TABLE OF CONTENTS

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

INTRODUCTION

by Milton Loyer

This *Instructor's Solutions Manual* contains detailed solutions to all the exercises in the text *Elementary Statistics Using Excel*, Third Edition, by Mario Triola. To aid in the comprehension of the calculations, worked problems typically include intermediate steps of algebraic or computer/calculator notation. When appropriate, additional hints and comments are included and prefaced by NOTE.

Many statistical problems are best solved using particular formats. Because recognizing and following these patterns promote understanding, this manual identifies and employs such formats whenever practicable.

The related *Student Solutions Manual* contains the solutions to the odd-numbered exercises for each section, and the solutions to all of the end-of-chapter exercises: the Statistical Literacy and Critical Thinking, the Review Exercises, and the Cumulative Review Exercises. This manual includes those solutions exactly as they appear in the student manual. Recognizing that the instructor's manual will also be used by grading assistants, the even-numbered solutions have been prepared with the same level of detail.

I would like to thank Mario Triola for writing an excellent elementary statistics textbook and for inviting me to prepare this solutions manual.

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

Chapter 1

Introduction to Statistics

1-2 Types of Data

1. A parameter is a numerical value describing a population, while a statistic is a numerical value describing a sample.
2. Qualitative data are values indicating to which non-numeric category items belong, while quantitative data are numerical values representing counts or measurements for the items.
3. Discrete data can take on only a countable number of specified values, while continuous data can take on an infinite number of values over some range.
4. If the "experiment results in data that are continuous in nature," the data must be quantitative – i.e., they must be quantifiable in some numeric manner. Even when measuring concepts like behavior and color that take on values over some continuum, the researcher must have some underlying numerical structure in order to record and report the results as continuous in nature.
5. Statistic, since the value 2.58 was obtained from a sample of households.
6. Parameter, since the value 42% was obtained from the population of all governors.
7. Parameter, since the value 706 was obtained from the population of all passengers.
8. Statistic, since the value 4.6 was obtained from a sample of Americans.
9. Discrete, since the number of letters arriving at the target must be an integer.
10. Discrete, since the number of walk buttons that do not work must be an integer.
11. Continuous, since weight can be any value on a continuum.
12. Discrete, since the percent must be equivalent to one of a countable number of specified values – in particular, one of the values 0/1059, 1/1059, 2/1059,...1059/1059.
13. Nominal, since the numbers are for identification only. Numbers on jerseys are merely numerical names for the players.
14. Ordinal, since the ratings are relative positions in a hierarchy.
15. Nominal, since the numbers are for identification only. Even though SS numbers are assigned chronologically within regions and can be placed in numerical order, there is no meaningful way to compare 208-34-3338 and 517-94-1438. If all the numbers had been assigned chronologically beginning with 000-00-0001, like the order of finishers in a race, then SS numbers would illustrate the ordinal level of measurement.
16. Ratio, since differences are meaningful and zero "yes" responses has a natural meaning.
17. Interval, since differences are meaningful but ratios are not.
18. Ratio, since differences are meaningful and zero dollars has a natural meaning.
19. Ordinal, since the ratings are relative positions in a hierarchy.
20. Interval, since differences are meaningful but ratios are not.

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

IMPORTANT NOTE for exercises 21-24: The population and sample are determined by the intent of the researcher, which must be clearly defined at the outset of any project. Unfortunately these exercises state only what the researcher did, and do not specifically identify the intent of the researcher. Consequently there may be differences of interpretation in some of the exercises, but some general principles apply. (1) The sample is a subset of the population of interest and must have the same units as the population. If the population of interest is all households, for example, then the sample must be a selection of households and not a selection of adults – as households with more adults would have a higher chance of being included in the study, thus creating a bias. (2) The problem of nonresponse must be addressed. If 500 persons are randomly selected and asked a personal question, for example, but only 400 choose to answer the question, what is the sample? Depending on the situation, the sample could be either the 500 people randomly selected (and the 100 "no response" answers be reported as part of the sample data) or the just 400 people who actually gave data in the form of a specific answer.

21. a. The sample is the 25 Senators selected at random.
    b. The population is all 100 US Senators.
    Yes. If the selections are made at random there is no bias in the design of the experiment that would cause the sample not to be representative.

22. a. The sample is the 5018 randomly selected households.
    b. The population is all US households.
    Yes. If the households were selected at random there is no reason why they would not be representative of the population.

