Variable	Minimum	Q1	Median	Q3	Maximum
Call Duration	65.0	138.3	228.0	276.8	1141.0

(c) Minitab Output:



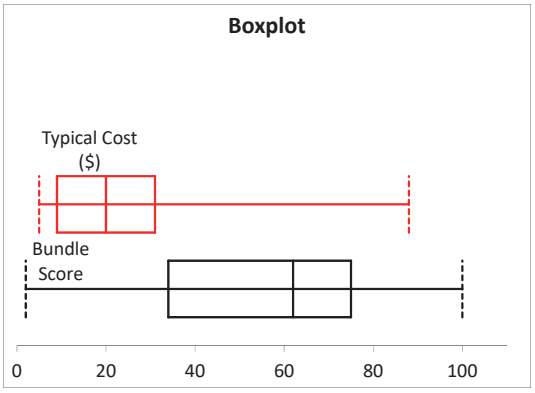
Disregarding the outlier, the distribution is left skewed.

(d) 75% of the call durations are less than 273 seconds. The target call duration of less than 240 seconds is only met for approximately 60% of calls. One could conclude the target is not met.

3.70 (a), (b)

	Bundle Score	*Typical Cost ($)*
Mean	54.775	24.175
Standard Error	4.367344951	2.866224064
Median	62	20
Mode	75	8
Standard Deviation	27.62151475	18.12759265
Sample Variance	762.9480769	328.6096154
Kurtosis	-0.845357193	2.766393511
Skewness	-0.48041728	1.541239625
Range	98	83
Minimum	2	5
Maximum	100	88
Sum	2191	967
Count	40	40
First Quartile	34	9
Third Quartile	75	31
Interquartile Range	41	22
CV	50.43%	74.98%

(c)



The typical cost is right-skewed, while the bundle score is left-skewed.

(d) $r = \dfrac{\text{cov}(X, Y)}{S_X S_Y} = 0.3465$

(e) The mean typical cost is \$24.18, with an average spread around the mean equaling \$18.13. The spread between the lowest and highest costs is \$83. The middle 50% of the typical cost fall over a range of \$22 from \$9 to \$31, while half of the typical cost is below \$20. The mean bundle score is 54.775, with an average spread around the mean equaling 27.6215. The spread between the lowest and highest scores is 98. The middle 50% of the scores fall over a range of 41 from 34 to 75, while half of the scores are below 62. The typical cost is right-skewed, while the bundle score is left-skewed. There is a weak positive linear relationship between typical cost and bundle score.

3.71 Excel output:

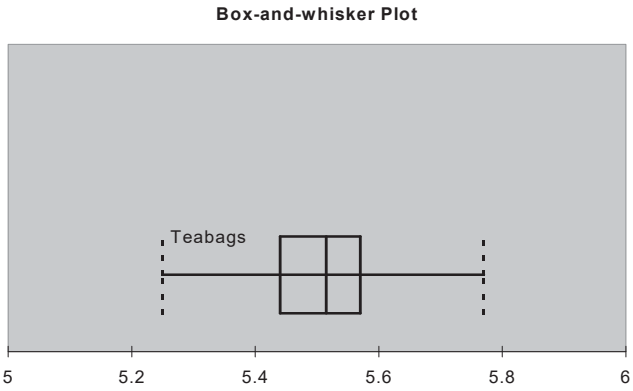
Teabags	
Mean	5.5014
Standard Error	0.014967
Median	5.515
Mode	5.53
Standard Deviation	0.10583
Sample Variance	0.0112
Kurtosis	0.127022
Skewness	–0.15249
Range	0.52
Minimum	5.25
Maximum	5.77
Sum	275.07
Count	50
First Quartile	5.44
Third Quartile	5.57
Interquartile Range	0.13
CV	1.9237%

(a) mean = 5.5014, median = 5.515, first quartile = 5.44, third quartile = 5.57

(b) Range = 0.52 Interquartile range = 0.13 Variance = 0.0112, Standard Deviation = 0.10583 Coefficient of Variation = 1.924%

(c) The mean weight of the tea bags in the sample is 5.5014 grams while the middle ranked weight is 5.515. The company should be concerned about the central tendency because that is where the majority of the weight will cluster around. The average of the squared differences between the weights in the sample and the sample mean is 0.0112 whereas the square-root of it is 0.106 gram. The difference between the lightest and the heaviest tea bags in the sample is 0.52. 50% of the tea bags in the sample weigh between 5.44 and 5.57 grams. According to the empirical rule, about 68% of the tea bags produced will have weight that falls within 0.106 grams around 5.5014 grams. The company producing the tea bags should be concerned about the variation because tea bags will not weigh exactly the same due to various factors in the production process, e.g. temperature and humidity inside the factory, differences in the density of the tea, etc. Having some idea about the amount of variation will enable the company to adjust the production process accordingly.

3.71 (d) cont.



The data is slightly left skewed.

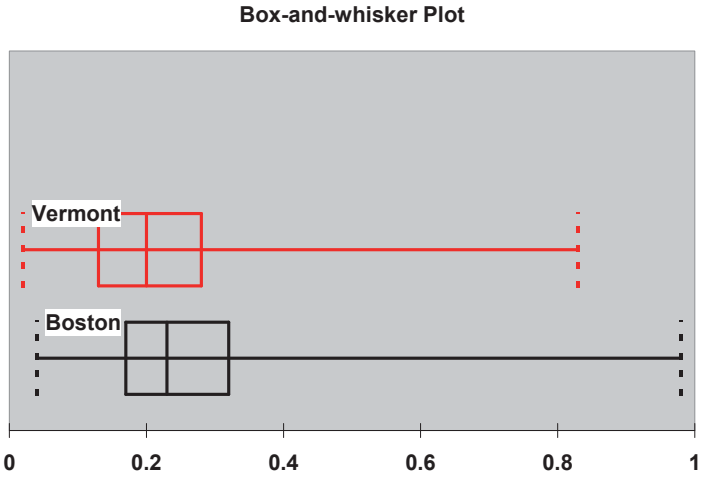
(e) On average, the weight of the teabags is quite close to the target of 5.5 grams. Even though the mean weight is close to the target weight of 5.5 grams, the standard deviation of 0.106 indicates that about 75% of the teabags will fall within 0.212 grams around the target weight of 5.5 grams. The interquartile range of 0.13 also indicates that half of the teabags in the sample fall in an interval 0.13 grams around the median weight of 5.515 grams. The process can be adjusted to reduce the variation of the weight around the target mean.

3.72 (a) Excel output:

Five-number Summary

	Boston	Vermont
Minimum	0.04	0.02
First Quartile	0.17	0.13
Median	0.23	0.2
Third Quartile	0.32	0.28
Maximum	0.98	0.83

(b)



Both distributions are right skewed.

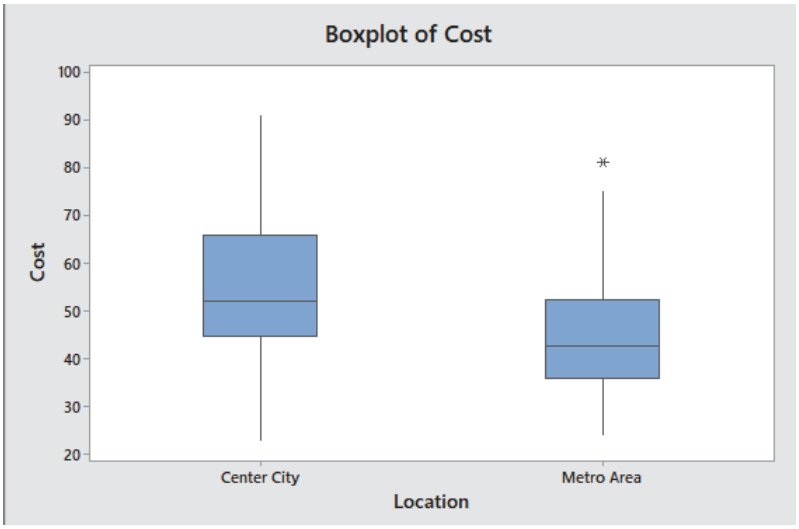
3.72 cont. (c) Both sets of shingles did quite well in achieving a granule loss of 0.8 gram or less. The Boston shingles had only two data points greater than 0.8 gram. The next highest to these was 0.6 gram. These two data points can be considered outliers. Only 1.176% of the shingles failed the specification. In the Vermont shingles, only one data point was greater than 0.8 gram. The next highest was 0.58 gram. Thus, only 0.714% of the shingles failed to meet the specification.

3.73 (a) Minitab Output:

Variable	Location	Minimum	Q1	Median	Q3	Maximum
Cost	Center City	23.00	44.75	52.00	66.00	91.00
	Metro Area	24.00	36.00	42.50	52.25	81.00

* Note Minitab uses a slightly different formula to calculate the quartiles 5 number summary is min, Q1, median, Q3, max

(b)



(c) Minitab Output:

Correlation: Cost, Summated Rating

Correlations

Pearson correlation	0.605
P-value	0.000

There is a positive correlation between cost of a meal and summated rating. The higher priced restaurants tend to receive higher rating than the lower priced restaurants.

(d) The median cost of a meal in the center city is $52 while the median cost of a meal in the metro area is $42.50. The range in costs of meals in the center city is greater than the range in costs of meals in the metro area.

3.74 (a), (b), (c)

	Calories	*Protein*	*Cholesterol*
Calories	1		
Protein	0.464411	1	
Cholesterol	0.177665	0.141673	1

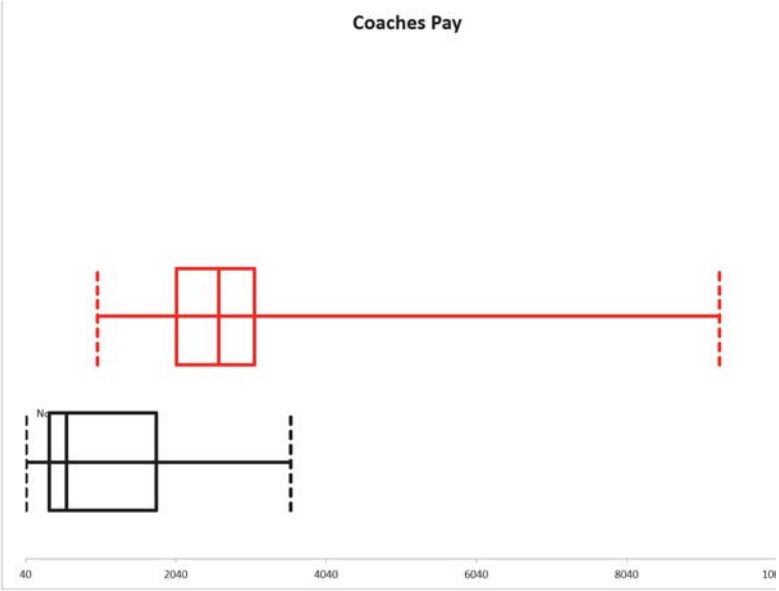
3.74 cont. (d) There is a rather weak positive linear relationship between calories and protein with a correlation coefficient of 0.46. The positive linear relationship between calories and cholesterol is quite weak at .178.

3.75 (a), (b), (d) Excel output:

	A	B	C
1	Coaches Pay		
2			
3	Five-Number Summary		
4		No Major Conference	Yes Major Conference
5	Minimum	50	1000
6	First Quartile	355	2050
7	Median	588.5	2609
8	Third Quartile	1780	3082
9	Maximum	3570	9277
10			

	A	B
1	Coaches Pay No Major Conference	
2		
3		Variable
4	Mean	1025.25
5	Median	588.5
6	Mode	#N/A
7	Minimum	50
8	Maximum	3570
9	Range	3520
10	Variance	1056629.2955
11	Standard Deviation	1027.9248
12	Coeff. of Variation	100.26%
13	Skewness	1.6449
14	Kurtosis	2.4704
15	Count	12
16	Standard Error	296.7363
17		

	A	B
1	Coaches Pay Yes Major Conference	
2		
3		Variable
4	Mean	2764.625
5	Median	2609
6	Mode	3200
7	Minimum	1000
8	Maximum	9277
9	Range	8277
10	Variance	1548829.4762
11	Standard Deviation	1244.5198
12	Coeff. of Variation	45.02%
13	Skewness	2.9268
14	Kurtosis	12.7876
15	Count	64
16	Standard Error	155.5650
17		



(c) The average coaches pay for major conference and non-major conference are $2,764,625 and $1,025,250 respectively. The middle rank pay is $2,609,000 and $588,500 for major conference and non-major conference coaches, respectively. The difference in coaches pay for major conference and non-major conference are $8,277,000 and $3,520,000, respectively. The differences in coaches pay among the middle 50% of coaches in each category pay are $1,032,000 and $1,425,000, respectively. The typical distances of coaches pay around the mean for major conference and non-major conference are $1,244,520 and $1,027,925, respectively. The amounts of average spread around the mean relative to the mean coaches pay for major conference and non-major conference are 45.02% and 100.26%, respectively.

(d) The coaches pay for both conferences are right-skewed.

3.75 cont. (e) On average, major conference coaches are paid more than $1.7 million more than non-major conference coaches. There is much more variation in pay among major conference coaches compared to non-major conference coaches.

3.76 (a), (b)

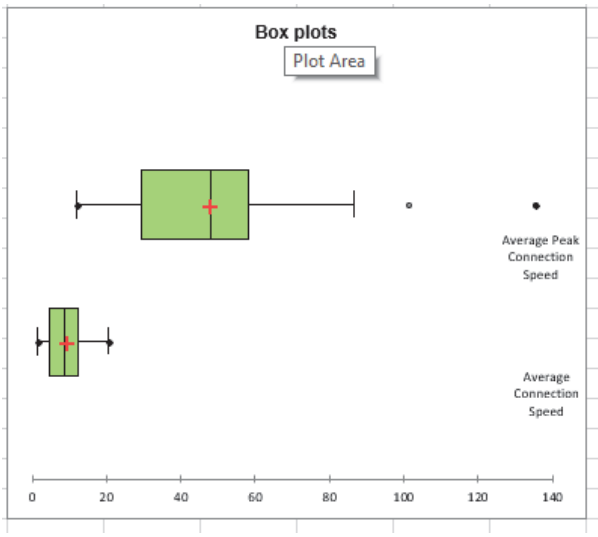
	Annual Taxes on $176 Home	*Median Home Value ($000)*
Mean	1,979.490196	195.6509804
Median	1,763	165.9
Mode	#NA	#NA
Minimum	489	100.2
Maximum	4,029	504.5
Range	3,540	404.3
Variance	11,065.8549	7,418.7265
Standard Deviation	900.5919	86.1320
Coefficient of variance	45.50%	44.02%
Skewness	0.6423	1.6988
Kurtosis	–0.5014	3.3069
Count	51	51
Standard Error	126.1081	12.0609

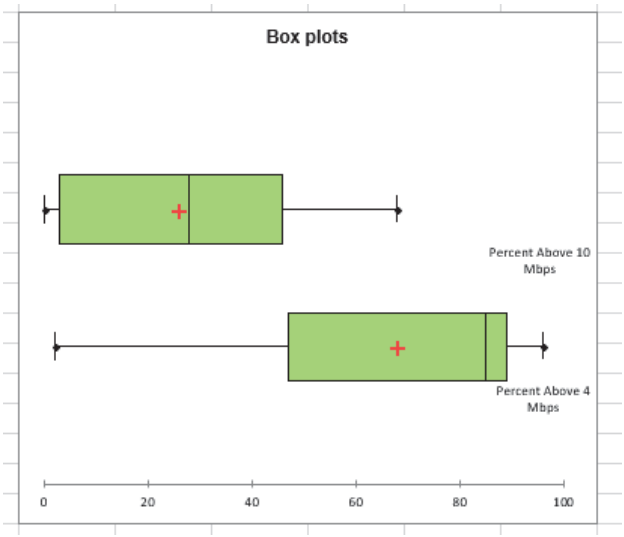
(c) The boxplot shows that taxes are right skewed and the median value of homes is highly right skewed.(d) The coefficient of correlation is –0.041. (e) There is a large variation in taxes and the median value of homes from state to state.

3.77 (a), (b)

Statistic	Average Connection Speed	Average Peak Connection Speed	Percent Above 4 Mbps	Percent Above 10 Mbps
Range	19.000	123.600	93.900	67.900
1st Quartile	4.650	29.550	47.000	2.900
Median	8.700	47.900	85.000	28.000
3rd Quartile	12.550	58.100	89.000	46.000
Mean	9.028	47.473	67.655	25.865
Variance (n-1)	22.511	535.193	837.039	478.819
Standard deviation (n-1)	4.745	23.134	28.932	21.882
Variation coefficient	0.521	0.483	0.424	0.838
IQR	7.900	28.550	42.000	43.100

3.77 cont. (c)





(c) The average connection speed is slightly right skewed. The average peak connection speed is interesting, the mean and median are approximately equal which suggest symmetry but the distance from Q1 to the median is larger than the distance from the median to Q3 which suggest that the data is left skewed however the distance from the smallest data value to the median is less than the distance from the median to the largest data value which suggest the data is right skewed. If the outliers were removed the mean would fall below the median and the data would be left-skewed.

The percent of time above 10 Mbps is slightly left skewed. The percent of time above 4 Mbps is extremely left skewed with the mean considerably less than the median.

(d) The correlation between mean connection speed and mean peak connection speed is 0.768.

The correlation between percent of time the speed is above 4 Mbps and the percent of the time the connection speed is above 10 Mbps is 0.800

(e) The average of the average connections speeds for the various countries surveyed is 9.028 Mbps. Half of the countries surveyed have average connection speeds below 8.7 Mbps. One-quarter of the countries surveyed have average connection speeds below 4.65 Mbps while another one-quarter have average connection speeds above 12.55 Mbps.

The range of average connection speeds is 19 Mbps. The middle 50% of the countries have average connection speed spread over 7.9 Mbps. The typical spread of the average connection speed around the mean is 4.745 Mbps.

The average of the average peak connections speeds for the various countries surveyed is 47.473 Mbps. Half of the countries surveyed have average peak connection speeds below 47.90 Mbps. One-quarter of the countries surveyed have average peak connection speeds below 29.550 Mbps while another one-quarter have average peak connection speeds above 58.1 Mbps. The range of average peak connections speeds is 123.6 Mbps. The middle 50% of the countries have average connection peak speed spread over 28.550 Mbps. The typical spread of the average peak connection speed around the mean is 23.134 Mbps.

For the countries surveyed, the average of the percent of time the connection speed is above 4Mbs is 67.655% while the average of the percent of time the connection speed is above 10Mbps is 25.865%.

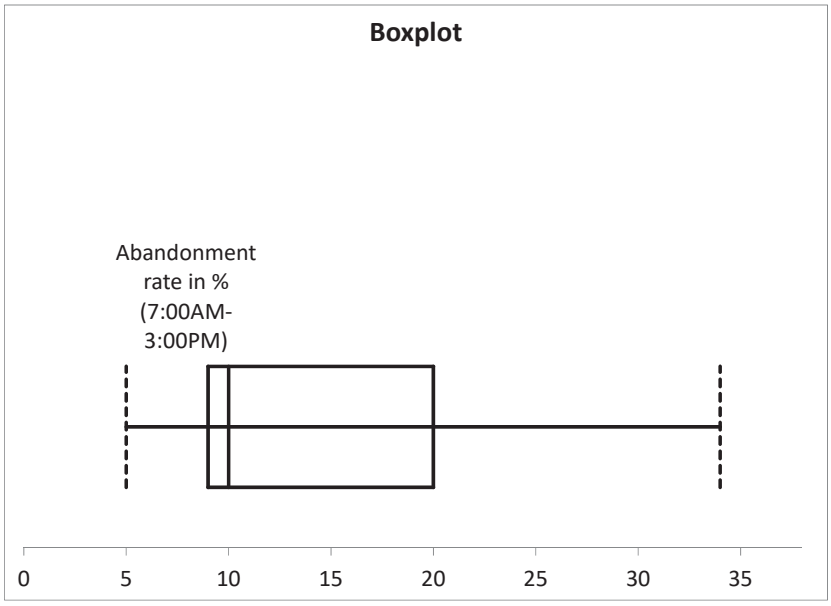
3.77 cont. (e) Because the coefficient of variation is 52.1% for average connection speed, 48.3% for average peak connection speed, 42.4% for percent of time above 4 Mbps and 83.8% for percent of time above 10 Mbps one concludes that relative to the mean, percent time above 10 Mbps is the much more variable than the other measures.

(f) There is a positive linear relationship between mean connection speed and mean peak connection speed and there is also a positive linear relationship between percent of time connection speed is above 4 Mbps and the percent of time connection speed is above 10 Mbps.

3.78 (a), (b)

	Abandonment rate in % (7:00AM-3:00PM)
Mean	13.86363636
Standard Error	1.625414306
Median	10
Mode	9
Standard Deviation	7.623868875
Sample Variance	58.12337662
Kurtosis	0.723568739
Skewness	1.180708144
Range	29
Minimum	5
Maximum	34
Sum	305
Count	22
First Quartile	9
Third Quartile	20
Interquartile Range	11
CV	54.99%

(c)



The data are right-skewed.

3.78 (d) $r = 0.7575$

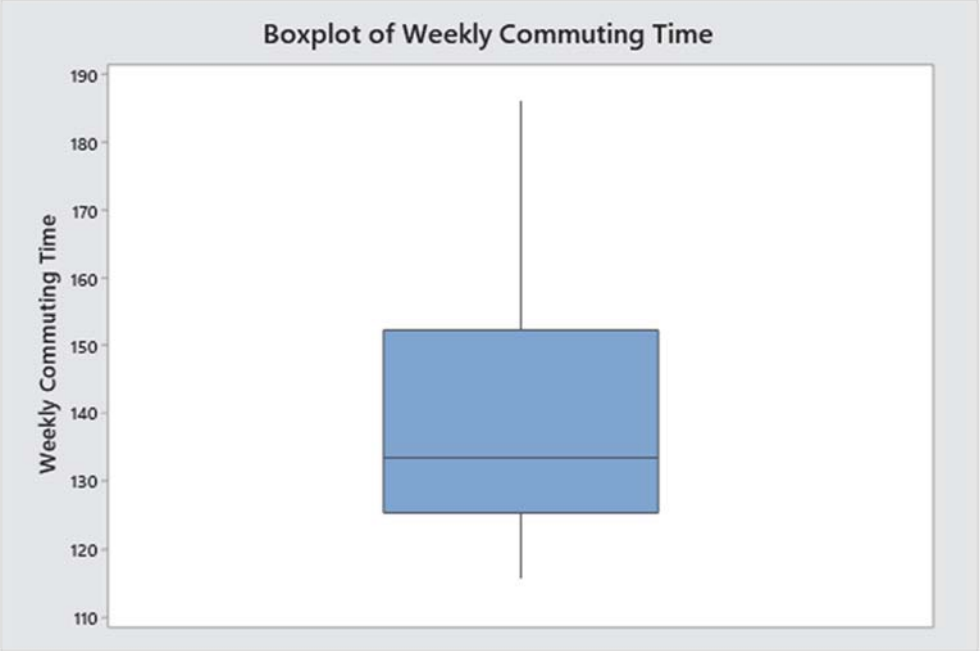
cont. (e) The average abandonment rate is 13.86%. Half of the abandonment rates are less than 10%. One-quarter of the abandonment rates are less than 9% while another one-quarter are more than 20%. The overall spread of the abandonment rates is 29%. The middle 50% of the abandonment rates are spread over 11%. The average spread of abandonment rates around the mean is 7.62%. The abandonment rates are right-skewed.

3.79 (a)–(c)

Descriptive Statistics: Weekly Commuting Time

Statistics

Variable	Mean	StDev	Variance	CoefVar	Q1	Median	Q3	Range	IQR
Weekly Commuting Time	138.83	17.51	306.63	12.61	125.50	133.50	152.25	70.00	26.75



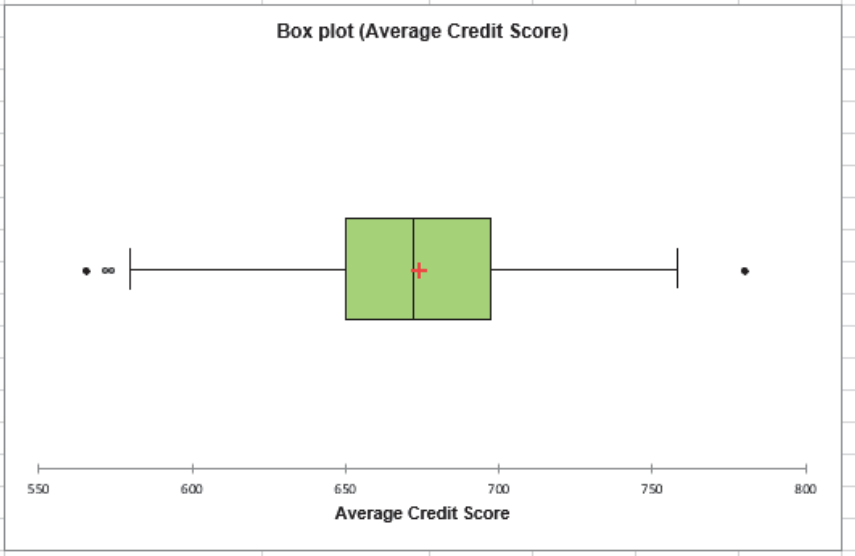
The average weekly commuting time is right-skewed.

(d) The average of the average weekly commuting time is 138.83 minutes. Half of the average weekly commuting time is less than 133.50 minutes. One-quarter of the average weekly commuting time is less than 125.5 minutes while another one-quarter is more than 152.25 minutes. The range of average weekly commuting time is 70 minutes. The middle 50% of the average weekly commuting time spreads over 26.75 minutes. The typical spread of average weekly commuting time around the mean is 17.51.

3.80 (a)–(c) Excel output:

Statistic	verage Credit Score
Range	214.510
1st Quartile	649.835
Median	672.015
3rd Quartile	697.195
Mean	673.244
Variance (n-1)	1005.878
Standard deviation (n-1)	31.716
IQR	47.360
Variation coefficient	0.047

Box plots:

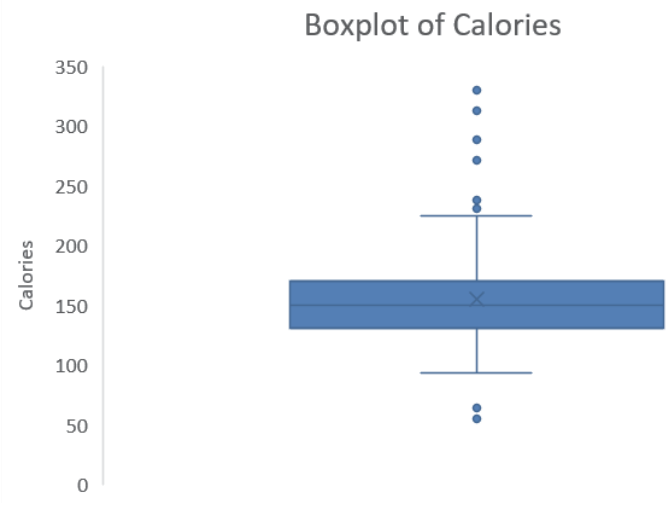


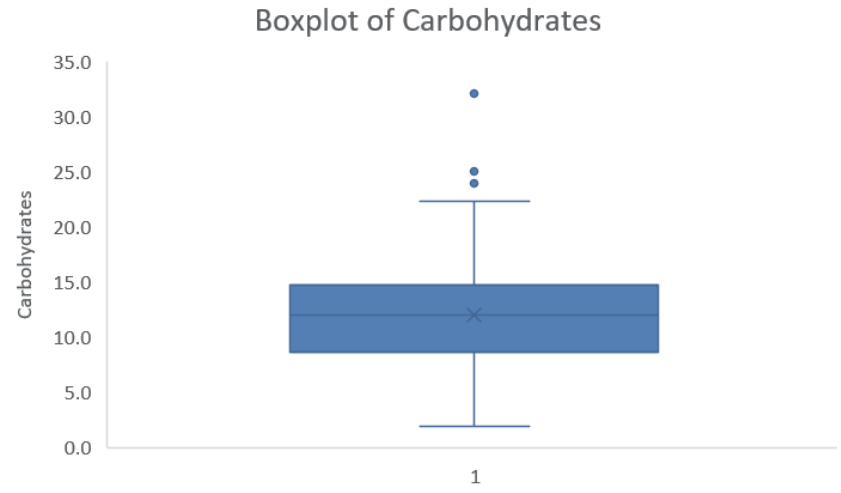
The data are symmetrical.

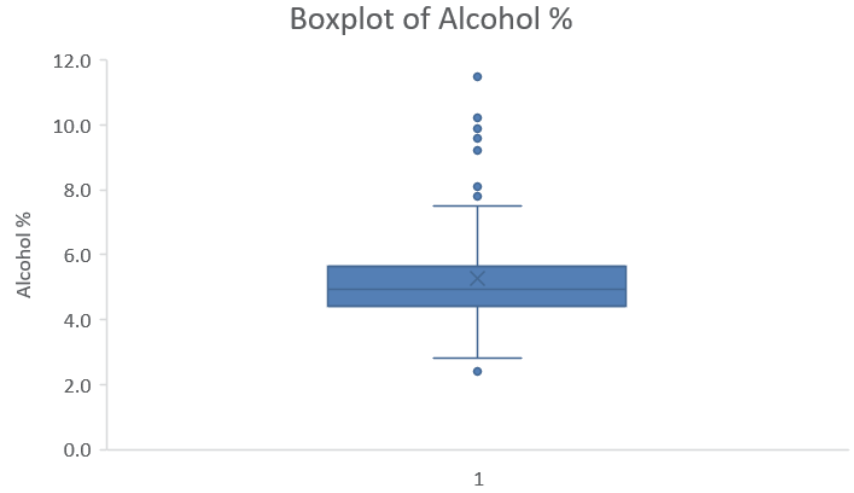
(d) The mean of the average credit scores is 673.24. Half of the average credit scores are less than 672.02. One-quarter of the average credit scores are less than 649.84 while another one-quarter is more than 697.2. The range of the average credit score is 214.51. The middle 50% of the average credit scores is spread over 47.36. The typical spread of the average credit scores around the mean is 31.716.

3.81 The variables "gender" and "major" are categorical and cannot be summarized with boxplots because boxplots are created using the data from numerical variables. Similarly, the mean is a static computed on numerical variables so is not appropriate for the categorical variables "gender" or "major". Pie charts are used for categorical variables, so they should not be created using data from the numerical variables "grade point average" and "height."

3.82 Minitab Output:







3.82
cont.

	A	B
1	Alcohol %	
2		
3	Five-Number Summary	
4	Minimum	2.4
5	First Quartile	4.4
6	Median	4.92
7	Third Quartile	5.65
8	Maximum	11.5
9	IQR	1.25
10	Range	9.1

	A	B
1	**Alcohol**	
2		
3		**Alcohol %**
4	**Mean**	5.270127389
5	**Median**	4.92
6	**Mode**	4.2
7	**Minimum**	2.4
8	**Maximum**	11.5
9	**Range**	9.1
10	**Variance**	1.8337
11	**Standard Deviation**	1.3541
12	**Coeff. of Variation**	25.69%
13	**Skewness**	1.8403
14	**Kurtosis**	4.5833
15	**Count**	157
16	**Standard Error**	0.1081
17		

	A	B
1	Carbohydrates	
2		
3	Five-Number Summary	
4	Minimum	1.9
5	First Quartile	8.65
6	Median	12
7	Third Quartile	14.7
8	Maximum	32.1
9	IQR	6.05
10	Range	30.2
11		

	A	B	C
1	**Carbohydrates**		
2			
3		**Carbohydrates**	
4	**Mean**	12.05235669	
5	**Median**	12	
6	**Mode**	12	
7	**Minimum**	1.9	
8	**Maximum**	32.1	
9	**Range**	30.2	
10	**Variance**	24.8119	
11	**Standard Deviation**	4.9812	
12	**Coeff. of Variation**	41.33%	
13	**Skewness**	0.4908	
14	**Kurtosis**	1.0801	
15	**Count**	157	
16	**Standard Error**	0.3975	
17			

	A	B
1	Calories	
2		
3	Five-Number Summary	
4	Minimum	55
5	First Quartile	130.5
6	Median	150
7	Third Quartile	170.5
8	Maximum	330
9	IQR	40
10	Range	275
11		
12		

	A	B	C
1	**Calories**		
2			
3		**Calories**	
4	**Mean**	155.5095541	
5	**Median**	150	
6	**Mode**	110	
7	**Minimum**	55	
8	**Maximum**	330	
9	**Range**	275	
10	**Variance**	1918.6746	
11	**Standard Deviation**	43.8027	
12	**Coeff. of Variation**	28.17%	
13	**Skewness**	1.2148	
14	**Kurtosis**	2.9712	
15	**Count**	157	
16	**Standard Error**	3.4958	
17			

3.82 cont. The amount of % alcohol is right skewed with an average at 5.278%. Half of the beers have % alcohol below 4.92%. The middle 50% of the beers have alcohol content spread over a range of 1.25%. The highest alcohol content is at 11.5% while the lowest is at 2.4%. The typical spread of alcohol content around the mean is 1.354%.

The number of calories is symmetric with an average at 155.51. Half of the beers have calories below 150.00. The middle 50% of the beers have calories spread over a range of 40. The highest number of calories is 330 while the lowest is 55. The typical spread of calories around the mean is 43.80.

The number of carbohydrates is symmetric with three high outliers from the boxplot with an average at 12.052, which is almost identical to the median at 12.000. Half of the beers have carbohydrates below 12.000. The middle 50% of the beers have carbohydrates spread over a range of 6.05. The highest number of carbohydrates is 32.10 while the lowest is 1.9. The typical spread of carbohydrates around the mean is 4.98.

Chapter 4

4.1 (a) Simple events include tossing a head or tossing a tail.
(b) Joint events include tossing three heads (HHH), a head followed by two tails (HTT), a tail followed by two heads (THH), and three tails (TTT).
(c) Tossing a tail on the first toss
(d) The sample space is the collection of (HHH), (HHT), (HTH), (THH), (TTH), (THT), (HTT), and (TTT).

4.2 (a) Simple events include selecting a red ball.
(b) Selecting a white ball
(c) The sample space is the collection of "a red ball being selected" and "a white ball being selected."

4.3 (a) $\frac{30}{90} = \frac{1}{3} = 0.33$

(b) $\frac{60}{90} = \frac{2}{3} = 0.67$

(c) $\frac{10}{90} = \frac{1}{9} = 0.11$

(d) $\frac{30}{90} + \frac{30}{90} - \frac{10}{90} = \frac{50}{90} = \frac{5}{9} = 0.556$

4.4 (a) $\frac{60}{100} = \frac{3}{5} = 0.6$

(b) $\frac{10}{100} = \frac{1}{10} = 0.1$

(c) $\frac{35}{100} = \frac{7}{20} = 0.35$

(d) $\frac{60}{100} + \frac{65}{100} - \frac{35}{100} = \frac{90}{100} = \frac{9}{10} = 0.9$

4.5 (a) *a priori*
(b) Subjective
(c) *a priori*
(d) Empirical

4.6 (a) Mutually exclusive, not collectively exhaustive.
(b) Not mutually exclusive, not collectively exhaustive.
(c) Mutually exclusive, not collectively exhaustive.
(d) Mutually exclusive, collectively exhaustive

4.7 (a) The joint probability of mutually exclusive events (being listed on the New York Stock Exchange and NASDAQ) is zero.
(b) The joint probability of the events (owning a smartphone and a tablet) is not zero because a consumer can own both a smartphone and a tablet at the same time.

