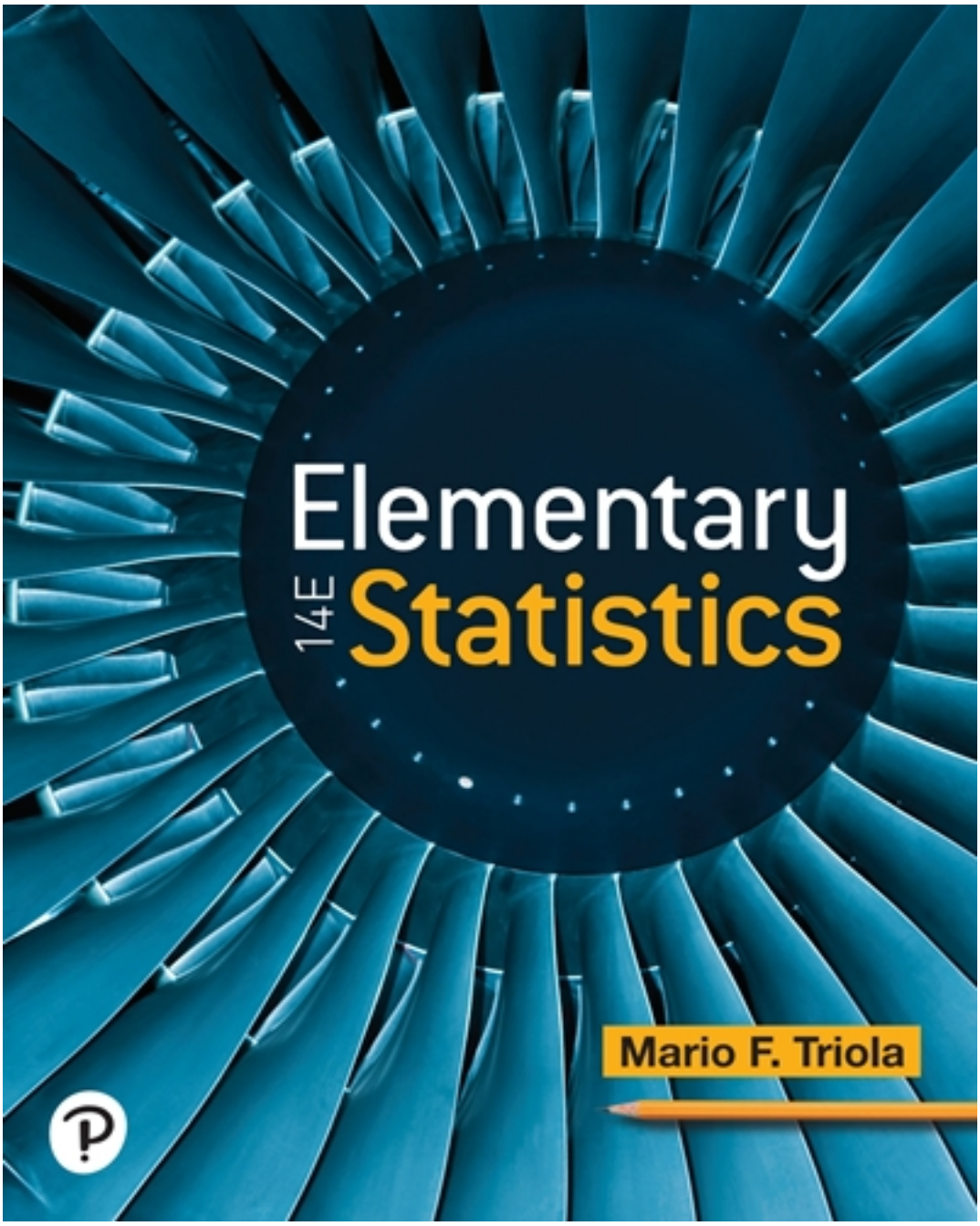
Solutions for Elementary Statistics 14th Edition by Triola

CLICK HERE TO ACCESS COMPLETE Solutions



Solutions

Insider's Guide

to Teaching with the Triola Statistics Series

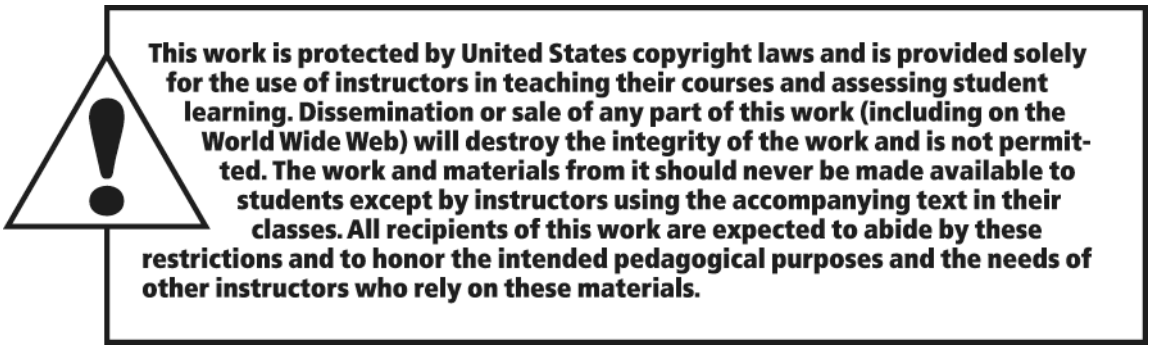
Triola Statistics Series:

Elementary Statistics, 14th Edition
Elementary Statistics Using Excel, 7th Edition
Essentials of Statistics, 7th Edition

Support Manual for Adjuncts and Full-Time Professors

Mario F. Triola

Pearson



ISBN-13: 978-0-13-733326-4
ISBN-10: 0-13-733326-9

Contents

(Click Page Number to View)

I. Introduction **2**

II. How Should Statistics Be Taught? **3**

III. Selecting a Technology **9**

IV. Getting Started with Statdisk **13**

V. Chapter and Section Comments, Activities, Surveys, and Examples

Chapter 1 17
Chapter 2 22
Chapter 3 26
Chapter 4 31
Chapter 5 37
Chapter 6 41
Chapter 7 45
Chapter 8 49
Chapter 9 55
Chapter 10 61
Chapter 11 66
Chapter 12 69
Chapter 13 72
Chapter 14 78
Chapter 15 81

VI. Supplements, Videos, and Technology Resources **82**

VII. General Teaching Tips **84**

VIII. Suggested Course Syllabi **92**

IX. Transition Guide from 13th Edition to 14th Edition **98**

Introduction

Dear Professor:

The Pearson team is delighted that you are using a book in the Triola Statistics Series. The author has personally visited many professors across the United States, and he is happy to see a wide range of extremely effective statistics courses. Some professors have many years of experience in teaching statistics, and they continue to improve their courses each year. Other professors are relatively new to teaching statistics. Some professors are adjuncts who do not teach on a full-time basis, and their contact with other teachers is minimal. In some cases, large sections are taught with teaching assistants who might be graduate students with little or no teaching experience. This *Insider's Guide to Teaching Statistics* is intended to provide some insight into effective teaching for such groups.

Please know that the recommendations and suggestions in this guide are not being made in an authoritarian spirit of "you *must* do it *this* way." The content of this guide should be tempered with personal teaching styles, individual course objectives, and individual student needs and goals.

In addition to teaching students fundamental concepts of statistics, the introductory statistics course is an excellent vehicle for doing much more than teaching course content. The introductory statistics course provides us with an ideal opportunity to foster student growth in these important areas:

Critical thinking

Technology usage

Public speaking

Working cooperatively in groups

This guide includes recommendations that are helpful in encouraging student growth in these important areas.

We hope that this guide is helpful, and any suggestions for improvement are most welcome. We sincerely wish you the best of luck!

II. How Should Statistics Be Taught?

One of the most important points to be made in this *Insider's Guide* is the basic approach to teaching the introductory statistics course. Here are some important principles:

1. The introductory statistics course should be taught in a way that is fundamentally different from the approach used in traditional mathematics courses. Arithmetic computations or algebraic manipulations are not nearly as important as the ability to *understand* results and to be able to *interpret* results in a meaningful way.

2. The introductory statistics course focuses on *real applications* instead of abstractions.

3. Textbooks in the Triola Statistics Series are full of real data. Examples, exercises, and test questions should involve students with real data as much as possible. Fabricated data have little use in the introductory statistics course.

4. There should not be a high priority placed on covering as many different topics as possible. It is much better to cover fewer topics well than to cover many topics poorly.

The following pages identify the GAISE recommendations. The author comments about the GAISE recommendations are designed to clarify the above points.

GAISE Recommendations

GAISE is an acronym for "Guidelines for Assessment and Instruction in Statistics Education." These guidelines are recommendations from a project sponsored by the American Statistical Association (ASA). Here are six GAISE recommendations for the teaching of introductory statistics:

1. **Teach statistical thinking.**
   - **Teach statistics as an investigative process of problem-solving and decision-making.**
   - **Give students experience with multivariable thinking.**

2. **Focus on conceptual understanding.**

3. **Integrate real data with context and purpose.**

4. **Foster active learning.**

5. **Use technology to explore concepts and analyze data.**

6. **Use assessments to improve and evaluate student learning.**

The author enthusiastically supports these recommendations, and much of the content of this *Insider's Guide* is devoted to implementation of these recommendations. Here are some comments about the six recommendations.

1. Teach statistical thinking.

The importance of sound sampling techniques should be introduced early and often throughout the introductory statistics course. Part of "literacy" is understanding the meaning of terms such as *simple random sample* and *voluntary response sample*. Statistical thinking is used when a student recognizes that results obtained from a poorly selected sample might be results without any real validity. For example, newspapers, magazines, television shows, and Internet sites often conduct surveys by asking people to respond to some question. However, the responses constitute a voluntary response sample, and students should know that any conclusions based on such a sample do not apply to the larger population. This is one simple example of the type of critical thinking that should be fostered throughout the course.

In teaching the introductory statistics course, it is not important to memorize formulas or the detailed mechanics of statistical methods. It is not important to be able to reproduce the formula for the standard deviation s, and it is not so important to be able to do the arithmetic required for manually computing values of standard deviations. Instead, it is important to *understand* what the standard deviation s

measures. On a very basic level, it is important for students to know quite well that the standard deviation is a measure of *variation*. It is *really* important that students develop an ability to *understand* and *interpret* values of the standard deviation *s.* The empirical rule and Chebyshev's theorem are commonly presented as tools that help students understand and interpret variation, but the author recommends skipping those two topics and focusing instead on the *range rule of thumb* presented in the book. It is easy to apply, and students generally understand it quite well, so it becomes a very effective tool that can help students understand and interpret values of standard deviations. This topic will be discussed further when measures of variation are discussed later in this guide. But this topic is excellent for making the point that we should emphasize statistical literacy and develop statistical thinking.