23. a. The sample is the 1059 randomly selected adults.
    b. The population is all US adults.
    Yes. If the adults were selected at random there is no reason why they would not be representative of the population.
    NOTE: This is a sample of adults, not of homes. The appropriate inference is that about 39% of the adults live in a home with a gun, not that there is a gun in about 39% of the homes. Since adults and not homes were selected at random, homes with more adults are more likely to be represented in the survey – and that bias would prevent the survey from being representative of all homes.

24. NOTE: It is not clear whether the goal of the graduate student is to estimate (1) the proportion of all American adults who prefer snail mail over e-mail, or (2) the proportion of the adults that she knows who prefer snail mail over e-mail.
    a. The sample is the 65 adults who responded to the survey. This is a voluntary response sample, as defined in the next section.
    b. If the goal is (1) above, the population is all American adults – and her 500 target adults would not necessarily be representative of the population. If the goal is (2) above, the population is all 500 adults that the graduate students knows.
    No. The sample includes only those who cared enough about the graduate student and/or the issue in question to offer a response. Those who do not care enough to respond will not be represented.

25. Temperature ratios are not meaningful because a temperature of 0° does not represent the absence of temperature in the same sense that $0 represents the absence of money. The zero temperature in the exercise (whether Fahrenheit or Centigrade) was determined by a criterion other than "the absence of temperature."

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

26. That value has no meaning.
    NOTE: When there are only two categories coded 0 and 1, the average gives the proportion of the cases falling in the category coded 1. When there are more than two categories, such calculations with nominal data have no meaning.

27. This is an example of ordinal data. It is not interval data because differences are not meaningful – i.e., the difference between the ratings +4 and +5 does not necessarily represent the same differential in the quality of food as the difference between 0 and +1.

1-3 Critical Thinking

1. A voluntary response sample is one in which the participants choose whether or not to participate. This is generally unsuitable for statistical purposes because it is persons with strong feelings and/or a personal interest that tend to respond – and the opinions of such people are not necessarily representative of the entire population.

2. No, the fact that there is a correlation between length of study time and grade received does not necessarily imply that more study time causes higher grades. There could be a third factor, for example, either genetic or environmental, that causes people to be more conscientious about spending time in study and that also causes people to get higher grades (whether or not they study).

3. No. This is a voluntary response sample. Most of the data in a voluntary response sample comes from people with strong feelings or a personal interest in the topic, and there is usually a lack of data from "typical" people.

4. No. The wording of the question encourages people to give an immediate negative response, while their true opinion of the candidate will presumably be based on thoughtful consideration of more information.

5. People who choose to play basketball are already taller than the general population when they choose to participate in the sport. It is not the participation that makes them taller.

6. College graduates tend to earn more money than non-graduates, and people who have more money are able to purchase better health care. Having more money (whether or not it resulted from having a college degree) and not studying more is the primary contributing factor toward longer life.

7. If the population of Orange County includes significantly more minority drivers than white drivers, one would expect more speeding tickets to be issued to minorities than to whites – even if the percentage of white drivers who violated the speed limit was greater than the percentage of minority drivers who did so. It is also possible that police tend to target minority drivers – so that the numbers of tickets issued to the various racial/ethnic groups does not correspond to the actual amount of speed limit violations occurring. The fact that more speeding tickets are given to minorities does not warrant the conclusion that minority persons are more likely to speed.

8. Common colds typically run their course in two weeks or less. Persons with a cold can be expected to show improvement in two weeks even if they do nothing about the ailment.

9. The fact that the study was financed by a company with vested interests in the outcome might have influenced the conducting and/or reporting to concentrate on aspects of the study favorable to the company. The fact that a product contains one ingredient associated with

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

positive health benefits may not outweigh other negative aspects of the product that were not reported.

10. The headline suggests a level of precision that is not warranted. While the national census attempts to be complete, various types of undercounts are known to occur. In addition, the data are collected over a period of time – during which people are constantly dying and being born. Even if the data collected were 100% correct, no single figure could claim to be the exact population at a specific given moment.

11. There are at least two reasons why her conclusions should not be applied to the general population of all women. First, the target sample was biased because she sent the 100,000 questionnaires to members of various women's groups – and women who join such groups are not necessarily representative of the entire population, and certainly do not include the "non-joiner" types. Secondly, the 4500 replies she received are a voluntary response sample (and a very small 4.5% response rate at that) – which means that the responses will include an over representation of those with strong feelings and/or a personal interest in the survey.

12. There are at least two reasons why the results cannot be used to reach any conclusion about how the general population feels about keeping the United Nations in the United States. First, only viewers of the ABC "Nightline" program, which are not necessarily representative of the general population, were aware of the poll. Secondly, the 186,000 persons who responded (and there is always the possibility that some people called more than once in order to give their opinion more weight) are a voluntary response sample – which means that the responses will include an over representation of those with strong feelings and/or a personal interest in the survey.