4.7 cont. (c) The joint probability of mutually exclusive events (being an Apple cellphone and a Samsung cellphone) is zero.

(d) The joint probability of the events (an automobile that is a Toyota and was manufactured in the U.S) is not zero because a Toyota can be manufactured in the U.S.

4.8 (a) Is a millennial.

(b) Is a millennial and has a retirement account.

(c) Does not have a retirement account.

(d) Is a millennial and has a retirement account is a joint event because it consists of two characteristics.

4.9 (a) $P(\text{has a retirement savings account}) = \dfrac{579+599}{579+628+599+532} = 0.5038$

(b) $P(\text{a millennial and has a retirement account}) = \dfrac{579}{2,338} = 0.2476$

(c) P(the American adult was a millennial *or* has a retirement account)
$= \dfrac{1,207+1,178-579}{2,338} = 0.7725$

(d) Question (b) asks for the intersection: both conditions being satisfied, the "and" composition. Question (c) asks for the "or" composition: one of the two conditions being satisfied. The difference between the two answers are the outcomes where one of the two conditions is satisfied, but not the other.

4.10 Answers will vary.

(a) A marketer who plans to increase use of LinkedIn.

(b) A B2B marketer who plans to increase use of LinkedIn.

(c) A marketer who does not plan to increase use of LinkedIn.

(d) A marketer who plans to increase use of LinkedIn and is a B2C marketer is a joint event because it consists of two characteristics, plans to increase use of LinkedIn and is a B2C marketer.

4.11 (a) $P(\text{plans to increase use of LinkedIn}) = \dfrac{2,891}{5,726} = 0.5049$

(b) $P(\text{B2C marketer}) = \dfrac{3,779}{5,726} = 0.6600$

(c) P(plans to increase use of LinkedIn or is a B2C marketer)
$= \dfrac{2,891+3,779-1,625}{5,726} = 0.8811$

(d) The probability of "plans to increase use of LinkedIn or is a B2C marketer" includes the probability of "plans to increase use of LinkedIn" and the probability of "is a B2C marketer" minus the probability of "plans to increase use of LinkedIn and is a B2C marketer."

4.12 (a) P(fully supports increased use of educational technologies in higher ed)
$= \dfrac{686}{1,803} = 0.3805$

4.12 (b) $P(\text{is a digital learning leader}) = \dfrac{206}{1{,}803} = 0.1143$

cont. (c) $P(\text{fully supports increased use of educational technologies } \textit{or} \text{ is a digital learning leader}) = \dfrac{686 + 206 - 175}{1{,}803} = \dfrac{717}{1{,}803} = 0.3977$

(d) The probability in (c) includes those who fully support increased use of educational technologies in higher education and those who are digital learning leaders.

4.13 (a) $P(\text{is concerned about limited access to money affecting the business}) = \dfrac{58}{296} = 0.1959$

(b) $P(\text{is a female SMB owner } \textit{and} \text{ is concerned about limited access to money affecting the business}) = \dfrac{38}{296} = 0.1284$

(c) $P(\text{is a female SMB owner } \textit{or} \text{ is concerned about limited access to money affecting the business}) = \dfrac{150 + 58 - 38}{296} = \dfrac{170}{296} = 0.5743$

(d) The probability of "is a female SMB owner *or* is concerned about limited access to money affecting the business" includes the probability of "is a female SMB owner," the probability of "is concerned about limited access to money affecting the business," minus the joint probability of "is a female SMB owner *and* is concerned about limited access to money affecting the business."

4.14

Important to Understand Privacy Policy	Older adults	Younger adults	Total
Yes	911	195	1106
No	90	65	155
Total	1001	260	1261

(a) $P(\text{is important to have a clear understanding of a company's privacy policy before signing up for its service online}) = \dfrac{1{,}106}{1{,}261} = 0.8771$

(b) $P(\text{is an older adult and indicates it is important to have a clear understanding of a company's privacy policy before signing up for its service online}) = \dfrac{911}{1{,}261} = 0.7224$

(c) $P(\text{is an older adult or indicates it is important to have a clear understanding of a company's privacy policy before signing up for its service online}) = \dfrac{1{,}001 + 1{,}106 - 911}{1{,}261} = \dfrac{1{,}196}{1{,}261} = 0.9485$

(d) $P(\text{is an older adult or a younger adult}) = \dfrac{1{,}261}{1{,}261} = 1.00$

4.15

Needs Warranty-Related Repair	U.S.	Non-U.S.	Total
Yes	0.025	0.015	0.04
No	0.575	0.385	0.96
Total	0.600	0.400	1.00

(a) P(needs warranty repair) = 0.04
(b) P(needs warranty repair *and* manufacturer based in U.S.) = 0.025
(c) P(needs warranty repair *or* manufacturer based in U.S.) = 0.615
(d) P(needs warranty repair *or* manufacturer not based in U.S.) = 0.425

4.16 (a) $P(A \mid B) = \frac{10}{30} = \frac{1}{3} = 0.33$

(b) $P(A \mid B') = \frac{20}{60} = \frac{1}{3} = 0.33$

(c) $P(A' \mid B') = \frac{40}{60} = \frac{2}{3} = 0.67$

(d) Since $P(A \mid B) = P(A) = \frac{1}{3}$, events A and B are statistically independent.

4.17 (a) $P(A \mid B) = \frac{10}{35} = \frac{2}{7} = 0.2857$

(b) $P(A' \mid B') = \frac{35}{65} = \frac{7}{13} = 0.5385$

(c) $P(A \mid B') = \frac{30}{65} = \frac{6}{13} = 0.4615$

(d) Since $P(A \mid B) = 0.2857$ and $P(A) = 0.40$, events A and B are not statistically independent.

4.18 $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.4}{0.8} = \frac{1}{2} = 0.5$

4.19 $P(A \text{ and } B) = P(A)\,P(B) = (0.7) \cdot (0.6) = 0.42$

4.20 Since $P(A \text{ and } B) = 0.20$ and $P(A)\,P(B) = 0.12$, events A and B are not statistically independent.

4.21 (a) P(the American adult is a millennial | the American adult has a retirement savings account) $= \frac{579}{1{,}178} = 0.4915$

(b) P(the American adult has a retirement savings account | the American adult is a millennial) $= \frac{579}{1{,}207} = 0.4797$

(c) The conditional events are reversed.

(d) Since P(the American adult is a millennial | the American adult has a retirement savings account) = 0.4915 is not equal to P(the American adult is a millennial) = 0.5163, "the American adult has a retirement savings account" and "age group" are not statistically independent.

4.22 (a) $P(\text{Increased use of LinkedIn} \mid \text{B2B marketer}) = \frac{1,266}{1,947} = 0.6502$

(b) $P(\text{Increased use of LinkedIn} \mid \text{B2C marketer}) = \frac{1,625}{3,779} = 0.4300$

(c) $P(\text{Increased use of LinkedIn}) = 0.5049$ and $P(\text{Increased use of LinkedIn} \mid \text{B2B marketer}) = 0.6502$ are not equal. Therefore, increased use of LinkedIn and business focus are not independent.

4.23 (a) $P(\text{Concerned about limited access to money affecting the business} \mid \text{Female}) = \frac{38}{150} = 0.2533$

(b) $P(\text{Concerned about limited access to money affecting the business} \mid \text{Male}) = \frac{20}{146} = 0.1370$

(c) Since $P(\text{Concerned about limited access to money affecting the business}) = 0.1959$ is not equal to $P(\text{Concerned about limited access to money affecting the business} \mid \text{Female}) = 0.2533$, concern about limited access to money affecting the business and gender are not statistically independent.

4.24 (a) $P(\text{fully supports increased use of educational technologies in higher ed} \mid \text{faculty member}) = \frac{511}{1,597} = 0.3200$

(b) $P(\text{does not fully supports increased use of educational technologies in higher ed} \mid \text{faculty member}) = \frac{1,086}{1,597} = 0.6800$

(c) $P(\text{fully supports increased use of educational technologies in higher ed} \mid \text{digital learning leader}) = \frac{175}{206} = 0.8495$

(d) $P(\text{does not fully supports increased use of educational technologies in higher ed} \mid \text{digital learning leader}) = \frac{31}{206} = 0.1505$

4.25

Important to Understand Privacy Policy	Older adults	Younger adults	Total
Yes	911	195	1106
No	90	65	155
Total	1001	260	1261

(a) $P(\text{important to understand privacy policy} \mid \text{older adult}) = \frac{911}{1,001} = 0.9101$

(b) $P(\text{younger adult} \mid \text{does not indicate that it is important to understand privacy policy}) = \frac{65}{155} = 0.4194$

4.25 cont. (c) Since P(younger adult | does not indicate that it is important to understand privacy policy) = 0.4194 is not equal to P(younger adult) $= \frac{260}{1,261} = 0.2062,$ indicates that it is important to understand privacy policy and adult age are not independent.

4.26

Needs Warranty-Related Repair	U.S.	Non-U.S.	Total
Yes	0.025	0.015	0.04
No	0.575	0.385	0.96
Total	0.600	0.400	1.00

(a) P(needs warranty repair | manufacturer based in U.S.) = $\frac{0.025}{0.6} = 0.0417$

(b) P(needs warranty repair | manufacturer not based in U.S.) = $\frac{0.015}{0.4} = 0.0375$

(c) Since P(needs warranty repair | manufacturer based in U.S.) = $\frac{0.025}{0.6} = 0.0417$ is not equal to P(needs warranty repair) = 0.04, the two events are not independent.

4.27 (a) P(higher for the year) $= \frac{36+8}{1,26136+8+12+12} = 0.6471$

(b) P(higher for the year | higher first week) $= \frac{36}{36+8} = 0.8182$

(c) Since P(higher for the year) = 0.6471 is not equal to P(higher for the year | higher first week) = 0.8182, the two events, first-week performance and annual performance, are not statistically independent.

(d) Answers will vary.

4.28 (a) P(both queens) = $\frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2,652} = \frac{1}{221} = 0.0045$

(b) P(10 followed by 5 or 6) = $\frac{4}{52} \cdot \frac{8}{51} = \frac{32}{2,652} = \frac{8}{663} = 0.012$

(c) P(both queens) = $\frac{4}{52} \cdot \frac{4}{52} = \frac{16}{2,704} = \frac{1}{169} = 0.0059$

(d) P(blackjack) = $\frac{16}{52} \cdot \frac{4}{51} + \frac{4}{52} \cdot \frac{16}{51} = \frac{128}{2,652} = \frac{32}{663} = 0.0483$

4.29 (a) P(2 red cellphones) = $\frac{2}{9} \cdot \frac{1}{8} = \frac{2}{72} = \frac{1}{36} = 0.0278$

(b) P(1 red cellphone *and* 1 black cellphone) = $\frac{7}{9} \cdot \frac{2}{8} + \frac{2}{9} \cdot \frac{7}{8} = \frac{28}{72} = \frac{7}{18} = 0.3889$

(c) P(3 red cellphones) = $\left(\frac{2}{9}\right)^3 = 0.01097$

4.29 (d) (a) $P(\text{2 red cellphones}) = \frac{2}{9} \cdot \frac{2}{9} = \frac{4}{81} = 0.0494$

cont. (b) $P(\text{1 red cellphone } \textit{and} \text{ 1 black cellphone}) = 2\left(\frac{7}{9}\right)\left(\frac{2}{9}\right) = 0.3457$

4.30 With *a priori* probability, the probability of success is based on prior knowledge of the process involved. With empirical probability, outcomes are based on observed data. Subjective probability refers to the chance of occurrence assigned to an event by a particular individual.

4.31 A simple event can be described by a single characteristic. Joint probability refers to phenomena containing two or more events.

4.32 The general addition rule is used by adding the probability of A and the probability of B and then subtracting the joint probability of A and B.

4.33 Events are mutually exclusive if both cannot occur at the same time. Events are collectively exhaustive if one of the events must occur.

4.34 If events A and B are statistically independent, the conditional probability of event A given B is equal to the probability of A.

4.35 When events A and B are independent, the probability of A and B is the product of the probability of event A and the probability of event B. When events A and B are not independent, the probability of A and B is the product of the conditional probability of event A given event B and the probability of event B.

4.36 (a)

	Generation		
Prefer Hybrid Advice	**Boomers**	**Millennials**	**Total**
Yes	140	320	460
No	360	180	540
Total	500	500	1,000

(b) A simple event is "prefers hybrid investment advice." A joint event is "being a baby boomer and preferring hybrid investment advice."

(c) $P(\text{prefers hybrid advice}) = \frac{460}{1,000} = 0.46$

(d) $P(\text{prefers hybrid advice } \textit{and} \text{ is a baby boomer}) = \frac{140}{1,000} = 0.14$

(e) They are not independent because baby boomers and millennials have different probabilities of preferring hybrid investment advice.

4.37 (a) $\text{P(is an HR employee)} = \frac{132}{400} = 0.33$

(b) $P(\text{is an HR employee } \textit{or} \text{ indicates that absenteeism is an important metric}) = \frac{132 + 126 - 54}{400} = \frac{204}{400} = 0.51$

4.37 (c) P(does not indicate that presenteeism is an important metric *and* is a non-HR employee)
cont. $= \frac{225}{400} = 0.5625$

(d) P(does indicate that presenteeism is an important metric *or* is a non-HR employee)
$= \frac{304 + 268 - 225}{400} = \frac{347}{400} = 0.8675$

(e) P(a non-HR employee given they indicate that presenteeism is an important metric)
$= \frac{43}{96} = 0.4479$

(f) They are not independent

(g) They are not independent

4.38 (a) $P(\text{creative display}) = \frac{82}{276} = 0.2971$

(b) $P(\text{creative display } or \text{ informational resources}) = \frac{82 + 33}{276} = \frac{115}{276} = 0.4167$

(c) $P(\text{business website } or \text{ online sales}) = \frac{129 + 45 - 32}{276} = \frac{142}{276} = 0.5145$

(d) $P(\text{business website } and \text{ online sales}) = \frac{32}{276} = 0.1159$

(e) $P(\text{online business presence} \mid \text{personal website}) = \frac{4}{147} = 0.0272$

4.39 (a) $P(\text{lack of process tools is a factor}) = \frac{125}{386} = 0.3238$

(b) $P(\text{lack of process tools is a factor} \mid \text{B2B firm}) = \frac{90}{272} = 0.3309$

(c) $P(\text{lack of process tools is a factor} \mid \text{B2C firm}) = \frac{35}{114} = 0.3070$

(d) $P(\text{lack of people who can link to practice}) = \frac{111}{386} = 0.2876$

(e) $P(\text{lack of people who can link to practice} \mid \text{B2B firm}) = \frac{75}{272} = 0.2757$

(f) $P(\text{lack of people who can link to practice} \mid \text{B2C firm}) = \frac{36}{114} = 0.3158$

(g) There is very little difference between B2B and B2C firms.

Chapter 5

5.1 PHStat output for Distribution A:

Probabilities & Outcomes:	P	X	Y
	0.5	0	
	0.2	1	
	0.15	2	
	0.1	3	
	0.05	4	
Statistics			
E(X)	1		
E(Y)	0		
Variance(X)	1.5		
Standard Deviation(X)	1.224745		
Variance(Y)	0		
Standard Deviation(Y)	0		
Covariance(XY)	0		
Variance(X+Y)	1.5		
Standard Deviation(X+Y)	1.224745		

PHStat output for Distribution B:

Probabilities & Outcomes:	P	X	Y
	0.05	0	
	0.1	1	
	0.15	2	
	0.2	3	
	0.5	4	
Statistics			
E(X)	3		
E(Y)	0		
Variance(X)	1.5		
Standard Deviation(X)	1.224745		
Variance(Y)	0		
Standard Deviation(Y)	0		
Covariance(XY)	0		
Variance(X+Y)	1.5		
Standard Deviation(X+Y)	1.224745		

(a)

	Distribution A			**Distribution B**	
X	*P*(*X*)	*X***P*(*X*)	*X*	*P*(*X*)	*X***P*(*X*)
0	0.50	0.00	0	0.50	0.00
1	0.20	0.20	1	0.20	0.20
2	0.15	0.30	2	0.15	0.30
3	0.10	0.30	3	0.10	0.30
4	0.05	0.20	4	0.05	0.20
	μ=	1.00		μ=	1.00

5.1 cont. (b)

Distribution A

X	$(X-\mu)^2$	$P(X)$	$(X-\mu)^2 * P(X)$
0	$(-1)^2$	0.50	0.50
1	$(0)^2$	0.20	0.00
2	$(1)^2$	0.15	0.15
3	$(2)^2$	0.10	0.40
4	$(3)^2$	0.05	0.45
		$\sigma^2 =$	1.50

$$\sigma = \sqrt{\sum (X-\mu)^2 \cdot P(X)} = 1.22$$

Distribution B

X	$(X-\mu)^2$	$P(X)$	$(X-\mu)^2 * P(X)$
0	$(-3)^2$	0.05	0.45
1	$(-2)^2$	0.10	0.40
2	$(-1)^2$	0.15	0.15
3	$(0)^2$	0.20	0.00
4	$(1)^2$	0.50	0.50
		$\sigma^2 =$	1.50

$$\sigma = \sqrt{\sum (X-\mu)^2 \cdot P(X)} = 1.22$$

(c) For distribution A, $P(X \geq 3) = 0.10 + 0.05 = 0.15$
For distribution B, $P(X \geq 3) = 0.20 + 0.50 = 0.70$

(d) The means are different, but the variances are the same.

5.2 PHStat output:

Probabilities & Outcomes:	P	X
	0.1	0
	0.2	1
	0.45	2
	0.15	3
	0.05	4
	0.05	5
Statistics		
E(X)	2	
E(Y)	0	
Variance(X)	1.4	
Standard Deviation(X)	1.183216	
Variance(Y)	0	
Standard Deviation(Y)	0	
Covariance(XY)	0	
Variance(X+Y)	1.4	
Standard Deviation(X+Y)	1.183216	

5.2 cont. (a)–(b)

X	$P(X)$	$X*P(X)$	$(X-\mu_X)^2$	$(X-\mu_X)^2*P(X)$
0	0.10	0.00	4	0.40
1	0.20	0.20	1	0.20
2	0.45	0.90	0	0.00
3	0.15	0.45	1	0.15
4	0.05	0.20	4	0.20
5	0.05	0.25	9	0.45
	(a) Mean =	2.00	Variance =	1.40
			(b) Stdev =	1.18321596

(c) $P(X \geq 2) = 0.45 + 0.15 + 0.05 + 0.05 = 0.70$

5.3 (a) Based on the fact that the odds of winning are expressed out with a base of 31,478, you would think that the automobile dealership sent out 31,478 fliers.

(b) $\mu = \Sigma_{i=1}^{N} X_i P(X_i) = \$\ 5.80$

(c) $\sigma = \sqrt{\Sigma_{i=1}^{N} \left[X_i - E(X_i) \right]^2 P(X_i)} = \$\ 140.88$

(d) The total cost of the prizes is \$25,000 + \$100 + 31,476 * \$5 = \$182,480. Assuming that the cost of producing the fliers is negligible, the cost of reaching a single customer is \$182,480/31,478 = \$5.80. The effectiveness of the promotion will depend on how many customers will show up in the show room.

5.4 (a)

X	$P(X)$
\$ – 1	21/36
\$ + 1	15/36

(b)

X	$P(X)$
\$ – 1	21/36
\$ + 1	15/36

(c)

X	$P(X)$
\$ – 1	30/36
\$ + 4	6/36

(d) \$ – 0.167 for each method of play

5.5 Excel Output:

	A	B	C	D	E	F
1	**Arrivals = X**	**Frequency**	**Probability = Frequency/n**	**[X-E(X)]^2**		
2	0	14	0.07	8.410	=(A2-B13)^2	
3	1	31	0.155	3.610	=(A3-B13)^2	
4	2	47	0.235	0.810	=(A4-B13)^2	
5	3	41	0.205	0.010	=(A5-B13)^2	
6	4	29	0.145	1.210	=(A6-B13)^2	
7	5	21	0.105	4.410	=(A7-B13)^2	
8	6	10	0.05	9.610	=(A8-B13)^2	
9	7	5	0.025	16.810	=(A9-B13)^2	
10	8	2	0.01	26.010	=(A10-B13)^2	
11		**Statistics**				
12	n	200	=SUM(B2:B10)			
13	E(x)	2.9	=SUMPRODUCT(A2:A10,C2:C10)			
14	Variance (X)	3.14	=SUMPRODUCT(C2:C10,D2:D10)			
15	Standard Deviation (X)	1.772004515	=SQRT(B14)			

(a) $\mu = E(X) = 2.9$

(b) $\sigma = 1.77$

(c) $P(X < 2) = 0.07 + 0.155 = 0.225$

5.6 Excel Output:

	A	B	C	D	E	F
1	**Arrivals = X**	**Frequency**	**Probability = Frequency/n**	**[X-E(X)]^2**		
2	0	13	0.125	4.434	=(A2-B12)^2	
3	1	25	0.240384615	1.223	=(A3-B12)^2	
4	2	32	0.307692308	0.011	=(A4-B12)^2	
5	3	17	0.163461538	0.800	=(A5-B12)^2	
6	4	9	0.086538462	3.588	=(A6-B12)^2	
7	5	6	0.057692308	8.377	=(A7-B12)^2	
8	6	1	0.009615385	15.165	=(A8-B12)^2	
9	7	1	0.009615385	23.953	=(A9-B12)^2	
10		**Statistics**				
11	n	104	=SUM(B2:B9)			
12	E(x)	2.105769231	=SUMPRODUCT(A2:A9,C2:C9)			
13	Variance (X)	2.152274408	=SUMPRODUCT(C2:C9,D2:D9)			
14	Standard Deviation (X)	1.467063192	=SQRT(B13)			
15	P(X>1)	0.634615385	=SUM(C4:C9)			

(a) $\mu = E(X) = 2.1058$

(b) $\sigma = 1.4671$

(c) $P(X > 1) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) = 0.6346$

5.7 PHStat output:

Probabilities & Outcomes:	P	X	Y
	0.1	–50	–100
	0.3	20	50
	0.4	100	130
	0.2	150	200

Statistics			
E(X)	71		
E(Y)	97		
Variance(X)	3829		
Standard Deviation(X)	61.87891		
Variance(Y)	7101		
Standard Deviation(Y)	84.26743		
Covariance(XY)	5113		
Variance(X+Y)	21156		
Standard Deviation(X+Y)	145.451		

(a) $E(X) = \$71$ $E(Y) = \$97$

(b) $\sigma_X = 61.88$ $\sigma_Y = 84.27$

(c) Stock Y gives the investor a higher expected return than stock X, but also has a higher standard deviation. Risk-averse investors would invest in stock X, whereas risk takers would invest in stock Y.

5.8 Excel Output:

	A	B	C	D	E
1	Corporate Bond Fund X	Probability	[X-E(X)]^2		
2	-300	0.01	128307.24	=(A2-B12)^2	
3	-70	0.09	16435.24	=(A3-B12)^2	
4	30	0.15	795.24	=(A4-B12)^2	
5	60	0.35	3.24	=(A5-B12)^2	
6	100	0.3	1747.24	=(A6-B12)^2	
7	120	0.1	3819.24	=(A7-B12)^2	
8					
9					
10					
11	Statistics				
12	E(X)	58.2	=SUMPRODUCT(A2:A7,B2:B7)		
13	Variance(X)	3788.76	=SUMPRODUCT(B2:B7,C2:C7)		
14	Standard Deviation (X)	61.55290407	=SQRT(B13)		
15					
16					

5.8
cont.

	A	B	C	D	E
1	Common Stock Fund X	Probability	[X-E(X)]^2		
2	-999	0.01	1127865.24	=(A2-B12)^2	
3	-300	0.09	131776.2601	=(A3-B12)^2	
4	-100	0.15	26572.2601	=(A4-B12)^2	
5	100	0.35	1368.2601	=(A5-B12)^2	
6	150	0.3	7567.2601	=(A6-B12)^2	
7	350	0.1	82363.2601	=(A7-B12)^2	
8					
9					
10					
11	Statistics				
12	E(X)	63.01	=SUMPRODUCT(A2:A7,B2:B7)		
13	Variance(X)	38109.7499	=SUMPRODUCT(B2:B7,C2:C7)		
14	Standard Deviation (X)	195.2171865	=SQRT(B13)		
15					

(a) E(Bond Fund) = $58.20; E(Common Stock Fund) = $63.01

(b) $\sigma_{\text{bond fund}} = \$61.55$; $\sigma_{\text{common stock fund}} = \$195.22$

(c) Based on the expected value criteria, you would choose the common stock fund. However, the common stock fund also has a standard deviation more than three times higher than that for the corporate bond fund. An investor should carefully weigh the increased risk.

(d) If you chose the common stock fund, you would need to assess your reaction to the small possibility that you could lose virtually all of your entire investment.

5.9 (a) 0.5997

(b) 0.0016

(c) 0.0439

(d) 0.4018

PHstat output for part (d):

Binomial Probabilities						
Data						
Sample size	6					
Probability of an event of interest	0.83					
Statistics						
Mean	4.98					
Variance	0.8466					
Standard deviation	0.920109					
Binomial Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	2.41E-05	2.41E-05	0	0.999976	1
	1	0.000707	0.000731	2.41E-05	0.999269	0.999976
	2	0.008631	0.009362	0.000731	0.990638	0.999269
	3	0.056184	0.065546	0.009362	0.934454	0.990638
	4	0.205732	0.271277	0.065546	0.728723	0.934454
	5	0.401782	0.67306	0.271277	0.32694	0.728723
	6	0.32694	1	0.67306	0	0.32694

5.10 (a) $\mu = 4(0.10) = 0.40$ $\sigma = \sqrt{4 \cdot 0.1 \cdot 0.9} = 0.60$

(b) $\mu = 4(0.40) = 1.60$ $\sigma = \sqrt{4 \cdot 0.4 \cdot 0.6} = 0.98$

(c) $\mu = 5(0.80) = 0.40$ $\sigma = \sqrt{5 \cdot 0.8 \cdot 0.2} = 0.894$

(d) $\mu = 3(0.5) = 1.50$ $\sigma = \sqrt{3 \cdot 0.5 \cdot 0.5} = 0.866$

5.11 Given $\pi = 0.5$ and $n = 5$, $P(X = 5) = 0.0312$.

5.12 Excel Output:

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**6**		
5	**Probability of an event of interest**	**0.47**		
6				
7	**Parameters**			
8	**Mean**	**2.82**		
9	**Variance**	**1.4946**		
10	**Standard deviation**	**1.22254**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0222**		
15	**1**	**0.1179**		
16	**2**	**0.2615**		
17	**3**	**0.3091**		
18	**4**	**0.2056**		
19	**5**	**0.0729**		
20	**6**	**0.0108**		
21				
22	Prob(at least 4) = P(X>=4)	0.2893	=SUM(B18:B20)	

(a) $P(X = 4) = 0.2056$

(b) $P(X = 6) = 0.0108$

(c) $P(X \geq 4) = 0.2056 + 0.0729 + 0.0108 = 0.2893$

(d) $\mu = 2.82$, $\sigma = 1.22254$

(e) That each American adult owns an iPhone or does not own an iPhone and that next six adults selected are independent.

5.13 PHStat output:

Data	
Sample size	5
Probability of an event of interest	0.25
Statistics	
Mean	1.25
Variance	0.9375
Standard deviation	0.968246

5.13 cont. PHStat output:

Binomial Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.237305	0.237305	0	0.762695	1
	1	0.395508	0.632813	0.237305	0.367188	0.762695
	2	0.263672	0.896484	0.632813	0.103516	0.367188
	3	0.087891	0.984375	0.896484	0.015625	0.103516
	4	0.014648	0.999023	0.984375	0.000977	0.015625
	5	0.000977	1	0.999023	0	0.000977

If $\pi = 0.25$ and $n = 5$,

(a) $P(X = 5) = 0.0010$

(b) $P(X \geq 4) = P(X = 4) + P(X = 5) = 0.0146 + 0.0010 = 0.0156$

(c) $P(X = 0) = 0.2373$

(d) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$
$= 0.2373 + 0.3955 + 0.2637 = 0.8965$

5.14 Excel Output:

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**10**		
5	**Probability of an event of interest**	**0.03**		
6				
7	**Parameters**			
8	**Mean**	**0.3**		
9	**Variance**	**0.291**		
10	**Standard deviation**	**0.53944**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.7374**		
15	**1**	**0.2281**		
16	**2**	**0.0317**		
17	**3**	**0.0026**		
18	**4**	**0.0001**		
19	**5**	**0.0000**		
20	**6**	**0.0000**		
21	**7**	**0.0000**		
22	**8**	**0.0000**		
23	**9**	**0.0000**		
24	**10**	**0.0000**		
25				
26	Prob(two or fewer) = P(X<=2)	0.9972	=SUM(B14:B16)	
27	Prob(Three or more) = P(X>=3)	0.0028	=SUM(B17:B24)	

(a) $P(X = 0) = 0.7374$

(b) $P(X = 1) = 0.2281$

5.14 cont. (c) $P(X \leq 2) = 0.9972$

(d) $P(X \geq 3) = 0.0028$

5.15 Partial PHStat output:

Binomial Probabilities						
Data						
Sample size	20					
Probability of an event of interest	0.08					
Statistics						
Mean	1.6					
Variance	1.4720					
Standard deviation	1.2133					
Binomial Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.1887	0.1887	0.0000	0.8113	1.0000
	1	0.3282	0.5169	0.1887	0.4831	0.8113
	2	0.2711	0.7879	0.5169	0.2121	0.4831

(a) $\mu = 1.6$ $\sigma = 1.2133$

(b) $P(X = 0) = 0.1887$

(c) $P(X = 1) = 0.3282$

(d) $P(X \geq 2) = 0.4831$

5.16 Excel Output:

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	3		
5	**Probability of an event of interest**	0.909		
6				
7	**Parameters**			
8	**Mean**	2.727		
9	**Variance**	0.24816		
10	**Standard deviation**	0.49815		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	0	0.0008		
15	1	0.0226		
16	2	0.2256		
17	3	0.7511		
18				
19	Prob(at least 2) = P(X>=2)	0.9767	=SUM(B16:B17)	

(a) $P(X = 3) = 0.7511$

(b) $P(X = 0) = 0.0008$

(c) $P(X \geq 2) = 0.9767$

5.16. (d) $\mu = 2.727$ $\sigma_X = 0.49815$

cont. On the average, over the long run, you theoretically expect 2.727 orders to be filled correctly in a sample of 3 orders with a standard deviation of 0.49815

(e) McDonald's has a slightly higher probability of filling orders correctly and Wendy's has a slightly lower probability

5.17 Excel Output

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**3**		
5	**Probability of an event of interest**	**0.929**		
6				
7	**Parameters**			
8	**Mean**	**2.787**		
9	**Variance**	**0.19788**		
10	**Standard deviation**	**0.44483**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0004**		
15	**1**	**0.0140**		
16	**2**	**0.1838**		
17	**3**	**0.8018**		
18				
19	Prob(at least 2) = P(X>=2)	0.9856	=SUM(B16:B17)	

(a) $P(X = 3) = 0.8018$

(b) $P(X = 0) = 0.0004$

(c) $P(X \geq 2) = 0.9856$

(d) $\mu = 2.787$ $\sigma_X = 0.44483$

On the average, over the long run, you theoretically expect 2.787 orders to be filled correctly in a sample of 3 orders with a standard deviation of 0.44483

(e) Out of all three fast-food restaurants, McDonald's has the highest probability of filling orders correctly.

5.18 (a) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:				2.5		
Poisson Probabilities Table						
	X	**P(X)**	**P(<=X)**	**P(<X)**	**P(>X)**	**P(>=X)**
	2	0.256516	0.543813	0.287297	0.456187	0.712703

Using the equation, if $\lambda = 2.5, P(X = 2) = \dfrac{e^{-2.5} \cdot (2.5)^2}{2!} = 0.2565$

5.18 cont. (b) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:				8		
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	8	0.139587	0.592547	0.452961	0.407453	0.547039

If $\lambda = 8.0$, $P(X = 8) = 0.1396$

(c) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:				0.5		
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.606531	0.606531	0.000000	0.393469	1.000000
	1	0.303265	0.909796	0.606531	0.090204	0.393469

If $\lambda = 0.5$, $P(X = 1) = 0.3033$

(d) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:				3.7		
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.024724	0.024724	0.000000	0.975276	1.000000

If $\lambda = 3.7$, $P(X = 0) = 0.0247$

5.19 (a) Partial PHStat output:

Poisson Probabilities						
Data						
Mean/Expected number of events of interest:				2		
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.135335	0.135335	0.000000	0.864665	1.000000
	1	0.270671	0.406006	0.135335	0.593994	0.864665
	2	0.270671	0.676676	0.406006	0.323324	0.593994

If $\lambda = 2.0$, $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - [0.1353 + 0.2707] = 0.5940$

5.19 cont. (b) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:			8			
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.000335	0.000335	0.000000	0.999665	1.000000
	1	0.002684	0.003019	0.000335	0.996981	0.999665
	2	0.010735	0.013754	0.003019	0.986246	0.996981
	3	0.028626	0.042380	0.013754	0.957620	0.986246

If $\lambda = 8.0$, $P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$
$= 1 - [0.0003 + 0.0027 + 0.0107] = 1 - 0.0137 = 0.9863$

(c) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:			0.5			
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.606531	0.606531	0.000000	0.393469	1.000000
	1	0.303265	0.909796	0.606531	0.090204	0.393469

If $\lambda = 0.5$, $P(X \leq 1) = P(X = 0) + P(X = 1) = 0.6065 + 0.3033 = 0.9098$

(d) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:			4			
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.018316	0.018316	0.000000	0.981684	1.000000
	1	0.073263	0.091578	0.018316	0.908422	0.981684

If $\lambda = 4.0$, $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.0183 = 0.9817$

5.19 cont. (e) Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:				5		
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.006738	0.006738	0.000000	0.993262	1.000000
	1	0.033690	0.040428	0.006738	0.959572	0.993262
	2	0.084224	0.124652	0.040428	0.875348	0.959572
	3	0.140374	0.265026	0.124652	0.734974	0.875348

If $\lambda = 5.0$, $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$
$= 0.0067 + 0.0337 + 0.0842 + 0.1404 = 0.2650$

5.20 PHStat output for (a) – (d)

Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.006738	0.006738	0.000000	0.993262	1.000000
	1	0.033690	0.040428	0.006738	0.959572	0.993262
	2	0.084224	0.124652	0.040428	0.875348	0.959572
	3	0.140374	0.265026	0.124652	0.734974	0.875348
	4	0.175467	0.440493	0.265026	0.559507	0.734974
	5	0.175467	0.615961	0.440493	0.384039	0.559507
	6	0.146223	0.762183	0.615961	0.237817	0.384039
	7	0.104445	0.866628	0.762183	0.133372	0.237817
	8	0.065278	0.931906	0.866628	0.068094	0.133372
	9	0.036266	0.968172	0.931906	0.031828	0.068094
	10	0.018133	0.986305	0.968172	0.013695	0.031828
	11	0.008242	0.994547	0.986305	0.005453	0.013695
	12	0.003434	0.997981	0.994547	0.002019	0.005453
	13	0.001321	0.999302	0.997981	0.000698	0.002019
	14	0.000472	0.999774	0.999302	0.000226	0.000698
	15	0.000157	0.999931	0.999774	0.000069	0.000226
	16	0.000049	0.999980	0.999931	0.000020	0.000069
	17	0.000014	0.999995	0.999980	0.000005	0.000020
	18	0.000004	0.999999	0.999995	0.000001	0.000005
	19	0.000001	1.000000	0.999999	0.000000	0.000001
	20	0.000000	1.000000	1.000000	0.000000	0.000000

Given λ=5.0,
(a) $P(X = 1) = 0.0337$
(b) $P(X < 1) = 0.0067$
(c) $P(X > 1) = 0.9596$
(d) $P(X \leq 1) = 0.0404$

5.21 Excel Output:

	A	B	C	D	E
1	Poisson Probabilities				
2					
3	Data				
4	Mean/Expected number of events of interest:				3.13
5					
6	Poisson Probabilities Table				
7	X	P(X)			
8	0	0.0437			
9	1	0.1368			
10	2	0.2141			
11	3	0.2234			
12	4	0.1748			
13	5	0.1094			
14	6	0.0571			
15	7	0.0255			
16	8	0.0100			

(a) $P(X = 0) = 0.0437$

(b) $P(X = 1) = 0.1368$

(c) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)] = 1 - [0.0437 + 0.1368] = 0.8194$

(d) $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 0.0437 + 0.1368 + 0.2141 = 0.3947$

5.22 (a)–(c) Portion of PHStat output

Data						
Average/Expected number of successes:				6		
Poisson Probabilities Table						
	X	P(X)	P(<=X)	P(<X)	P(>X)	P(>=X)
	0	0.002479	0.002479	0.000000	0.997521	1.000000
	1	0.014873	0.017351	0.002479	0.982649	0.997521
	2	0.044618	0.061969	0.017351	0.938031	0.982649
	3	0.089235	0.151204	0.061969	0.848796	0.938031
	4	0.133853	0.285057	0.151204	0.714943	0.848796
	5	(b) 0.160623	0.445680	(a) 0.285057	0.554320	(c) 0.714943
	6	0.160623	0.606303	0.445680	0.393697	0.554320
	7	0.137677	0.743980	0.606303	0.256020	0.393697
	8	0.103258	0.847237	0.743980	0.152763	0.256020
	9	0.068838	0.916076	0.847237	0.083924	0.152763
	10	0.041303	0.957379	0.916076	0.042621	0.083924
	11	0.022529	0.979908	0.957379	0.020092	0.042621
	12	0.011264	0.991173	0.979908	0.008827	0.020092
	13	0.005199	0.996372	0.991173	0.003628	0.008827
	14	0.002228	0.998600	0.996372	0.001400	0.003628
	15	0.000891	0.999491	0.998600	0.000509	0.001400
	16	0.000334	0.999825	0.999491	0.000175	0.000509
	17	0.000118	0.999943	0.999825	0.000057	0.000175

5.22 (a) $P(X < 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

cont. $= \frac{e^{-6}(6)^0}{0!} ++ \frac{e^{-6}(6)^2}{2!} + \frac{e^{-6}(6)^3}{3!} + \frac{e^{-6}(6)^4}{4!}$

$= 0.002479 + 0.014873 + 0.044618 + 0.089235 + 0.133853 = 0.2851$

(b) $P(X = 5) = \frac{e^{-6}(6)^5}{5!} = 0.1606$

(c) $P(X \geq 5) = 1 - P(X < 5) = 1 - 0.2851 = 0.7149$

(d) $P(X = 4 \text{ or } X = 5) = P(X = 4) + P(X = 5) = \frac{e^{-6}(6)^4}{4!} + \frac{e^{-6}(6)^5}{5!} = 0.2945$

$= \frac{e^{-6}(6)^1}{1!}$

5.23 Partial PHStat output:

Poisson Probabilities						
Data						
Average/Expected number of successes:				6		
Poisson Probabilities Table						
	X	**P(X)**	**P(<=X)**	**P(<X)**	**P(>X)**	**P(>=X)**
	3	0.089235	0.151204	0.061969	0.848796	0.938031

If $\lambda = 6.0$, $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$= 0.0025 + 0.0149 + 0.0446 + 0.0892 = 0.1512$

$n \cdot P(X \leq 3) = 100 \cdot (0.1512) = 15.12$, so 15 or 16 cookies will probably be discarded.