When teaching an introductory calculus course, the author might give a test question that asks students to write the definition of the derivative of a function $f(x)$, and he might ask students to compute the derivative of $f(x) = x^2$ while showing all of the steps involved. Calculus students should know the definition of the derivative and they should be able to apply it. However, the author would never ask statistics students to write the formula for the standard deviation or to calculate the standard deviation of a list of values while showing all work. Instead, the author encourages questions that test *understanding*. Here are examples of good and bad test questions:

Bad test question: Write the formulas for the mean and standard deviation, then compute the mean and standard deviation of the values 23.7, 11.2, 43.5, 77.2, 49.0, 27.3, and show all work.

Good test question: Listed below are weights (in grams) of newly minted quarters. (a) Find the mean. (b) Find the standard deviation. (c) In the context of the given weights, is a weight of 5.23 g *significantly low*? Explain your choice. (d) What is an adverse consequence of minting quarters with weights that vary too much?

5.71 5.71 5.59 5.61 5.63

When students find the mean $\bar{x}$ = 5.650 g and standard deviation of s = 0.057 g, they should be encouraged to use some technology, such as a TI-83 Plus or TI-84 Plus calculator. There is little to be gained by requiring that such statistics be calculated manually. A good answer to part (c) of the preceding question is the statement that yes, a weight of 5.23 g would be significantly low because it is more than two standard deviations below the mean. One of several good answers to part (d) would be a statement that if weights of minted quarters vary too much, vending machines will reject too many valid coins. Part (d) is designed to emphasize the point that methods of statistics have real, important, and meaningful applications instead of being abstract concepts that might not have any real applications.

2. Focus on conceptual understanding.

A good illustration of this point can be seen in the data from eruptions of the Old Faithful geyser (top table) and data from actual low temperatures and forecast low temperatures (bottom table):

Duration (sec)	240	120	178	234	235	269	255	220
Interval After (min)	92	65	72	94	83	94	101	87

Actual low temp (°F)	54	54	55	60	64
Low temp forecast five days earlier	56	57	59	56	64

When discussing correlation/regression, we might present the top table and ask if there is a correlation between the duration of an eruption and the time interval after the eruption to the next eruption. When discussing matched data, we might present the bottom table given above, and we might ask if the differences between the actual and forecast temperatures are from a population with a mean of 0. But instead of focusing too much on the details of the computations involved, we should stress the fundamental difference between the two sets of data summarized in the preceding tables. Students should learn how to ask the best questions. Given the top Old Faithful table, students should see that the issue is one of a *relationship* between the two variables. Given the bottom temperature table, students should see that a key element is the list of *differences* between the actual and forecast temperatures, and a mean difference equal to zero is evidence that the forecast temperatures are accurate. It's not the structure of paired data that determines the method that is most appropriate; it is the *context* of the data.

3. Use real data with a context and purpose.

George Cobb is a leader in statistics education. He wrote an article about evaluating introductory statistics textbooks (see "Introductory Textbooks: A Framework for evaluation", *Journal of the American Statistical Association*, Vol. 82, No. 397) and he included the following statement:

> ***"Are the Data Sets Real or Fake? Not that many years ago, all it took was this first question to dispatch most books to the morgue. Fortunately, that is changing. It is true that there are still books on the market whose examples have been bled white of vital detail, but it is now easier to shun them. I hope that soon we will have seen the last of the infamous XYZ Corporation and Hospitals A, B, C, ..."***

In the 14th edition of *Elementary Statistics,* 94% of the examples involve real data, and 93% of the exercises involve real data. With real data, students see how statistical concepts have meaningful applications. It is very likely that they will encounter data from the discipline that they might be considering as a major.

4. Foster active learning.

The following is so true when considered in the context of teaching an introductory statistics course:

Tell me something, and I will forget.
Show me and I will remember.
Involve me, and I will learn.

If you want your students to have a learning experience that will affect them for their entire lives, *involve* them with active learning. The Triola Statistics Series textbooks, including *Elementary Statistics,* 14th Edition, have end-of-chapter Cooperative Group Activities.

Some statistics professors believe that the entire course should be based on activities, and some other statistics professors do not include any activities at all. Somewhere between these extremes is a balance that allows active involvement along with enough time for teaching concepts using traditional methods.

Recommendation: If you do no activities at all, begin with just one or two activities to see how well they work. Then, assuming that all goes well, include more activities in future courses.

5. Use technology to explore concepts and analyze data.

Many statistics professors teach an effective course by allowing students to use any one of a variety of different scientific calculators. The author recommends that a specific technology be used. Triola statistics books include displays from Statdisk, Minitab, Excel, the TI-83/84 Plus calculator, and StatCrunch. There are also supplements for SPSS and SAS.