13. Using the telephone directory as a source of survey subjects eliminates persons without telephones and persons with unlisted numbers, and it also over-includes persons who live in a house with more than one telephone listing (as some households have different numbers and/or listings for each spouse as well as the children). A more subtle problem is that telephone listings are usually by household, not by individual – which means that a person in a one-adult household has a greater chance of being included in the survey than a person in a multi-adult household.

14. The volume of the smaller box is 1 cubic inch, and the volume of the larger box is 8 cubic inches. This gives the visual impression of an eight-fold increase and does not correctly depict the relationship between the taxes ten years ago and the taxes now.

15. People who were killed in motorcycle crashes, when a helmet may have saved their lives, were not present to testify.

16. Persons will be reluctant to give their financial advisor a poor rating for fear that he will treat treat poorly and/or make their investments less of a priority in the future. At the very least it would create an awkward working relationship. The company should forward comments to the individual financial advisors without revealing which customers made which comments.

17. No, this is not likely to be a good estimate of the average of all wages earned in the United States. States with small populations but high wages (due to the cost of living) like Alaska and Hawaii, for example, contribute 2/50 = 4% of the average for the 50 states – but since they do not contain 4% of the US wage earners, they inflate the estimate to make it higher than the true average of all US wage earners. In general, the average of averages does not give the overall average. Imagine, for example, two states: one where all 4 wage earners earned $100,000 each, and one where all 96 wage earners earned $50,000 each. The true average wage for the

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

country would be [(4)(100,000) + (96)(50,000)]/100 = 5,200,000/100 = $52,000, but the average of the state averages would be [100,000 + 50,000]/2 = 150,000/2 = $75,000.

18. One problem would be in obtaining the correct value: some people may not know their exact heights, some people may think they know their heights but be mistaken, and some people may deliberately exaggerate their heights. Another problem would be in recording the correct value: some people may record 6'2" and have it misread as 62", some people who are not mathematical and/or not used to speaking in terms of inches only may mistakenly convert 6'2" into 62", and some people of other cultures who know their exact height in another system may not be familiar with inches – can you give your height in centimeters?

19. No, the results will not be representative of all households in the state. This approach violates the basic principle that the population and the sample have the same units – if the population of interest is all *households*, then you cannot survey *children*. This creates a bias because households with fewer children have a smaller chance of being included in the data, and households with no children will be missed entirely.

20. As medical knowledge and treatment improve, we expect death rates from most causes – including SIDS – to decrease. The fact that the death rate decreased while the supine position was being advised may be just a coincidence, as it might have decreased regardless of the position being advised.

21. a. 3/20 = (3/20)(100%) = 15%
    b. 56.7% = 56.7/100 = 0.567
    c. (34%)(500) = (34/100)(500) = 170
    d. 0.789 = (0.789)(100%) = 78.9%

22. a. (15%)(620) = (15/100)(620) = 93
    b. 5% = 5/100 = 0.05
    c. 0.01 = (0.01)(100%) = 1%
    d. 987/1068 = (987/1068)(100%) = 92.4%

23. a. (52%)(1038) = (52/100)(1038) = 540
    b. (52/1038)(100%) = 5%

24. a. (19/270)(100%) = 7.0%
    b. (3.0%)(270) = (3.0/100)(270) = 8

25. NOTE: The number who <u>actually committed</u> a campus crime while under the influence of alcohol or drugs cannot be determined from the survey, but the number who <u>say</u> that have committed such crimes can be determined as shown below. This distinction is an important one.
    (8%)(1875) = (8/100)(1875) = 150 say they've committed a campus crime
    (62%)(150) = (62/100)(150) = 93 say they've committed a campus crime while under the influence of alcohol or drugs

26. a. When plaque is reduced by 100%, there is none remaining. To reduce plaque by 300% would be to eliminate 3 times as much plaque as there is to begin with – which is not physically possible.
    b. When foreign investment falls 100%, there is none remaining. For foreign investment (funds flowing into a country from the outside) to fall 500%, there would be a loss of five times what the investment was to begin with – assuming that "negative foreign investment" (funds flowing from a country into the outside?) is not the concept involved, it is not possible for the amount of something that falls away to exceed the amount actually present.

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

c. When the odds of the occurrence of some event are reduced by 100%, the event becomes impossible. No further reduction is possible. The ad probably intended to say that "The Club makes car theft four times less likely."

