CHAPTER 1 ANSWERS

Section 1.1
Statistical Literacy and Critical Thinking

1 A population is the complete set of people or things being studied, while a sample is a subset of the population. The difference is that the sample is only a part of the complete population.

2 The two uses do not have the same meaning. The term *baseball statistics* refers to measurements or data that summarize past results. The other use of *statistics* refers to the science of using statistical methods for analyzing the effectiveness of the drug.

3 A *sample statistic* is a characteristic of a *sample* found by consolidating or summarizing raw data. A *population parameter* is a characteristic of an entire *population*. Since it is not usually practical to obtain raw data for entire large populations, it is also not likely that population parameters can be directly measured. For that reason, we use measured sample statistics to make inferences about the values of population parameters.

4 The margin of error is used to help describe the range of values likely to contain the value of a population parameter of interest. In many cases, that range of values is found by simply adding and subtracting the margin of error from the value of the sample statistic obtained in the study.

5 This statement does not make sense. Population parameters are inferred from sample statistics, so it's not possible to have the former without the latter. The only way to determine a population parameter is to obtain raw data for every individual in the population, in which case there is no error at all.

6 This statement is sensible. It suggests that Smith had a substantial lead two weeks before the election, but leads can certainly evaporate in two weeks. It is also possible that the poll was not conducted carefully enough to ensure that the sample was representative of the population. In this case, the 70% figure could have badly misrepresented the population proportion that would vote for Smith, leading to incorrect conclusions about his chances of winning.

7 This statement does not make sense. The poll makes it seem like Johnson should win the election because the confidence interval for the percent of voters voting for Johnson runs from 54% - 3% to 54% + 3% (51% to 57%), suggesting that he should have obtained more than half of the votes, enough to win. However, in most cases such as this, the margin of error is defined to mean that we can be 95% confident that the true percent of votes lies in the range from 51% to 57%. That means that there is also a 5% chance that the actual percent of votes is outside that range – above 57% or below 51%. If, in fact, it does lies below 51%, there is a chance that it also lies below 50%, in which case Johnson loses the election.

8 This statement does not make sense. A larger margin of error means a less certain result; networks would not pay the same amount of money for less certain results.

9 This statement does not make sense. The population of interest is people who have suffered a family tragedy, not people who may have been sick in the past month. The sample must be taken from the population of interest

10 This statement makes sense. The purpose of using statistical methods is to help with decision-making. If the survey were well-conducted, a sample of size 1000 makes it possible to draw conclusions with a high level of confidence, and it makes sense to follow the guidance of the results of the survey. Of course, the results of the survey cannot guarantee the results of the advertising campaign, which has yet to be designed.

Concepts and Applications

11 The sample consists of the 1002 adults in the U.S. who were surveyed. The population consists of all adults in the U.S. The sample statistic is the 48% who said that they were in favor the use of federal tax dollars to fund medical research using stem cells obtained from human embryos. The value of the population parameter is not known, but it is the percentage of all adults in the U.S who say that they are in favor.

12 The sample consists of the 1236 adults who were surveyed. The population consists of all adults. The sample statistic is the 9% who said that bad luck occurs on a Friday that is the 13^{th} of the month. The value of the population parameter is not known, but it is the percentage of all adults who say that bad luck occurs on a Friday that is the 13^{th} of the month.

13 The sample consists of the distances to the selected stars. The population consists of the distances to all starts in the galaxy. The sample statistic is the mean of the distances to the selected stars in the galaxy. The population parameter is unknown, but it is the mean of the distances to all of the stars in the galaxy.

14 The sample consists of the 374 children aged 3 to 11 years who were treated with 100mcg of Nasonex. The population consists of all children aged 3 to 11 years old. The sample statistic is the 17% who experienced headaches. The population parameter is not known, but it is the percentage of all children 3 to 11 years old who experience headaches when treated with Nasonex.

15 The range of values likely to contain the true value of the population parameter is from 48% - 3% to 48% + 3% or from 45% to 51%.

16 The range of values likely to contain the true value of the population parameter is from 45% - 3% to 45% + 3% or from 42% to 48%.

17 The range of values likely to contain the true value of the population parameter (mean BMI) is from 26.0 – 3.4 to 26.0 + 3.4 or from 22.6 to 29.4.