The author's personal preference is to require that Statdisk and a second different technology be used. Each student should do several software projects using a technology. However, choosing a technology to be used for an introductory statistics course is a complex decision that must take several factors into account. Some colleges have adopted a decision to use Excel because so many students use Excel in their work after graduation. Some colleges avoid Excel because its statistics functions are not as good as they should be. Some colleges use Minitab, and the latest release includes features that make it a perfectly good choice. Some statistics professors prefer to require TI-83/84 Plus calculators because they can do so much statistical number crunching and they can be used in class and on tests. Some statistics professors would like to require TI-83/84 Plus calculators, but are reluctant to do so because of the calculator's cost. The author had that same concern the first time that he required those calculators, so he announced that any student could sell him their calculator at the end of the course. At the end of that semester, *no* students wanted to give up their calculators. Their desire to keep their calculators instead of turning them in for cash was a strong indication about how they perceived the usefulness of those calculators.

Statdisk Statdisk (www.Statdisk.com) is a free and easy-to-use online statistics package designed specifically for the Triola Statistics Series textbooks. Millions of students have used Statdisk, so it is among the most widely used statistics programs. The latest version of Statdisk is one that the author is proud to have as a major and important supplement. Because Statdisk can do almost all of the functions described in the textbook, it can be used as the technology in the introductory statistics course. If another technology, such as Excel or SPSS, is used as the primary technology, it would be really helpful to have students use Statdisk as a supplement to the main technology being used. By getting results from Statdisk along with results from another technology, students are more likely to confirm that their results are correct. Access Statdisk at www.Statdisk.com.

Technology for New Approaches While the technology can do the statistics number crunching, it should also be used to explore concepts and new approaches. When considering the effects of an outlier, for example, a hypothesis test could be conducted with the outlier included and again with the outlier excluded. Probability can be better studied with simulations. The resampling methods of bootstrapping and randomization can sometimes be used when traditional methods should not be used. For ideas about how to include technology, see the Technology Project at the end of each chapter in the Triola textbook

6. Use assessments to improve and evaluate student learning.

Traditional tests and quizzes are one important method of assessment, but there are others. The author favors the use of activities and at least one major project. The author favors a capstone group project conducted near the end of the course. Students can work together in groups of four (more or less), and each group should conduct a project that involves the planning of an experiment or a method for collecting data in an observational study. After collecting original data, the group will make an inference by using the methods learned in the course.

A group presentation should involve each member speaking for at least one or two minutes. Software results should also be included, along with a brief written report. Assessment is an important component of such a project. How do you assess the work of individual members that participate in a group project? Here is one method that the author found to be effective: Survey each group member and ask him or her to assess the work done by the other group members. For example, ask each group member to submit a separate form for each of the other group members, and that form should include an assessment of the other team members' work, such as "was a major contributor to the project," "did an average amount of work on the project," "did some but little work on the project," or "did not participate in any meaningful way." Students are quite honest about the work of their peers, and they are generally quite satisfied with this process of assessment.

The author favors four or five tests given during the semester, along with a comprehensive final examination. Activities and projects should also be part of the assessment plan.

Chapter 1: Introduction to Statistics

Section 1-1: Statistical and Critical Thinking

1. The respondents are a voluntary response sample or a self-selected sample. Because those with strong interests in the topic are more likely to respond, it is very possible that their responses do not reflect the opinions or behavior of the general population.
2. a. The sample consists of the 1046 adults who were surveyed. The population consists of all adults.
   b. When asked, respondents might be inclined to avoid the shame of the unhealthy habit of not washing their hands, so the reported rate of 70% might well be much higher than it is in reality. It is generally better to observe or measure human behavior than to ask subjects about it.
3. Statistical significance is indicated when methods of statistics are used to reach a conclusion that a treatment is effective, but common sense might suggest that the treatment does not make enough of a difference to justify its use or to be practical. Yes, it is possible for a study to have statistical significance, but not practical significance.
4. No. Correlation does not imply causation. The example illustrates a correlation that is clearly not the result of any interaction or cause effect relationship between per capita consumption of margarine and the divorce rate in Maine.
5. Yes, there does appear to be a potential to create a bias.
6. No, there does not appear to be a potential to create a bias.
7. No, there does not appear to be a potential to create a bias.
8. Yes, there does appear to be a potential to create a bias.
9. The sample is a voluntary response sample and has strong potential to be flawed.
10. The samples are voluntary response samples and have potential for being flawed, but this approach might be necessary due to ethical considerations involved in randomly selecting subjects and somehow imposing treatments on them.
11. The sampling method appears to be sound.
12. The sampling method appears to be sound.
13. The Ornish weight loss program has statistical significance, because the results are so unlikely (3 chances in 1000) to occur by chance. It does not have practical significance because the amount of lost weight (3.3 lb) is so small.
14. Because there is only one chance in a thousand of getting such success rates by chance, the difference does appear to have statistical significance. The 92% success rate for surgery appears to be substantially better than the 72% success rate for splints, so the difference does appear to have practical significance.
15. The difference between Mendel's 25% rate and the result of 26% is not statistically significant. According to Mendel's theory, 145 of the 580 peas would have yellow pods, but the results consisted of 152 peas with yellow pods. The difference of 7 peas with yellow pods among the 580 offspring does not appear to be statistically significant. The difference does not appear to have practical significance.
16. Because there is a 25% chance of getting such results with a program that has no effect, the program does not appear to have statistical significance. Because the average increase is only 3 IQ points, the program does not appear to have practical significance.
17. With 40 out of 41 ballots having the Democrat first, it appears that the result is statistically significant. Because of the great advantage enjoyed by Democrats, the results also have practical significance.
18. Because it is so unlikely (0.3%) to get these results by chance, the results have statistical significance. With about 57% (from 235/414) of the coin toss winners going on to win the game, the result appears to have practical significance.
19. There appears to be statistical significance given the large discrepancy between 79.1% and 39%. Because the results are so far from yielding a jury of peers, it appears that the results have practical significance.
20. With only a 0.0000006% chance of getting such results, it appears that the results are statistically significant. The discrepancy between the 61% rate for voters who actually did vote and the 70% rate of those who said that they voted is a fairly large discrepancy, and the results appear to have practical significance.