27. If the researcher started with 20 mice in each group, and none of the mice dropped out of the experiment for any reason, then the proportions of success could only be fractions like 0/20, 1/20, 2/20, etc. – and the success rates in percents could only be multiples of 5 like 0%, 5%, 10%, etc. A success rate of 53%, for example, would not be possible.

28. The following are reasons why the report should not be taken as a good representation of public opinion.
    (1) Only people who read that particular newspaper were included in the survey. The survey did not reach the general public.
    (2) Those who responded constituted a voluntary response sample, which would not necessarily be representative of the population.
    (3) The question was worded in a biased way that would tend to elicit a negative response.
    (4) If 20 readers responded, the percent of those responding "no" would have to be a multiple of 5 – since 0/20 = 0%, 1/20 = 5%, 2/20 = 10%, etc.

1-4 Design of Experiments

1. In a random sample, every individual member has an equal chance of being selected; while in a simple random sample of size n, every possible sample of size n has an equal chance of being selected.

2. In a observational study, measurements are taken from subjects as they are and without any attempt to modify them; while in an experiment, measurements are taken from subjects at least some of which have been modified in order to assess the effects of the modification.

3. Blinding occurs when the subject (or evaluator) does not know whether the real vaccine or a placebo has been given. Double-blinding occurs when both the subject and the evaluator don't know whether the real vaccine or a placebo has been given. Blinding is important in situations when the data involve subjective responses (e.g., whether there is less pain, more mobility, etc.) that may be influenced (deliberately or subconsciously) by knowing whether the real vaccine or the placebo has been given.

4. If there are differences in the way men and women respond, blocking helps the experiment by
    (1) ensuring that there are some of each gender receiving both the real vaccine and the placebo. If by chance the men received only the real vaccine and the women received only the placebo, for example, the affect of the vaccine would be confounded with gender – and it would not be possible to tell whether any observed results were due to the vaccine or to gender.
    (2) allowing the subsequent statistical analysis to take gender into consideration and identify the specific differences in the way men and women respond.

5. Observational study, since the researcher (Emily) merely measured whether the therapist could identify the chosen hand and did nothing to modify the therapist.

6. Experiment, since the effect of an applied treatment (in this case a zero dose of the medicine) was measured.
    NOTE: There is room for disagreement – some might argue that this is an observational study, since specific characteristics are measured on unmodified subjects. If the patients were given a

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

placebo and/or were part of a larger formal study in which other persons received the proper medicine, then the given statement describes an experiment. Even if there was no such larger formal study, the spirit of the statement is that the subjects were selected and monitored for the purposes of seeing what particular effect their "treatment" had and/or comparing this effect to known results when the proper treatment was given.

7. Observational study, since specific characteristics are measured on unmodified tablets.

8. Experiment, since the effect of an applied treatment (in this case wearing magnetic bracelets) was measured.

9. Prospective, since the data are to be collected in the future from a group sharing a common factor.

10. Retrospective, since the researcher is going back in time to collect data about a characteristic of concern. The concern regards the crash, and he is going back in time to see whether those involved had consumed alcohol prior to the crash.
NOTE: The study is not retrospective because he is examining reports from the past years. Even if he were examining crashes from today only, it would be retrospective because he is going back in time (before the crash) to gather data.

11. Cross-sectional, since the data are collected for a specific point in time.

12. Retrospective, since the researcher is going back in time to collect data about a characteristic of concern. The concern regards present income, and he is going back in time to see whether persons had taken a course in statistics.

13. Systematic, since every 5^{th} driver was stopped.

14. Cluster, since all the voters at selected polling stations were interviewed.

15. Stratified, since the population of interest was divided into activity subgroups from which the actual sampling was done.

16. Convenience, since the sample is those who happen to be in his family.

17. Cluster, since all the wait staff at selected restaurants were surveyed.

18. Stratified, since the population of interest will be divided into age groups from which the actual sampling will be done.

19. Convenience, since the sample is those who happen to be in the author's classes.

20. Systematic, since every 100^{th} name is selected.

21. Random, since each adult has an equal chance of being selected.
NOTE: This is a complex situation. The above answer ignores the fact that there is not a 1 to 1 correspondence between adults and phone numbers. An adult with more than one number will have a higher chance of being selected. An adult who shares a phone number with other adults will have a lower chance of being selected. And, of course, adults with no phone numbers have no chance of being selected at all.

22. Stratified, since the population of interest was divided into employment subgroups from which the actual sampling was done.