5.24 Excel Output:

	A	B	C	D	E
1	**Poisson Probabilities**				
2					
3	**Data**				
4	**Mean/Expected number of events of interest:**				**0.32**
5					
6	**Poisson Probabilities Table**				
7	***X***	***P(X)***			
8	**0**	**0.7261**			
9	**1**	**0.2324**			
10	**2**	**0.0372**			
11	**3**	**0.0040**			
12	**4**	**0.0003**			
13	**5**	**0.0000**			
14	**6**	**0.0000**			

(a) $P(X = 0) = 0.7261$

(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.7261 = 0.2739$

(c) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - [0.7261 + 0.2324] = 0.0415$

5.25 Excel Output:

	A	B	C	D	E
1	Poisson Probabilities				
2					
3	Data				
4	Mean/Expected number of events of interest:				0.9
5					
6	Poisson Probabilities Table				
7	X	P(X)			
8	0	0.4066			
9	1	0.3659			
10	2	0.1647			
11	3	0.0494			
12	4	0.0111			
13	5	0.0020			
14	6	0.0003			
15	7	0.0000			

(a) $P(X = 0) = 0.4066$
(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.4066 = 0.5934$
(c) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - [0.4066 + 0.3659] = 0.2275$

5.26 Excel Output:

	A	B	C	D	E	F	G
1	Poisson Probabilities						
2							
3	Data						
4	Mean/Expected number of events of interest:				3.5		
5							
6	Poisson Probabilities Table						
7	X	P(X)		(c) P(more than 1) = 1- P(X < =1)		0.8641	
8	0	0.0302		(d) P(fewer than 2) = P(X <=1)		0.1359	
9	1	0.1057					
10	2	0.1850					
11	3	0.2158					
12	4	0.1888					
13	5	0.1322					
14	6	0.0771					
15	7	0.0385					
16	8	0.0169					
17	9	0.0066					
18	10	0.0023					
19	11	0.0007					
20	12	0.0002					
21	13	0.0001					

(a) $P(X = 0) = 0.0302$
(b) $P(X = 1) = 0.1057$
(c) $P(X > 1) = 1 - P(X \leq 1) = 1 - [0.0302 + 0.1057] = 0.8641$
(d) $P(X < 2) = P(X \leq 1) = 0.0302 + 0.1057 = 0.1359$

5.27 Excel Output:

	A	B	C	D	E
1	Poisson Probabilities				
2					
3	Data				
4	Mean/Expected number of events of interest:				1.48
5					
6	Poisson Probabilities Table				
7	*X*	*P(X)*			
8	0	0.2276			
9	1	0.3369			
10	2	0.2493			
11	3	0.1230			
12	4	0.0455			
13	5	0.0135			
14	6	0.0033			
15	7	0.0007			
16	8	0.0001			
17	9	0.0000			

(a) For the number of problems with 2019 model Ford to be distributed as a Poisson random variable, we need to assume that (i) the probability that a problem occurs in a given Ford is the same for any other new Ford, (ii) the number of problems that a Ford has is independent of the number of problems any other Ford has, (iii) the probability that two or more problems will occur in some area of a Ford approaches zero as the area becomes smaller. Yes, these assumptions are reasonable in this problem.

(b) $P(X = 0) = 0.2276$

(c) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.8139$

(d) An operational definition for problem can be "a specific feature in the car that is not performing according to its intended designed function." The operational definition is important in interpreting the initial quality score because different customers can have different expectations of what function a feature is supposed to perform.

5.28 Excel Output:

	A	B	C	D	E
1	Poisson Probabilities				
2					
3	Data				
4	Mean/Expected number of events of interest:				1.08
5					
6	Poisson Probabilities Table				
7	*X*	*P(X)*			
8	0	0.3396			
9	1	0.3668			
10	2	0.1981			
11	3	0.0713			
12	4	0.0193			
13	5	0.0042			
14	6	0.0007			
15	7	0.0001			
16	8	0.0000			

(a) $P(X = 0) = 0.3396$

5.28 (b) $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.9044$

cont. (c) Because Ford had a higher mean rate of problems per car than Toyota, the probability of a randomly selected Ford having zero problems and the probability of no more than two problems are both lower than for Toyota.

5.29 Partial PHStat output:

Poisson Probabilities						
Data						
Mean/Expected number of events of interest:				0.8		
Poisson Probabilities Table						
	X	**P(X)**	**P(<=X)**	**P(<X)**	**P(>X)**	**P(>=X)**
	0	0.449329	0.449329	0.000000	0.550671	1.000000
	1	0.359463	0.808792	0.449329	0.191208	0.550671
	2	0.143785	0.952577	0.808792	0.047423	0.191208
	3	0.038343	0.990920	0.952577	0.009080	0.047423
	4	0.007669	0.998589	0.990920	0.001411	0.009080
	5	0.001227	0.999816	0.998589	0.000184	0.001411
	6	0.000164	0.999979	0.999816	0.000021	0.000184
	7	0.000019	0.999998	0.999979	0.000002	0.000021
	8	0.000002	1.000000	0.999998	0.000000	0.000002

(a) For the number of phone calls received in a 1-minute period to be distributed as a Poisson random variable, we need to assume that (i) the probability that a phone call is received in a given 1-minute period is the same for all the other 1-minute periods, (ii) the number of phone calls received in a given 1-minute period is independent of the number of phone calls received in any other 1-minute period, (iii) the probability that two or more phone calls received in a time period approaches zero as the length of the time period becomes smaller.

(b) $\lambda = 0.8$, $P(X = 0) = 0.4493$

(c) $\lambda = 0.8$, $P(X \ge 3) = 0.0474$

(d) $\lambda = 0.8$, $P(X \le 6) = 0.999979$. A maximum of 6 phone calls will be received in a 1-minute period 99.99% of the time.

5.30 The expected value is the average of a probability distribution. It is the value that can be expected to occur on the average, in the long run.

5.31 The four properties of a situation that must be present in order to use the binomial distribution are
(i) the sample consists of a fixed number of observations, n,
(ii) each observation can be classified into one of two mutually exclusive and collectively exhaustive categories, usually called "an event of interest" and "not an event of interest",
(iii) the probability of an observation being classified as "an event of interest", π, is constant from observation to observation and
(iv) the outcome (i.e., "an event of interest" or "not an event of interest") of any observation is independent of the outcome of any other observation.

5.32 The four properties of a situation that must be present in order to use the Poisson distribution are
(i) you are interested in counting the number of times a particular event occurs in a given area of opportunity (defined by time, length, surface area, and so forth),

5.32 cont. (ii) the probability that an event occurs in a given area of opportunity is the same for all of the areas of opportunity,
(iii) the number of events that occur in one area of opportunity is independent of the number of events that occur in other areas of opportunity and
(iv) the probability that two or more events will occur in an area of opportunity approaches zero as the area of opportunity becomes smaller.

5.33 (a) PHStat output:

Covariance Analysis

Probabilities & Outcomes:	P	X	Y
	0.001	-1000000	
	0.999	4000	

Statistics	
E(X)	2996
E(Y)	0
Variance(X)	1.01E+09
Standard Deviation(X)	31733.39
Variance(Y)	0
Standard Deviation(Y)	0
Covariance(XY)	0
Variance(X+Y)	1.01E+09
Standard Deviation(X+Y)	31733.39

The expected value of the profit made by the insurance company is $2996.

(b) On average, the promoter will have to pay $4000 while the insurance company will make a profit of $2996. This is not a win-win opportunity.

5.34 (a) 0.664
(b) 0.664
Excel Output

	A	B
1	Binomial Probabilities	
2		
3	Data	
4	Sample size	5
5	Probability of an event of interest	0.664
6		
7	Parameters	
8	Mean	3.32
9	Variance	1.11552
10	Standard deviation	1.05618
11		
12	Binomial Probabilities Table	
13	*X*	*P(X)*
14	0	0.0043
15	1	0.0423
16	2	0.1672
17	3	0.3305
18	4	0.3266
19	5	0.1291

$\pi = 0.664,\ n = 5$

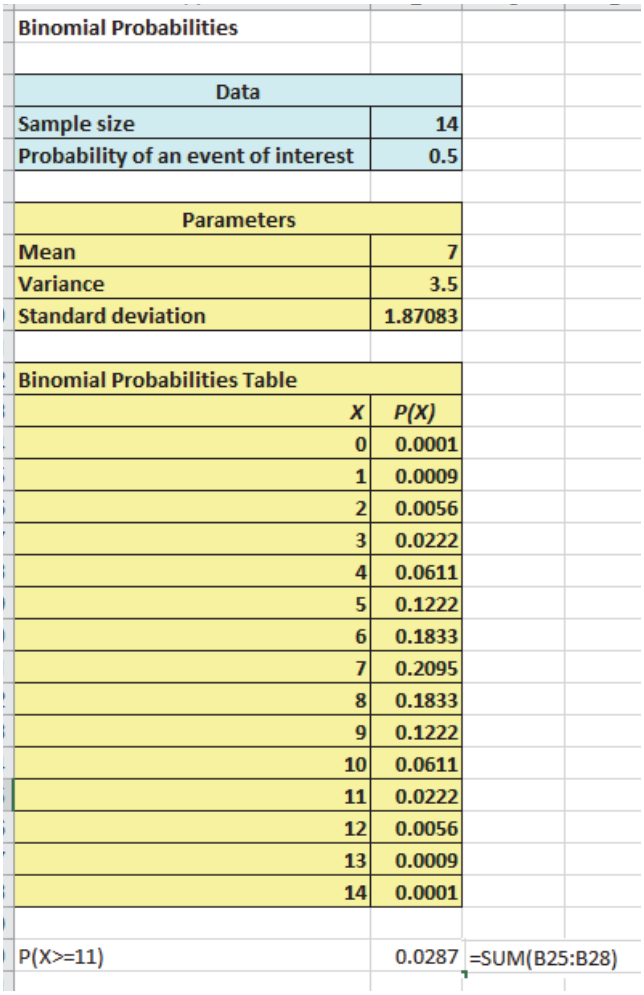
5.34 cont.
(c) $P(X = 4) = 0.3266$
(d) $P(X = 0) = 0.0043$
(e) Stock prices tend to rise in the years when the economy is expanding and fall in the years of recession or contraction. Hence, the probability that the price will rise in one year is not independent from year to year.

5.35 Excel Output

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**10**		
5	**Probability of an event of interest**	**0.91**		
6				
7	**Parameters**			
8	**Mean**	**9.1**		
9	**Variance**	**0.819**		
10	**Standard deviation**	**0.90499**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0000**		
15	**1**	**0.0000**		
16	**2**	**0.0000**		
17	**3**	**0.0000**		
18	**4**	**0.0001**		
19	**5**	**0.0009**		
20	**6**	**0.0078**		
21	**7**	**0.0452**		
22	**8**	**0.1714**		
23	**9**	**0.3851**		
24	**10**	**0.3894**		
25				
26	P(at least 8) = P(X>=8)	0.9460	=SUM(B22:B24)	
27	P(atmost 6) = P(X<=6)	0.0088	=SUM(B14:B20)	

(a) $P(X = 8) = 0.1714$
(b) $P(X \geq 8) = 0.9460$
(c) $P(X \leq 6) = 0.0088$
(d) The probability that only three respondents use two or more social media channels is /or 0. If the probability that a retail brand uses two or more social media channels for business is 0.91 it is essentially impossible that only three business in 10 would use two or more social media channels. We might conclude that this geographical area has very limited internet access and it is not appropriate to use the model in this area.

5.36 Excel Output



Binomial Probabilities		
Data		
Sample size	14	
Probability of an event of interest	0.5	
Parameters		
Mean	7	
Variance	3.5	
Standard deviation	1.87083	
Binomial Probabilities Table		
X	*P(X)*	
0	0.0001	
1	0.0009	
2	0.0056	
3	0.0222	
4	0.0611	
5	0.1222	
6	0.1833	
7	0.2095	
8	0.1833	
9	0.1222	
10	0.0611	
11	0.0222	
12	0.0056	
13	0.0009	
14	0.0001	
P(X>=11)	0.0287	=SUM(B25:B28)

(a) $P(X \geq 11) = 0.0287$

5.36 cont. (b) Excel Output

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**14**		
5	**Probability of an event of interest**	**0.75**		
6				
7	**Parameters**			
8	**Mean**	**10.5**		
9	**Variance**	**2.625**		
10	**Standard deviation**	**1.62019**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0000**		
15	**1**	**0.0000**		
16	**2**	**0.0000**		
17	**3**	**0.0000**		
18	**4**	**0.0003**		
19	**5**	**0.0018**		
20	**6**	**0.0082**		
21	**7**	**0.0280**		
22	**8**	**0.0734**		
23	**9**	**0.1468**		
24	**10**	**0.2202**		
25	**11**	**0.2402**		
26	**12**	**0.1802**		
27	**13**	**0.0832**		
28	**14**	**0.0178**		
29				
30	P(X>=11)	0.5213	=SUM(B25:B28)	
31				

$P(X \geq 11) = 0.5213$

5.37 Excel Output:

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**10**		
5	**Probability of an event of interest**	**0.75**		
6				
7	**Parameters**			
8	**Mean**	**7.5**		
9	**Variance**	**1.875**		
10	**Standard deviation**	**1.36931**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0000**		
15	**1**	**0.0000**		
16	**2**	**0.0004**		
17	**3**	**0.0031**		
18	**4**	**0.0162**		
19	**5**	**0.0584**		
20	**6**	**0.1460**		
21	**7**	**0.2503**		
22	**8**	**0.2816**		
23	**9**	**0.1877**		
24	**10**	**0.0563**		
25				
26	P(x >5)	0.9219	=SUM(B20:B24)	

(a) $P(X = 0) = 0.000$
(b) $P(X = 5) = 0.0584$
(c) $P(X > 5) = 0.9219$
(d) $\mu = 7.5,\ \sigma = 1.3693$

5.38 Excel Output:

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**10**		
5	**Probability of an event of interest**	**0.4**		
6				
7	**Parameters**			
8	**Mean**	**4**		
9	**Variance**	**2.4**		
10	**Standard deviation**	**1.54919**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0060**		
15	**1**	**0.0403**		
16	**2**	**0.1209**		
17	**3**	**0.2150**		
18	**4**	**0.2508**		
19	**5**	**0.2007**		
20	**6**	**0.1115**		
21	**7**	**0.0425**		
22	**8**	**0.0106**		
23	**9**	**0.0016**		
24	**10**	**0.0001**		
25				
26	P(x >5)	0.1662	=SUM(B20:B24)	

(a) $P(X = 0) = 0.0060$

(b) $P(X = 5) = 0.2007$

(c) $P(X > 5) = 0.1662$

(d) $\mu = 4,\ \sigma = 1.54919$

(e) Since the percentage of bills containing an error is lower in this problem, the probability is higher in (a) and (b) of this problem and lower in (c).

5.39 Excel Output:
One or two word searches

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**10**		
5	**Probability of an event of interest**	**0.46**		
6				
7	**Parameters**			
8	**Mean**	**4.6**		
9	**Variance**	**2.484**		
10	**Standard deviation**	**1.576071064**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.002108**		
15	**1**	**0.017960**		
16	**2**	**0.068846**		
17	**3**	**0.156391**		
18	**4**	**0.233138**		
19	**5**	**0.238319**		
20	**6**	**0.169177**		
21	**7**	**0.082351**		
22	**8**	**0.026306**		
23	**9**	**0.004980**		
24	**10**	**0.000424**		
25				
26	P(more than 5) = P(X> 5)	0.2832	=SUM(B20:B24)	
27				

5.39 cont. Five or six word searches

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**10**		
5	**Probability of an event of interest**	**0.21**		
6				
7	**Parameters**			
8	**Mean**	**2.1**		
9	**Variance**	**1.659**		
10	**Standard deviation**	**1.288021739**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.094683**		
15	**1**	**0.251688**		
16	**2**	**0.301070**		
17	**3**	**0.213417**		
18	**4**	**0.099279**		
19	**5**	**0.031669**		
20	**6**	**0.007015**		
21	**7**	**0.001066**		
22	**8**	**0.000106**		
23	**9**	**0.000006**		
24	**10**	**0.000000**		
25				
26	P(more than 5) = P(X> 5)	0.0082	=SUM(B20:B24)	

(a) $n = 10$, $\pi = 0.46$ $P(X > 5) = 0.2832$

(b) $n = 10$, $\pi = 0.21$ $P(X > 5) = 0.0082$

(c) $n = 10$, $\pi = 0.46$ $P(X = 0) = 0.0021$

(d) The assumptions needed are (i) there are only two mutually exclusive and collectively exhaustive outcomes – "one or two word searches" or "not one or two word searches" and "five or six word searches" or "not five or six word searches", (ii) the probabilities are constant, and (iii) the outcome are independent

5.40 Excel Output:

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**20**		
5	**Probability of an event of interest**	**0.46**		
6				
7	**Parameters**			
8	**Mean**	**9.2**		
9	**Variance**	**4.968**		
10	**Standard deviation**	**2.2289**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0000**		
15	**1**	**0.0001**		
16	**2**	**0.0006**		
17	**3**	**0.0031**		
18	**4**	**0.0113**		
19	**5**	**0.0309**		
20	**6**	**0.0658**		
21	**7**	**0.1122**		
22	**8**	**0.1553**		
23	**9**	**0.1763**		
24	**10**	**0.1652**		
25	**11**	**0.1280**		
26	**12**	**0.0818**		
27	**13**	**0.0429**		
28	**14**	**0.0183**		
29	**15**	**0.0062**		
30	**16**	**0.0017**		
31	**17**	**0.0003**		
32	**18**	**0.0000**		
33	**19**	**0.0000**		
34	**20**	**0.0000**		
35				
36	P(No more than 5) = P(X<=5)	0.0461	=SUM(B14:B19)	
37	P(5 or more) = P(X>=5)	0.9848	=SUM(B19:B34)	

(a) $E(X) = \mu = 9.2$

(b) $\sigma = 2.2289$

(c) $P(X = 10) = 0.1652$

(d) $P(X \leq 5) = 0.0461$

(e) $P(X \geq 5) = 0.9848$

5.41 Excel Output

	A	B	C	D
1	**Binomial Probabilities**			
2				
3	**Data**			
4	**Sample size**	**20**		
5	**Probability of an event of interest**	**0.24**		
6				
7	**Parameters**			
8	**Mean**	**4.8**		
9	**Variance**	**3.648**		
10	**Standard deviation**	**1.91**		
11				
12	**Binomial Probabilities Table**			
13	***X***	***P(X)***		
14	**0**	**0.0041**		
15	**1**	**0.0261**		
16	**2**	**0.0783**		
17	**3**	**0.1484**		
18	**4**	**0.1991**		
19	**5**	**0.2012**		
20	**6**	**0.1589**		
21	**7**	**0.1003**		
22	**8**	**0.0515**		
23	**9**	**0.0217**		
24	**10**	**0.0075**		
25	**11**	**0.0022**		
26	**12**	**0.0005**		
27	**13**	**0.0001**		
28	**14**	**0.0000**		
29	**15**	**0.0000**		
30	**16**	**0.0000**		
31	**17**	**0.0000**		
32	**18**	**0.0000**		
33	**19**	**0.0000**		
34	**20**	**0.0000**		
35				
36	P(No more than 2) = P(X<=2)	0.1085	=SUM(B14:B16)	
37	P(3 or more) = P(X>=3)	0.8915	=SUM(B17:B34)	

(a) $E(X) = \mu = 4.8$

(b) $\sigma = 1.91$

(c) $P(X = 0) = 0.0041$

(d) $P(X \leq 2) = 0.1085$

(e) $P(X \geq 3) = 0.8915$

5.42 Excel Output:

(a)	0.0000	=1-BINOM.DIST(35,44,0.5,1)
(b)	0.0564	=1-BINOM.DIST(35,44,0.7,1)
(c)	0.9720	=1-BINOM.DIST(35,44,0.9,1)

(a) $\pi = 0.5$, $P(X \geq 36) = 0.0000$

(b) $\pi = 0.7$, $P(X \geq 36) = 0.0564$

5.42 cont. (c) $\pi = 0.9$, $P(X \geq 36) = 0.9720$

(d) Based on the results in (a)–(c), the probability that the Standard & Poor's 500 index will increase if there is an early gain in the first five trading days of the year is very likely to be close to 0.90 because that yields a probability of 97.20% that at least 36 of the 44 years the Standard & Poor's 500 index will increase the entire year.

5.43 Excel Output:

(a)	0.000152932	=1-BINOM.DIST(37,50,0.5,1)

(a) $\pi = 0.5$, $P(X \geq 38) = 0.000152932$ which is essentially zero.

(b) It is ludicrous to believe that there is a correlation between the performance of the stock market and the winner of a Super Bowl. If the indicator is a random event, the probability of making a correct prediction 38 or more times out of 50 trials is virtually zero.

5.44 Excel Output

	A	B	C	D	E
1	**Poisson Probabilities**				
2					
3		**Data**			
4	**Mean/Expected number of events of interest:**				**7**
5					
6	**Poisson Probabilities Table**				
7	***X***	***P(X)***			
8	**0**	**0.0009**			
9	**1**	**0.0064**			
10	**2**	**0.0223**			
11	**3**	**0.0521**			
12	**4**	**0.0912**			
13	**5**	**0.1277**			
14	**6**	**0.1490**			
15	**7**	**0.1490**			
16	**8**	**0.1304**			
17	**9**	**0.1014**			
18	**10**	**0.0710**			
19	**11**	**0.0452**			
20	**12**	**0.0263**			
21	**13**	**0.0142**			
22	**14**	**0.0071**			
23	**15**	**0.0033**			
24	**16**	**0.0014**			
25	**17**	**0.0006**			
26	**18**	**0.0002**			
27	**19**	**0.0001**			
28	**20**	**0.0000**			
29					
30	P(10 or fewer)	0.9015	=SUM(B8:B18)		
31	P(11 or more))	0.0985	=SUM(B19:B28)		

5.44 cont.

(a) The assumptions needed are (i) the probability that a questionable claim is referred by an investigator is constant, (ii) the probability that a questionable claim is referred by an investigator approaches 0 as the interval gets smaller, and (iii) the probability that a questionable claim is referred by an investigator is independent from interval to interval.

(b) $P(X = 5) = 0.1277$

(c) $P(X \leq 10) = 0.9015$

(d) $P(X \geq 11) = 0.0985$

Chapter 6

6.1 PHStat output:

Normal Probabilities				
Common Data				
Mean	0			
Standard Deviation	1			
			Probability for a Range	
Probability for X <=			From X Value	1.57
X Value	1.57		To X Value	1.84
Z Value	1.57		Z Value for 1.57	1.57
P(X<=1.57)	0.9417924		Z Value for 1.84	1.84
			P(X<=1.57)	0.9418
Probability for X >			P(X<=1.84)	0.9671
X Value	1.84		P(1.57<=X<=1.84)	0.0253
Z Value	1.84			
P(X>1.84)	0.0329		Find X and Z Given Cum. Pctage.	
			Cumulative Percentage	95.00%
Probability for X<1.57 or X >1.84			Z Value	1.644854
P(X<1.57 or X >1.84)	0.9747		X Value	1.644854

(a) $P(Z < 1.57) = 0.9418$
(b) $P(Z > 1.84) = 1 - 0.9671 = 0.0329$
(c) $P(1.57 < Z < 1.84) = 0.9671 - 0.9418 = 0.0253$
(d) $P(Z < 1.57) + P(Z > 1.84) = 0.9418 + (1 - 0.9671) = 0.9747$

6.2 PHStat output:

Normal Probabilities				
Common Data				
Mean	0			
Standard Deviation	1			
			Probability for a Range	
Probability for X <=			From X Value	1.57
X Value	–1.57		To X Value	1.84
Z Value	–1.57		Z Value for 1.57	1.57
P(X<=–1.57)	0.0582076		Z Value for 1.84	1.84
			P(X<=1.57)	0.9418
Probability for X >			P(X<=1.84)	0.9671
X Value	1.84		P(1.57<=X<=1.84)	0.0253
Z Value	1.84			
P(X>1.84)	0.0329		Find X and Z Given Cum. Pctage.	
			Cumulative Percentage	84.13%
Probability for X<–1.57 or X >1.84			Z Value	0.999815
P(X<–1.57 or X >1.84)	0.0911		X Value	0.999815

(a) $P(-1.57 < Z < 1.84) = 0.9671 - 0.0582 = 0.9089$
(b) $P(Z < -1.57) + P(Z > 1.84) = 0.0582 + 0.0329 = 0.0911$

6.2 (c) If $P(Z > A) = 0.025$, $P(Z < A) = 0.975$. $A = +1.96$.

cont. (d) If $P(-A < Z < A) = 0.6826$, $P(Z < A) = 0.8413$. So 68.26% of the area is captured between $-A = -1.00$ and $A = +1.00$.

6.3 (a) Partial PHStat output:

Normal Probabilities				
Common Data				
Mean	0			
Standard Deviation	1			
			Probability for a Range	
Probability for X <=			From X Value	1.57
X Value	1.08		To X Value	1.84
Z Value	1.08		Z Value for 1.57	1.57
P(X<=1.08)	0.8599289		Z Value for 1.84	1.84
			P(X<=1.57)	0.9418
Probability for X >			P(X<=1.84)	0.9671
X Value	–0.21		P(1.57<=X<=1.84)	0.0253
Z Value	–0.21			
P(X>–0.21)	0.5832		Find X and Z Given Cum. Pctage.	
			Cumulative Percentage	84.13%
Probability for X<1.08 or X >–0.21			Z Value	0.999815
P(X<1.08 or X >–0.21)	1.4431		X Value	0.999815

$P(Z < 1.08) = 0.8599$

(b) $P(Z > -0.21) = 1.0 - 0.4168 = 0.5832$

(c) Partial PHStat output:

Probability for X<–0.21 or X >0	
P(X<–0.21 or X >0)	0.9168

$P(Z < -0.21) + P(Z > 0) = 0.4168 + 0.5 = 0.9168$

(d) Partial PHStat output:

Probability for X<–0.21 or X >1.08	
P(X<–0.21 or X >1.08)	0.5569

$P(Z < -0.21) + P(Z > 1.08) = 0.4168 + (1 - 0.8599) = 0.5569$

6.4 PHStat output:

Normal Probabilities				
Common Data				
Mean	0			
Standard Deviation	1			
			Probability for a Range	
Probability for X <=			From X Value	–1.96
X Value	–0.21		To X Value	–0.21
Z Value	–0.21		Z Value for –1.96	–1.96
P(X<=–0.21)	0.4168338		Z Value for –0.21	–0.21
			P(X<=–1.96)	0.0250
Probability for X >			P(X<=–0.21)	0.4168
X Value	1.08		P(–1.96<=X<=–0.21)	0.3918
Z Value	1.08			
P(X>1.08)	0.1401		Find X and Z Given Cum. Pctage.	
			Cumulative Percentage	84.13%
Probability for X<–0.21 or X >1.08			Z Value	0.999815
P(X<–0.21 or X >1.08)	0.5569		X Value	0.999815

(a) $P(Z > 1.08) = 1 - 0.8599 = 0.1401$

(b) $P(Z < -0.21) = 0.4168$

(c) $P(-1.96 < Z < -0.21) = 0.4168 - 0.0250 = 0.3918$

(d) $P(Z > A) = 0.1587$, $P(Z < A) = 0.8413$. $A = +1.00$.

6.5 (a) Partial PHStat output:

Common Data	
Mean	100
Standard Deviation	10
Probability for X <=	
X Value	70
Z Value	–3
P(X<=70)	0.0013499
Probability for X >	
X Value	75
Z Value	–2.5
P(X>75)	0.9938

$$Z = \frac{X - \mu}{\sigma} = \frac{75 - 100}{10} = -2.50$$

$P(X > 75) = P(Z > -2.50) = 1 - P(Z < -2.50) = 1 - 0.0062 = 0.9938$

(b) $$Z = \frac{X - \mu}{\sigma} = \frac{70 - 100}{10} = -3.00$$

$P(X < 70) = P(Z < -3.00) = 0.00135$

6.5 cont. (c) Partial PHStat output:

Common Data	
Mean	100
Standard Deviation	10
Probability for X <=	
X Value	80
Z Value	–2
P(X<=80)	0.0227501
Probability for X >	
X Value	110
Z Value	1
P(X>110)	0.1587
Probability for X<80 or X >110	
P(X<80 or X >110)	0.1814

$$Z = \frac{X - \mu}{\sigma} = \frac{80 - 100}{10} = -2.00 \qquad Z = \frac{X - \mu}{\sigma} = \frac{110 - 100}{10} = 1.00$$

$P(X < 80) = P(Z < -2.00) = 0.0228$

$P(X > 110) = P(Z > 1.00) = 1 - P(Z < 1.00) = 1.0 - 0.8413 = 0.1587$

$P(X < 80) + P(X > 110) = 0.0228 + 0.1587 = 0.1815$

(d) $P(X_{\text{lower}} < X < X_{\text{upper}}) = 0.80$

$P(-1.28 < Z) = 0.10$ and $P(Z < 1.28) = 0.90$

$$Z = -1.28 = \frac{X_{\text{lower}} - 100}{10} \qquad Z = +1.28 = \frac{X_{\text{upper}} - 100}{10}$$

$X_{\text{lower}} = 100 - 1.28(10) = 87.20$ and $X_{\text{upper}} = 100 + 1.28(10) = 112.80$

6.6 (a) Partial PHStat output:

Common Data				
Mean	50			
Standard Deviation	4			
			Probability for a Range	
Probability for X <=			From X Value	42
X Value	42		To X Value	43
Z Value	–2		Z Value for 42	–2
P(X<=42)	0.0227501		Z Value for 43	–1.75
			P(X<=42)	0.0228
Probability for X >			P(X<=43)	0.0401
X Value	43		P(42<=X<=43)	0.0173
Z Value	–1.75			
P(X>43)	0.9599		Find X and Z Given Cum. Pctage.	
			Cumulative Percentage	5.00%
Probability for X<42 or X >43			Z Value	–1.644854
P(X<42 or X >43)	0.9827		X Value	43.42059

$P(X > 43) = P(Z > -1.75) = 1 - 0.0401 = 0.9599$

6.6 cont. (b) $P(X < 42) = P(Z < -2.00) = 0.0228$

(c) $P(X < A) = 0.05$,

$$Z = -1.645 = \frac{A - 50}{4} \qquad A = 50 - 1.645(4) = 43.42$$

(d) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	80.00%
Z Value	0.841621
X Value	53.36648

$P(X_{lower} < X < X_{upper}) = 0.60$

$P(Z < -0.84) = 0.20$ and $P(Z < 0.84) = 0.80$

$$Z = -0.84 = \frac{X_{lower} - 50}{4} \qquad Z = +0.84 = \frac{X_{upper} - 50}{4}$$

$X_{lower} = 50 - 0.84(4) = 46.64$ and $X_{upper} = 50 + 0.84(4) = 53.36$

6.7 (a) $P(X > 33) = P(Z > -0.91) = 0.8186$

Probability for X >	
X Value	33
Z Value	-0.91
P(X>33)	0.8186

(b) $P(10 < X < 20) = P(-3.21 < Z < -2.21) = 0.0129$

Probability for a Range	
From X Value	10
To X Value	20
Z Value for 10	-3.21
Z Value for 20	-2.21
P(X<=10)	0.0007
P(X<=20)	0.0136
P(10<=X<=20)	0.0129

(c) $P(X < 10) = P(Z < -3.21) = 0.0007$

Probability for X <=	
X Value	10
Z Value	-3.21
P(X<=10)	0.0007

(d) $P(X < A) = 0.99 \qquad Z = 2.33 = \frac{A - 42.1}{10} \qquad A = 65.36$

Find X and Z Given a Cum. Pctage.	
Cumulative Percentage	99.00%
Z Value	2.33
X Value	65.36

Note: The above answers are obtained using Excel **COMPUTE worksheet** of the **Normal workbook**. They may be slightly different when Table E.2 is used.