21. Yes. Each column of 8 AM and 12 AM temperatures is recorded from the same subject, so each pair is matched.
22. No. The source is from university researchers who do not appear to gain from distorting the data.
23. The data can be used to address the issue of whether there is a correlation between body temperatures at 8 AM and at 12 AM. Also, the data can be used to determine whether there are differences between body temperatures at 8 AM and at 12 AM.
24. Because the differences could easily occur by chance (with a 64% chance), the differences do not appear to have statistical significance.
25. No. The lemon imports are weights in metric tons and the crash fatality rates are fatalities per 100,000 population, so their differences are meaningless.
26. The issue that can be addressed is whether there is a correlation, or association, between lemon imports and crash fatality rates.
27. No. The author of an article for the *Journal of Chemical Information and Modeling* has no reason to collect or present the data in a way that is biased.
28. No. Correlation does not imply causation, so a statistical correlation between lemon imports and crash fatality rates should not be used to conclude that lemon imports are the cause of fatal crashes.
29. It is questionable that the sponsor is the Idaho Potato Commission and the favorite vegetable is potatoes.
30. The sample is a voluntary response sample, so there is a good chance that the results do not reflect the larger population of people who have a water preference.
31. The correlation, or association, between two variables does not mean that one of the variables is the cause of the other. Correlation does not imply causation. Clearly, sour cream consumption is not directly related in any way to motorcycle fatalities.
32. The sponsor of the poll is an electronic cigarette maker, so the sponsor does have an interest in the poll results. The source is questionable.
33. The correlation, or association, between two variables does not mean that one of the variables is the cause of the other. Correlation does not imply causation.
34. The correlation, or association, between two variables does not mean that one of the variables is the cause of the other. Correlation does not imply causation.
35. The sample is a voluntary response sample, so there is a good chance that the results do not accurately reflect the larger population.
36. Because the nutritionists are paid such large amounts of money, they might be more inclined to find favorable results. It is very possible that the results represent desired outcomes instead of actual outcomes.
37. a. 700 adults
    b. 55%
38. a. 253.31 subjects
    b. No. Because the result is a count of people among the 347 who were surveyed, the result must be a whole number.
    c. 253 subjects
    d. 32%
39. a. 559.2 respondents
    b. No. Because the result is a count of respondents among the 1165 engaged or married women who were surveyed, the result must be a whole number.
    c. 559 respondents
    d. 8%
40. a. 847.56 drivers
    b. No. Because the result is a count of respondents saying that they text while driving, the result must be a whole number.
    c. 848 drivers
    d. Given that texting while driving is extremely dangerous, the result of 42% of drivers who text while driving is far too high. The result suggests that steps should be taken to substantially lower that rate.

41. Because a reduction of 100% would eliminate all of the size, it is not possible to reduce the size by 100% or more.
42. In an editorial criticizing the statement, the *New York Times* correctly interpreted the 100% improvement to mean that no baggage is being lost, which was not true.
43. Because a reduction of 100% would eliminate all plaque, it is not possible to reduce it by more than 100%.
44. Because a reduction of 100% would eliminate all car thefts, it is not possible to reduce it by more than 100%.
45. If one subgroup receives a 4% raise and another subgroup receives a 4% raise, the combined group will receive a 4% raise, not an 8% raise. The percentages should not be added in this case.
46. The wording of the question is biased and tends to encourage negative responses. The sample size of 20 is too small. Survey respondents are self-selected instead of being randomly selected by the newspaper. If 20 readers respond, the percentages should be multiples of 5, so 87% and 13% are not possible results.
47. All percentages of success should be multiples of 5. The given percentages cannot be correct.

Section 1-2: Types of Data

1. The population consists of all adults in the United States, and the sample is the 1001 adults who were surveyed. Because the value of 69% refers to the sample, it is a statistic.
2. a. quantitative
   b. categorical
   c. categorical
   d. quantitative
3. Only part (b) describes discrete data.
4. a. The sample is the 36,000 adults who were surveyed. The population is all adults in the United States.
   b. statistic
   c. ratio
   d. discrete
5. statistic
6. parameter
7. parameter
8. statistic
9. statistic
10. statistic
11. parameter
12. parameter
13. continuous
14. continuous
15. discrete
16. discrete
17. continuous
18. discrete
19. discrete
20. continuous
21. nominal
22. ordinal
23. ordinal
24. ratio
25. interval
26. nominal
27. ratio
28. interval
29. The numbers are not counts or measures of anything. They are at the nominal level of measurement, and it makes no sense to compute the average (mean) of them.
30. The digits are not counts or measures of anything. They are at the nominal level of measurement and it makes no sense to calculate their average (mean).
31. The temperatures are at the interval level of measurement. Because there is no natural starting point with 0°F representing "no heat," ratios such as "twice" make no sense, so it is wrong to say that it is twice as warm in Paris as it is in Anchorage.
32. The ranks are at the ordinal level of measurement. Differences between the universities cannot be determined, so there is no way to know whether the difference between Harvard and MIT is the same as the difference between Stanford and the University of California at Berkeley.
33. a. Continuous, because the number of possible values is infinite and not countable.
    b. Discrete, because the number of possible values is finite.
    c. Discrete, because the number of possible values is finite.
    d. Discrete, because the number of possible values is infinite and countable.