23. Cluster, since all the students in selected classes were interviewed.
NOTE: Ideally the division of the population should place each subject in one and only one cluster, and every subject should have an equal chance to be included in the sample. In this

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

scenario there are at least two problems: (1) students carrying more classes will have a greater chance of being included in the survey and (2) a student may be included in more than one cluster. While some techniques may eliminate those two problems, they could introduce others – for example, selecting the 10 classes from among all 1 pm MWF classes, means that a student not taking a class then has zero chance of being included in the survey.

24. Random, since each subject has an equal chance of being selected. This is also a simple random sample, since every possible group of 12 subjects has an equal chance of being selected.

25. Yes, this results in a random sample because each return has an equal chance of being included in the sample.
No, this does not result in a simple random sample of size 48 because some groupings of 48 returns are not possible – for example, a grouping with 24 Monday returns and 24 Tuesday returns would not be possible.

26. No, this does not result in a random because each student does not have an equal chance of being included in the sample. Students who typically arrive last because of a previous class, for example, are less likely to be included.
No, this does not result in a simple random sample of size 10 because not all possible groupings of size 10 have an equal chance to occur. Groupings involving students who typically arrive last because of a previous class, for example, are less likely to occur.

27. Whether the sample is a random sample depends on how the first selection is made. If the engineer chooses the first one at random from 1 to 10,000 and every $10{,}000^{th}$ one thereafter, then every M&M has an equal chance of being selected (namely 1 in 10,000) and the sample is a random sample. If the engineer determines to start with #1 and choose every $10{,}000^{th}$ one thereafter, then some M&M's have no chance of being selected (e.g., #2) and the sample is not a random sample.
No, no matter how the first selection is made the sample will not be a simple random sample of size n. All possible groupings of size n are not possible – any grouping containing #1 and #2, for example, could not occur.

28. Yes, this results in a random sample because every driver whether male or female has an equal chance of being selected.
No, this does not result in a simple random sample of size 40 because not all possible groupings of size 40 have an equal chance to occur. Groupings with 21 men and 19 women, for example, cannot occur.

29. Yes, this results in a random sample because every student has an equal chance of being selected.
Yes, this results in a simple random sample of size 6 because all possible groupings of size 6 have an equal chance to occur.

30. Yes, this results in a random sample because every pill has an equal chance of being selected.
Yes, this results in a simple random sample of size 10 because all possible groupings of size 10 have an equal chance to occur.

31. Assume that "students who use this book" refers to a particular semester, or other point in time. Before attempting any sampling procedure below, two preliminary steps are necessary: (1) obtain from the publisher a list of each school that uses the book and (2) obtain from each school the number of students enrolled in courses requiring the text. Note that using only the number of books sold to the school by the publisher for the semester of will miss those

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

students purchasing used copies turned in from previous semesters, buying textbooks on the Internet, etc. "Students" in parts (a)-(e) below is understood to mean students using this book.

a. Random: Conceptually (i.e., without actually doing so) make a list of all N students by assigning to each school one place on the list for each of its students. Pick 100 random numbers from 1 to N. Each number identifies a particular school and a conceptual student. For each particular school identified select a student at random (or more than one student, if that school was identified more than once). In this manner every student has the same chance of being selected. In fact every possible grouping of 100 students has the same chance of being selected, and the sample is also a simple random sample.
b. Systematic. Place the n schools in a list. Pick a random number between 1 and (n/10) to determine a starting selection. Select that school and every $(n/10)^{th}$ school thereafter. At each of the 10 schools thus systematically selected, randomly (or systematically) select and survey 10 students.
c. Convenience. Randomly (or conveniently) select and survey 100 students from your school.
d. Stratified. Divide the schools into four categories: 2 year, 4 year public, 4 year private, proprietary. Randomly select one school from each category. For each school randomly (or in some stratified manner) select and survey the number of students equal to the percent of the N students that are in that category.
e. Cluster. Select a school at random and survey all the students, or cluster sample again by selecting a class within that school and surveying all the students in that class. Repeat as necessary until the desired sample of 100 is reached.

32. Confounding occurs when the researcher is unable to determine which factor (typically one planned and one unplanned) produced an observed effect. If a restaurant tries adding a buffet to its evening menu for one week and it happens to be the week that a nearby theater shows a real blockbuster that attracts extra crowds to the area, the restaurant cannot know whether any increased business was due to the new buffet or to the extra traffic created by the theater.