(e) The per capita consumption of bottled water in China is much lower than the per capita consumption of bottled water in the United states.

6.8 Partial PHStat output:

Common Data	
Mean	50
Standard Deviation	12

Probability for X <=	
X Value	30
Z Value	–1.666667
P(X<=30)	0.0477904

Probability for X >	
X Value	60
Z Value	0.8333333
P(X>60)	0.2023

Probability for X<30 or X >60	
P(X<30 or X >60)	0.2501

Probability for a Range	
From X Value	34
To X Value	50
Z Value for 34	–1.333333
Z Value for 50	0
P(X<=34)	0.0912
P(X<=50)	0.5000
P(34<=X<=50)	0.4088

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	20.00%
Z Value	–0.841621
X Value	39.90055

(a) $P(34 < X < 50) = P(-1.33 < Z < 0) = 0.4088$

(b) $P(X < 30) + P(X > 60) = P(Z < -1.67) + P(Z > 0.83)$
$= 0.0475 + (1.0 - 0.7967) = 0.2508$

(c) $P(X > \text{A}) = 0.80 \qquad P(Z < -0.84) \cong 0.20 \qquad Z = -0.84 = \dfrac{A - 50}{12}$

$A = 50 - 0.84(12) = 39.92$ thousand miles or 39,920 miles

(d) Partial PHStat output:

Common Data	
Mean	50
Standard Deviation	10

Probability for X <=	
X Value	30
Z Value	–2
P(X<=30)	0.0227501

Probability for X >	
X Value	60
Z Value	1
P(X>60)	0.1587

Probability for X<30 or X >60	
P(X<30 or X >60)	0.1814

Probability for a Range	
From X Value	34
To X Value	50
Z Value for 34	–1.6
Z Value for 50	0
P(X<=34)	0.0548
P(X<=50)	0.5000
P(34<=X<=50)	0.4452

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	20.00%
Z Value	–0.841621
X Value	41.58379

The smaller standard deviation makes the Z-values larger.

(a) $P(34 < X < 50) = P(-1.60 < Z < 0) = 0.4452$

(b) $P(X < 30) + P(X > 60) = P(Z < -2.00) + P(Z > 1.00)$
$= 0.0228 + (1.0 - 0.8413) = 0.1815$

(c) $A = 50 - 0.84(10) = 41.6$ thousand miles or 41,600 miles

6.9 (a) $P(X > 100) = P(Z > -2.52) = .9941$

Probability for X >	
X Value	100
Z Value	-2.52
P(X>100)	0.9941

(b) What is the probability thart a randomly selected millennial speent between \$100 and \$200 per month?

$P(100 < X < 200) = P(-2.52 < Z < 1.48) = .9247$

Probability for a Range	
From X Value	100
To X Value	200
Z Value for 100	-2.52
Z Value for 200	1.48
P(X<=100)	0.0059
P(X<=200)	0.9306
P(100<=X<=200)	0.9247

(c) $P(X_{lower} < X < X_{upper}) = 0.95$

$P(Z < -1.96) = 0.0250$ and $P(Z < 1.96) = 0.9750$

$$Z = -1.96 = \frac{X_{lower} - 163}{25} \qquad Z = +1.96 = \frac{X_{upper} - 163}{25}$$

$X_{lower} = (-1.96)(25) + 163 = 114$

$X_{upper} = (1.96)(25) + 163 = 212$

Find X Values Given a Percentage	
Percentage	95.00%
Z Value	-1.96
Lower X Value	114.00
Upper X Value	212.00

6.10 PHStat output:

Common Data	
Mean	73
Standard Deviation	8

Probability for X <=	
X Value	91
Z Value	2.25
P(X<=91)	0.9877755

Probability for X >	
X Value	81
Z Value	1
P(X>81)	0.1587

Probability for X<91 or X >81	
P(X<91 or X >81)	1.1464

Probability for a Range	
From X Value	65
To X Value	89
Z Value for 65	–1
Z Value for 89	2
P(X<=65)	0.1587
P(X<=89)	0.9772
P(65<=X<=89)	0.8186

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	95.00%
Z Value	1.644854
X Value	86.15883

6.10 cont. (a) $P(X < 91) = P(Z < 2.25) = 0.9878$

(b) $P(65 < X < 89) = P(-1.00 < Z < 2.00) = 0.9772 - 0.1587 = 0.8185$

(c) $P(X > A) = 0.05$ $P(Z < 1.645) = 0.9500$

$$Z = 1.645 = \frac{A - 73}{8} \qquad A = 73 + 1.645(8) = 86.16\%$$

(d) Option 1: $P(X > A) = 0.10$ $P(Z < 1.28) \cong 0.9000$

$$Z = \frac{81 - 73}{8} = 1.00$$

Since your score of 81% on this exam represents a Z-score of 1.00, which is below the minimum Z-score of 1.28, you will not earn an "A" grade on the exam under this grading option.

Option 2: $Z = \frac{68 - 62}{3} = 2.00$

Since your score of 68% on this exam represents a Z-score of 2.00, which is well above the minimum Z-score of 1.28, you will earn an "A" grade on the exam under this grading option. You should prefer Option 2.

6.11 (a) $P(X > 33) = P(Z > 0.9425) = 0.1730$

Probability for X <=	
X Value	10
Z Value	-1.9325
P(X<=10)	0.0266

(b) $P(10 < X < 20) = P(-1.9325 < Z < -0.6825) = 0.2208$

Probability for a Range	
From X Value	10
To X Value	20
Z Value for 10	-1.9325
Z Value for 20	-0.6825
P(X<=10)	0.0266
P(X<=20)	0.2475
P(10<=X<=20)	0.2208

(c) $P(X < 10) = P(Z < -1.9325) = 0.0266$

Probability for X <=	
X Value	10
Z Value	-1.9325
P(X<=10)	0.0266

(d) $P(X < A) = 0.99$ $Z = 2.33 = \frac{A - 25.46}{8}$ $A = 44.07$

Find X and Z Given a Cum. Pctage.	
Cumulative Percentage	99.00%
Z Value	2.33
X Value	44.07

(e) The per capita consumption of bottled water in China is much lower than the per capita consumption of bottled water in the United states.

6.12 (a) $P(X > 50) = P(Z > 1.5625) = 0.0591$

Probability for X >	
X Value	50
Z Value	1.5625
P(X>50)	0.0591

(b) $P(25 < X < 40) = P(-1.5625 < Z < 0.3125) = 0.5636$

Probability for a Range	
From X Value	25
To X Value	40
Z Value for 25	-1.5625
Z Value for 40	0.3125
P(X<=25)	0.0591
P(X<=40)	0.6227
P(25<=X<=40)	0.5636

(c) $P(X < 10) = P(Z < -3.4375) = 0.0003$

Probability for X <=	
X Value	10
Z Value	-3.4375
P(X<=10)	0.0003

(d) $P(X < A) = 0.99$ $Z = 2.33$ $A = 56.11$

Find X and Z Given a Cum. Pctage.	
Cumulative Percentage	99.00%
Z Value	2.33
X Value	56.11

6.13 (a) Partial PHStat output:

Probability for a Range	
From X Value	21.99
To X Value	22
Z Value for 21.99	–2.4
Z Value for 22	–0.4
P(X<=21.99)	0.0082
P(X<=22)	0.3446
P(21.99<=X<=22)	0.3364

$P(21.99 < X < 22.00) = P(-2.4 < Z < -0.4) = 0.3364$

6.13 cont. (b) Partial PHStat output:

Probability for a Range	
From X Value	21.99
To X Value	22.01
Z Value for 21.99	–2.4
Z Value for 22.01	1.6
P(X<=21.99)	0.0082
P(X<=22.01)	0.9452
P(21.99<=X<=22.01)	0.9370

$P(21.99 < X < 22.01) = P(-2.4 < Z < 1.6) = 0.9370$

(c) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	98.00%
Z Value	2.05375
X Value	22.0123

$P(X > A) = 0.02 \qquad Z = 2.05 \qquad A = 22.0123$

(d) (a) Partial PHStat output:

Probability for a Range	
From X Value	21.99
To X Value	22
Z Value for 21.99	–3
Z Value for 22	–0.5
P(X<=21.99)	0.0013
P(X<=22)	0.3085
P(21.99<=X<=22)	0.3072

$P(21.99 < X < 22.00) = P(-3.0 < Z < -0.5) = 0.3072$

(b) Partial PHStat output:

Probability for a Range	
From X Value	21.99
To X Value	22.01
Z Value for 21.99	–3
Z Value for 22.01	2
P(X<=21.99)	0.0013
P(X<=22.01)	0.9772
P(21.99<=X<=22.01)	0.9759

$P(21.99 < X < 22.01) = P(-3.0 < Z < 2) = 0.9759$

(c) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	98.00%
Z Value	2.05375
X Value	22.0102

$P(X > A) = 0.02 \qquad Z = 2.05 \qquad A = 22.0102$

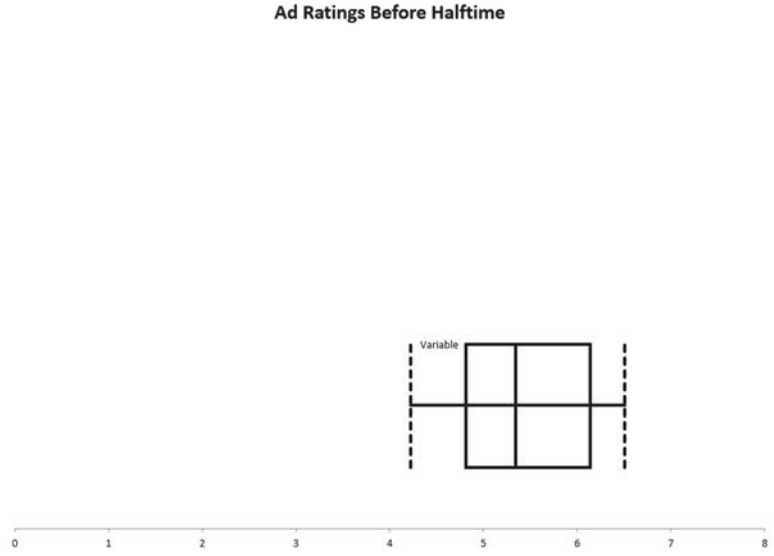
6.14 With 39 values, the smallest of the standard normal quantile values covers an area under the normal curve of 0.025. The corresponding Z value is –1.96. The middle (20th) value has a cumulative area of 0.50 and a corresponding Z value of 0.0. The largest of the standard normal quantile values covers an area under the normal curve of 0.975, and its corresponding Z value is +1.96.

6.15 Area under normal curve covered: 0.1429 0.2857 0.4286 0.5714 0.7143 0.8571
Standardized normal quantile value: –1.07 –0.57 –0.18 + 0.18 + 0.57 + 1.07

6.16 (a) Excel output:
Before Halftime Ad Ratings

	A	B	C	D	E
1	**Before Halftime Ad Ratings**				
2					
3		**Variable**			
4	**Mean**	5.382903226			
5	**Median**	5.34			
6	**Mode**	5.34			
7	**Minimum**	4.22			
8	**Maximum**	6.51			
9	**Range**	2.29		**6*stand dev**	4.291523
10	**Variance**	0.5116			
11	**Standard Deviation**	0.7153			
12	**Coeff. of Variation**	13.29%			
13	**Skewness**	0.0358			
14	**Kurtosis**	-1.2567			
15	**Count**	31			
16	**Standard Error**	0.1285			
17					
18	Five-Number Summary				
19	Minimum	4.22			
20	First Quartile	4.81			
21	Median	5.34			
22	Third Quartile	6.14			
23	Maximum	6.51			
24	IQR	1.33		**1.33*stand dev**	0.951288
25					

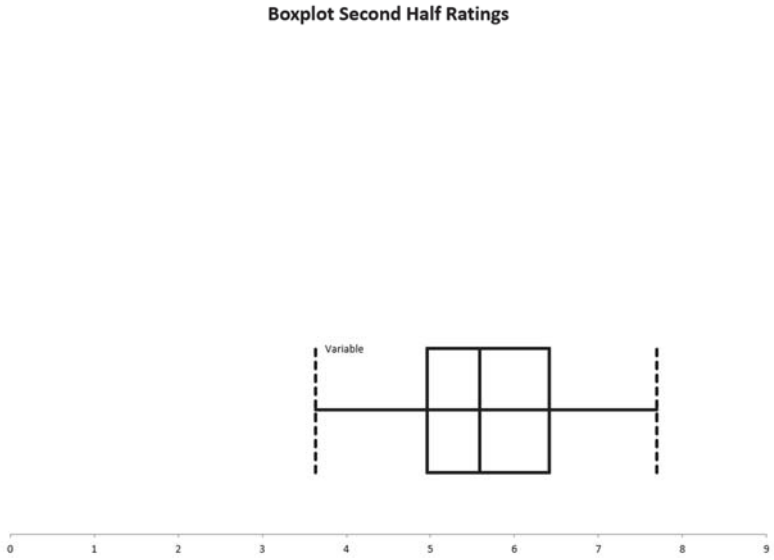
6.16 (a)
cont.



Halftime and after Ad Ratings

	A	B	C	D	E	F
1	**Descriptive Summary**					
2						
3		**Rating**				
4	**Mean**	5.65				
5	**Median**	5.59				
6	**Mode**	4.96				
7	**Minimum**	3.63				
8	**Maximum**	7.69				
9	**Range**	4.06	**6*stand dev**	6.246562952		
10	**Variance**	1.0839	**mean + 1 SD**	6.6918	63%	within 1 SD
11	**Standard Deviation**	1.0411	**mean - 1SD**	4.6096		
12	**Coeff. of Variation**	18.42%	**mean +1.28SD**	6.98	77.78%	within 1.28SD
13	**Skewness**	0.1406	**mean - 1.28SD**	4.3181		
14	**Kurtosis**	-0.5910	**mean +2SD**	7.7329	all values within 2SD	
15	**Count**	27	**mean - 2SD**	3.56855309		
16	**Standard Error**	0.2004				
17	**First Quartile**	4.96				
18	**Third Quartile**	6.41				
19	**IQR**	1.45	**1.33*stand dev**	1.384654788		
20						

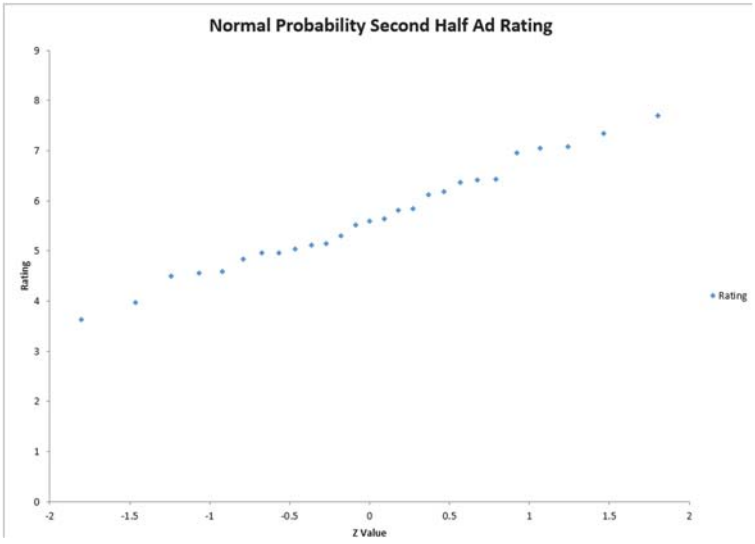
6.16 cont. (a)



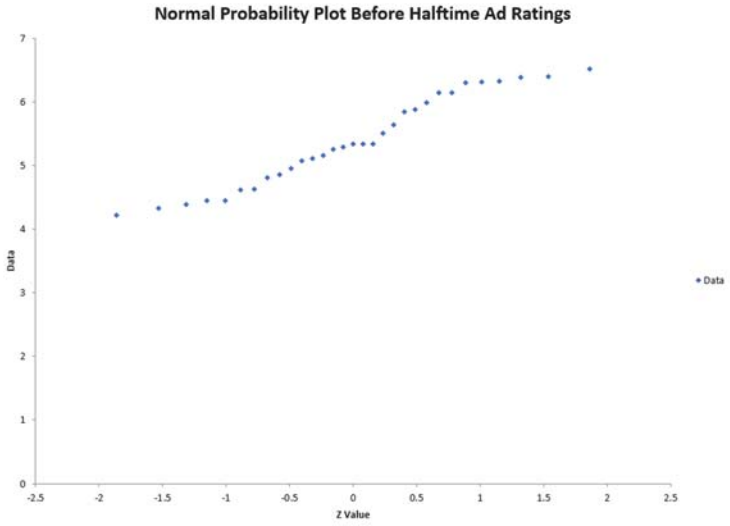
Super Bowl Ad Ratings First and Second Period: (a) Mean = 5.38, median = 5.34, S = 0.7153, range = 2.29, 6S = 4.2918, interquartile range = 1.33, 1.33(0.7153) = 0.9513. The mean is approximately the same as the median. The range is much less than 6S, and the interquartile range is more than 1.33S. The skewness statistic is 0.0358, indicating a symmetric distribution, and the kurtosis statistic is –1.2567, indicating a platykurtic distribution.

Super Bowl Ad Ratings Halftime and Afterward: (a) Mean = 5.65, median = 5.59, S = 1.0411, range = 4.06, 6S = 6.2466, interquartile range = 1.45, 1.33(1.0411) = 1.3847. The mean is approximately the same as the median. The range is much less than 6S, and the interquartile range is slightly more than 1.33S. The skewness statistic is 0.1406, indicating an approximately symmetric distribution, and the kurtosis statistic is –0.5910, indicating a platykurtic distribution.

(b)



6.16 (b)
cont.

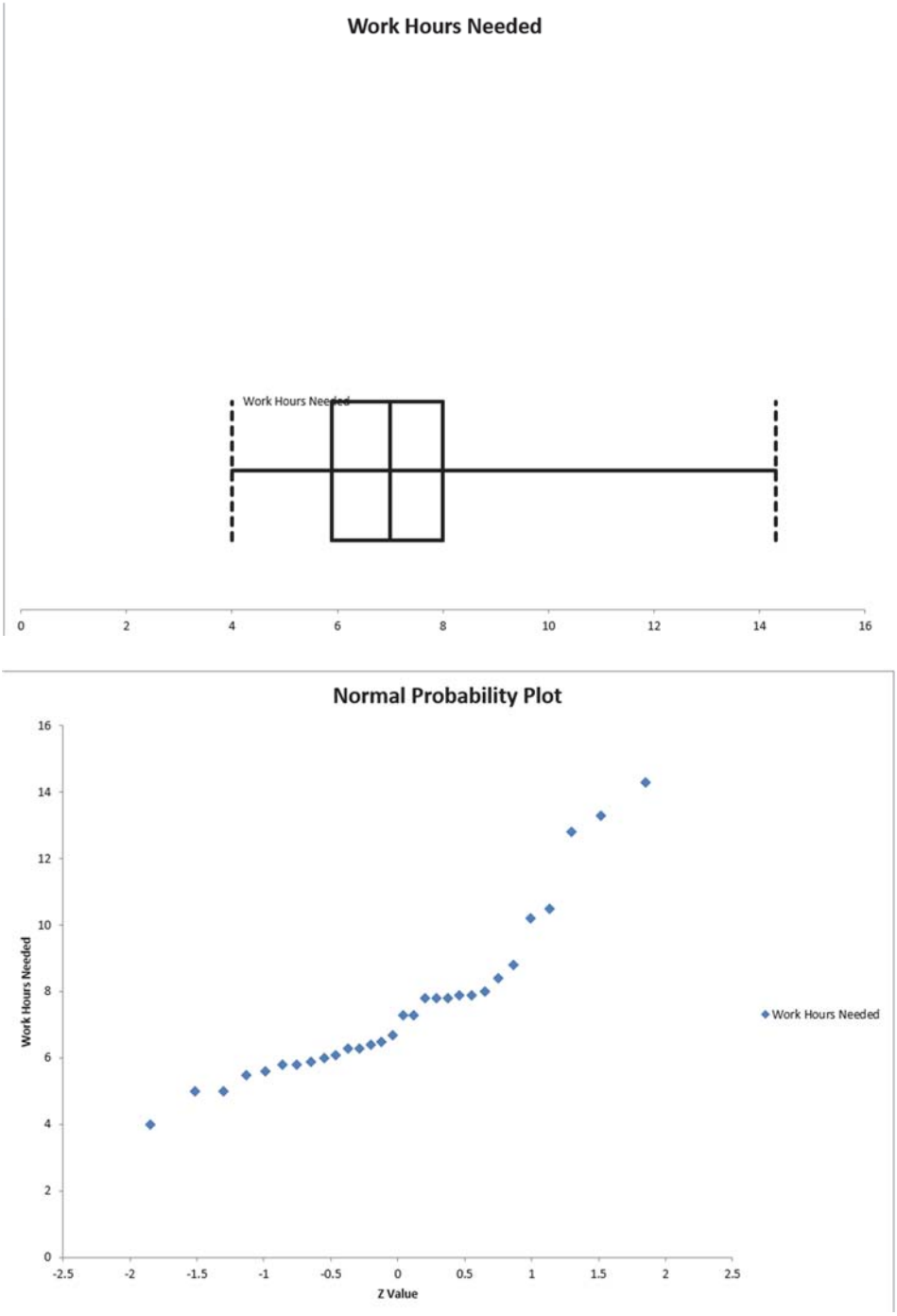


The data for the second half appear to follow a normal distribution.

6.17 Excel Output:

	A	B	C	D	E
1	**Work Hours Needed**				
2					
3		**Work Hours Needed**			
4	**Mean**	7.566666667			
5	**Median**	7			
6	**Mode**	7.8			
7	**Minimum**	4			
8	**Maximum**	14.3			
9	**Range**	10.3		**6*stand dev**	14.8727
10	**Variance**	6.1444			
11	**Standard Deviation**	2.4788			
12	**Coeff. of Variation**	32.76%			
13	**Skewness**	1.3521			
14	**Kurtosis**	1.5387			
15	**Count**	30			
16	**Standard Error**	0.4526			
17					
18	Five-Number Summary				
19	Minimum	4			
20	First Quartile	5.9			
21	Median	7			
22	Third Quartile	8			
23	Maximum	14.3			
24	IQR	2.1		**1.33*stand dev**	3.296782
25					

6.17
cont.

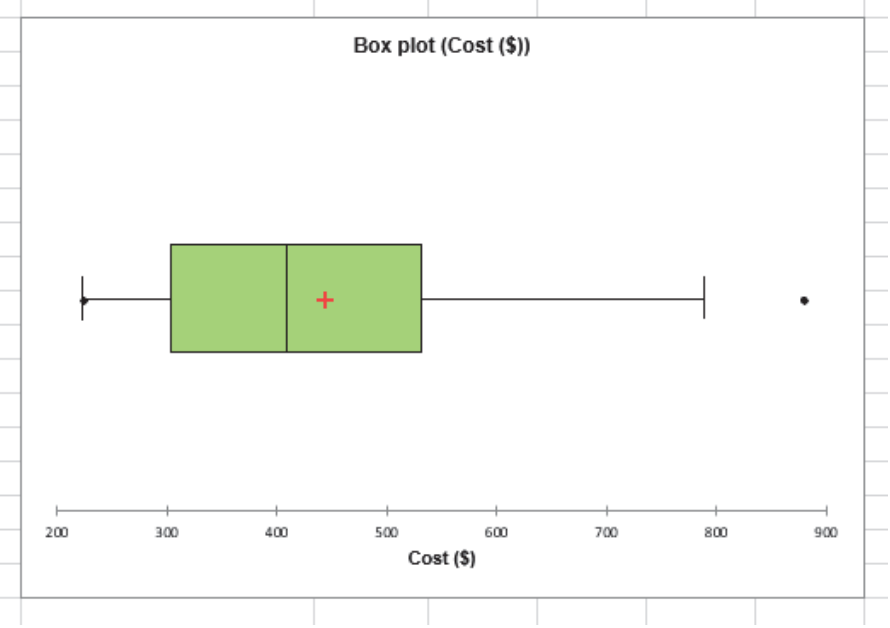


(a)(b) The mean is above the median. The range is smaller than 6 times the standard deviation and the interquartile range is smaller than 1.33 times the standard deviation. The boxplot is right-skewed. The normal probability plot indicates departure from the normal distribution. The skewness of 1.4525 indicates a highly right-skewed distribution.

6.18 (a)(b) Excel Output:

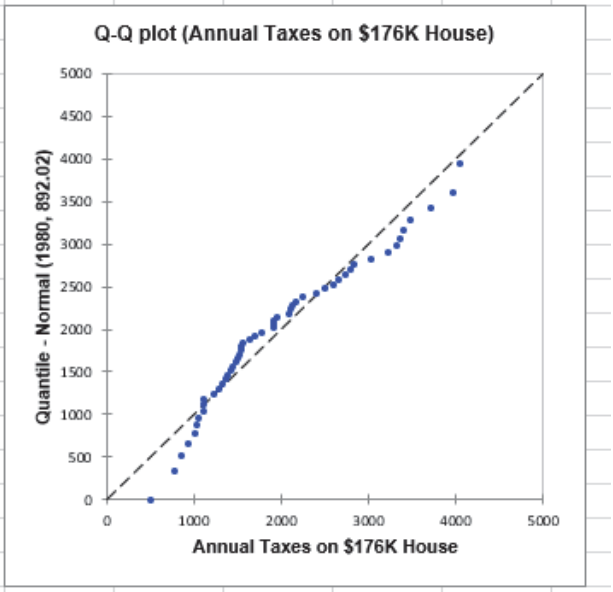
Descriptive statistics (Quantitative data):

Statistic	Taxes on $176K House		
Nbr. of observations	51		
Minimum	489.000		
Maximum	4029.000		
Range	3540.000	6S	5405.362
1st Quartile	1334.000		
Median	1763.000		
3rd Quartile	2614.500		
Mean	1979.941		
IQR	1280.500	1.33S	1198.189
Variance (n-1)	811609.376		
Standard deviation (n-1)	900.894		
Variation coefficient	0.451		
Skewness (Fisher)	0.641		
Skewness (Bowley)	0.330		
Kurtosis (Fisher)	-0.506		



6.18 (a)(b) cont.

Normal Q-Q plots:

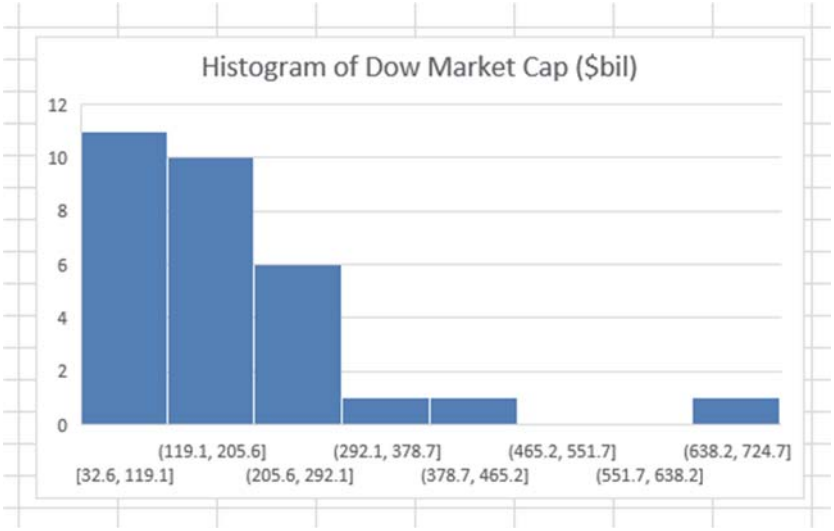


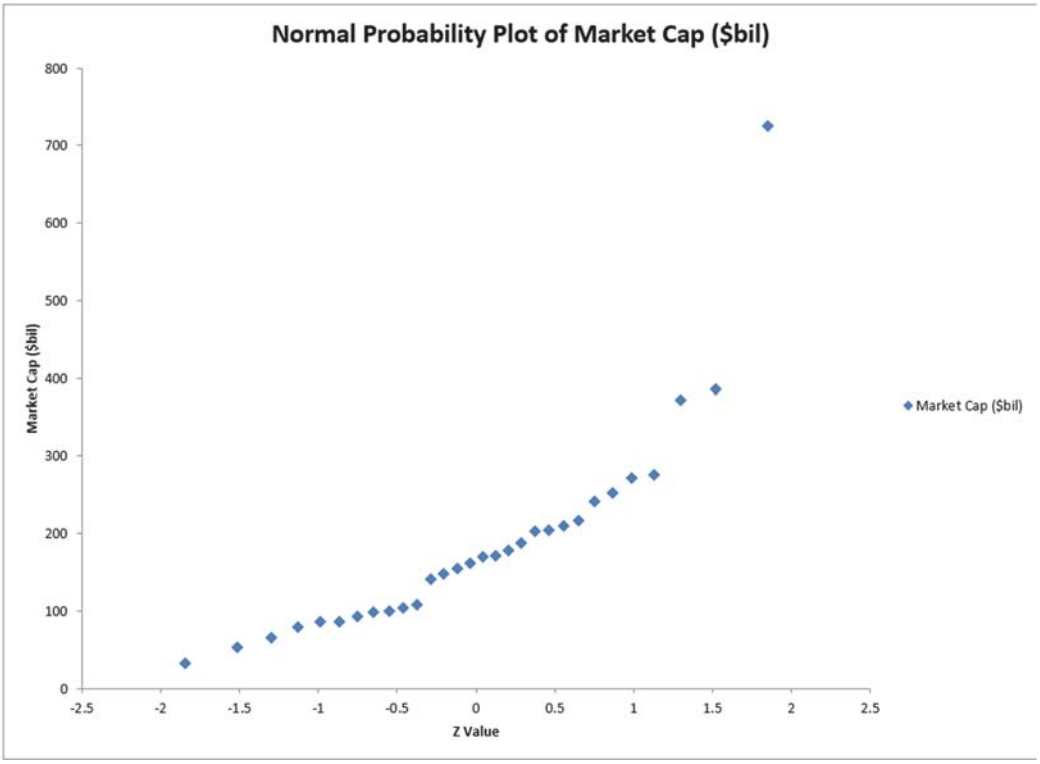
The mean is greater than the median. The range is much less than 6S, and the IQR is more than 1.33S. The box plot is right skewed. The normal probability plot along with the skewness and kurtosis statistics indicate a departure from the normal distribution.

6.19 (a) Excel output:

	A	B	C	D	E
1	**Descriptive Summary**				
2					
3		**Market Cap ($bil)**			
4	**Mean**	185.7566667			
5	**Median**	166.2			
6	**Mode**	#N/A			
7	**Minimum**	32.6			
8	**Maximum**	724.7			
9	**Range**	692.1		6*SD	802.2522
10	**Variance**	17878.0156			
11	**Standard Deviation**	133.7087			
12	**Coeff. of Variation**	71.98%			
13	**Skewness**	2.4498			
14	**Kurtosis**	8.4684			
15	**Count**	30			
16	**Standard Error**	24.4118			
17					
18	Five-Number Summary				
19	Minimum	32.6			
20	First Quartile	98.8			
21	Median	166.2			
22	Third Quartile	217.2			
23	Maximum	724.7			
24	IQR	118.4		1.33*SD	177.8326
25					

6.19 cont. (a)





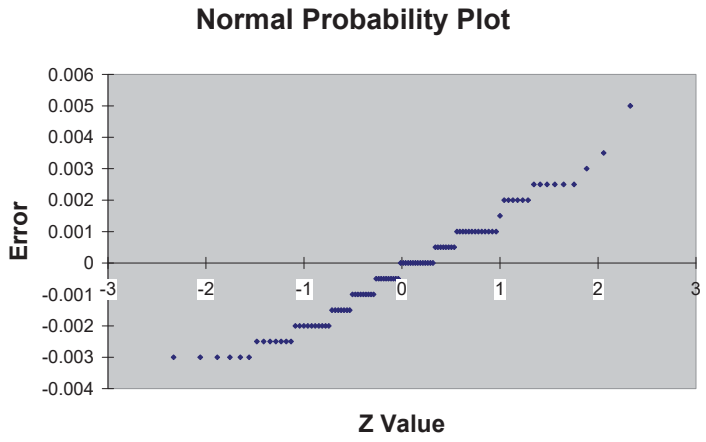
(c) From parts (a) – (c) Conclude that the range is much less than 6S and the IQR is less than 1.33S, the mean is larger than the median, the normal probability plot appears right skewed, the histogram appears right-skewed and both the skewness and kurtosis statistics indicate a departure from a normal distribution.

6.20 Excel output:

Error	
Mean	–0.00023
Median	0
Mode	0
Standard Deviation	0.001696
Sample Variance	2.88E-06
Range	0.008
Minimum	–0.003
Maximum	0.005
First Quartile	–0.0015
Third Quartile	0.001
1.33 Std Dev	0.002255
Interquartile Range	0.0025
6 Std Dev	0.010175

(a) Because the interquartile range is close to 1.33S and the range is also close to 6S, the data appear to be approximately normally distributed.

(b)



The normal probability plot suggests that the data appear to be approximately normally distributed.

6.21 Excel Output:

	A	B	C	D	E
1	**Descriptive Summary**				
2					
3		**One-Year CD**			
4	**Mean**	1.66			
5	**Median**	1.96			
6	**Mode**	0.1			
7	**Minimum**	0.01			
8	**Maximum**	2.86			
9	**Range**	2.85	6*SD	6.059406	
10	**Variance**	1.0199			
11	**Standard Deviation**	1.0099			
12	**Coeff. of Variation**	60.85%			
13	**Skewness**	-0.4851			
14	**Kurtosis**	-1.2933			
15	**Count**	46			
16	**Standard Error**	0.1489			
17					
18					
19	Five-Number Summary				
20	Minimum	0.01			
21	First Quartile	0.5			
22	Median	1.96			
23	Third Quartile	2.5			
24	Maximum	2.86			
25	IQR	2	1.33*SD	1.343168	
26					

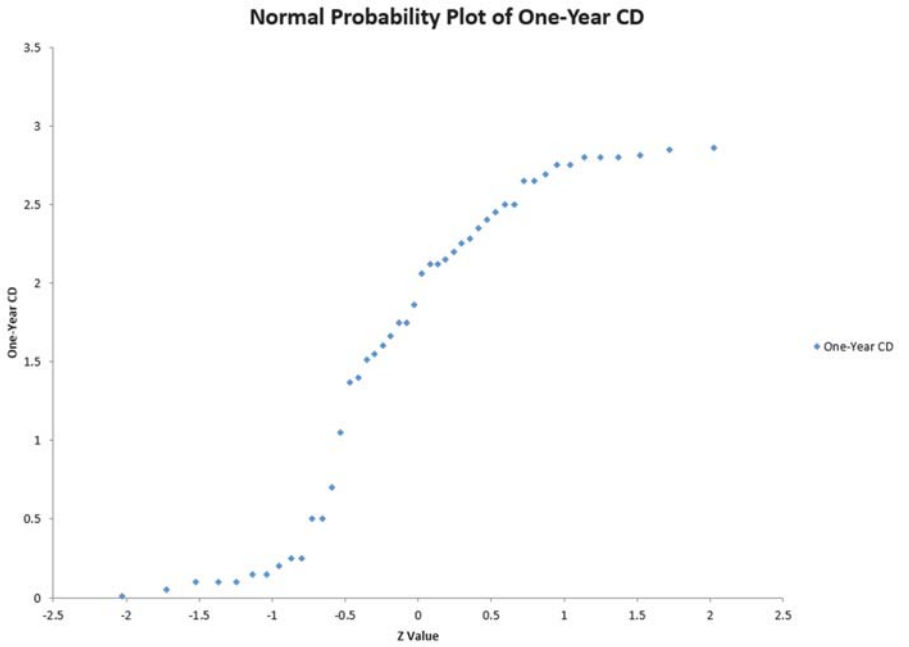
6.21
cont.