34. Interval level of measurement. The direction of north represented by $0°$ is arbitrary, and $0°$ does not represent "no direction." Differences between degrees are meaningful; the difference between $30°$ and $60°$ is the same as the difference between $150°$ and $180°$. But ratios are not meaningful; the ratio of $60°$ to $30°$ does not result in twice some direction. (These degree measurements are directions, not amounts of rotation.)

Section 1-3: Collecting Sample Data

1. The study is an experiment because subjects were given treatments.
2. The subjects in the study did not know whether they were given the magnet treatment or the sham treatment, and those who administered the treatments also did not know.
3. The group sample sizes are large enough so that the researchers could see the effects of the two treatments, but it would have been better to have larger samples.
4. The sample appears to be a convenience sample. Given that the subjects were all patients at a Veterans Affairs hospital, it is not likely that the sample is representative of the population, so it is questionable whether the results can be generalized for the population of subjects with chronic low back pain.
5. The sample appears to be a convenience sample. By e-mailing the survey to a readily available group of Internet users, it was easy to obtain results. Although there is a real potential for getting a sample group that is not representative of the population, indications of which ear is used for cell phone calls and which hand is dominant do not appear to be factors that would be distorted much by a sample bias.
6. The study is an observational study because the subjects were not given any treatment.
7. With 717 responses, the response rate is 14%, which does appear to be quite low. In general, a very low response rate creates a serious potential for getting a biased sample that consists of those with a special interest in the topic.
8. Answers vary, but the following are good possibilities.
   a. Obtain a printed copy of the class roster, assign consecutive numbers (integers), then use a computer to randomly generate six of those numbers.
   b. Select every third student leaving class until six students are chosen.
   c. Randomly select three males and three females.
   d. Randomly select a row, and then select the students in that row. (Use only the first six to meet the requirement of a sample of size six.)
   e. Select the first six students who enter the class.
9. systematic
10. convenience
11. random
12. stratified
13. cluster
14. random
15. stratified
16. systematic
17. random
18. cluster
19. convenience
20. systematic
21. Observational study. The sample is a convenience sample consisting of subjects who decided themselves to respond. Such voluntary response samples have a high chance of not being representative of the larger population, so the sample may well be biased. The question was posted in an electronic edition of a newspaper, so the sample is biased from the beginning.
22. Experiment. The sample subjects consist of male physicians only. It would have been better to include females. Also, it would be better to include male and females who are not physicians.
23. Experiment. This experiment would create an *extremely* dangerous and illegal situation that has a real potential to result in injury or death. It's difficult enough to drive in New York City while being completely sober.
24. Observational study. The sample of only three students is too small.
25. Experiment. The biased sample created by using a small sample of college students cannot be fixed by using a larger sample. The larger sample will still be a biased sample that is not representative of the population of all adults.
26. Experiment. Calling the subjects and asking them to report their weights has a high risk of getting results that do not reflect the actual weights. It would have been much better to somehow measure the weights instead of asking the subjects to report them.

27. Observational study. Respondents who have been convicted of felonies are not likely to respond honestly to the second question. The survey will suffer from a "social desirability bias" because subjects will tend to respond in ways that will be viewed favorably by those conducting the survey.
28. Observational study. The number of responses is very small, and the response rate of only 1.52% is far too small. With such a low response rate, there is a real possibility that the sample of respondents is biased and consists only of those with special interests in the survey topic.
29. prospective study
30. retrospective study
31. cross-sectional study
32. prospective study
33. matched pairs design
34. randomized block design
35. completely randomized design
36. matched pairs design
37. a. Not a simple random sample, but it is a random sample.
    b. Simple random sample and also a random sample.
    c. Not a simple random sample and not a random sample.

Quick Quiz

1. No. The numbers do not measure or count anything.
2. nominal
3. continuous
4. quantitative data
5. ratio
6. statistic
7. no
8. observational study
9. The subjects did not know whether they were getting aspirin or the placebo.
10. simple random sample

Review Exercises

1. The respondents are a voluntary response sample or a self-selected sample. Because those with strong interests in the topic are more likely to respond, it is very possible that their responses do not reflect the opinions or behavior of the general population.
2. a. The sample is a voluntary response sample, so the results are questionable.
   b. statistic
   c. observational study
3. Randomized: Subjects were assigned to the different groups through a process of random selection, whereby they had the same chance of belonging to each group. Double-blind: The subjects did not know which of the three groups they were in, and the people who evaluated results did not know either.
4. No. Correlation does not imply causality.
5. a. systematic
   b. stratified
   c. simple random sample
   d. Convenience
   e. Cluster
6. Yes. The two questions give the false impression that they are addressing very different issues. Most people would be in favor of defending marriage, so the first question is likely to receive a substantial number of "yes" responses. The second question better describes the issue and subjects are much more likely to have varied responses.
7. a. discrete
   b. ratio
   c. The mailed responses would be a voluntary response sample, so those with strong opinions or greater interest in the topics are more likely to respond. It is very possible that the results do not reflect the true opinions of the population of all state residents.
   d. stratified
   e. cluster