33. The study may be described as
    a. prospective because the data are to be collected in the future, after the patients are assigned to different treatment groups.
    b. randomized because the patients were assigned to one of the three treatment groups at random.
    c. double-blind because neither the patient nor the evaluating physician knew which treatment had been administered.
    d. placebo-controlled because there was a treatment with no drug that allowed the researchers to distinguish between real effects of the drugs in question and (1) the psychological effects of being "treated" and/or (2) the effects (e.g., natural improvement over time) of not receiving any drug.

1-5 **Introduction to Excel**

1. A spreadsheet is a collection of data organized into rows and columns.
2. In Excel, a spreadsheet is called a worksheet.
3. An Excel workbook is an Excel file containing one or more worksheets.
4. An Excel worksheet is a single spreadsheet.

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

5. a. Items too big for the allotted cells are accepted, but partially hidden behind adjacent occupied cells to the right. Alpha-numeric items are automatically left justified within each cell, and numerical values are automatically right justified within each cell.
   b. Click on the **save** icon (diskette) on the tool bar and follow the path to the correct disk drive.
   c. Click on the **print** icon (printer) on the tool bar. The printed worksheet is produced in the same format it appeared on the screen – i.e., with the same amount of material showing and the same within-cell justification, but the cell lines do not appear.
   d. Click on the lower **X** in the upper right hand corner.
   e. Click on the **open** icon (opening file folder) on the tool bar.
   f. When the **delete** key is pressed the highlighted cells become empty – i.e., the items that occupied the cells have been deleted. When the **undo** arrow is clicked, the items reappear.
6. same as #5
7. same as #5
8. same as #5
9. There will be 8 columns, with 1 title row followed by 36 rows of data.
10. There will be 6 columns, with 1 title row followed by from 8 to 27 rows of data.
11. There will be 5 columns, with 1 title row followed by
    15 row labels in column 1.
    15 row of data in columns 2-4.
    10 rows of data in column 5.
12. There will be 9 columns, with 1 title row followed by 54 rows of data.
13. Clicking on the **help** cue on the top line activates the interactive help icon that asks for a question. Typing the word **save** leads to a series of verbal prompts preceded by a question mark. Clicking on **save a file** and then **save a copy of a file** leads to the desired information – viz., click on the **file** cue on the top line, then click on the **save as** option and follow the directions. NOTE: There is more than one possible technique, and the exact procedure and cues depend upon the particular version of Excel being used.
14. a. There will be 9 columns, with 1 title row and 40 rows of data.
    b. Several choices appear, in alphabetical order from Anova to z-Test.
    c. Several input and output options appear.
    d. Entering a2:a41 for the input range produces, among others, the following values:
       mean = 342672.5. The average selling price is $342,672.50
       range = 310000. The difference between the highest and lowest selling prices is $310,000.
       minimum = 210000. The lowest selling price is $210,000.
       maximum = 520000. The highest selling price is $520,000.
       sum = 13706900. The total of all the selling prices is $13,706,900.
       count = 40. There were 40 selling prices in the data set.

Statistical Literacy and Critical Thinking

1. A large sample is not necessarily a good sample. As in many other situations, quality is more important than quantity. More important than the size of the sample is whether it was selected in such as way as to be representative of the population of interest.

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

2. a. Height data are quantitative because height is an actual numerical measurement and not merely a categorization.
   b. Height data are continuous because they can take on any value over a specified range.
   c. The population is all runners who finish the New York Marathon.

3. This sample would not necessarily be representative of the general population. It is a voluntary response sample, which typically includes only those with certain motivation or characteristics – in this case that would probably be people who either (1) think they have a high IQ and/or enjoy such tests or (2) need the $50.

4. This would not be a good estimate for at least two reasons. First, she appears to be depending on values provided by the owners. An owner might not know a car's true value or might tend to exaggerate its value. Secondly, she is counting each state equally. Alaska has very few cars, but they probably include many relatively expensive (even considering that state's overall high cost of items) four-wheel drive utility vehicles. California has thousands of times as many cars as Alaska, and they would probably tend to be less expensive smaller vehicles. The average for 10 cars from each state would likely not be a good estimate for the average of all the cars in the nation.

Review Exercises

1. The results cannot be considered representative of the population of the United States for at least two reasons. First, only AOL Internet subscribers were polled – and they are not necessarily representative racially, socio-economically, geographically, etc. Secondly, the responders constitute a voluntary response sample – and such responders tend to be those with strong opinions and/or a personal stake in the issue.