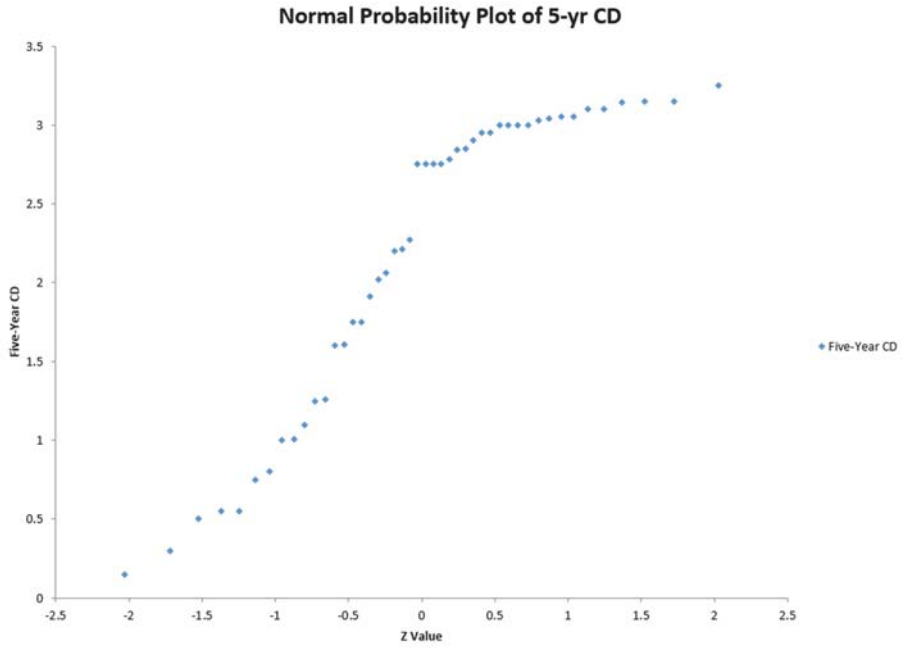
	A	B	C	D
1	**Descriptive Summary**			
2				
3		**Five-Year CD**		
4	**Mean**	2.17		
5	**Median**	2.75		
6	**Mode**	3		
7	**Minimum**	0.15		
8	**Maximum**	3.25		
9	**Range**	3.1	6*SD	5.802443
10	**Variance**	0.9352		
11	**Standard Deviation**	0.9671		
12	**Coeff. of Variation**	44.52%		
13	**Skewness**	-0.6822		
14	**Kurtosis**	-0.9819		
15	**Count**	46		
16	**Standard Error**	0.1426		
17				
18	Five-Number Summary			
19	Minimum	0.15		
20	First Quartile	1.26		
21	Median	2.75		
22	Third Quartile	3		
23	Maximum	3.25		
24	IQR	1.74	1.33*SD	1.286208

(a) For the One-year CD the mean is smaller than the median; the range is smaller than 6 times the standard deviation and the interquartile range is larger than 1.33 times the standard deviation. The data do not appear to be normally distributed.

For the Five-Year CD the mean is smaller than the median; the range is smaller than 6 times the standard deviation and the interquartile range is larger than 1.33 times the standard deviation. The data appear to deviate from the normal distribution.

6.21 (b)
cont.

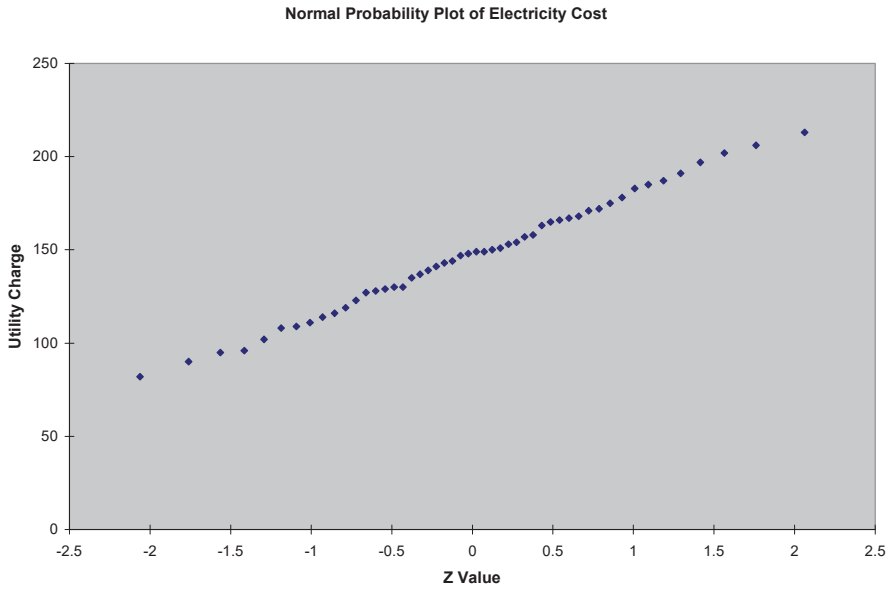




6.22 (a) Five-number summary: 82 127 148.5 168 213 mean = 147.06
range = 131 interquartile range = 41 standard deviation = 31.69

The mean is very close to the median. The five-number summary suggests that the distribution is quite symmetrical around the median. The interquartile range is very close to 1.33 times the standard deviation. The range is about $50 below 6 times the standard deviation. In general, the distribution of the data appears to closely resemble a normal distribution.
Note: The quartiles are obtained using PHStat without any interpolation.

6.22 (b) cont.



The normal probability plot confirms that the data appear to be approximately normally distributed.

6.23 (a) $P(5 < X < 7) = (7 - 5)/10 = 0.2$

(b) $P(2 < X < 3) = (3 - 2)/10 = 0.1$

(c) $\mu = \dfrac{0+10}{2} = 5$

(d) $\sigma = \sqrt{\dfrac{(10-0)^2}{12}} = 2.8868$

6.24 (a) $P(0 < X < 20) = (20 - 0)/120 = 0.1667$

(b) $P(10 < X < 30) = (30 - 10)/120 = 0.1667$

(c) $P(35 < X < 120) = (120 - 35)/120 = 0.7083$

(d) $\mu = \dfrac{0+120}{2} = 60 \qquad \sigma = \sqrt{\dfrac{(120-0)^2}{12}} = 34.6410$

6.25 (a) $P(25 < X < 30) = (30 - 25)/(40 - 20) = 0.25$

(b) $P(X < 35) = (35 - 20)/(40 - 20) = 0.75$

(c) $\mu = \dfrac{20+40}{2} = 30 \qquad \sigma = \sqrt{\dfrac{(40-20)^2}{12}} = 5.7735$

6.26 (a) $P(X < 4.5) = (4.5 - 3.5)/(5.5 - 3.5) = 0.5$

(b) $P(X > 4) = (5.5 - 4)/(5.5 - 3.5) = 0.75$

(c) $P(4.0 < X < 4.5) = (4.5 - 4.0)/(5.5 - 3.5) = 0.25$

(d) $\mu = \dfrac{3.5+5.5}{2} = 4.5 \qquad \sigma = \sqrt{\dfrac{(5.5-3.5)^2}{12}} = 0.5774$

6.27 (a) $P(X < 70) = (70 - 64)/(74 - 64) = 0.6$

(b) $P(65 < X < 70) = (70 - 65)/(74 - 64) = 0.5$

(c) $P(X > 65) = (74 - 65)/(74 - 64) = 0.9$

(d) $\mu = \dfrac{74 + 64}{2} = 69 \qquad \sigma = \sqrt{\dfrac{(74 - 64)^2}{12}} = 2.8868$

6.28 Using Table E.2, first find the cumulative area up to the larger value, and then subtract the cumulative area up to the smaller value.

6.29 Find the Z value corresponding to the given percentile and then use the equation $X = \mu + z\sigma$.

6.30 The normal distribution is bell-shaped; its measures of central tendency are all equal; its middle 50% is within 1.33 standard deviations of its mean; and 99.7% of its values are contained within three standard deviations of its mean.

6.31 Both the normal distribution and the uniform distribution are symmetric but the uniform distribution has a bounded range while the normal distribution ranges from negative infinity to positive infinity. The exponential distribution is right-skewed and ranges from zero to infinity.

6.32 If the distribution is normal, the plot of the Z values on the horizontal axis and the original values on the vertical axis will be a straight line.

6.33 (a) Partial PHStat output:

Probability for a Range	
From X Value	**0.75**
To X Value	**0.753**
Z Value for 0.75	**–0.75**
Z Value for 0.753	**0**
P(X<=0.75)	**0.2266**
P(X<=0.753)	**0.5000**
P(0.75<=X<=0.753)	**0.2734**

$P(0.75 < X < 0.753) = P(-0.75 < Z < 0) = 0.2734$

(b) Partial PHStat output:

Probability for a Range	
From X Value	**0.74**
To X Value	**0.75**
Z Value for 0.74	**–3.25**
Z Value for 0.75	**–0.75**
P(X<=0.74)	**0.0006**
P(X<=0.75)	**0.2266**
P(0.74<=X<=0.75)	**0.2261**

$P(0.74 < X < 0.75) = P(-3.25 < Z < -0.75) = 0.2266 - 0.00058 = 0.2260$

6.33 cont. (c) Partial PHStat output:

Probability for X >	
X Value	0.76
Z Value	1.75
P(X>0.76)	0.0401

$P(X > 0.76) = P(Z > 1.75) = 1.0 - 0.9599 = 0.0401$

(d) Partial PHStat output:

Probability for X <=	
X Value	0.74
Z Value	–3.25
P(X<=0.74)	0.000577

$P(X < 0.74) = P(Z < -3.25) = 0.00058$

(e) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	7.00%
Z Value	–1.475791
X Value	0.747097

$P(X < A) = P(Z < -1.48) = 0.07 \qquad A = 0.753 - 1.48(0.004) = 0.7471$

6.34 (a) Partial PHStat output:

Probability for a Range	
From X Value	1.9
To X Value	2
Z Value for 1.9	–2
Z Value for 2	0
P(X<=1.9)	0.0228
P(X<=2)	0.5000
P(1.9<=X<=2)	0.4772

$P(1.90 < X < 2.00) = P(-2.00 < Z < 0) = 0.4772$

(b) Partial PHStat output:

Probability for a Range	
From X Value	1.9
To X Value	2.1
Z Value for 1.9	–2
Z Value for 2.1	2
P(X<=1.9)	0.0228
P(X<=2.1)	0.9772
P(1.9<=X<=2.1)	0.9545

$P(1.90 < X < 2.10) = P(-2.00 < Z < 2.00) = 0.9772 - 0.0228 = 0.9544$

(c) Partial PHStat output:

Probability for X<1.9 or X >2.1	
P(X<1.9 or X >2.1)	0.0455

$P(X < 1.90) + P(X > 2.10) = 1 - P(1.90 < X < 2.10) = 0.0456$

6.34 cont. (d) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	1.00%
Z Value	–2.326348
X Value	1.883683

$P(X > A) = P(Z > -2.33) = 0.99$ $A = 2.00 - 2.33(0.05) = 1.8835$

(e) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	99.50%
Z Value	2.575829
X Value	2.128791

$P(A < X < B) = P(-2.58 < Z < 2.58) = 0.99$

$A = 2.00 - 2.58(0.05) = 1.8710$ $B = 2.00 + 2.58(0.05) = 2.1290$

6.35 (a) Partial PHStat output:

Probability for a Range	
From X Value	1.9
To X Value	2
Z Value for 1.9	–2.4
Z Value for 2	–0.4
P(X<=1.9)	0.0082
P(X<=2)	0.3446
P(1.9<=X<=2)	0.3364

$P(1.90 < X < 2.00) = P(-2.40 < Z < -0.40) = 0.3446 - 0.0082 = 0.3364$

(b) Partial PHStat output:

Probability for a Range	
From X Value	1.9
To X Value	2.1
Z Value for 1.9	–2.4
Z Value for 2.1	1.6
P(X<=1.9)	0.0082
P(X<=2.1)	0.9452
P(1.9<=X<=2.1)	0.9370

$P(1.90 < X < 2.10) = P(-2.40 < Z < 1.60) = 0.9452 - 0.0082 = 0.9370$

(c) Partial PHStat output:

Probability for a Range	
From X Value	1.9
To X Value	2.1
Z Value for 1.9	–2.4
Z Value for 2.1	1.6
P(X<=1.9)	0.0082
P(X<=2.1)	0.9452
P(1.9<=X<=2.1)	0.9370

$P(X < 1.90) + P(X > 2.10) = 1 - P(1.90 < X < 2.10) = 0.0630$

6.35 cont. (d) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	1.00%
Z Value	–2.326348
X Value	1.903683

$P(X > A) = P(Z > -2.33) = 0.99$ $A = 2.02 - 2.33(0.05) = 1.9035$

(e) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	99.50%
Z Value	2.575829
X Value	2.148791

$P(A < X < B) = P(-2.58 < Z < 2.58) = 0.99$
$A = 2.02 - 2.58(0.05) = 1.8910$ $B = 2.02 + 2.58(0.05) = 2.1490$

6.36 (a) Partial PHStat output:

Probability for X <=	
X Value	210
Z Value	-2
P(X<=210)	0.0228

$P(X < 210) = P(Z < -2) = 0.0228$

(b)

Probability for a Range	
From X Value	270
To X Value	300
Z Value for 270	1
Z Value for 300	2.5
P(X<=270)	0.8413
P(X<=300)	0.9938
P(270<=X<=300)	0.1524

$P(270 < X < 300) = P(1.0 < Z < 2.5) = 0.1524$

(c)

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	90.00%
Z Value	1.2816
X Value	275.6310

$P(X < A) = P(Z < 1.2816) = 0.90 A = 250 + 20(1.2816) = \$275.63$

(d)

Find X Values Given a Percentage	
Percentage	80.00%
Z Value	-1.28
Lower X Value	224.37
Upper X Value	275.63

6.36 cont. (d) $P(A < X < B) = P(-1.2816 < Z < 1.2816) = 0.80$

$A = 250 - 1.28(500) = \$224.37$

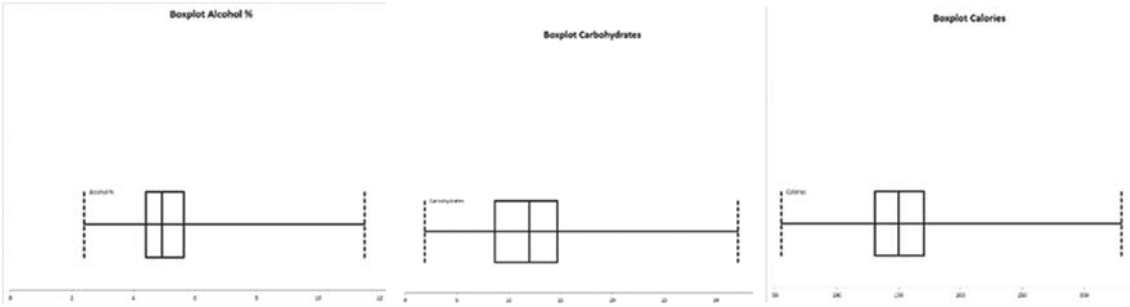
$B = 250 + 1.28(500) = \$275.63$

6.37 Excel Output:

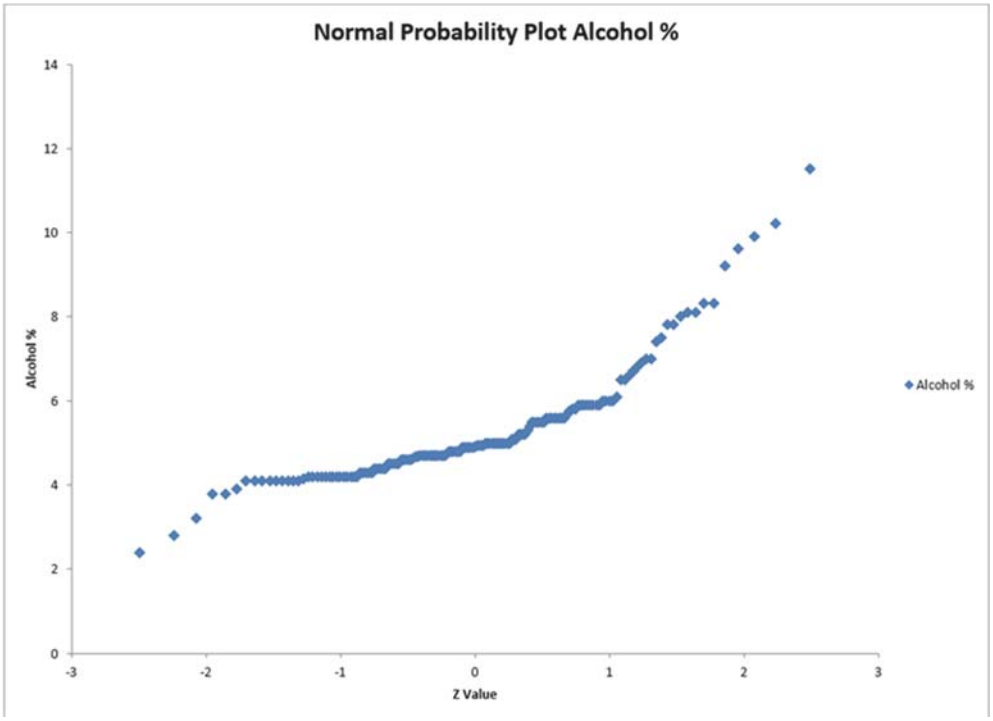
	A	B	C	D
1	**Descriptive Summary**			
2				
3		**Alcohol %**	**Carbohydrates**	**Calories**
4	**Mean**	5.270127389	12.05235669	155.5095541
5	**Median**	4.92	12	150
6	**Mode**	4.2	12	110
7	**Minimum**	2.4	1.9	55
8	**Maximum**	11.5	32.1	330
9	**Range**	9.1	30.2	275
10	**Variance**	1.8337	24.8119	1918.6746
11	**Standard Deviation**	1.3541	4.9812	43.8027
12	**Coeff. of Variation**	25.69%	41.33%	28.17%
13	**Skewness**	1.8403	0.4908	1.2148
14	**Kurtosis**	4.5833	1.0801	2.9712
15	**Count**	157	157	157
16	**Standard Error**	0.1081	0.3975	3.4958
17				
18	Minimum	2.4	1.9	55
19	First Quartile	4.4	8.65	130.5
20	Median	4.92	12	150
21	Third Quartile	5.65	14.7	170.5
22	Maximum	11.5	32.1	330
23	IQR	1.25	6.05	40
24	6*standard deviation	8.124822329	29.88690617	262.8160672
25	1.33*standard deviation	1.801002283	6.624930867	58.25756155

Alcohol %:

The mean is greater than the median; the range is larger than 6 times the standard deviation and the interquartile range is smaller than 1.33 times the standard deviation. The data appear to deviate from the normal distribution.



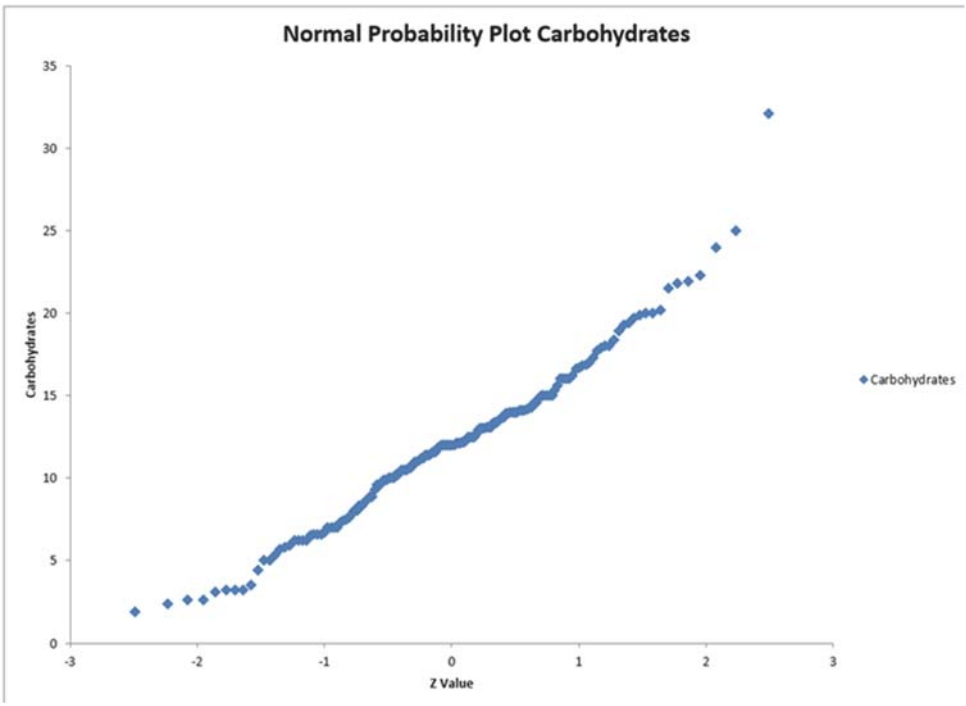
6.37
cont.



The normal probability plot suggests that data are not normally distributed. The kurtosis is 4.583 indicating a distribution that is more peaked than a normal distribution, with more values in the tails. The skewness of 1.840 suggests that the distribution is right-skewed.

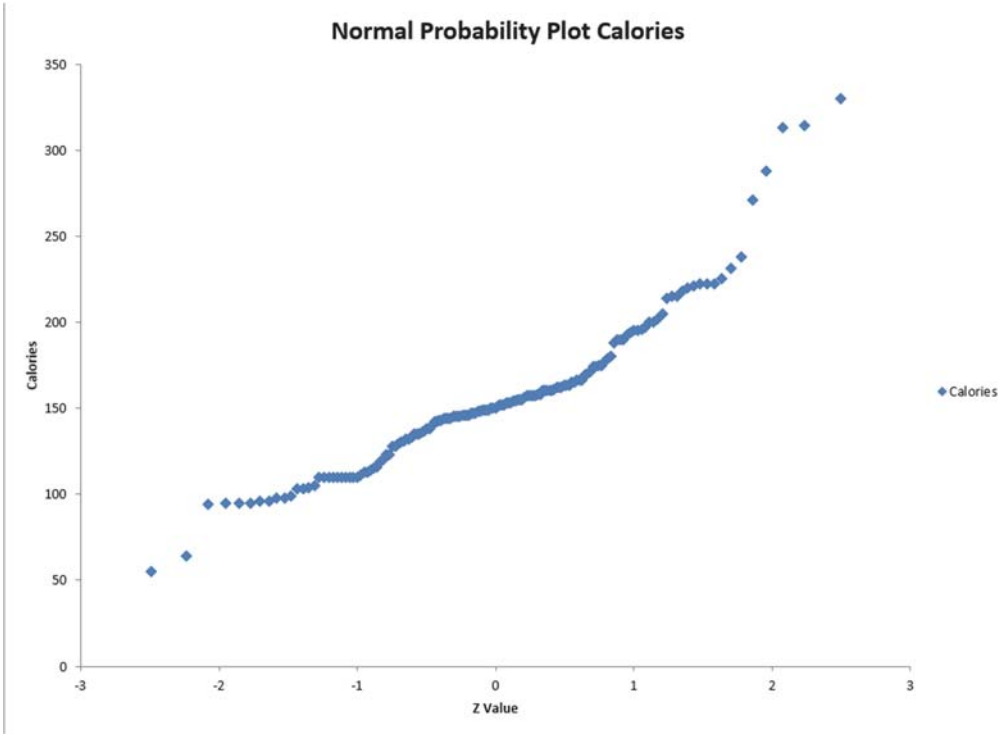
6.37 cont. **Carbohydrates:**

The mean is approximately equal to the median; the range is approximately equal to 6 times the standard deviation and the interquartile range is slightly smaller than 1.33 times the standard deviation. The data appear to be normally distributed.



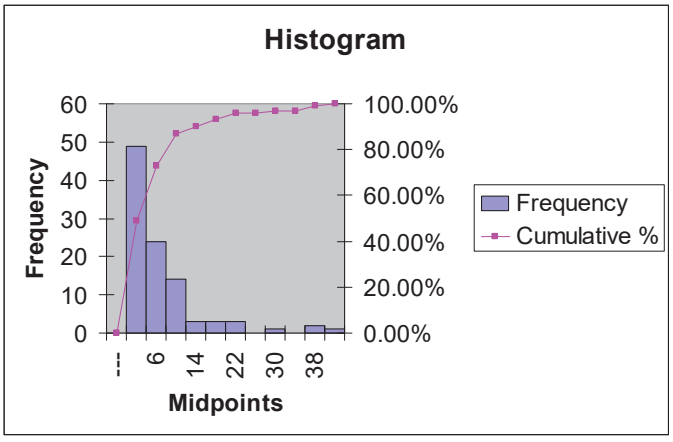
The normal probability plot suggests that the data are approximately normally distributed. The kurtosis is 1.0801 indicating a distribution that is slightly more peaked than a normal distribution, with more values in the tails. The skewness of 0.4908 indicates that the distribution deviates slightly from the normal distribution.

Calories:
The mean is greater than the median; the range is greater than 6 times the standard deviation and the interquartile range is smaller than 1.33 times the standard deviation. The data appear to deviate away from the normal distribution.

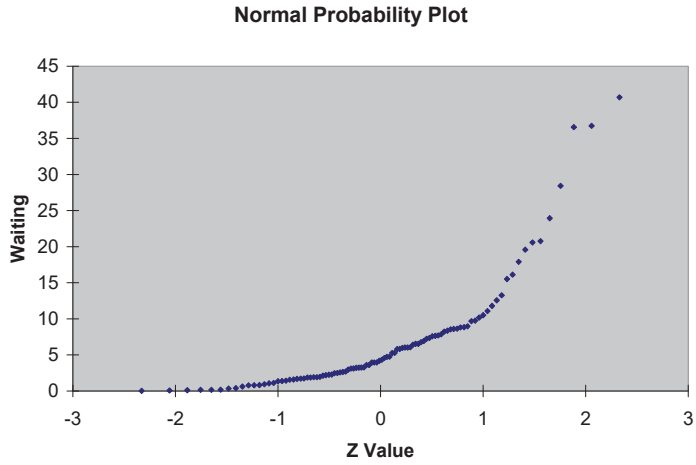


The normal probability plot suggests that the data are somewhat right-skewed. The kurtosis is 2.9712 indicating a distribution that is more peaked than a normal distribution, with more values in the tails. The skewness of 1.2148 suggests that the distribution is right-skewed.

6.38 (a) Waiting time will more closely resemble an exponential distribution.
(b) Seating time will more closely resemble a normal distribution.
(c)

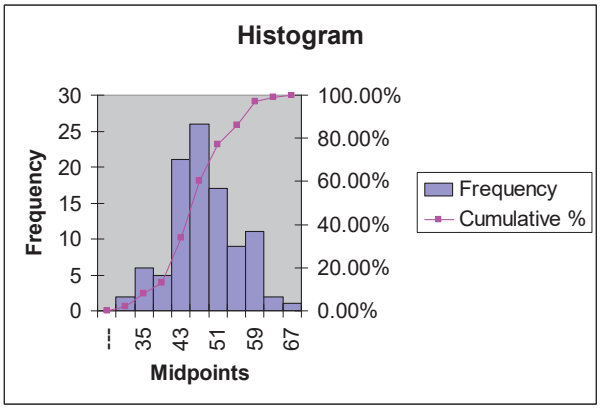


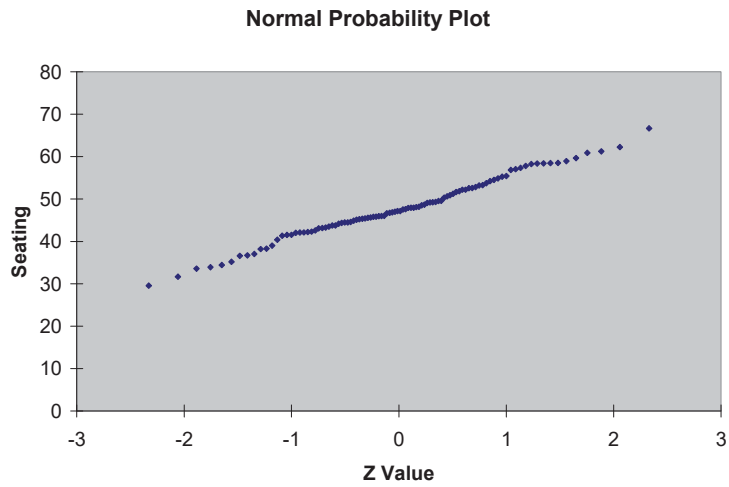
6.38 (c)
cont.



Both the histogram and normal probability plot suggest that waiting time more closely resembles an exponential distribution.

(d)





Both the histogram and normal probability plot suggest that seating time more closely resembles a normal distribution.

6.39 (a) Excel Output:

12	Probability for X >	
13	X Value	0
14	Z Value	0.312
15	P(X>0)	0.3775

$P(X > 0) = P(Z > 0.312) = 0.3775$

(b) Excel Output:

12	Probability for X >	
13	X Value	10
14	Z Value	0.812
15	P(X>10)	0.2084

$P(X > 10) = P(Z > 0.812) = 0.2084$

(c) Excel Output:

7	Probability for X <=	
8	X Value	-20
9	Z Value	-0.688
10	P(X<=-20)	0.2457

$P(X < -20) = P(Z < -0.688) = 0.2457$

(d) Excel Output

7	Probability for X <=	
8	X Value	-30
9	Z Value	-1.188
10	P(X<=-30)	0.1174

$P(X < -30) = P(Z < -1.188) = 0.1174$

(e) (a) Excel Output:

2	Probability for X >	
3	X Value	0
4	Z Value	0.1293333
5	P(X>0)	0.4485

$P(X > 0) = P(Z > 0.129) = 0.4485$

(b) Excel Output:

12	Probability for X >	
13	X Value	10
14	Z Value	0.4626667
15	P(X>10)	0.3218

$P(X > 10) = P(Z > 0.463) = 0.3218$

6.39 cont. (e) (c) Excel Output:

	Probability for X <=	
7		
8	X Value	-20
9	Z Value	-0.537333
10	P(X<=-20)	0.2955

$P(X < -20) = P(Z < -0.537) = 0.2955$

(d) Excel Output

	Probability for X <=	
7		
8	X Value	-30
9	Z Value	-0.870667
10	P(X<=-30)	0.1920

$P(X < -30) = P(Z < -0.871) = 0.1920$

(f) The probability that a S&P 500 stock gained value in 2018 is 0.3775. The probability that a NASDAQ stock gained value in 2018 is 0.4485. The probability that a S&P 500 stock gained 10% or more value in 2018 is 0.2084. The probability that a NASDAQ stock gained 10% or more value in 2018 is 0.3218. The probability that a S&P 500 stock lost 20% or more value in 2018 is 0.2457. The probability that a NASDAQ stock lost 20% or more value in 2018 is 0.2955. The probability that a S&P 500 stock lost 30% or more value in 2018 is 0.1174. The probability that a NASDAQ stock lost 30% or more value in 2018 is 0.1920. The larger standard deviation of the NASDAQ is associated with higher risk.

6.40 (a) Excel Output:

Probability for X <=	
X Value	5900
Z Value	-0.1
P(X<=5900)	0.4602

$P(X < \$5900) = P(Z < -0.1) = 0.4602$

(b) Excel Output:

Probability for a Range	
From X Value	5700
To X Value	6100
Z Value for 5700	-0.6
Z Value for 6100	0.4
P(X<=5700)	0.2743
P(X<=6100)	0.6554
P(5700<=X<=6100)	0.3812

$P(\$5700 < X < \$6100) = P(-0.6 < Z < 0.4) = 0.3812$

6.40 cont. (c) Excel Output:

Probability for X >	
X Value	6500
Z Value	1.4
P(X>6500)	0.0808

$P(X > \$6500) = P(Z > 1.4) = 0.0808$

(d) Excel Output:

Find X and Z Given a Cum. Pctage.	
Cumulative Percentage	1.00%
Z Value	-2.33
X Value	5009.46

$P(A < X) = 0.01$ $Z = -2.33$ $A = \$5009.46$

(e) Excel Output:

Find X Values Given a Percentage	
Percentage	95.00%
Z Value	-1.96
Lower X Value	5156.01
Upper X Value	6723.99

$P(A < X < B) = 0.95$ $Z = -1.9600$ $A = \$5156.01$
$Z = 1.96$ $B = \$6723.99$

6.41 (a) Excel Output:

Probability for X <=	
X Value	5900
Z Value	-1.378
P(X<=5900)	0.0841

$P(X < \$5900) = P(Z < -1.378) = 0.0841$

(b) Excel Output:

Probability for a Range	
From X Value	5700
To X Value	6100
Z Value for 5700	-1.778
Z Value for 6100	-0.978
P(X<=5700)	0.0377
P(X<=6100)	0.1640
P(5700<=X<=6100)	0.1263

$P(\$5700 < X < \$6100) = P(-1.778 < Z < -0.978) = 0.1263$

6.41 cont. (c) Excel Output:

Probability for X >	
X Value	6500
Z Value	-0.178
P(X>6500)	0.5706

$P(X > \$6500) = P(Z > -0.178) = 0.5706$

(d) Excel Output:

Find X and Z Given a Cum. Pctage.	
Cumulative Percentage	1.00%
Z Value	-2.33
X Value	5425.83

$P(A < X) = 0.01$ $Z = -2.33$ $A = \$5425.83$

(e) Excel Output:

Find X Values Given a Percentage	
Percentage	95.00%
Z Value	-1.96
Lower X Value	5609.02
Upper X Value	7568.98

$P(A < X < B) = 0.95$ $Z = -1.96$ $A = \$5609.02$
$Z = 1.96$ $B = \$7568.98$

(f) The probability an intern at Intel will earn less than \$5900 per month is 0.4602 while the probability that a Facebook intern will earn less than \$5900 is 0.0841. 99% of the interns at Intel will earn at least \$5009.46 while 99% of the Facebook interns will earn at least \$5425.83. 95% of the interns at Intel will earn between \$5156.01 and \$6723.99 while 95% of the interns at Facebook will earn between \$5609.02 and \$7568.98.