8. a. If they have no fat at all, they have 100% less than any other amount with fat, so the 125% figure cannot be correct.
   b. 686
   c. 28%
9. a. interval data; systematic sample
   b. nominal data; stratified sample
   c. ordinal data; convenience sample
10. Because there is less than a 1% chance of getting the results by chance, the method does appear to have statistical significance. The result of 239 boys in 291 births is a rate of 82% so it is above the 50% rate expected by chance, and it does appear to be high enough to have practical significance. The procedure appears to have both statistical significance and practical significance.

Cumulative Review Exercises

1. The mean is $\frac{135+149+145+129+118+119+115+133+107+188+127+131}{12} = 133.0$. The IQ score of 188 appears to be substantially higher than the other IQ scores.
2. $0.5^{13} = 0.000122$
3. $\frac{203-176}{6} = 4.5,$ which is an unusually high value.
4. $\frac{98.2-98.6}{\frac{0.62}{\sqrt{106}}} = -6.64$
5. $\frac{1.95996^2 \cdot 0.25}{0.03^2} = 1068$
6. $\frac{188-107}{4} = 20.25$
7. $\frac{(135-133.0)^2}{11} = 0.364$
8. $\sqrt{\frac{(98.4-98.6)^2+(98.6-98.6)^2+(98.8-98.6)^2}{3-1}} = \sqrt{0.04} = 0.20$
9. $0.3^6 = 0.000729$
10. $8^{12} = 68{,}719{,}476{,}736$ (or about 68,719,477,000)
11. $85^6 = 377{,}149{,}515{,}625$ (or about 377,149,520,000)
12. $0.2^{12} = 0.000000004096$

Chapter 2: Exploring Data with Tables and Graphs

Section 2-1: Frequency Distributions for Organizing and Summarizing Data

1. The table summarizes 1000 commute times. It is not possible to identify the exact values of all of the original times.
2. The classes of 0–30, 30–60, …, 120–150 overlap, so it is not always clear which class we should put a value in. For example, the commute time value of 30 minutes could go in the first class or the second class. The classes should be mutually exclusive, meaning that there is no overlap.
3.

Daily Commute Time in Boston (minutes)	Relative Frequency
0–29	46.8%
30–59	42.2%
60–89	9.2%
90–119	1.0%
120–149	0.8%

4. The sum of the relative frequencies is 125%, but it should be 100%, with a small round off error. All of the relative frequencies appear to be roughly the same, but if they are from a normal distribution, they should start low, reach a maximum, and then decrease.
5. Class width: 10
   Class midpoints: 24.5, 34.5, 44.5, 54.5, 64.5, 74.5, 84.5
   Class boundaries: 19.5, 29.5, 39.5, 49.5, 59.5, 69.5, 79.5, 89.5
   Number: 91
6. Class width: 10
   Class midpoints: 24.5, 34.5, 44.5, 54.5, 64.5, 74.5
   Class boundaries: 19.5, 29.5, 39.5, 49.5, 59.5, 69.5, 79.5
   Number: 91
7. Class width: 100
   Class midpoints: 49.5, 149.5, 249.5, 349.5, 449.5, 549.5, 649.5
   Class boundaries: –0.5, 99.5, 199.5, 299.5, 399.5, 499.5, 599.5, 699.5
   Number: 153
8. Class width: 100
   Class midpoints: 149.5, 249.5, 349.5, 449.5, 549.5
   Class boundaries: 99.5, 199.5, 299.5, 399.5, 499.5, 599.5
   Number: 147
9. No. The maximum frequency is in the second class instead of being near the middle, so the frequencies below the maximum do not mirror those above the maximum.
10. Yes. The frequencies start low, reach a maximum of 38, and then decrease. The values below the maximum are very roughly a mirror image of those above it.
11. Yes. Except for the single value that lies between 600 and 699, the frequencies start low, reach a maximum of 90, and then decrease. The values below the maximum are very roughly a mirror image of those above it. (That single value between 600 and 699 is an outlier that makes the determination of a normal distribution somewhat questionable, but using a *loose interpretation* of the criteria for normality, it is reasonable to conclude that the distribution is normal.)
12. Yes. Except for two values that lie between 500 and 599, there is a low frequency of 25, then a maximum frequency of 92, and then a low frequency of 28. The values below and above the maximum are roughly a mirror image. (Those two values between 500 and 599 are outliers that make the determination of a normal distribution somewhat questionable, but using a *loose interpretation* of the criteria for normality, it is reasonable to conclude that the distribution is normal.)

13. The data amounts do not appear to have a normal distribution. The distribution does not appear to be symmetric because the frequencies preceding the maximum frequency of 16 are far outweighed by the frequencies following the maximum.

Daily Commute Time in Chicago (Minutes)	Frequency
0–14	5
15–29	16
30–44	14
45–59	9
60–74	5
75–89	1

14. The ages do appear to have a normal distribution.

Age (years)	Frequency
40–44	2
45–49	7
50–54	10
55–59	10
60–64	6
65–69	3
70–74	1

15.