2. Let N be the total number of full-time students and n be the desired sample size.
   a. Random. Obtain a list of all N full-time students, and number the students from 1 to N. Select n random numbers from 1 to N, and survey each student whose number on the list is one of the random numbers selected.
   b. Systematic. Obtain a list of all N full-time students, and number the students from 1 to N. Let m be the largest integer less than the fraction N/n, select a random number between 1 and m, and survey that student and every m^{th} student thereafter.
   c. Convenience. Select a location by which most of the students usually pass and survey the first n full-time students that pass by.
   d. Stratified. Obtain a list of all N full-time students and the gender of each. Divide the list by gender and randomly survey n/2 students of each gender.
   e. Cluster. Obtain a list of all the classes meeting at a popular time, select at random one of the classes (or as many of the classes as needed to obtain a sample of n students), and survey all the students in that class.

3. a. Ratio, since differences are valid and there is a meaningful 0.
   b. Ordinal, since there is a hierarchy but differences are not valid.
   c. Nominal, since the categories have no meaningful inherent order.
   d. Interval, since differences are valid but the 0 is arbitrary.

4. a. Discrete, since the number of shares held must be an integer.
   NOTE: Even if partial shares are allowed (e.g., 5½ shares), the number of shares must be some fractional value and cannot be any value on a continuum – e.g., a person cannot own π shares..

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

b. Ratio, since differences between values are valid and there is a meaningful 0.
c. Stratified, since the population of interest (all stockholders) was divided into subpopulations (by states) from which the actual sampling was done.
d. Statistic, since it is calculated from a sample and not from the entire population.
e. There is no unique correct answer, but the following are reasonable possibilities. (1) The proportion of stockholders holding above that certain number of shares that would make them "influential." (2) The proportion of stockholders holding below that certain number of shares that would make them "insignificant." (3) The numbers of shares held by the largest stockholders.
f. There is no unique correct answer, but the following are reasonable possibilities. (1) The results would be from a voluntary response sample, and hence not necessarily be representative of all stockholders. (2) If the questionnaire did not include information on the number of shares owned, the views of small stockholders could not be distinguished from the views of large stockholders (whose views should carry more weight).

5. NOTE: The second part of the problem involves deciding "whether the sampling scheme is likely to result in a sample that is representative of the population." This is subjective, since the term "representative" has not been well-defined. Mathematically, any random sample is representative in the sense that there is no bias and a series of such samples can be expected to average out to the true population values – even though any one particular sample may not be representative. And so we expect a random sample of size n=50 to be representative. But how about a random sample of size n=2? It has all the mathematical properties of a random sample of size n=50, but common sense suggests that any one sample of size n=2 is not necessarily likely to be representative of the population.
   a. Systematic, since the selections were made at regular intervals. Yes, there is no reason why every 500^{th} stockholder should have some characteristic that would introduce a bias.
   b. Convenience, since those selected were the ones who happened to attend. No, those who attend would tend to be the more interested and/or more well-off stockholders.
   c. Cluster, since the stockholders were organized into groups (by stockbroker) and all the stockholders in the selected groups were surveyed. Yes, since every stockholder has the same chance of being included in the sample (assuming each stockholder works through a single stockbroker), the sample will be a random sample – and considering the large number of stockbrokers involved (see the NOTE at the beginning of the problem), it is reasonable to expect the sample to be representative of the population.
   d. Random, since each stockholder has the same chance of being selected. Yes, since every stockholder has the same chance of being included in the sample, the scheme is likely to result in a representative sample.
   e. Stratified, since the stockholders were divided into subpopulations (zip codes) from which the actual sampling was done. No, since all the zip codes were given equal weights (5 stockholders from each) – because "significant" zip codes (with large numbers of large stockholders) are counted equal with "insignificant" zip codes (with few numbers of small stockholders), the sample will probably be biased in favor of small stockholders.

6. a. Blinding occurs when those involved in an experiment (either as subjects or evaluators) do not know whether they are dealing with a treatment or a placebo. It might be used in this experiment by (1) not telling the subjects whether they are receiving Sleepeze or the placebo and/or (2) not telling any post-experiment interviewers or evaluators which subjects received Sleepeze and which ones received the placebo. Double-blinding

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

occurs when neither the subjects nor the evaluators know whether they are dealing with a treatment or a placebo.

b. The data reported will probably involve subjective assessments (e.g., "on a scale of 1 to 10, how well did it work?") that may be influenced by whether the subject was known to have received Sleepeze or the placebo.

c. In a completely randomized design, subjects are assigned to the treatments (in this case Sleepeze or the placebo) at random.

d. In a rigorously controlled design, the subjects are carefully placed into treatment groups (in this case Sleepeze or the placebo) according to characteristics that might affect the outcome of the experiment. For example: if the Sleepeze group contains an athletic male under 20, then the placebo group should contain an athletic male under 20 – since gender, physical condition, and age are all characteristics that might reasonably affect the outcome of the experiment. This makes the treatment groups as similar as possible, so that any observed differences are likely to be due to the treatments and not to extraneous factors that might affect the experiment.

e. Replication involves performing the experiment on a sample of subjects large enough to ensure (1) that atypical responses of a few subjects will not give a distorted view of the true situation and (2) that repeating the experiment would likely produce similar results.