6.42 *Class project solutions may vary.*

Chapter 7

7.1 PHstat output:

Common Data	
Mean	100
Standard Deviation	2

Probability for X <=	
X Value	95
Z Value	–2.5
P(X<=95)	0.0062097

Probability for X >	
X Value	102.2
Z Value	1.1
P(X>102.2)	0.1357

Probability for X<95 or X >102.2	
P(X<95 or X >102.2)	0.1419

Probability for a Range	
From X Value	95
To X Value	97.5
Z Value for 95	–2.5
Z Value for 97.5	–1.25
P(X<=95)	0.0062
P(X<=97.5)	0.1056
P(95<=X<=97.5)	0.0994

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	35.00%
Z Value	–0.38532
X Value	99.22936

(a) $P(\overline{X} < 95) = P(Z < -2.50) = 0.0062$

(b) $P(95 < \overline{X} < 97.5) = P(-2.50 < Z < -1.25) = 0.1056 - 0.0062 = 0.0994$

(c) $P(\overline{X} > 102.2) = P(Z > 1.10) = 1.0 - 0.8643 = 0.1357$

(d) $P(\overline{X} > A) = P(Z > -0.39) = 0.65 \qquad \overline{X} = 100 - 0.39\left(\frac{10}{\sqrt{25}}\right) = 99.22$

7.2 PHStat output:

Common Data	
Mean	50
Standard Deviation	0.5

Probability for X <=	
X Value	47
Z Value	–6
P(X<=47)	9.866E-10

Probability for X >	
X Value	51.5
Z Value	3
P(X>51.5)	0.0013

Probability for X<47 or X >51.5	
P(X<47 or X >51.5)	0.0013

Probability for X >	
X Value	51.1
Z Value	2.2
P(X>51.1)	0.0139

Probability for a Range	
From X Value	47
To X Value	49.5
Z Value for 47	–6
Z Value for 49.5	–1
P(X<=47)	0.0000
P(X<=49.5)	0.1587
P(47<=X<=49.5)	0.1587

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	65.00%
Z Value	0.38532
X Value	50.19266

7.2 cont.

(a) $P(\overline{X} < 47) = P(Z < -6.00) =$ virtually zero

(b) $P(47 < \overline{X} < 49.5) = P(-6.00 < Z < -1.00) = 0.1587 - 0.00 = 0.1587$

(c) $P(\overline{X} > 51.1) = P(Z > 2.20) = 1.0 - 0.9861 = 0.0139$

(d) $P(\overline{X} > A) = P(Z > 0.39) = 0.35 \qquad \overline{X} = 50 + 0.39(0.5) = 50.195$

7.3

(a) For samples of 25 customer receipts for a supermarket for a year, the sampling distribution of sample means is the distribution of means from all possible samples of 25 customer receipts for a supermarket for that year.

(b) For samples of 25 insurance payouts in a particular geographical area in a year, the sampling distribution of sample means is the distribution of means from all possible samples of 25 insurance payouts in that particular geographical area in that year.

(c) For samples of 25 Call Center logs of inbound calls tracking handling time for a credit card company during the year, the sampling distribution of sample means is the distribution of means from all possible samples of 25 Call Center logs of inbound calls tracking handling time for a credit card company during that year.

7.4 (a) Sampling Distribution of the Mean for $n = 2$ (without replacement)

Sample Number	Outcomes	Sample Means $\overline{X}_i$
1	1, 3	$\overline{X}_1 = 2$
2	1, 6	$\overline{X}_2 = 3.5$
3	1, 7	$\overline{X}_3 = 4$
4	1, 9	$\overline{X}_4 = 5$
5	1, 10	$\overline{X}_5 = 5.5$
6	3, 6	$\overline{X}_6 = 4.5$
7	3, 7	$\overline{X}_7 = 5$
8	3, 9	$\overline{X}_8 = 6$
9	3, 10	$\overline{X}_9 = 6.5$
10	6, 7	$\overline{X}_{10} = 6.5$
11	6, 9	$\overline{X}_{11} = 7.5$
12	6, 10	$\overline{X}_{12} = 8$
13	7, 9	$\overline{X}_{13} = 8$
14	7, 10	$\overline{X}_{14} = 8.5$
15	9, 10	$\overline{X}_{15} = 9.5$

Mean of All Possible Sample Means: $\mu_{\overline{X}} = \dfrac{90}{15} = 6$

Mean of All Population Elements: $\mu = \dfrac{1+3+6+7+9+10}{6} = 6$

Both means are equal to 6. This property is called unbiasedness.

7.4 cont. (b) Sampling Distribution of the Mean for $n = 3$ (without replacement)

Sample Number	Outcomes	Sample Means $\overline{X}_i$
1	1, 3, 6	$\overline{X}_1 = 3\ 1/3$
2	1, 3, 7	$\overline{X}_2 = 3\ 2/3$
3	1, 3, 9	$\overline{X}_3 = 4\ 1/3$
4	1, 3, 10	$\overline{X}_4 = 4\ 2/3$
5	1, 6, 7	$\overline{X}_5 = 4\ 2/3$
6	1, 6, 9	$\overline{X}_6 = 5\ 1/3$
7	1, 6, 10	$\overline{X}_7 = 5\ 2/3$
8	3, 6, 7	$\overline{X}_8 = 5\ 1/3$
9	3, 6, 9	$\overline{X}_9 = 6$
10	3, 6, 10	$\overline{X}_{10} = 6\ 1/3$
11	6, 7, 9	$\overline{X}_{11} = 7\ 1/3$
12	6, 7, 10	$\overline{X}_{12} = 7\ 2/3$
13	6, 9, 10	$\overline{X}_{13} = 8\ 1/3$
14	7, 9, 10	$\overline{X}_{14} = 8\ 2/3$
15	1, 7, 9	$\overline{X}_{15} = 5\ 2/3$
16	1, 7, 10	$\overline{X}_{16} = 6$
17	1, 9, 10	$\overline{X}_{17} = 6\ 2/3$
18	3, 7, 9	$\overline{X}_{18} = 6\ 1/3$
19	3, 7, 10	$\overline{X}_{19} = 6\ 2/3$
20	3, 9, 10	$\overline{X}_{20} = 7\ 1/3$

$\mu_{\overline{X}} = \dfrac{120}{20} = 6$ This is equal to μ, the population mean.

(c) The distribution for $n = 3$ has less variability. The larger sample size has resulted in sample means being closer to μ.

(d) (a) Sampling Distribution of the Mean for $n = 2$ (with replacement)

Sample Number	Outcomes	Sample Means $\overline{X}_i$
1	1, 1	$\overline{X}_1 = 1$
2	1, 3	$\overline{X}_2 = 2$
3	1, 6	$\overline{X}_3 = 3.5$
4	1, 7	$\overline{X}_4 = 4.5$
5	1, 9	$\overline{X}_5 = 5$
6	1, 10	$\overline{X}_6 = 5.5$
7	3,1	$\overline{X}_7 = 2$
8	3, 3	$\overline{X}_8 = 3$
9	3, 6	$\overline{X}_9 = 4.5$

(table continues on next page)

7.4 cont. (d) (a)

Sample Number	Outcomes	Sample Means $\overline{X}_i$
10	3, 7	$\overline{X}_{10} = 5$
11	3, 9	$\overline{X}_{11} = 6$
12	3, 10	$\overline{X}_{12} = 6.5$
13	6, 1	$\overline{X}_{13} = 3.5$
14	6, 3	$\overline{X}_{14} = 4.5$
15	6, 6	$\overline{X}_{15} = 6$
16	6, 7	$\overline{X}_{16} = 6.5$
17	6, 9	$\overline{X}_{17} = 7.5$
18	6, 10	$\overline{X}_{18} = 8$
19	7, 1	$\overline{X}_{19} = 4$
20	7,3	$\overline{X}_{20} = 5$
21	7, 6	$\overline{X}_{21} = 6.5$
22	7, 7	$\overline{X}_{22} = 7$
23	7, 9	$\overline{X}_{23} = 8$
24	7, 10	$\overline{X}_{24} = 8.5$
25	9, 1	$\overline{X}_{25} = 5$
26	9, 3	$\overline{X}_{26} = 6$
27	9, 6	$\overline{X}_{27} = 7.5$
28	9, 7	$\overline{X}_{28} = 8$
29	9, 9	$\overline{X}_{29} = 9$
30	9, 10	$\overline{X}_{30} = 9.5$
31	10, 1	$\overline{X}_{31} = 5.5$
32	10, 3	$\overline{X}_{32} = 6.5$
33	10, 6	$\overline{X}_{33} = 8$
34	10, 7	$\overline{X}_{34} = 8.5$
35	10, 9	$\overline{X}_{35} = 9.5$
36	10, 10	$\overline{X}_{36} = 10$

Mean of All Possible Sample Means:

$$\mu_{\overline{X}} = \frac{216}{36} = 6$$

Mean of All Population Elements:

$$\mu = \frac{1+3+6+7+7+12}{6} = 6$$

Both means are equal to 6. This property is called unbiasedness.

(b) Repeat the same process for the sampling distribution of the mean for $n = 3$ (with replacement). There will be $6^3 = 216$ different samples.

$\mu_{\overline{X}} = 6$ This is equal to μ, the population mean.

(c) The distribution for $n = 3$ has less variability. The larger sample size has resulted in more sample means being close to μ.

7.5 (a) $P(X < 2.03) = P(Z < -0.8) = 0.2119$

Excel Output:

	A	B
4	Mean	2.05
5	Standard Deviation	0.025
6		
7	Probability for X <=	
8	X Value	2.03
9	Z Value	-0.8
10	P(X<=2.03)	0.2119

(b) Because the amount of water in a two-liter bottle is approximately normally distributed, the sampling distribution of samples of 4 will also be approximately normal with a mean of $\mu_{\bar{X}} = \mu = 2.05$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 0.0125.$

$P(\bar{X} < 2.03) = P(Z < -1.6) = 0.0548$

Excel Output:

	A	B
4	Mean	2.05
5	Standard Deviation	0.0125
6		
7	Probability for X <=	
8	X Value	2.03
9	Z Value	-1.6
10	P(X<=2.03)	0.0548

(c) Because the amount of water in a two-liter bottle is approximately normally distributed, the sampling distribution of samples of 25 will also be approximately normal with a mean of $\mu_{\bar{X}} = \mu = 2.05$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 0.005.$

$P(\bar{X} < 2.03) = P(Z < -4) = 0.000$

Excel Output:

	A	B
4	Mean	2.05
5	Standard Deviation	0.005
6		
7	Probability for X <=	
8	X Value	2.03
9	Z Value	-4
10	P(X<=2.03)	0.0000

(d) (a) refers to the amount of water in an individual two-liter bottle while (c) refers to the mean amount of water in a sample of 25 two-liter water bottles. There is a 21.19% chance that an individual water bottle will contain less than 2.03 liters but a zero chance that the mean amount of water in 25 water bottles will be less than 2.03 liters.

(e) Increasing the sample size from four to 25 reduced the probability that the mean amount of water will be less than 2.03 liters from 5.48% to 0%.

7.6 (a) $P(X < 42.035) = P(Z < -0.6) = 0.2743$

Excel Output:

	A	B
4	Mean	42.05
5	Standard Deviation	0.025
6		
7	Probability for X <=	
8	X Value	42.035
9	Z Value	-0.6
10	P(X<=42.035)	0.2743

(b) Because the weight of an energy bar is approximately normally distributed, the sampling distribution of samples of 4 will also be approximately normal with a mean of

$\mu_{\bar{X}} = \mu = 42.05$ and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = 0.0125.$

$P(\bar{X} < 42.035) = P(Z < -1.2) = 0.1151$

Excel Output:

	A	B
4	Mean	42.05
5	Standard Deviation	0.0125
6		
7	Probability for X <=	
8	X Value	42.035
9	Z Value	-1.2
10	P(X<=42.035)	0.1151

(c) Because the weight of an energy bar is approximately normally distributed, the sampling distribution of samples of 25 will also be approximately normal with a mean of

$\mu_{\bar{X}} = \mu = 42.05$ and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = 0.005.$

$P(\bar{X} < 42.035) = P(Z < -3) = 0.0013$

Excel Output:

	A	B
4	Mean	42.05
5	Standard Deviation	0.005
6		
7	Probability for X <=	
8	X Value	42.035
9	Z Value	-3
10	P(X<=42.035)	0.0013
11		

(d) (a) refers to an individual energy bar while (c) refers to the mean of a sample of 25 energy bars. There is a 27.43% chance that an individual energy bar will have a weight below 42.05 grams but only a chance of 0.135% that a mean of 25 energy bars will have a weight below 42.05 grams.

(e) Increasing the sample size from four to 25 reduced the probability the mean will have a weight below 42.05 grams from 11.51% to 0.135%.

7.7 (a) Because the population diameter of tennis balls is approximately normally distributed, the sampling distribution of samples of 9 will also be approximately normal with a mean of $\mu_{\bar{X}} = \mu = 2.63$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 0.01$.

(b) $P(\bar{X} < 2.61) = P(Z < -2.00) = 0.0228$

Probability for X <=	
X Value	2.61
Z Value	–2
P(X<=2.61)	0.0227501

(c) $P(2.62 < \bar{X} < 2.64) = P(-1.00 < Z < 1.00) = 0.6827$

Probability for a Range	
From X Value	2.61
To X Value	2.64
Z Value for 2.62	–1
Z Value for 2.64	1
P(X<=2.62)	0.1587
P(X<=2.64)	0.8413
P(2.62<=X<=2.64))	0.6827

(d) $P(A < \bar{X} < B) = P(-1.000 < Z < 1.000) = 0.68$

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	20.00%
Z Value	–0.841621
X Value	2.621584

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	80.00%
Z Value	0.841621
X Value	2.638416

Lower bound: $\bar{X} = 2.6216$
Upper bound: $\bar{X} = 2.6384$

7.8 (a) When $n = 4$, the shape of the sampling distribution of $\overline{X}$ should closely resemble the shape of the distribution of the population from which the sample is selected. Because the mean is larger than the median, the distribution of the sales price of new houses is skewed to the right, and so is the sampling distribution of $\overline{X}$ although it will be less skewed than the population.

(b) If you select samples of $n = 100$, the shape of the sampling distribution of the sample mean will be very close to a normal distribution with a mean of \$382,700 and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \$9{,}000$.

7.8 cont. (c) $P(\overline{X} < 370{,}000) = P(Z < -1.411) = .0791$

Excel Output:

Normal Probabilities	
Common Data	
Mean	382700
Standard Deviation	9000
Probability for X <=	
X Value	370000
Z Value	-1.411111
P(X<=370000)	0.0791

(d) $P(350{,}000 < \overline{X} < 365{,}000) = P(-3.63 < Z < -1.97) = 0.0245$

Excel Output:

	Probability for a Range	
20		
21	From X Value	350000
22	To X Value	365000
23	Z Value for 350000	-3.633333
24	Z Value for 365000	-1.966667
25	P(X<=350000)	0.0001
26	P(X<=365000)	0.0246
27	P(350000<=X<=365000)	0.0245

7.9 (a) Because the number of apps used per month by smartphone owners is assumed to be normally distributed, the sampling distribution of samples of 25 will also be approximately normal with a mean of $\mu_{\overline{X}} = \mu = 30$ and $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = 1.$

$P(29 < \overline{X} < 31) = P(-1 < Z < 1) = 0.6827$

Excel Output:

	Probability for a Range	
20		
21	From X Value	29
22	To X Value	31
23	Z Value for 29	-1
24	Z Value for 31	1
25	P(X<=29)	0.1587
26	P(X<=31)	0.8413
27	P(29<=X<=31)	0.6827

7.9 cont. (b) $P(28 < \overline{X} < 32) = P(-2 < Z < 2) = 0.9545$

Excel Output:

20	Probability for a Range	
21	From X Value	28
22	To X Value	32
23	Z Value for 28	-2
24	Z Value for 32	2
25	P(X<=28)	0.0228
26	P(X<=32)	0.9772
27	P(28<=X<=32)	0.9545

(c) Because the number of apps used per month by smartphone owners is assumed to be normally distributed, the sampling distribution of samples of 100 will also be approximately normal with a mean of $\mu_{\overline{X}} = \mu = 30$ and $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = 0.5$.

$P(29 < \overline{X} < 31) = P(-2 < Z < 2) = 0.9545$

Excel Output:

20	Probability for a Range	
21	From X Value	29
22	To X Value	31
23	Z Value for 29	-2
24	Z Value for 31	2
25	P(X<=29)	0.0228
26	P(X<=31)	0.9772
27	P(29<=X<=31)	0.9545

(d) With the sample size increasing from $n = 25$ to $n = 100$, more sample means will be closer to the distribution mean. The standard error of the sampling distribution of size 100 is much smaller than that of size 25, so the likelihood that the sample mean will fall within ±1 apps of the mean is much higher for samples of size 100 (probability = 0.9545) than for samples of size 25 (probability = 0. 6827).

7.10 (a) $\mu_{\overline{X}} = \mu = 15$ and $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = 1$

$P(\overline{X} > 14) = P(Z > -1) = 0.8413$

Excel Output:

12	Probability for X >	
13	X Value	14
14	Z Value	-1
15	P(X>14)	0.8413

7.10 cont. (b) $P(\overline{X} < A) = P(Z < 1.04) = 0.85$ $\overline{X} = 15 + 1.04\,(1) = 16.04$

Excel Output:

29	Find X and Z Given a Cum. Pctage.	
30	Cumulative Percentage	85.00%
31	Z Value	1.04
32	X Value	16.04

(c) To be able to use the standardized normal distribution as an approximation for the area under the curve, you must assume that the population is approximately symmetrical.

(d) $P(\overline{X} < A) = P(Z < 1.04) = 0.85$ $\overline{X} = 15 + 1.04\,(0.5) = 15.52$

Excel Output:

29	Find X and Z Given a Cum. Pctage.	
30	Cumulative Percentage	85.00%
31	Z Value	1.04
32	X Value	15.52

7.11 (a) $p = \dfrac{48}{64} = 0.75$

(b) $\sigma_p = \sqrt{\dfrac{0.70(0.30)}{64}} = 0.0573$

7.12 (a) $p = \dfrac{20}{50} = 0.40$

(b) $\sigma_p = \sqrt{\dfrac{(0.45)(0.55)}{50}} = 0.0704$

7.13 (a) $p = 14/40 = 0.35$

(b) $\sigma_p = \sqrt{\dfrac{0.30(0.70)}{40}} = 0.0725$

7.14 (a) $\mu_p = \pi = 0.501$, $\sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}} = \sqrt{\dfrac{0.501(1-0.501)}{100}} = 0.05$

Partial PHstat output:

Probability for X >	
X Value	0.55
Z Value	0.98
P(X>0.55)	0.1635

$P(p > 0.55) = P(Z > 0.98) = 1 - 0.8365 = 0.1635$

7.14 (b) $\mu_p = \pi = 0.60$, $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.6(1-0.6)}{100}} = 0.04899$

Partial PHstat output:

Probability for X >	
X Value	0.55
Z Value	–1.020621
P(X>0.55)	0.8463

$P(p > 0.55) = P(Z > -1.021) = 1 - 0.1539 = 0.8461$

(c) $\mu_p = \pi = 0.49$, $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.49(1-0.49)}{100}} = 0.05$

Partial PHstat output:

Probability for X >	
X Value	0.55
Z Value	1.2002401
P(X>0.55)	0.1150

$P(p > 0.55) = P(Z > 1.20) = 1 - 0.8849 = 0.1151$

(d) Increasing the sample size by a factor of 4 decreases the standard error by a factor of 2.

(a) Partial PHstat output:

Probability for X >	
X Value	0.55
Z Value	1.9600039
P(X>0.55)	0.0250

$P(p > 0.55) = P(Z > 1.96) = 1 - 0.9750 = 0.0250$

(b) Partial PHstat output:

Probability for X >	
X Value	0.55
Z Value	–2.041241
P(X>0.55)	0.9794

$P(p > 0.55) = P(Z > -2.04) = 1 - 0.0207 = 0.9793$

(c) Partial PHstat output:

Probability for X >	
X Value	0.55
Z Value	2.4004801
P(X>0.55)	0.0082

$P(p > 0.55) = P(Z > 2.40) = 1 - 0.9918 = 0.0082$

If the sample size is increased to 400, the probably in (a), (b) and (c) is smaller, larger, and smaller, respectively because the standard error of the sampling distribution of the sample proportion becomes smaller and, hence, the sampling distribution is more concentrated around the true population proportion.

7.15 (a) Partial PHstat output:

Probability for a Range	
From X Value	0.5
To X Value	0.6
Z Value for 0.5	0
Z Value for 0.6	2.828427
P(X<=0.5)	0.5000
P(X<=0.6)	0.9977
P(0.5<=X<=0.6)	0.4977

$P(0.50 < p < 0.60) = P(0 < Z < 2.83) = 0.4977$

(b) Partial PHstat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	95.00%
Z Value	1.644854
X Value	0.558154

$P(-1.645 < Z < 1.645) = 0.90$

$p = .50 - 1.645(0.0354) = 0.4418$ $p = .50 + 1.645(0.0354) = 0.5582$

(c) Partial PHstat output:

Probability for X >	
X Value	0.65
Z Value	4.2426407
P(X>0.65)	0.0000

$P(p > 0.65) = P(Z > 4.24)$ = virtually zero

(d) Partial PHstat output:

Probability for X >	
X Value	0.6
Z Value	2.8284271
P(X>0.6)	0.0023

If $n = 200$, $P(p > 0.60) = P(Z > 2.83) = 1.0 - 0.9977 = 0.0023$

Probability for X >	
X Value	0.55
Z Value	3.1622777
P(X>0.55)	0.00078

If $n = 1000$, $P(p > 0.55) = P(Z > 3.16) = 1.0 - 0.99921 = 0.00079$

More than 60% correct in a sample of 200 is more likely than more than 55% correct in a sample of 1000.

7.16 $\mu_p = \pi = 0.65,\ \sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}} = \sqrt{\dfrac{0.65(1-0.65)}{100}} = 0.0477$

(a) $P(p < 0.7) = P(Z < 1.0483) = 0.8527$

7.16 cont. (a) Excel Output:

7	Probability for X <=	
8	X Value	0.7
9	Z Value	1.0482848
10	P(X<=0.7)	0.8527

(b) $P(0.6 < p < 0.7) = P(-1.05 < Z < 1.05) = 0.7055$
Excel Output:

20	Probability for a Range	
21	From X Value	0.6
22	To X Value	0.7
23	Z Value for 0.6	-1.048285
24	Z Value for 0.7	1.0482848
25	P(X<=0.6)	0.1473
26	P(X<=0.7)	0.8527
27	P(0.6<=X<=0.7)	0.7055

(c) $P(p > 0.7) = P(Z > 1.048) = 0.1473$
Excel Output:

2	Probability for X >	
3	X Value	0.7
4	Z Value	1.0482848
5	P(X>0.7)	0.1473

(d) $\mu_p = \pi = 0.65,\ \sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}} = \sqrt{\dfrac{0.65(1-0.65)}{400}} = 0.0238$

$P(p < 0.7) = P(Z < 2.0966) = 0.982$
$P(0.6 < p < 0.7) = P(-2.0966 < Z < 2.0966) = 0.9640$
$P(p > 0.7) = P(Z > 2.0966) = 0.0180$
Excel Output:

1	Normal Probabilities	
2		
3	Common Data	
4	Mean	0.65
5	Standard Deviation	0.0238485
6		
7	Probability for X <=	
8	X Value	0.7
9	Z Value	2.0965697
10	P(X<=0.7)	0.9820
11		
12	Probability for X >	
13	X Value	0.7
14	Z Value	2.0965697
15	P(X>0.7)	0.0180

7.16 cont. (d)

	Probability for a Range	
21	From X Value	0.6
22	To X Value	0.7
23	Z Value for 0.6	-2.09657
24	Z Value for 0.7	2.0965697
25	P(X<=0.6)	0.0180
26	P(X<=0.7)	0.9820
27	P(0.6<=X<=0.7)	0.9640

7.17 $\mu_p = \pi = 0.33,\ \sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}} = \sqrt{\dfrac{0.33(1-0.33)}{100}} = 0.0470$

Excel Output:

	A	B
1	Normal Probabilities	
2		
3	Common Data	
4	Mean	0.33
5	Standard Deviation	0.0470213
6		
7	Probability for X <=	
8	X Value	0.3
9	Z Value	-0.638009
10	P(X<=0.3)	0.2617
11		
12	Probability for X >	
13	X Value	0.38
14	Z Value	1.0633485
15	P(X>0.38)	0.1438

	Probability for a Range	
21	From X Value	0.28
22	To X Value	0.38
23	Z Value for 0.28	-1.063349
24	Z Value for 0.38	1.0633485
25	P(X<=0.28)	0.1438
26	P(X<=0.38)	0.8562
27	P(0.28<=X<=0.38)	0.7124

(a) $P(p < 0.3) = P(Z < -0.638) = 0.2617$

(b) $P(0.28 < p < 0.38) = P(-1.063 < Z < 1.063) = 0.7124$

(c) $P(p > 0.38) = P(Z > 1.0633) = 0.1438$

7.17 (d) $\mu_p = \pi = 0.33,\ \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.33(1-0.33)}{400}} = 0.0235$

cont. (a) $P(p < 0.3) = P(Z < -1.276) = 0.1010$

(b) $P(0.28 < p < 0.38) = P(-2.127 < Z < 2.127) = 0.9666$

(c) $P(p > 0.38) = P(Z > 2.127) = 0.0167$

(c) Excel Output:

	A	B
1	Normal Probabilities	
2		
3	Common Data	
4	Mean	0.33
5	Standard Deviation	0.0235106
6		
7	Probability for X <=	
8	X Value	0.3
9	Z Value	-1.276018
10	P(X<=0.3)	0.1010
11		
12	Probability for X >	
13	X Value	0.38
14	Z Value	2.1266971
15	P(X>0.38)	0.0167
20	Probability for a Range	
21	From X Value	0.28
22	To X Value	0.38
23	Z Value for 0.28	-2.126697
24	Z Value for 0.38	2.1266971
25	P(X<=0.28)	0.0167
26	P(X<=0.38)	0.9833
27	P(0.28<=X<=0.38)	0.9666

7.18 (a) $\mu_p = \pi = 0.392,\ \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.392(1-0.392)}{200}} = 0.0345207$

$P(0.30 < p < 0.38) = P(-2.665 < Z < -0.3476) = 0.3602$

Excel Output:

Probability for a Range	
From X Value	0.3
To X Value	0.38
Z Value for 0.3	-2.665066
Z Value for 0.38	-0.347617
P(X<=0.3)	0.0038
P(X<=0.38)	0.3641
P(0.3<=X<=0.38)	0.3602

7.18 cont. (b) The probability is 90% that the sample percentage will be contained between 0.3352 and 0.4488.

Excel Output:

Find X Values Given a Percentage	
Percentage	90.00%
Z Value	-1.64
Lower X Value	0.3352
Upper X Value	0.4488

(c) The probability is 95% that the sample percentage will be contained between 0.3243 and 0.4597.

Excel Output:

	Find X Values Given a Percentage	
35	Percentage	95.00%
36	Z Value	-1.96
37	Lower X Value	0.3243
38	Upper X Value	0.4597

7.19 (a) $\mu_p = \pi = 0.65,\ \sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}} = \sqrt{\dfrac{0.65(1-0.65)}{100}} = 0.0477$

$P(0.64 < p < 0.69) = P(-0.209657 < Z < 0.8386279) = 0.3822$

Excel Output:

	Probability for a Range	
21	From X Value	0.64
22	To X Value	0.69
23	Z Value for 0.64	-0.209657
24	Z Value for 0.69	0.8386279
25	P(X<=0.64)	0.4170
26	P(X<=0.69)	0.7992
27	P(0.64<=X<=0.69)	0.3822

(b) The probability is 90% that the sample percentage will be contained between 0.5715 and 0.7285.

Excel Output:

Find X Values Given a Percentage	
Percentage	90.00%
Z Value	-1.6449
Lower X Value	0.5715
Upper X Value	0.7285

(c) The probability is 95% that the sample percentage will be contained between 0.5565 and 0.7435.

7.19 cont. (c) Excel Output:

Find X Values Given a Percentage	
Percentage	95.00%
Z Value	-1.9600
Lower X Value	0.5565
Upper X Value	0.7435

(d) $\mu_p = \pi = 0.65,\ \sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}} = \sqrt{\dfrac{0.65(1-0.65)}{400}} = 0.0238$

$P(0.64 < p < 0.69) = P(-0.209657 < Z < 0.8386279) = 0.6158$

Excel Output:

	Probability for a Range	
21	From X Value	0.64
22	To X Value	0.69
23	Z Value for 0.64	-0.419314
24	Z Value for 0.69	1.6772557
25	P(X<=0.64)	0.3375
26	P(X<=0.69)	0.9533
27	P(0.64<=X<=0.69)	0.6158

The probability is 90% that the sample percentage will be contained between 0.6108 and 0.6892.

Excel Output:

Find X Values Given a Percentage	
Percentage	90.00%
Z Value	-1.6449
Lower X Value	0.6108
Upper X Value	0.6892

The probability is 95% that the sample percentage will be contained between 0.6033 and 0.6967.

Find X Values Given a Percentage	
Percentage	95.00%
Z Value	-1.9600
Lower X Value	0.6033
Upper X Value	0.6967

7.20 $\mu_p = \pi = 0.80,\ \sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}} = \sqrt{\dfrac{0.80(1-0.80)}{200}} = 0.02828$

(a) $P(0.70 < p < 0.78) = P(-3.5355 < Z < -0.7071) = 0.2395$

Probability for a Range	
From X Value	0.7
To X Value	0.78
Z Value for 0.7	-3.535534
Z Value for 0.78	-0.707107
P(X<=0.7)	0.0002
P(X<=0.78)	0.2398
P(0.7<=X<=0.78)	0.2395

(b) The probability is 90% that the sample percentage will be contained between 0.7535 and 0.8465

	Find X Values Given a Percentage	
34		
35	Percentage	90.00%
36	Z Value	-1.64
37	Lower X Value	0.7535
38	Upper X Value	0.8465

(c) The probability is 95% that the sample percentage will be contained between 0.7446 and 0.8554

Find X Values Given a Percentage	
Percentage	95.00%
Z Value	-1.96
Lower X Value	0.7446
Upper X Value	0.8554

7.21 Because the average of all the possible sample means of size n is equal to the population mean.

7.22 The standard error of the sample means becomes smaller as larger sample sizes are taken. This is due to the fact that an extreme observation will have a smaller effect on the mean in a larger sample than in a small sample. Thus, the sample means will tend to be closer to the population mean as the sample size increases.

7.23 As larger sample sizes are taken, the effect of extreme values on the sample mean becomes smaller and smaller. With large enough samples, even though the population is not normally distributed, the sampling distribution of the mean will be approximately normally distributed.

7.24 The population distribution is the distribution of a particular variable of interest, while the sampling distribution represents the distribution of a statistic.

7.25 When the items of interest and the items not of interest are at least 5, the normal distribution can be used to approximate the binomial distribution.

7.26 $\mu_{\overline{X}} = 0.753$ $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.004}{5} = 0.0008$

PHStat output:

Common Data	
Mean	0.753
Standard Deviation	0.0008

Probability for X <=	
X Value	0.74
Z Value	–16.25
P(X<=0.74)	1.117E-59

Probability for X >	
X Value	0.76
Z Value	8.75
P(X>0.76)	0.0000

Probability for X<0.74 or X >0.76	
P(X<0.74 or X >0.76)	0.0000

Probability for a Range	
From X Value	0.74
To X Value	0.75
Z Value for 0.74	–16.25
Z Value for 0.75	–3.75
P(X<=0.74)	0.0000
P(X<=0.75)	0.0001
P(0.74<=X<=0.75)	0.00009

Probability for a Range	
From X Value	0.75
To X Value	0.753
Z Value for 0.75	–3.75
Z Value for 0.753	0
P(X<=0.75)	0.0001
P(X<=0.753)	0.5000
P(0.75<=X<=0.753)	0.4999

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	7.00%
Z Value	–1.475791
X Value	0.751819

(a) $P(0.75 < \overline{X} < 0.753) = P(-3.75 < Z < 0) = 0.5 - 0.00009 = 0.4999$

(b) $P(0.74 < \overline{X} < 0.75) = P(-16.25 < Z < -3.75) = 0.00009$

(c) $P(\overline{X} > 0.76) = P(Z > 8.75) =$ virtually zero

(d) $P(\overline{X} < 0.74) = P(Z < -16.25) =$ virtually zero

(e) $P(\overline{X} < A) = P(Z < -1.48) = 0.07$ $X = 0.753 - 1.48(0.0008) = 0.7518$

7.27 $\mu_{\bar{X}} = 2.0$ $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.05}{5} = 0.01$

PHStat output:

Common Data	
Mean	2
Standard Deviation	0.01

Probability for X <=	
X Value	1.98
Z Value	–2
P(X<=1.98)	0.0227501

Probability for X >	
X Value	2.01
Z Value	1
P(X>2.01)	0.1587

Probability for X<1.98 or X >2.01	
P(X<1.98 or X >2.01)	0.1814

Probability for a Range	
From X Value	1.99
To X Value	2
Z Value for 1.99	–1
Z Value for 2	0
P(X<=1.99)	0.1587
P(X<=2)	0.5000
P(1.99<=X<=2)	0.3413

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	1.00%
Z Value	–2.326348
X Value	1.976737

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	99.50%
Z Value	2.575829
X Value	2.025758

(a) $P(1.99 < \overline{X} < 2.00) = P(-1.00 < Z < 0) = 0.5 - 0.1587 = 0.3413$

(b) $P(\overline{X} < 1.98) = P(Z < -2.00) = 0.0228$

(c) $P(\overline{X} > 2.01) = P(Z > 1.00) = 1.0 - 0.8413 = 0.1587$

(d) $P(\overline{X} > A) = P(Z > -2.33) = 0.99$ $A = 2.00 - 2.33(0.01) = 1.9767$

(e) $P(A < X < B) = P(-2.58 < Z < 2.58) = 0.99$

$A = 2.00 - 2.58(0.01) = 1.9742$ $B = 2.00 + 2.58(0.01) = 2.0258$

7.28 $\mu_{\bar{X}} = 4.7$ $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{0.40}{5} = 0.08$

PHstat output:

Common Data	
Mean	4.7
Standard Deviation	0.08
Probability for X >	
X Value	4.6
Z Value	–1.25
P(X>4.6)	0.8944

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	23.00%
Z Value	–0.738847
X Value	4.640892

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	15.00%
Z Value	–1.036433
X Value	4.6170853

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	85.00%
Z Value	1.036433
X Value	4.782915

7.28 (a) $P(4.60 < \overline{X}) = P(-1.25 < Z) = 1 - 0.1056 = 0.8944$

cont. (b) $P(A < \overline{X} < B) = P(-1.04 < Z < 1.04) = 0.70$

$A = 4.70 - 1.04(0.08) = 4.6168$ ounces $X = 4.70 + 1.04(0.08) = 4.7832$ ounces

(c) $P(\overline{X} > A) = P(Z > -0.74) = 0.77$ $A = 4.70 - 0.74(0.08) = 4.6408$

7.29 $\mu_{\overline{X}} = 5.0$ $\sigma_{\overline{X}} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{0.40}{5} = 0.08$

(a) Partial PHStat output:

Probability for X >	
X Value	4.6
Z Value	–5
P(X>4.6)	1.0000

$P(4.60 < \overline{X}) = P(-5 < Z)$ = essentially 1.0

(b) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	15.00%
Z Value	–1.036433
X Value	4.917085

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	85.00%
Z Value	1.036433
X Value	5.082915

$P(A < \overline{X} < B) = P(-1.0364 < Z < 1.0364) = 0.70$

$A = 5.0 - 1.0364(0.08) = 4.9171$ ounces

$X = 5.0 + 1.0364(0.08) = 5.0829$ ounces

(c) Partial PHStat output:

Find X and Z Given Cum. Pctage.	
Cumulative Percentage	23.00%
Z Value	–0.738847
X Value	4.940892

$P(\overline{X} > A) = P(Z > -0.7388) = 0.77$ $A = 5.0 - 0.7388(0.08) = 4.9409$

7.30 $\mu_{\bar{X}} = 21.19$ $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{20}{4} = 5$

Excel Output:

1	Normal Probabilities	
2		
3	Common Data	
4	Mean	21.19
5	Standard Deviation	5
6		
7	Probability for X <=	
8	X Value	0
9	Z Value	-4.238
10	P(X<=0)	0.0000
11		
12	Probability for X >	
13	X Value	10
14	Z Value	-2.238
15	P(X>10)	0.9874

20	Probability for a Range	
21	From X Value	0
22	To X Value	10
23	Z Value for 0	-4.238
24	Z Value for 10	-2.238
25	P(X<=0)	0.0000
26	P(X<=10)	0.0126
27	P(0<=X<=10)	0.0126

(a) $P(\bar{X} < 0) = P(Z < -4.238) = 0.0000$

(b) $P(0 < \bar{X} < 10) = P(-4.238 < Z < -2.238) = 0.0126$

(c) $P(\bar{X} > 10) = P(Z > -2.238) = 0.9874$

7.31 Excel Output:

1	Normal Probabilities	
2		
3	Common Data	
4	Mean	-12.53
5	Standard Deviation	10
6		
7	Probability for X <=	
8	X Value	0
9	Z Value	1.253
10	P(X<=0)	0.8949
11		
12	Probability for X >	
13	X Value	-5
14	Z Value	0.753
15	P(X>-5)	0.2257

0	Probability for a Range	
1	From X Value	-20
2	To X Value	-10
3	Z Value for -20	-0.747
4	Z Value for -10	0.253
5	P(X<=-20)	0.2275
6	P(X<=-10)	0.5999
7	P(-20<=X<=-10)	0.3723

(a) $P(\bar{X} < 0) = P(Z < 1.25) = 0.8949$

(b) $P(-20 < \bar{X} < -10) = P(-0.747 < Z < 0.253) = 0.3723$

(c) $P(\bar{X} > -5) = P(Z > 0.753) = 0.2257$

7.31 $\mu_{\bar{X}} = -12.53 \qquad \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{10}{2} = 5$

cont. Excel Output:

	A	B
1	Normal Probabilities	
2		
3	Common Data	
4	Mean	-12.53
5	Standard Deviation	5
6		
7	Probability for X <=	
8	X Value	0
9	Z Value	2.506
10	P(X<=0)	0.9939
11		
12	Probability for X >	
13	X Value	-5
14	Z Value	1.506
15	P(X>-5)	0.0660

20	Probability for a Range	
21	From X Value	-20
22	To X Value	-10
23	Z Value for -20	-1.494
24	Z Value for -10	0.506
25	P(X<=-20)	0.0676
26	P(X<=-10)	0.6936
27	P(-20<=X<=-10)	0.6260

(d) $P(\overline{X} < 0) = P(Z < 2.506) = 0.9939$

(e) $P(-20 < \overline{X} < -10) = P(-1.494 < Z < 0.506) = 0.6260$

(f) $P(\overline{X} > -5) = P(Z > 1.506) = 0.0660$

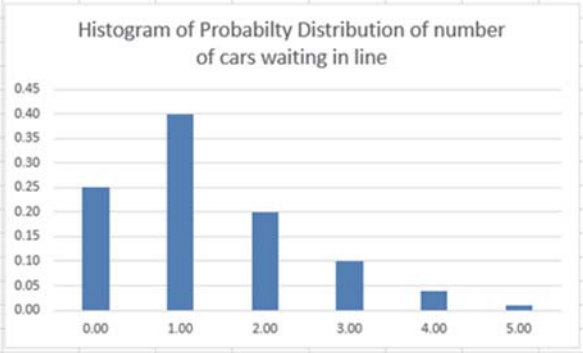
(g) Since the sample mean of returns of a sample of stocks is distributed closer to the population mean than the return of a single stock, the probabilities in (a) and (b) are lower than those in (d) and (e) while the probability in (c) is higher than that in (f).