Duration (sec)	Frequency
125–149	1
150–174	0
175–199	0
200–224	3
225–249	34
250–274	12

16. The intensities do not appear to have a normal distribution.

Tornado F-Scale	Frequency
0	24
1	16
2	2
3	2
4	1

17.

Burger King Lunch Service Times (sec)	Frequency
70–109	11
110–149	23
150–189	7
190–229	6
230–269	3
230–269	6

18.

Burger King Dinner Service Times (sec)	Frequency
30–69	1
70–109	6
110–149	26
150–189	7
190–229	3
230–269	6
270–309	1

19. The distribution does appear to be a normal distribution.

Weight (kg)	Frequency
40–49	2
50–59	22
60–69	23
70–79	13
80–89	3
90–99	4

20. The distribution does appear to be a normal distribution.

Weight (g)	Frequency
4.300–4.399	11
4.400–4.499	17
4.500–4.599	28
4.600–4.699	16
4.700–4.799	3

21. Because there are disproportionately more 0s and 5s, it appears that the heights were reported instead of measured. Consequently, it is likely that the results are not very accurate.

x	Frequency
0	9
1	2
2	1
3	3
4	1
5	15
6	2
7	0
8	3
9	1

22. Because there are disproportionately more 0s and 5s, it appears that the heights were reported instead of measured. There does appear to be a gap due to the tendency of respondents to round their heights to values ending in 0 or 5. Because the results appear to be reported instead of measured, it is likely that the results are not very accurate.

x	Frequency
0	26
1	1
2	1
3	2
4	2
5	12
6	1
7	0
8	4
9	1

23. The actresses appear to be generally younger than the actors.

Age When Oscar Was Won	Relative Frequency (Actresses)	Relative Frequency (Actors)
20–29	34.1%	1.1%
30–39	37.4%	31.9%
40–49	16.5%	41.8%
50–59	3.3%	17.6%
60–69	6.6%	6.6%
70–79	1.1%	1.1%
80–89	1.1%	0.0%

24. There do appear to be differences, but overall they are not very substantial differences.

Blood Platelet Count	Males	Females
0–99	0.7%	0.0%
100–199	33.3%	17.0%
200–299	58.8%	62.6%
300–399	6.5%	19.0%
400–499	0.0%	0.0%
500–599	0.0%	1.4%
600–699	0.7%	0.0%

25.

Age (years) of Best Actress When Oscar Was Won	Cumulative Frequency
Less than 30	31
Less than 40	65
Less than 50	80
Less than 60	83
Less than 70	89
Less than 80	90
Less than 90	91

26.

Age (years) of Best Actor When Oscar Was Won	Cumulative Frequency
Less than 30	1
Less than 40	30
Less than 50	68
Less than 60	84
Less than 70	90
Less than 80	91

27. No. The United States has 37.1% of the cost of piracy for only the five countries listed, not the total cost of piracy for all countries. Because only the top five costs of piracy are listed, we know only that any other country must have a cost less than $1.9 billion.

Country	Relative
United States	37.1%
China	35.5%
India	11.0%
France	8.6%
United Kingdom	7.8%

28. Yes, it appears that births occur on the days of the week with frequencies that are about the same.

Day	Relative Frequency
Monday	13.0%
Tuesday	16.5%
Wednesday	18.0%
Thursday	14.3%
Friday	14.3%
Saturday	10.8%
Sunday	13.3%

29. It is very similar to Table 2-2. Both frequency distributions begin with a low frequency in the first class, followed by the maximum frequency in the second class, and the frequencies are generally lower as you progress from top to bottom in the table. (TI data: Frequencies are 79, 148, 157, 43, 49, 6, 9, 0, 0, 9.)

Daily Commute Time in Los Angeles (minutes)	Frequency
0–14	157
15–29	324
30–44	282
45–59	103
60–74	98
75–89	7
90–104	18
105–119	0
120–134	0
135–149	11

30. It is very similar to the frequency distribution for Exercise 1. Both frequency distributions begin with the maximum frequency, and then the frequencies decrease moving from top to bottom in the table. (TI data: Frequencies are 289, 154, 43, 11, 0, 3.)

Daily Commute Time in Dallas (minutes)	Frequency
0–29	547
30–59	339
60–89	92
90–119	15
120–149	0
150–179	7

31. Yes, the frequency distribution appears to be a normal distribution.

Systolic Blood Pressure (mm Hg)	Frequency
80–99	11
100–119	116
120–139	131
140–159	34
160–179	7
180–199	1

32. Yes, the frequency distribution appears to be a normal distribution.

Diastolic Blood Pressure (mm Hg)	Frequency
40–54	27
55–69	107
70–84	133
85–99	31
100–114	2

33. Yes, the frequency distribution appears to be a normal distribution.

Magnitude	Frequency
1.00–1.49	19
1.50–1.99	97
2.00–2.49	187
2.50–2.99	147
3.00–3.49	100
3.50–3.99	38
4.00–4.49	8
4.50–4.99	4

34. No, the frequency distribution does not appear to be a normal distribution.

Depth (km)	Frequency
0.0–9.9	539
10.0–19.9	49
20.0–29.9	10
30.0–39.9	1
40.0–49.9	1