7. a. The figure 39.82% is a parameter, since it refers to the entire population.
   b. Discrete, since the numbers of votes cast must be integers.
   c. (39.82%)(1,865,908) = (39.82/100)(1,865,908) = 743,005

8. a. To contain 100% less fat would be to contain no fat at all – 100% of the fat has been removed. It is not physically possible to contain 125% less fat.
   b. (19%)(237) = (19/100)(237) = 45
   c. (30/186)(100%) = 16.1%

Cumulative Review Exercises

NOTE: Throughout the text intermediate mathematical steps will be shown to aid those who may be having difficulty with the calculations. In practice, most of the work can be done continuously on calculators or using Excel, and the intermediate values are unnecessary. Even when the calculations are not done continuously, DO NOT WRITE AN INTERMEDIATE VALUE ON YOUR PAPER AND THEN RE-ENTER IT. That practice can introduce round-off and copying errors. Store any intermediate values in the calculator memory or an Excel data cell. In general, the degree of accuracy appropriate for the final answer depends upon the particular problem – and while guidelines for this will be given as needed in subsequent chapters, three decimal accuracy is a good rule of thumb. In exercises #1-8, the one-step Excel format and answer is given following the usual algebraic format.

1. $$\frac{3.0630+3.0487+2.9149+3.1358+2.9753}{5}=\frac{15.1377}{5}=3.02754 \text{ grams}$$

   (3.0630 + 3.0487 + 2.1949 + 3.1358 + 2.9753)/5 = 3.02754 grams

2. $$\frac{98.20-98.60}{0.62}=\frac{-0.40}{0.62}=-0.645$$

   (98.20 – 98.60)/0.62 = –0.645161

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley

3. $\dfrac{98.20-98.60}{\dfrac{0.62}{\sqrt{106}}} = \dfrac{-0.40}{0.0602} = -6.642$

(98.20-98.60)/(0.62/SQRT(106)) = –6.642342

4. $\left[\dfrac{1.96 \cdot 0.25}{0.03}\right]^2 = [16.333]^2 = 266.778$

(1.96*0.25/0.03)^2 = 266.7778

5. $\dfrac{(50-45)^2}{45} = \dfrac{(5)^2}{45} = \dfrac{25}{45} = 0.556$

((50-45)^2)/45 = 0.555556

6. $\dfrac{(2-4)^2+(3-4)^2+(7-4)^2}{3-1} = \dfrac{(-2)^2+(-1)^2+(3)^2}{2} = \dfrac{4+1+9}{2} = \dfrac{14}{2} = 7$

((2-4)^2+(3-4)^4+(7-4)^2)/(3-1) = 7

7. $\sqrt{\dfrac{(2-4)^2+(3-4)^2+(7-4)^2}{3-1}} = \sqrt{\dfrac{(-2)^2+(-1)^2+(3)^2}{2}} = \sqrt{\dfrac{4+1+9}{2}} = \sqrt{\dfrac{14}{2}} = \sqrt{7} = 2.646$

SQRT(((2-4)^2+(3-4)^2+(7-4)^2)/(3-1)) = 2.645751

8. $\dfrac{8(151,879)-(516.5)(2176)}{\sqrt{8(34,525.75)-516.5^2}\sqrt{8(728,520)-2176^2}} = \dfrac{91128}{\sqrt{9433.75}\sqrt{1093184}} = 0.89735$

(8*151879-516.5*2176)/(SQRT(8*34525.75-516.5^2)*SQRT(8*728520-2176^2)) = 0.897352

9. $0.5^{10} = 0.5\text{^}10 = 0.000976563$

10. $2^{40} = 2\text{^}40 = 1.09951\text{E+}12 = 1,099,510,000,000$

NOTE: The final 7 digits are not really zero, but Excel does not provide all 13 digits of the exact answer.

11. $7^{12} = 7\text{^}12 = 13,841,287,201$

12. $0.8^{50} = 0.8\text{^}50 = 1.42725\text{E–}05 = 0.0000142725$

Copyright © 2012 Pearson Education, Inc. Publishing as Addison-Wesley