7.32 Class Project answers will vary. The mean of the uniform distribution is $\dfrac{a+b}{2}$, since the random numbers in the table range from 0 to 9 the mean is 4.5. When $n = 2$, the frequency distribution of the sample means for the class should be centered around 4.5 and have a shape similar to column B with $n = 2$ in Figure 7.4 page 232 of text. As the sample size increases the frequency distribution of the sample means should have the shape similar to a normal distribution centered around 4.5.

7.33 Class Project answers will vary. This scenario simulates a binomial distribution with $\pi = 0.5$, the mean is $n\pi = 10(0.5) = 5$. The frequency distribution of the entire class should have the shape similar to a normal distribution centered around 5.

7.34 Class Project answers will vary. Population mean is 1.310 and population standard deviation is 1.13

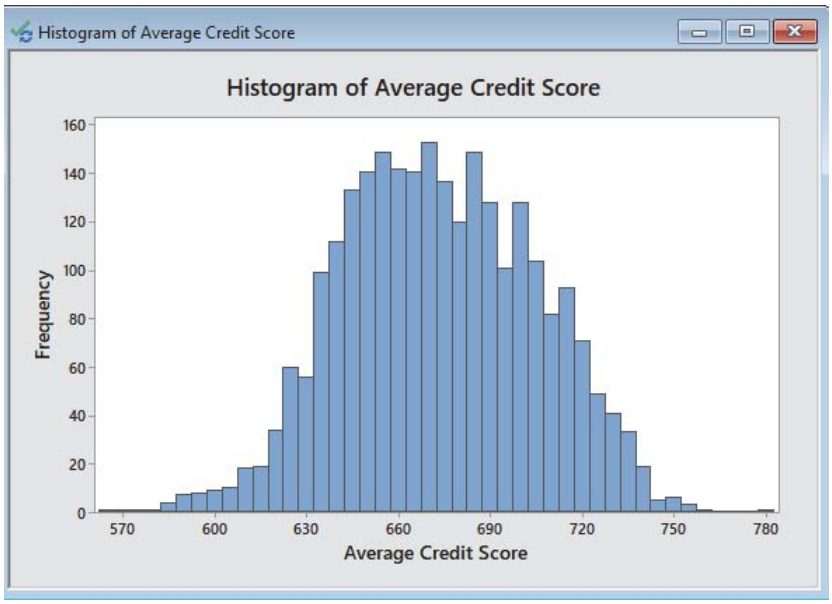
X	P(X=x)	x*P(X=x)	(x-mean)^2*P(X=x)
0.00	0.25	0.000	0.429025
1.00	0.40	0.400	0.03844
2.00	0.20	0.400	0.09522
3.00	0.10	0.300	0.28561
4.00	0.04	0.160	0.289444
5.00	0.01	0.050	0.136161
mean	1.310		
standard deviation	1.12867179		



Depending on class results, one should expect similar results to example 7.5 on page 233 of text. The frequency distributions of the sample means for each sample size should progress from a skewed population toward a bell-shaped distribution as the sample size increases.

7.35 (a)–(b) Class Project answers will vary.

(c) Since the population histogram of average credit score is fairly symmetrical, one can expect the class created frequency distributions of the sample means (for each sample size) to be approximately normal for $n = 5$, $n = 15$, and $n = 30$.



Chapter 8

8.1 $\overline{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}} = 85 \pm 1.96 \cdot \frac{8}{\sqrt{64}}$ $83.04 \leq \mu \leq 86.96$

8.2 $\overline{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}} = 125 \pm 2.58 \cdot \frac{24}{\sqrt{36}}$ $114.68 \leq \mu \leq 135.32$

8.3 Since the results of only one sample are used to indicate whether something has gone wrong in the production process, the manufacturer can never know with 100% certainty that the specific interval obtained from the sample includes the true population mean. In order to have 100% confidence, the entire population (sample size *N*) would have to be selected.

8.4 Yes, it is true since 5% of intervals will not include the population mean.

8.5 If all possible samples of the same size $n = 100$ are taken, 95% of them will include the true population mean time spent on the site per day. Thus you are 95 percent confident that this sample is one that does correctly estimate the true mean time spent on the site per day.

8.6 (a) You would compute the mean first because you need the mean to compute the standard deviation. If you had a sample, you would compute the sample mean. If you had the population mean, you would compute the population standard deviation.

(b) If you have a sample, you are computing the sample standard deviation not the population standard deviation needed in Equation 8.1. If you have a population, and have computed the population mean and population standard deviation, you don't need a confidence interval estimate of the population mean since you already have computed it.

8.7 If the population mean time spent on the site is 46 minutes a day, the confidence interval estimate stated in Problem 8.5 is correct because it contains the value 46 minutes.

8.8 Equation (8.1) assumes that you know the population standard deviation. Because you are selecting a sample of 100 from the population, you are computing a sample standard deviation, not the population standard deviation.

8.9 (a) $\overline{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}} = 0.993 \pm 2.58 \cdot \frac{0.02}{\sqrt{50}}$ $0.9857 \leq \mu \leq 1.0003$

(b) Since the value of 1.0 is included in the interval, there is no reason to believe that the mean is different from 1.0 gallon.

(c) No. Since σ is known and $n = 50$, from the Central Limit Theorem, we may assume that the sampling distribution of $\overline{X}$ is approximately normal.

(d) The reduced confidence level narrows the width of the confidence interval.

$\overline{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}} = 0.993 \pm 1.96 \cdot \frac{0.02}{\sqrt{50}}$ $0.9875 \leq \mu \leq 0.9985$

Since the value of 1.0 is not included in the interval, this changes the answer to part (b) and the distributor does have a right to complain.

8.10 (a) $\bar{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}} = 49{,}875 \pm 1.96 \cdot \frac{1500}{\sqrt{64}}$ $49{,}507.51 \le \mu \le 50{,}242.49$

Excel Output:

	A	B
1	**Confidence Estimate for the Mean**	
2		
3	**Data**	
4	**Population Standard Deviation**	**1500**
5	**Sample Mean**	**49875**
6	**Sample Size**	**64**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	187.5
11	Z Value	-1.9600
12	Interval Half Width	367.4932
13		
14	**Confidence Interval**	
15	**Interval Lower Limit**	**49507.5068**
16	**Interval Upper Limit**	**50242.4932**

(b) Yes, because the confidence interval includes 50,000 hours the manufacturer can support a claim that the bulbs have a mean of 50,000 hours.

(c) No. Because σ is known and n = 64, from the Central Limit Theorem, you know that the sampling distribution of $\bar{X}$ is approximately normal.

(d) $\bar{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}} = 49{,}875 \pm 1.96 \cdot \frac{500}{\sqrt{64}}$ $49{,}752.50 \le \mu \le 49{,}997.50$

The confidence interval is narrower, based on the population standard deviation of 500 hours and the confidence interval no longer includes 50,000 so the manufacturer could not state that the LED bulbs have a mean life of 50,000 hours.

(d) Excel Output:

	A	B
1	**Confidence Estimate for the Mean**	
2		
3	**Data**	
4	**Population Standard Deviation**	**500**
5	**Sample Mean**	**49875**
6	**Sample Size**	**64**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	62.5
11	Z Value	-1.9600
12	Interval Half Width	122.4977
13		
14	**Confidence Interval**	
15	**Interval Lower Limit**	**49752.5023**
16	**Interval Upper Limit**	**49997.4977**

8.11 $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 75 \pm 2.0301 \cdot \frac{24}{\sqrt{36}}$ $66.8796 \leq \mu \leq 83.1204$

8.12 (a) $df = 9,\ \alpha = 0.05,\ t_{\alpha/2} = 2.2622$

(b) $df = 9,\ \alpha = 0.01,\ t_{\alpha/2} = 3.2498$

(c) $df = 31,\ \alpha = 0.05,\ t_{\alpha/2} = 2.0395$

(d) $df = 64,\ \alpha = 0.05,\ t_{\alpha/2} = 1.9977$

(e) $df = 15,\ \alpha = 0.1,\ t_{\alpha/2} = 1.7531$

8.13 Set 1: $4.5 \pm 2.3646 \cdot \frac{3.7417}{\sqrt{8}}$ $1.3719 \leq \mu \leq 7.6281$

Set 2: $4.5 \pm 2.3646 \cdot \frac{2.4495}{\sqrt{8}}$ $2.4522 \leq \mu \leq 6.5478$

The data sets have different confidence interval widths because they have different values for the standard deviation.

8.14 Original data: $5.8571 \pm 2.4469 \cdot \frac{6.4660}{\sqrt{7}}$ $-0.1229 \leq \mu \leq 11.8371$

Altered data: $4.00 \pm 2.4469 \cdot \frac{2.1602}{\sqrt{7}}$ $2.0022 \leq \mu \leq 5.9978$

The presence of an outlier in the original data increases the value of the sample mean and greatly inflates the sample standard deviation.

8.15 (a) PHStat output:

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	**200**
Sample Mean	**1500**
Sample Size	**100**
Confidence Level	**95%**

Intermediate Calculations	
Standard Error of the Mean	20
Degrees of Freedom	99
t Value	1.9842
Interval Half Width	39.6843

Confidence Interval	
Interval Lower Limit	**1460.32**
Interval Upper Limit	**1539.68**

$\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 1500 \pm 1.9842 \cdot \frac{200}{\sqrt{100}}$ $\$\ 1460.32 \leq \mu \leq \$\ 1539.68$

(b) You can be 95% confident that the population mean spending for all Amazon Prime member shoppers is somewhere between $1460.32 and $1539.68.

8.16 (a) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 87 \pm 1.9781 \cdot \frac{9}{\sqrt{133}}$ $85.46 \le \mu \le 88.54$

3	**Data**	
4	Sample Standard Deviation	9
5	Sample Mean	87
6	Sample Size	133
7	Confidence Level	95%
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	0.7804
11	Degrees of Freedom	132
12	*t* Value	1.9781
13	Interval Half Width	1.5437
14		
15	**Confidence Interval**	
16	Interval Lower Limit	85.46
17	Interval Upper Limit	88.54

(b) You can be 95% confident that the population mean one-time gift donations is somewhere between \$85.46 and \$88.54.

8.17 (a) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 195.3 \pm 2.1098 \cdot \frac{21.4}{\sqrt{18}}$ $184.6581 \le \mu \le 205.9419$

(b) No, a grade of 200 is in the interval.

(c) It is not unusual to have an observed tread wear index of 210, which is outside the 95% confidence interval for the population mean tread wear index, because the standard error of the sample mean $\sigma / \sqrt{n}$ is smaller than the standard deviation of the population σ of the tread wear index for a single observed treat wear. Hence, the value of a single observed tread wear index varies around the population mean more than a sample mean does.

8.18 (a) $6.32 \le \mu \le 7.87$

Minitab Output:

One-Sample T: Cost ($)

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
15	7.093	1.406	0.363	(6.315, 7.872)

μ: mean of Cost ($)

(b) You can be 95% confident that the population mean amount spent for lunch at a fast-food restaurant is between \$6.31 and \$7.87.

(c) That the population distribution is normally distributed.

(d) The assumption of normality is not seriously violated and with a sample of 15, the validity of the confidence interval is not seriously impacted.

8.19 (a) Commuting Time $132.29 \leq \mu \leq 145.37$

	A	B
3	**Data**	
4	**Sample Standard Deviation**	**17.51075203**
5	**Sample Mean**	**138.8333333**
6	**Sample Size**	**30**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	3.1970
11	Degrees of Freedom	29
12	*t* Value	2.0452
13	Interval Half Width	6.5386
14		
15	**Confidence Interval**	
16	**Interval Lower Limit**	**132.29**
17	**Interval Upper Limit**	**145.37**

(b) You can be 95% confident that the population mean commuting time is somewhere between 132.29 minutes and 145.37 minutes.

(c) That the population distributions are normally distributed

(d) The assumption of normality is not seriously violated with sample sizes of 30. The validity of the confidence interval is not seriously impacted.

8.20 (a) For First and Second Quarter ads: $5.12 \leq \mu \leq 5.65$

For Halftime and Second half ads: $5.24 \leq \mu \leq 6.06$

Excel Output:

	A	B	C
1	**Estimate for the Mean Ad Ratings First and Second Quarter**		
2			
3	**Data**		
4	**Sample Standard Deviation**	**0.715253771**	
5	**Sample Mean**	**5.38**	
6	**Sample Size**	**31**	
7	**Confidence Level**	**95%**	
8			
9	Intermediate Calculations		
10	Standard Error of the Mean	0.1285	
11	Degrees of Freedom	30	
12	*t* Value	2.0423	
13	Interval Half Width	0.2624	
14			
15	**Confidence Interval**		
16	**Interval Lower Limit**	**5.12**	
17	**Interval Upper Limit**	**5.65**	

	A	B	C
1	**Estimate for the Mean Ad Ratings halftime and Second half**		
2			
3	**Data**		
4	**Sample Standard Deviation**	**1.041093825**	
5	**Sample Mean**	**5.65**	
6	**Sample Size**	**27**	
7	**Confidence Level**	**95%**	
8			
9	Intermediate Calculations		
10	Standard Error of the Mean	0.2004	
11	Degrees of Freedom	26	
12	*t* Value	2.0555	
13	Interval Half Width	0.4118	
14			
15	**Confidence Interval**		
16	**Interval Lower Limit**	**5.24**	
17	**Interval Upper Limit**	**6.06**	

(b) You are 95% confident that the mean rating for first and second quarter ads is between 5.12 and 5.65. You are 95% confident that the mean rating for halftime and second half ads is between 5.24 and 6.06.

(c) The confidence intervals for the two groups of ads are similar.

(d) You need to assume that the distributions of the rating for the two groups of ads are normally distributed.

(e) The distribution of each group of ads appears approximately normally distributed.

8.21 Excel Output:

	A	B
1	Estimate for the Mean yield for One-Year CD	
2		
3	Data	
4	Sample Standard Deviation	1.009900961
5	Sample Mean	1.66
6	Sample Size	46
7	Confidence Level	95%
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	0.1489
11	Degrees of Freedom	45
12	*t* Value	2.0141
13	Interval Half Width	0.2999
14		
15	Confidence Interval	
16	Interval Lower Limit	1.36
17	Interval Upper Limit	1.96
18		

	A	B
1	Estimate for the Mean yield for Five-Year CD	
2		
3	Data	
4	Sample Standard Deviation	0.967073902
5	Sample Mean	2.17
6	Sample Size	46
7	Confidence Level	95%
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	0.1426
11	Degrees of Freedom	45
12	*t* Value	2.0141
13	Interval Half Width	0.2872
14		
15	Confidence Interval	
16	Interval Lower Limit	1.89
17	Interval Upper Limit	2.46

(a) One Year CD $1.36 \le \mu \le 1.96$

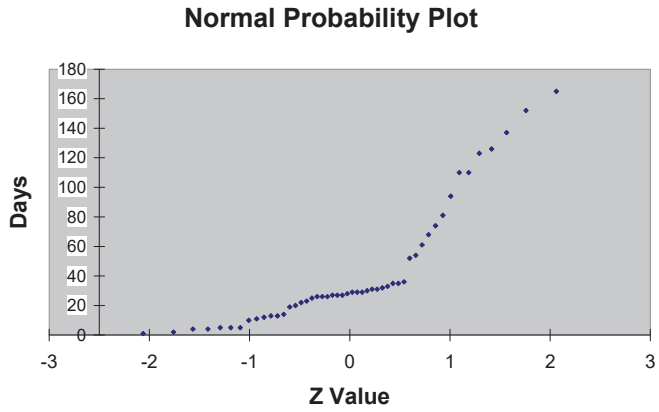
(b) Five Year CD $1.89 \le \mu \le 2.46$

(c) The mean yield for a one-year CD is somewhere between 1.36 and 1.96 with 95% confidence and the mean yield for a five-year CD is somewhere between 1.89 and 2.46 with 95% confidence.

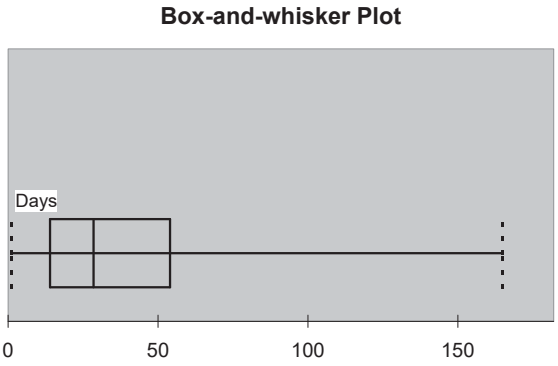
8.22 (a) $\bar{X} \pm t \cdot \dfrac{S}{\sqrt{n}} = 43.04 \pm 2.0096 \cdot \dfrac{41.9261}{\sqrt{50}}$ $31.12 \le \mu \le 54.96$

(b) The population distribution needs to be normally distribution.

(c)



8.22 (c) cont.



Both the normal probability plot and the boxplot suggest that the distribution is skewed to the right.

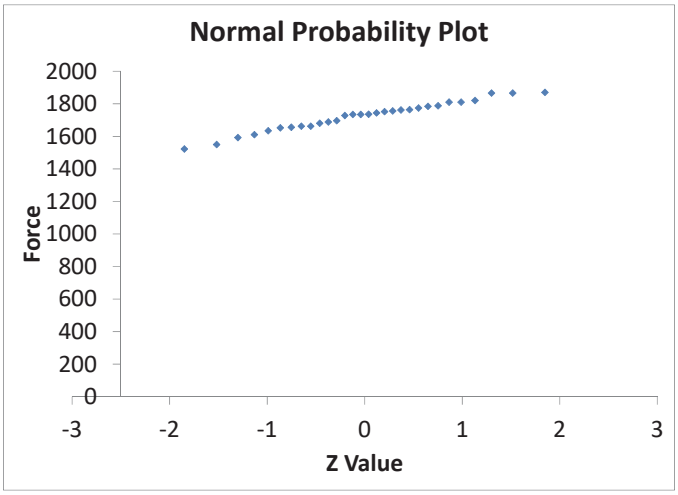
(d) Even though the population distribution is not normally distributed, with a sample of 50, the *t* distribution can still be used due to the Central Limit Theorem.

8.23 (a) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 1725.07 \pm 2.0452 \cdot \frac{88.55115}{\sqrt{30}}$ $1692.00 \leq \mu \leq 1758.13$

	A	B
1	**Estimate for the Mean Force**	
2		
3	**Data**	
4	**Sample Standard Deviation**	**88.55115**
5	**Sample Mean**	**1725.067**
6	**Sample Size**	**30**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	16.1672
11	Degrees of Freedom	29
12	*t* Value	2.0452
13	Interval Half Width	33.0655
14		
15	**Confidence Interval**	
16	**Interval Lower Limit**	**1692.00**
17	**Interval Upper Limit**	**1758.13**

(b) The population distribution needs to be normally distributed.

8.23 (c)
cont.

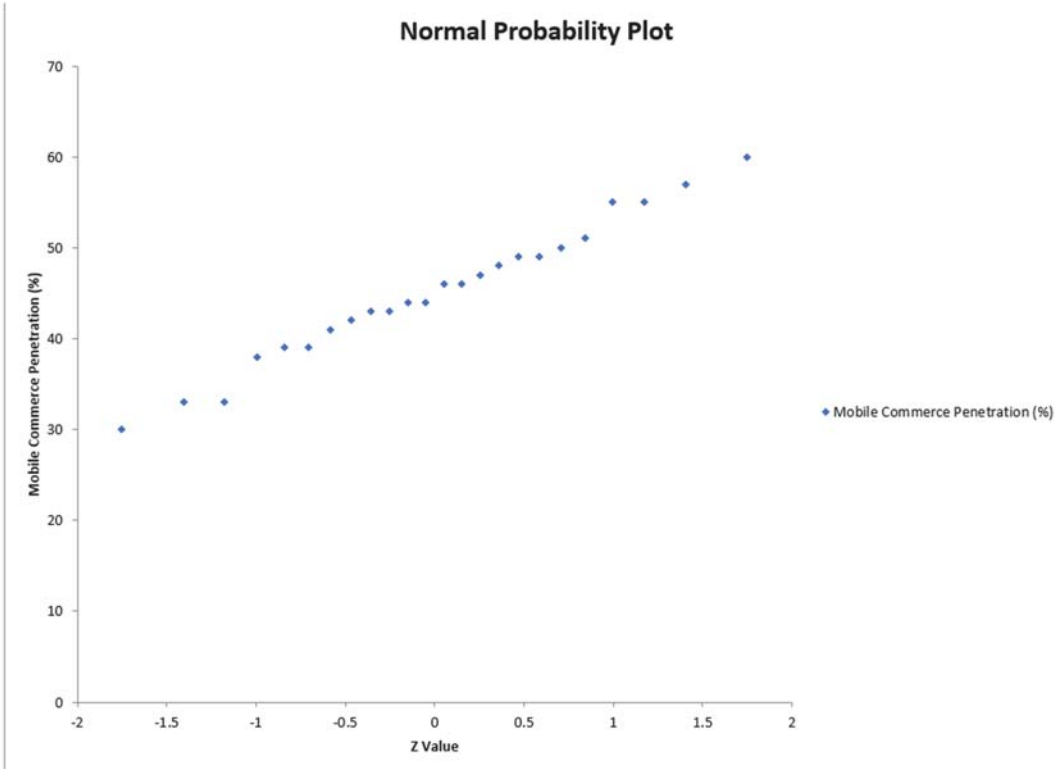


The normal probability plot indicates that the population distribution is normally distributed.

8.24 Excel Output:

	A	B	C
1	**Estimate for the Mean Mobile Commerce Penetration %**		
2			
3	**Data**		
4	**Sample Standard Deviation**	**7.649476633**	
5	**Sample Mean**	**45.08**	
6	**Sample Size**	**24**	
7	**Confidence Level**	**95%**	
8			
9	Intermediate Calculations		
10	Standard Error of the Mean	1.5614	
11	Degrees of Freedom	23	
12	*t* Value	2.0687	
13	Interval Half Width	3.2301	
14			
15	**Confidence Interval**		
16	**Interval Lower Limit**	**41.85**	
17	**Interval Upper Limit**	**48.31**	

8.24
cont.



(a) $41.85 \leq \mu \leq 48.31$

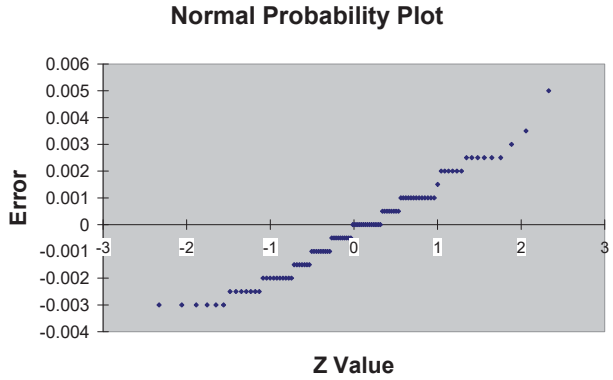
(b) The populations distribution is normally distributed.

(c) The normal probability plot appears to be approximately normally distributed

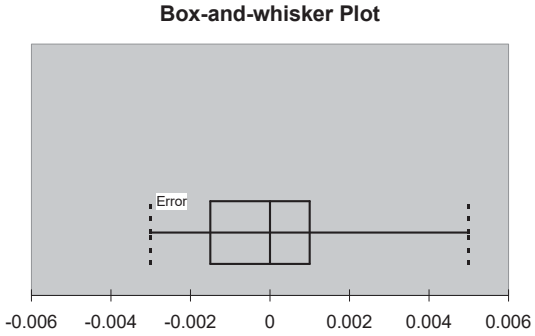
8.25 (a) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = -0.00023 \pm 1.9842 \cdot \frac{0.0017}{\sqrt{100}}$ $-0.000566 \leq \mu \leq 0.000106$

(b) The population distribution needs to be normally distributed. However, with a sample of 100, the *t* distribution can still be used as a result of the Central Limit Theorem even if the population distribution is not normal.

(c)



8.25 (c)
cont.



Both the normal probability plot and the boxplot suggest that the distribution is skewed to the right.

(d) We are 95% confident that the mean difference between the actual length of the steel part and the specified length of the steel part is between –0.000566 and 0.000106 inch , which is narrower than the plus or minus 0.005 inch requirement. The steel mill is doing a good job at meeting the requirement. This is consistent with the finding in Problem 2.43.

8.26 $p = \frac{X}{n} = \frac{50}{200} = 0.25$ $\quad p \pm Z \cdot \sqrt{\frac{p(1-p)}{n}} = 0.25 \pm 1.96\sqrt{\frac{0.25(0.75)}{200}}$

$$0.19 \le \pi \le 0.31$$

8.27 $p = \frac{X}{n} = \frac{25}{400} = 0.0625$ $\quad p \pm Z \cdot \sqrt{\frac{p(1-p)}{n}} = 0.0625 \pm 2.58\sqrt{\frac{0.0625(0.9375)}{400}}$

$$0.0313 \le \pi \le 0.0937$$

8.28 (a)

	A	B
1	**Purchase Additional Telephone Line**	
2		
3	**Sample Size**	**500**
4	**Number of Successes**	**135**
5	**Confidence Level**	**99%**
6	Sample Proportion	0.27
7	Z Value	-2.57583451
8	Standard Error of the Proportion	0.019854471
9	Interval Half Width	0.05114183
10	**Interval Lower Limit**	**0.21885817**
11	**Interval Upper Limit**	**0.32114183**

$p = \frac{X}{n} = \frac{135}{500} = 0.27$ $\quad p \pm Z \cdot \sqrt{\frac{p(1-p)}{n}} = 0.27 \pm 2.5758\sqrt{\frac{0.27(1-0.27)}{500}}$

$0.22 \le \pi \le 0.32$

(b) The manager in charge of promotional programs concerning residential customers can infer that the proportion of households that would purchase a new cellphone if it were made available at a substantially reduced installation cost is between 0.22 and 0.32 with a 99% level of confidence.

8.29 (a) Excel Output:

	A	B
1	Proportion of Males who reported window tinting	
2		
3	Data	
4	Sample Size	502
5	Number of Successes	46
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Sample Proportion	0.091633
10	Z Value	-1.9600
11	Standard Error of the Proportion	0.012877
12	Interval Half Width	0.0252
13		
14	Confidence Interval	
15	Interval Lower Limit	0.0664
16	Interval Upper Limit	0.1169
17		

$$p = 0.09 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.09 \pm 1.96\sqrt{\frac{0.09(1-0.09)}{502}}$$

$$0.0664 \le \pi \le 0.1169$$

(b) Excel Output:

	A	B	C
1	Proportion of Females who reported window tinting		
2			
3	Data		
4	Sample Size	502	
5	Number of Successes	71	
6	Confidence Level	95%	
7			
8	Intermediate Calculations		
9	Sample Proportion	0.141434	
10	Z Value	-1.9600	
11	Standard Error of the Proportion	0.015553	
12	Interval Half Width	0.0305	
13			
14	Confidence Interval		
15	Interval Lower Limit	0.1110	
16	Interval Upper Limit	0.1719	
17			

$$p = 0.14 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.14 \pm 1.96\sqrt{\frac{0.14(1-0.14)}{502}}$$

$$0.1110 \le \pi \le 0.1719$$

(c) One can infer that the proportion of males who state window tinting is their preferred luxury upgrade is somewhere between 0.07 and 0.12 with 95% confidence and the proportion of females who state window tinting is their preferred luxury upgrade is somewhere between 0.11 and 0.17 with 95% confidence.

8.30 (a) Excel Output:

	Data	
4	Sample Size	1000
5	Number of Successes	260
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Sample Proportion	0.26
10	Z Value	-1.9600
11	Standard Error of the Proportion	0.013871
12	Interval Half Width	0.0272
13		
14	Confidence Interval	
15	Interval Lower Limit	0.2328
16	Interval Upper Limit	0.2872
17		

$$p = 0.26 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.26 \pm 1.96\sqrt{\frac{0.26(1-0.26)}{1000}}$$

$$0.2328 \le \pi \le 0.2872$$

(b) No, you cannot because the interval estimate includes 0.25 (25%).

(c) Excel Output:

	Data	
4	Sample Size	10000
5	Number of Successes	2600
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Sample Proportion	0.26
10	Z Value	-1.9600
11	Standard Error of the Proportion	0.004386
12	Interval Half Width	0.0086
13		
14	Confidence Interval	
15	Interval Lower Limit	0.2514
16	Interval Upper Limit	0.2686

$$p = 0.26 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.26 \pm 1.96\sqrt{\frac{0.26(1-0.26)}{10,000}}$$

$$0.2514 \le \pi \le 0.2686$$

Yes, you can claim more than a quarter of all consumers value personalized experience most when shopping in retail store because the interval is above 0.25 (25%)

(d) The larger the sample size, the narrower the confidence interval, holding everything else constant.

8.31 (a) Excel Output

	A	B	C	D
1	Proportion of males who feel overloaded with too much information			
2				
3	Data			
4	Sample Size	1216		
5	Number of Successes	651		
6	Confidence Level	95%		
7				
8	Intermediate Calculations			
9	Sample Proportion	0.535362		
10	Z Value	-1.9600		
11	Standard Error of the Proportion	0.014303		
12	Interval Half Width	0.0280		
13				
14	Confidence Interval			
15	Interval Lower Limit	0.5073		
16	Interval Upper Limit	0.5634		
17				

$$p = 0.54 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.54 \pm 1.96\sqrt{\frac{0.54(1-0.54)}{1216}}$$

$$0.5073 \le \pi \le 0.5634$$

(b) Excel Output:

	A	B	C	D	E
1	Proportion of females who feel overloaded with too much information				
2					
3	Data				
4	Sample Size	304			
5	Number of Successes	134			
6	Confidence Level	95%			
7					
8	Intermediate Calculations				
9	Sample Proportion	0.440789			
10	Z Value	-1.9600			
11	Standard Error of the Proportion	0.028475			
12	Interval Half Width	0.0558			
13					
14	Confidence Interval				
15	Interval Lower Limit	0.3850			
16	Interval Upper Limit	0.4966			

$$p = 0.44 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.44 \pm 1.96\sqrt{\frac{0.44(1-0.44)}{304}}$$

$$0.3850 \le \pi \le 0.4966$$

(c) One can infer that the proportion of males who feel overloaded with too much information is somewhere between 0.51 and 0.56 with 95% confidence and the proportion of females who feel overloaded with too much information is somewhere between 0.39 and 0.50 with 95% confidence.

8.32 (a) Excel Output:

Data	
Sample Size	4787
Number of Successes	4178
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.87278
Z Value	-1.9600
Standard Error of the Proportion	0.004816
Interval Half Width	0.0094
Confidence Interval	
Interval Lower Limit	0.8633
Interval Upper Limit	0.8822

$$p = 0.87 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.87 \pm 1.96\sqrt{\frac{0.87(1-0.87)}{4787}}$$

$$0.8633 \leq \pi \leq 0.8822$$

(b) Excel Output:

Data	
Sample Size	4178
Number of Successes	789
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.188846
Z Value	-1.9600
Standard Error of the Proportion	0.006055
Interval Half Width	0.0119
Confidence Interval	
Interval Lower Limit	0.1770
Interval Upper Limit	0.2007

$$p = 0.19 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.19 \pm 1.96\sqrt{\frac{0.19(1-0.19)}{4178}}$$

$$0.1770 \leq \pi \leq 0.2007$$

(c) Because almost 90% of adults have purchased something online, but only about 20% are weekly online shoppers, the director of e-commerce sales may want to focus on those adults who are weekly online shoppers.

8.33 (a) Excel Output:

Data	
Sample Size	114
Number of Successes	57
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.5
Z Value	-1.9600
Standard Error of the Proportion	0.046829
Interval Half Width	0.0918
Confidence Interval	
Interval Lower Limit	0.4082
Interval Upper Limit	0.5918

$$p = 0.5 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.5 \pm 1.96\sqrt{\frac{0.5(1-0.5)}{114}}$$

$$0.4082 \le \pi \le 0.5918$$

(b) Excel Output:

Data	
Sample Size	114
Number of Successes	22
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.192982
Z Value	-1.9600
Standard Error of the Proportion	0.036961
Interval Half Width	0.0724
Confidence Interval	
Interval Lower Limit	0.1205
Interval Upper Limit	0.2654

$$p = 0.19 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.19 \pm 1.96\sqrt{\frac{0.19(1-0.19)}{114}}$$

$$0.1205 \le \pi \le 0.2654$$

(c) You can be 95% confident that the population proportion of U.S. CEOs who are extremely concerned about cyberthreats is somewhere between 0.4082 and 0.5918, and the population proportion of U.S. CEOs who are extremely concerned about lack of trust in business is somewhere between 0.1205 and 0.2654.

8.34 $n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot 15^2}{5^2} = 34.57$ Use $n = 35$

8.35 $n = \frac{Z^2\sigma^2}{e^2} = \frac{2.58^2 \cdot 100^2}{20^2} = 166.41$ Use $n = 167$

8.36 $n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{2.58^2(0.5)(0.5)}{(0.04)^2} = 1{,}040.06$ Use $n = 1{,}041$

8.37 $n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{1.96^2(0.4)(0.6)}{(0.02)^2} = 2{,}304.96$ Use $n = 2{,}305$

8.38 (a) $n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot 400^2}{50^2} = 245.86$ Use $n = 246$

(b) $n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot 400^2}{25^2} = 983.41$ Use $n = 984$

8.39 $n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot (0.02)^2}{(0.004)^2} = 96.04$ Use $n = 97$

8.40 Excel Output:

3	**Data**	
4	**Population Standard Deviation**	**1500**
5	**Sampling Error**	**400**
6	**Confidence Level**	**95%**
7		
8	Intemediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	54.0205
11		
12	**Result**	
13	**Sample Size Needed**	**55**
14		

$n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot 1500^2}{400^2} = 54.0225$ Use $n = 55$

8.41 $n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot (0.05)^2}{(0.01)^2} = 96.04$ Use $n = 97$

8.42 (a) $n = \frac{Z^2\sigma^2}{e^2} = \frac{2.5758^2 \cdot 20^2}{5^2} = 106.1583$ Use $n = 107$

(b) $n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot 20^2}{5^2} = 61.4633$ Use $n = 62$

8.43 (a) $n = \dfrac{Z^2\sigma^2}{e^2} = \dfrac{1.645^2 \cdot 45^2}{5^2} = 219.19$ Use $n = 220$

(b) $n = \dfrac{Z^2\sigma^2}{e^2} = \dfrac{2.5758^2 \cdot 45^2}{5^2} = 537.4266$ Use $n = 538$

8.44 Note: All the answers are computed using PHStat. Answers computed otherwise may be slightly different due to rounding.

(a) $n = \dfrac{Z^2\sigma^2}{e^2} = \dfrac{1.96^2 \cdot 2^2}{0.25^2} = 245.85$ Use $n = 246$

(b) $n = \dfrac{Z^2\sigma^2}{e^2} = \dfrac{1.96^2 \cdot 2.5^2}{0.25^2} = 384.15$ Use $n = 385$

(c) $n = \dfrac{Z^2\sigma^2}{e^2} = \dfrac{1.96^2 \cdot 3.0^2}{0.25^2} = 553.17$ Use $n = 554$

(d) When there is more variability in the population, a larger sample is needed to accurately estimate the mean.

8.45 (a) Excel Output:

3	**Data**	
4	**Estimate of True Proportion**	**0.2**
5	**Sampling Error**	**0.04**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	384.1459
11		
12	**Result**	
13	**Sample Size Needed**	**385**

$n = \dfrac{Z^2\pi(1-\pi)}{e^2} = \dfrac{1.96^2 \cdot (0.2)(1-0.2)}{0.04^2} = 384.16$ Use $n = 385$

(b) Excel Output:

3	**Data**	
4	**Estimate of True Proportion**	**0.2**
5	**Sampling Error**	**0.04**
6	**Confidence Level**	**99%**
7		
8	Intermediate Calculations	
9	Z Value	-2.5758
10	Calculated Sample Size	663.4897
11		
12	**Result**	
13	**Sample Size Needed**	**664**
14		

$n = \dfrac{Z^2\pi(1-\pi)}{e^2} = \dfrac{2.5758^2 \cdot (0.2)(1-0.2)}{0.04^2} = 663.4746$ Use $n = 664$

8.45 cont. (c) Excel Output:

3	**Data**	
4	**Estimate of True Proportion**	**0.2**
5	**Sampling Error**	**0.02**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	1536.5835
11		
12	**Result**	
13	**Sample Size Needed**	**1537**
14		

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.2)(1-0.2)}{0.02^2} = 1536.64 \qquad \text{Use } n = 1537$$

(d) Excel Output:

3	**Data**	
4	**Estimate of True Proportion**	**0.2**
5	**Sampling Error**	**0.02**
6	**Confidence Level**	**99%**
7		
8	Intermediate Calculations	
9	Z Value	-2.5758
10	Calculated Sample Size	2653.9586
11		
12	**Result**	
13	**Sample Size Needed**	**2654**
14		

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{2.5758^2 \cdot (0.2)(1-0.2)}{0.02^2} = 2653.898 \qquad \text{Use } n = 2654$$

(e) The higher the level of confidence desired, the larger is the sample size required. The smaller the sampling error desired, the larger is the sample size required.

8.46 (a) Excel Output:

	Data	
	Sample Size	**115**
	Number of Successes	**81**
	Confidence Level	**95%**
	Intermediate Calculations	
	Sample Proportion	0.704348
0	Z Value	-1.9600
1	Standard Error of the Proportion	0.042553
2	Interval Half Width	0.0834
3		
4	**Confidence Interval**	
5	**Interval Lower Limit**	**0.6209**
6	**Interval Upper Limit**	**0.7878**
7		

8.46 (a) $p = 0.70 \quad p \pm Z\sqrt{\dfrac{p(1-p)}{n}} = 0.70 \pm 1.96\sqrt{\dfrac{0.70(1-0.70)}{115}}$

cont. $0.6209 \leq \pi \leq 0.7878$

(b) Excel Output

Data	
Sample Size	115
Number of Successes	68
Confidence Level	95%

Intermediate Calculations	
Sample Proportion	0.591304
Z Value	-1.9600
Standard Error of the Proportion	0.045841
Interval Half Width	0.0898

Confidence Interval	
Interval Lower Limit	0.5015
Interval Upper Limit	0.6812

$p = 0.59 \quad p \pm Z\sqrt{\dfrac{p(1-p)}{n}} = 0.59 \pm 1.96\sqrt{\dfrac{0.59(1-0.59)}{115}}$

$0.5015 \leq \pi \leq 0.6812$

(c) Excel Output:

Data	
Sample Size	115
Number of Successes	16
Confidence Level	95%

Intermediate Calculations	
Sample Proportion	0.13913
Z Value	-1.9600
Standard Error of the Proportion	0.032272
Interval Half Width	0.0633

Confidence Interval	
Interval Lower Limit	0.0759
Interval Upper Limit	0.2024

$p = 0.14 \quad p \pm Z\sqrt{\dfrac{p(1-p)}{n}} = 0.14 \pm 1.96\sqrt{\dfrac{0.14(1-0.14)}{115}}$

$0.0759 \leq \pi \leq 0.2024$

8.46 cont. (d) (a) Excel Output:

Data	
Estimate of True Proportion	0.7
Sampling Error	0.02
Confidence Level	95%
Intermediate Calculations	
Z Value	-1.9600
Calculated Sample Size	2016.7659
Result	
Sample Size Needed	2017

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.7)(1-0.7)}{0.02^2} = 2016.84 \quad \text{Use } n = 2017$$

(b) Excel Output:

	Data	
4	Estimate of True Proportion	0.59
5	Sampling Error	0.02
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	2323.1222
11		
12	Result	
13	Sample Size Needed	2324

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.59)(1-0.59)}{0.02^2} = 2323.208 \text{ Use } n = 2324$$

(c) Excel Output:

	Data	
4	Estimate of True Proportion	0.14
5	Sampling Error	0.02
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	1156.2791
11		
12	Result	
13	Sample Size Needed	1157

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.14)(1-0.14)}{0.02^2} = 1156.322 \text{ Use } n = 1157$$

8.47 (a) Excel Output:

Data	
Sample Size	443
Number of Successes	130
Confidence Level	95%

Intermediate Calculations	
Sample Proportion	0.293454
Z Value	-1.9600
Standard Error of the Proportion	0.021634
Interval Half Width	0.0424

Confidence Interval	
Interval Lower Limit	0.2511
Interval Upper Limit	0.3359

$$p = 0.29 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.29 \pm 1.96\sqrt{\frac{0.29(1-0.29)}{443}}$$

$0.2511 \le \pi \le 0.3359$

(b) You are 95% confident that the population proportion of nonprofits nationwide indicating that the greatest diversity challenge they face is retaining younger staff (under 30) is somewhere between 0.2511 and 0.3359.

(c) Excel Output

Data	
Estimate of True Proportion	0.29
Sampling Error	0.01
Confidence Level	95%

Intermediate Calculations	
Z Value	-1.9600
Calculated Sample Size	7909.5637

Result	
Sample Size Needed	7910

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.29)(1-0.29)}{0.01^2} = 7909.854 \qquad \text{Use } n = 7910$$

8.48 (a) If you conducted a follow-up study, you would use $\pi = 0.38$ in the sample size formula because it is based on past information on the proportion.

8.48 cont. (b) Excel Output:

3	**Data**	
4	**Estimate of True Proportion**	**0.38**
5	**Sampling Error**	**0.03**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	1005.6086
11		
12	**Result**	
13	**Sample Size Needed**	**1006**
14		

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.38)(1-0.38)}{0.03^2} = 1005.646 \qquad \text{Use } n = 1006$$

8.49 (a) Excel Output:

3	**Data**	
4	**Estimate of True Proportion**	**0.4**
5	**Sampling Error**	**0.03**
6	**Confidence Level**	**99%**
7		
8	Intermediate Calculations	
9	Z Value	-2.5758
10	Calculated Sample Size	1769.3058
11		
12	**Result**	
13	**Sample Size Needed**	**1770**
14		

$$n = \frac{Z^2\pi(1-\pi)}{e^2} = \frac{2.5758^2 \cdot (0.4)(1-0.4)}{0.03^2} = 1769.266 \qquad \text{Use } n = 1770$$

(b) Excel Output:

3	**Data**	
4	**Estimate of True Proportion**	**0.4**
5	**Sampling Error**	**0.05**
6	**Confidence Level**	**99%**
7		
8	Intermediate Calculations	
9	Z Value	-2.5758
10	Calculated Sample Size	636.9501
11		
12	**Result**	
13	**Sample Size Needed**	**637**
14		

8.49 (b) $n = \dfrac{Z^2\pi(1-\pi)}{e^2} = \dfrac{2.5758^2 \cdot (0.4)(1-0.4)}{0.05^2} = 636.9356$ Use $n = 637$

cont. (c) A smaller sampling error requires a larger sample size.

8.50 The only way to have 100% confidence is to obtain the parameter of interest, rather than a sample statistic. From another perspective, the range of the normal and *t* distribution is infinite, so a *Z* or *t* value that contains 100% of the area cannot be obtained.

8.51 The *t* distribution is used for obtaining a confidence interval for the mean when σ is unknown.

8.52 If the confidence level is increased, a greater area under the normal or *t* distribution needs to be included. This leads to an increased value of *Z* or *t*, and thus a wider interval.

8.53 The term $\pi(1-\pi)$ reaches its largest value when the population proportion is at 0.5. Hence, the sample size $n = \dfrac{Z^2\pi(1-\pi)}{e^2}$ needed to determine the proportion is smaller when the population proportion is 0.20 than when the population proportion is 0.50.

8.54 (a) PC/laptop Excel Output:

3	**Data**	
4	**Sample Size**	**1000**
5	**Number of Successes**	**840**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Sample Proportion	0.84
10	Z Value	-1.9600
11	Standard Error of the Proportion	0.011593
12	Interval Half Width	0.0227
13		
14	**Confidence Interval**	
15	**Interval Lower Limit**	**0.8173**
16	**Interval Upper Limit**	**0.8627**
17		

$$p = 0.84 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.84 \pm 1.96\sqrt{\frac{0.84(1-0.84)}{1000}}$$

$$0.8173 \le \pi \le 0.8627$$

8.54 cont. (a) Smartphone Excel Output:

Data	
Sample Size	1000
Number of Successes	910
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.91
Z Value	-1.9600
Standard Error of the Proportion	0.00905
Interval Half Width	0.0177
Confidence Interval	
Interval Lower Limit	0.8923
Interval Upper Limit	0.9277

$$p = 0.91 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.91 \pm 1.96\sqrt{\frac{0.91(1-0.91)}{1000}}$$

$$0.8923 \leq \pi \leq 0.9277$$

Tablet Excel Output:

Data	
Sample Size	1000
Number of Successes	500
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.5
Z Value	-1.9600
Standard Error of the Proportion	0.015811
Interval Half Width	0.0310
Confidence Interval	
Interval Lower Limit	0.4690
Interval Upper Limit	0.5310

$$p = 0.5 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.5 \pm 1.96\sqrt{\frac{0.5(1-0.5)}{1000}}$$

$$0.469 \leq \pi \leq 0.5310$$

8.54 cont. (a) Smart Watch Excel Output:

Data	
Sample Size	1000
Number of Successes	100
Confidence Level	95%

Intermediate Calculations	
Sample Proportion	0.1
Z Value	-1.9600
Standard Error of the Proportion	0.009487
Interval Half Width	0.0186

Confidence Interval	
Interval Lower Limit	0.0814
Interval Upper Limit	0.1186

$$p = 0.1 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.1 \pm \sqrt{\frac{0.1(1-0.1)}{100}}$$

$0.0814 \leq \pi \leq 0.1186$

(b) Most adults have a PC/laptop and a smartphone. Some adults have a tablet computer and very few have a smart watch.

8.55 (a) Digital Coupons Excel Output:

Data	
Sample Size	731
Number of Successes	358
Confidence Level	95%

Intermediate Calculations	
Sample Proportion	0.48974
Z Value	-1.9600
Standard Error of the Proportion	0.018489
Interval Half Width	0.0362

Confidence Interval	
Interval Lower Limit	0.4535
Interval Upper Limit	0.5260

$$p = 0.49 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.49 \pm \sqrt{\frac{0.49(1-0.49)}{731}}$$

$0.4535 \leq \pi \leq 0.5260$

8.55 cont. (a) Look up recipes Excel Output:

Data	
Sample Size	731
Number of Successes	355
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.485636
Z Value	-1.9600
Standard Error of the Proportion	0.018486
Interval Half Width	0.0362
Confidence Interval	
Interval Lower Limit	0.4494
Interval Upper Limit	0.5219

$$p = 0.485 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.485 \pm \sqrt{\frac{0.485(1-0.485)}{731}}$$

$$0.4494 \leq \pi \leq 0.5219$$

Read product reviews Excel Output:

Data	
Sample Size	731
Number of Successes	234
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.320109
Z Value	-1.9600
Standard Error of the Proportion	0.017255
Interval Half Width	0.0338
Confidence Interval	
Interval Lower Limit	0.2863
Interval Upper Limit	0.3539

$$p = 0.32 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.32 \pm \sqrt{\frac{0.32(1-0.32)}{731}}$$

$$0.2863 \leq \pi \leq 0.3539$$

8.55 cont. (a) Locate in-store items Excel Output:

3	**Data**	
4	**Sample Size**	**731**
5	**Number of Successes**	**154**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Sample Proportion	0.21067
10	Z Value	-1.9600
11	Standard Error of the Proportion	0.015082
12	Interval Half Width	0.0296
13		
14	**Confidence Interval**	
15	**Interval Lower Limit**	**0.1811**
16	**Interval Upper Limit**	**0.2402**
17		

$$p = 0.21 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.21 \pm \sqrt{\frac{0.21(1-0.21)}{731}}$$

$0.1811 \leq \pi \leq 0.2402$

(b) About half of smartphone owners use their phone to access digital coupons or look up recipes while shopping in a grocery store. Fewer smartphone owners use their phone to read product reviews or locate items in the store while shopping in a grocery store.

8.56 Note: All the answers are computed using PHStat. Answers computed otherwise may be slightly different due to rounding.

(a) PHStat output:

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	**3.5**
Sample Mean	**51**
Sample Size	**40**
Confidence Level	**95%**

Intermediate Calculations	
Standard Error of the Mean	0.553398591
Degrees of Freedom	39
t Value	2.0227
Interval Half Width	1.1194

Confidence Interval	
Interval Lower Limit	**49.88**
Interval Upper Limit	**52.12**

$49.88 \leq \mu \leq 52.12$

8.56 cont. (b) PHStat output:

Confidence Interval Estimate for the Proportio

Data	
Sample Size	**40**
Number of Successes	**32**
Confidence Level	**95%**

Intermediate Calculations	
Sample Proportion	0.8000
Z Value	-1.9600
Standard Error of the Proportion	0.0632
Interval Half Width	0.1240

Confidence Interval	
Interval Lower Limit	**0.6760**
Interval Upper Limit	**0.9240**

$$0.6760 \le \pi \le 0.9240$$

(c) $$n = \frac{Z^2 \cdot \sigma^2}{e^2} = \frac{1.96^2 \cdot 5^2}{2^2} = 24.01$$ Use $n = 25$

(d) $$n = \frac{Z^2 \cdot \pi \cdot (1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.5) \cdot (0.5)}{(0.06)^2} = 266.7680$$ Use $n = 267$

(e) If a single sample were to be selected for both purposes, the larger of the two sample sizes ($n = 267$) should be used.

8.57 (a) Excel Output:

2		
3	**Data**	
4	Sample Standard Deviation	8
5	Sample Mean	42
6	Sample Size	50
7	Confidence Level	99%
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	1.1314
11	Degrees of Freedom	49
12	*t* Value	2.6800
13	Interval Half Width	3.0320
14		
15	**Confidence Interval**	
16	Interval Lower Limit	38.97
17	Interval Upper Limit	45.03
18		

$$\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 42 \pm 2.680 \cdot \frac{8}{\sqrt{50}} \qquad 38.97 \le \mu \le 45.03$$

8.57 cont. (b) Excel Output:

3	**Data**	
4	**Sample Size**	**50**
5	**Number of Successes**	**13**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Sample Proportion	0.26
10	Z Value	-1.9600
11	Standard Error of the Proportion	0.062032
12	Interval Half Width	0.1216
13		
14	**Confidence Interval**	
15	**Interval Lower Limit**	**0.1384**
16	**Interval Upper Limit**	**0.3816**
17		

$$p = 0.26 \quad p \pm Z\sqrt{\frac{p(1-p)}{n}} = 0.26 \pm \sqrt{\frac{0.26(1-0.26)}{50}}$$

$$0.1384 \leq \pi \leq 0.3816$$

8.58 (a) PHStat output:

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	**7.3**
Sample Mean	**6.2**
Sample Size	**25**
Confidence Level	**95%**

Intermediate Calculations	
Standard Error of the Mean	1.46
Degrees of Freedom	24
t Value	2.0639
Interval Half Width	3.0133

Confidence Interval	
Interval Lower Limit	**3.19**
Interval Upper Limit	**9.21**

$$3.19 \leq \mu \leq 9.21$$

8.58 cont. (b) PHStat output:

Confidence Interval Estimate for the Proportio

Data	
Sample Size	**25**
Number of Successes	**13**
Confidence Level	**95%**

Intermediate Calculations	
Sample Proportion	0.52
Z Value	-1.9600
Standard Error of the Proportion	0.0999
Interval Half Width	0.1958

Confidence Interval	
Interval Lower Limit	**0.3242**
Interval Upper Limit	**0.7158**

$$0.3241 \le \pi \le 0.7158$$

(c) $$n = \frac{Z^2 \cdot \sigma^2}{e^2} = \frac{1.96^2 \cdot 8^2}{1.5^2} = 109.2682$$ Use $n = 110$

(d) $$n = \frac{Z^2 \cdot \pi \cdot (1-\pi)}{e^2} = \frac{1.645^2 \cdot (0.5) \cdot (0.5)}{(0.075)^2} = 120.268$$ Use $n = 121$

(e) If a single sample were to be selected for both purposes, the larger of the two sample sizes ($n = 121$) should be used.

8.59 Note: All the answers are computed using PHStat. Answers computed otherwise may be slightly different due to rounding.

(a) PHStat output:

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	**1000**
Sample Mean	**12500**
Sample Size	**100**
Confidence Level	**95%**

Intermediate Calculations	
Standard Error of the Mean	100
Degrees of Freedom	99
t Value	1.9842
Interval Half Width	198.4217

Confidence Interval	
Interval Lower Limit	**12301.58**
Interval Upper Limit	**12698.42**

$\$12{,}301.58 \le \mu \le \$12{,}698.42$

8.59 cont. (b) PHStat output:

Confidence Interval Estimate for the Proportio

Data	
Sample Size	**100**
Number of Successes	**30**
Confidence Level	**95%**

Intermediate Calculations	
Sample Proportion	0.3000
Z Value	-1.9600
Standard Error of the Proportion	0.0458
Interval Half Width	0.0898

Confidence Interval	
Interval Lower Limit	**0.2102**
Interval Upper Limit	**0.3898**

$0.2102 \le \pi \le 0.3898$

(c) $n = \dfrac{Z^2 \cdot \sigma^2}{e^2} = \dfrac{2.58^2 \cdot 1000^2}{250^2} = 106.1583$ Use $n = 107$

(d) $n = \dfrac{Z^2 \cdot \pi \cdot (1-\pi)}{e^2} = \dfrac{1.645^2 \cdot (0.3) \cdot (1-0.3)}{(0.045)^2} = 280.5749$ Use $n = 281$

If a single sample were to be selected for both purposes, the larger of the two sample sizes ($n = 281$) should be used.

8.60 (a) $p \pm Z \cdot \sqrt{\dfrac{p(1-p)}{n}} = 0.31 \pm 1.645 \cdot \sqrt{\dfrac{0.31(0.69)}{200}}$ $0.2562 \le \pi \le 0.3638$

(b) $\bar{X} \pm t \cdot \dfrac{S}{\sqrt{n}} = 3.5 \pm 1.9720 \cdot \dfrac{2}{\sqrt{200}}$ $3.22 \le \mu \le 3.78$

(c) $\bar{X} \pm t \cdot \dfrac{S}{\sqrt{n}} = 18000 \pm 1.9720 \cdot \dfrac{3000}{\sqrt{200}}$ $\$17{,}581.68 \le \mu \le \$18{,}418.32$

8.61 (a) $\bar{X} \pm t \cdot \dfrac{S}{\sqrt{n}} = \$21.34 \pm 1.9949 \cdot \dfrac{\$9.22}{\sqrt{70}}$ $\$19.14 \le \mu \le \$23.54$

(b) $p \pm Z \cdot \sqrt{\dfrac{p(1-p)}{n}} = 0.3714 \pm 1.645 \cdot \sqrt{\dfrac{0.3714(0.6286)}{70}}$

$0.2764 \le \pi \le 0.4664$

(c) $n = \dfrac{Z^2 \sigma^2}{e^2} = \dfrac{1.96^2 \cdot 10^2}{1.5^2} = 170.74$ Use $n = 171$

(d) $n = \dfrac{Z^2 \cdot \pi \cdot (1-\pi)}{e^2} = \dfrac{1.645^2 \cdot (0.5) \cdot (0.5)}{(0.045)^2} = 334.08$ Use $n = 335$

(e) If a single sample were to be selected for both purposes, the larger of the two sample sizes ($n = 335$) should be used.

8.62 (a) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = \$38.54 \pm 2.0010 \cdot \frac{\$7.26}{\sqrt{60}}$ $\$36.66 \le \mu \le \$40.42$

(b) $p \pm Z \cdot \sqrt{\frac{p(1-p)}{n}} = 0.30 \pm 1.645 \cdot \sqrt{\frac{0.30(0.70)}{60}}$ $0.2027 \le \pi \le 0.3973$

(c) $n = \frac{Z^2\sigma^2}{e^2} = \frac{1.96^2 \cdot 8^2}{1.5^2} = 109.27$ Use $n = 110$

(d) $n = \frac{Z^2 \cdot \pi \cdot (1-\pi)}{e^2} = \frac{1.645^2 \cdot (0.5) \cdot (0.5)}{(0.04)^2} = 422.82$ Use $n = 423$

(e) If a single sample were to be selected for both purposes, the larger of the two sample sizes ($n = 423$) should be used.

8.63 (a) $n = \frac{Z^2 \cdot \pi \cdot (1-\pi)}{e^2} = \frac{1.96^2 \cdot (0.5) \cdot (0.5)}{(0.05)^2} = 384.16$ Use $n = 385$

If we assume that the population proportion is only 0.50, then a sample of 385 would be required. If the population proportion is 0.90, the sample size required is cut to 138.

(b) $p \pm Z \cdot \sqrt{\frac{p(1-p)}{n}} = 0.84 \pm 1.96 \cdot \sqrt{\frac{0.84(0.16)}{50}}$

$0.7384 \le \pi \le 0.9416$

(c) The representative can be 95% confidence that the actual proportion of bags that will do the job is between 74.5% and 93.5%. He/she can accordingly perform a cost-benefit analysis to decide if he/she wants to sell the Ice Melt product.

8.64 (a)

Confidence Interval Estimate for the Proportio

Data	
Sample Size	90
Number of Successes	51
Confidence Level	95%

Intermediate Calculations	
Sample Proportion	0.566666667
Z Value	-1.9600
Standard Error of the Proportion	0.0522
Interval Half Width	0.1024

Confidence Interval	
Interval Lower Limit	0.4643
Interval Upper Limit	0.6690

$p \pm Z \cdot \sqrt{\frac{p(1-p)}{n}} = 0.5667 \pm 1.96 \cdot \sqrt{\frac{0.5667(1-0.5667)}{90}}$

$0.4643 \le \pi \le 0.6690$

8.64 (b) cont.

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	1103.6491
Sample Mean	563.38
Sample Size	51
Confidence Level	95%

Intermediate Calculations	
Standard Error of the Mean	154.5417855
Degrees of Freedom	50
t Value	2.0086
Interval Half Width	310.4063

Confidence Interval	
Interval Lower Limit	252.97
Interval Upper Limit	873.79

$$\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 563.38 \pm 2.0086\left(\frac{1103.6491}{\sqrt{51}}\right) \qquad \$252.97 \le \mu \le \$873.79$$

8.65 (a) $$\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 5.5014 \pm 2.6800 \cdot \frac{0.1058}{\sqrt{50}} \qquad 5.46 \le \mu \le 5.54$$

(b) Since 5.5 grams is within the 99% confidence interval, the company can claim that the mean weight of tea in a bag is 5.5 grams with a 99% level of confidence.

(c) The assumption is valid as the weight of the tea bags is approximately normally distributed.

8.66 (a) MiniTab Output

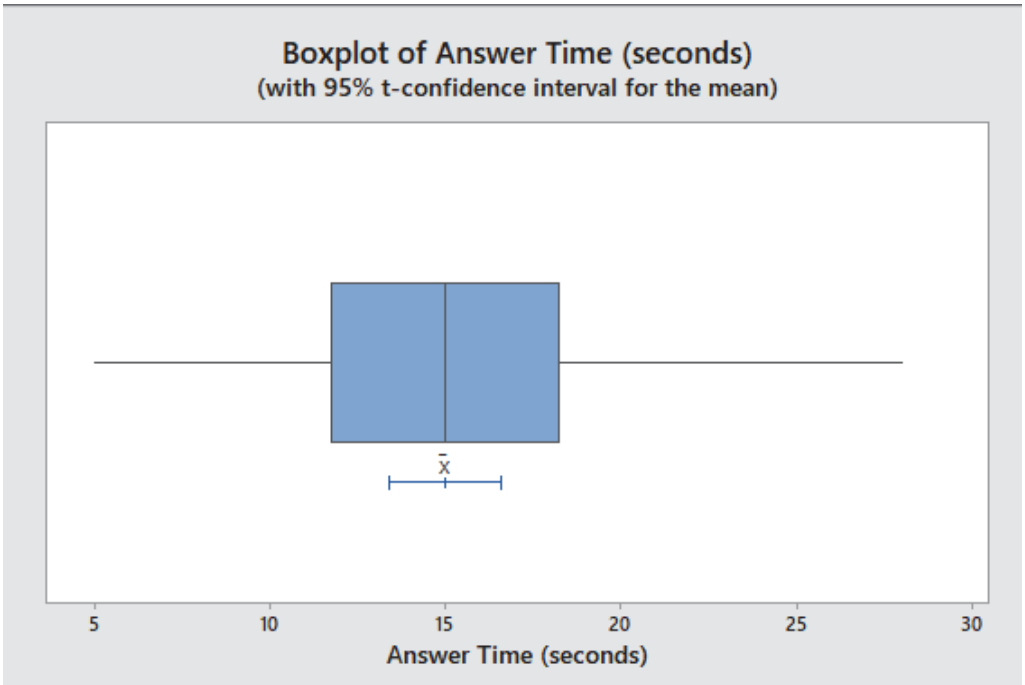
One-Sample T: Answer Time (seconds)

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
50	14.980	5.557	0.786	(13.401, 16.559)

μ: mean of Answer Time (seconds)

8.66 (a) cont.



$13.4001 \le \mu \le 16.559$

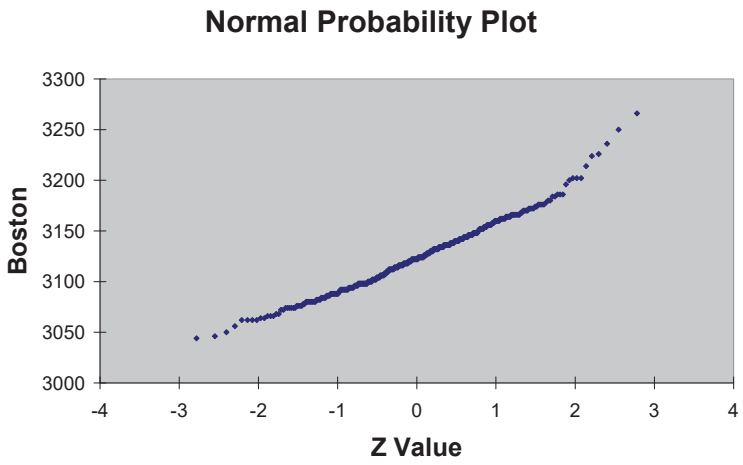
(b) With 95% confidence, the population mean answer time is somewhere between 13.40 and 16.56 seconds.

(c) The assumption is valid as the answer time is approximately normally distributed.

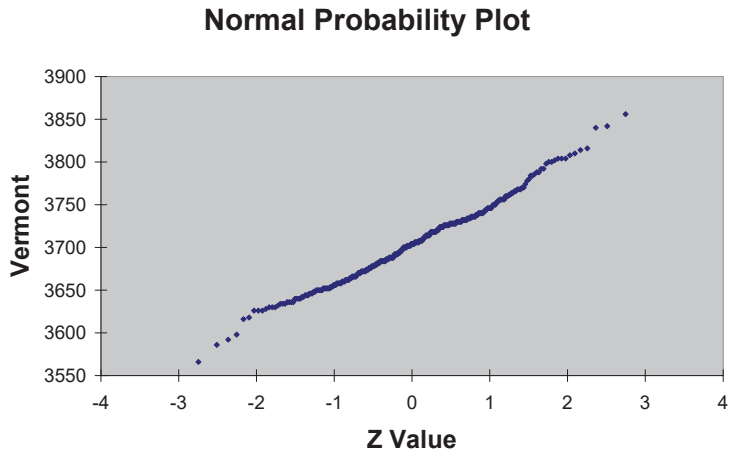
8.67 (a) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 3124.2147 \pm 1.9665 \cdot \frac{34.713}{\sqrt{368}}$ $3120.66 \le \mu \le 3127.77$

(b) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 3704.0424 \pm 1.9672 \cdot \frac{46.7443}{\sqrt{330}}$ $3698.98 < \mu < 3709.10$

(c)



8.67 (c) cont.



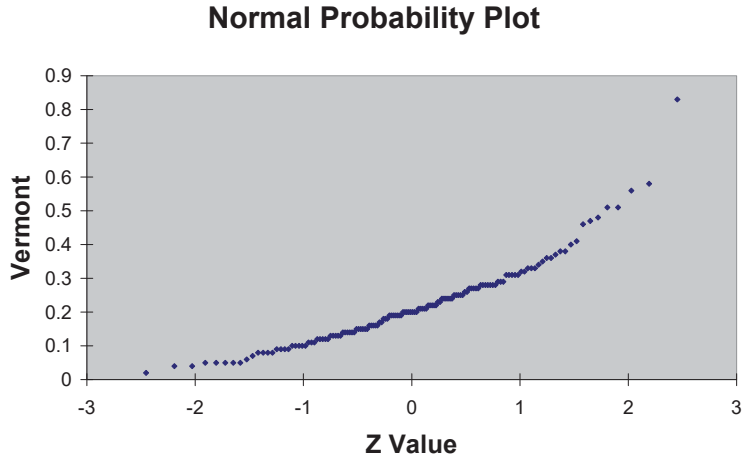
The weight for Boston shingles is slightly skewed to the right while the weight for Vermont shingles appears to be slightly skewed to the left.

(d) Since the two confidence intervals do not overlap, the mean weight of Vermont shingles is greater than the mean weight of Boston shingles.

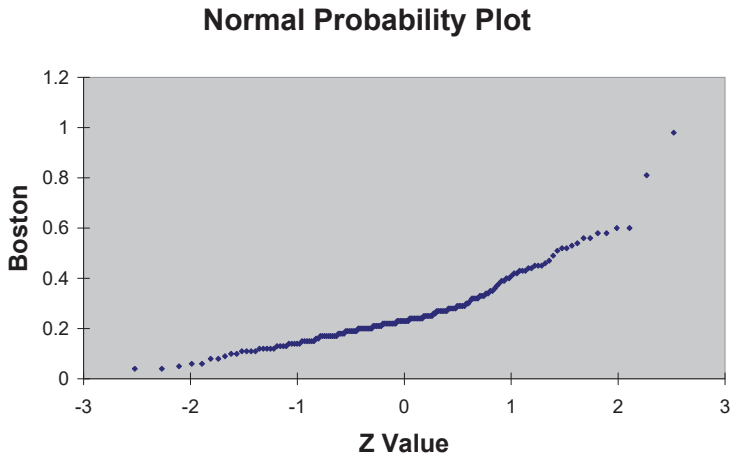
8.68 (a) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 0.2641 \pm 1.9741 \cdot \frac{0.1424}{\sqrt{170}}$ $0.2425 \le \mu \le 0.2856$

(b) $\bar{X} \pm t \cdot \frac{S}{\sqrt{n}} = 0.218 \pm 1.9772 \cdot \frac{0.1227}{\sqrt{140}}$ $0.1975 \le \mu \le 0.2385$

(c)



8.68 (c)
cont.



The amount of granule loss for both brands are skewed to the right but the sample sizes are large enough so the violation of the normality assumption is not critical.

(d) Because the two confidence intervals do not overlap, you can conclude that the mean granule loss of Boston shingles is higher than that of Vermont shingles

8.69 Report Writing Exercise answers will vary. An example of a report would be as follows:
One can conclude with 95% confidence that the mean time for a human agent to answer a call to the financial service center is between 13.401 and 16.559 seconds. The validity of this confidence interval estimate depends on the assumption that the processing time is normally distributed. From the box plot below the answer time appears approximately symmetric so the validity of the confidence interval is not in serious doubt.