18 The range of values likely to contain the true value of the population parameter (mean body temperature) is 98.2° F – 0.1° F to 98.2° F + 0.1° F or from 98.1° F to 98.3° F degrees.

19 Yes. Since the likely range of values of the percentage of votes for the Republican candidate is from 58% - 3% to 58% + 3% (or 55% to 61%), the results suggest that the Republican candidate will get more than half of the votes. However, there is no guarantee. If, for example, the polling technique is flawed or the candidate makes a comment the night before the election that irritates enough voters, the results of the survey may not correctly reflect the actual vote.

20 No. We cannot infer that the majority of Budweiser drinkers prefer Michelob when given tastes of Michelob and Schlitz. The range of population percentages preferring Michelob is from 52% - 10% to 52% + 10% (or from 42% to 62%). It is quite possible that less than half of the population would prefer Michelob when given the same taste test.

21 With a sample statistic of 70% and a margin of error of 3 percentage points, we are 95% confident that the interval from 67% to 73% contains the population parameter that is the true percentage of the voters who would say that they voted in the recent presidential election. This entire range is, however, somewhat higher than the actual 61% who voted according to the voting records. This suggests that there were some people in the sample who did not actually vote, but said that they did when polled. While it is still possible (as always) that this particular sample is unusual and everyone told the truth, the lower end of the range (67%) is quite far from 61%, making this an unlikely possibility.

22 It appears that the men who were surveyed may have been influenced by the gender of the interviewer. When they were interviewed by women, they may have been more inclined to respond in a way that they thought was more

favorable to the female interviewers.

23 a) The goal was to determine how satisfied executives were with their career choices. The population consists of all executives and the population parameter is the percentage of all executives who say that they would choose a different field if they could start their careers again.

b) The sample consists of the 1733 executives selected for the study. The raw data consist of the individual responses to the survey. The sample statistic is the 51% who would choose a different field.

c) The range of values likely to contain the population parameter is from 51% - 3% to 51% + 3% (or from 48% to 54%).

24 a) The goal was to determine the percentage of adults who would choose the ability to be invisible as the superpower they would most prefer. The population consists of all adults, and the population parameter is the percentage of all adults who say that they would choose the ability to become invisible as the superpower they would most prefer.

b) The sample consists of the 1018 adults selected for the study. The raw data consist of the individual responses to the survey. The sample statistic is the 11% who would choose the ability to become invisible as their preferred superpower.

c) The range of values likely to contain the population parameter is from 11% - 3.2% to 11% + 3.2% (or from 7.8% to 14.2%).

25 a) The goal is to determine the unemployment rate of adults. The population consists of all people employed or seeking employment in the U.S., and the population parameter is the percentage of people in the population who are unemployed.

b) The sample consists of the 60,000 households surveyed. The raw data consist of the individual responses to the question for each household. The sample statistic is the percentage of people in the sample who are unemployed, 4.6%.

c) The range of values likely to contain the population parameter is 4.6% $\pm$ 0.2% or 4.4% to 4.8%.

26 a) The goal is to determine the percentage of adults in the U.S. who keep money in regular savings accounts. The population consists of all adults in the U.S., and the population parameter is the percentage of all adults in the U.S. who keep money in regular savings accounts.

b) The sample consists of the 2000 adults surveyed. The raw data are the individual responses of the adults in the sample. The sample statistic is the percentage of adults in the sample who keep money in regular savings accounts

c) The range of values likely to contain the population parameter is 64% $\pm$ 2.0% or 62% to 66%.

27 There are two possible scenarios for this study. In the first, we look for the percentage of drivers who are using cell phones while driving at the time of the survey. In the second, we look for the percentage of drivers who admit that they do use cell phones while driving (even if they are not using them at the time they are surveyed). The second percentage is likely to be higher than the first since there will always be some drivers who admit that they use a cell phone while driving even if they are not using one at the time they are surveyed. [Note: The results of both types of surveys can be greatly affected by the geographical area and time of day that samples are taken. Obviously, no one will be using a cell phone in an area where there is no service. Few drivers will use them while driving in the middle of the night. Many may be using them while caught in a rush hour traffic jam. These and other concerns point out that much care must be taken in designing such a study and in formulating any conclusions.]

First scenario:

Step 1: Goal: Determine the percentage of all drivers who are using cell

phones while they are driving.
Step 2: Choose a sample of drivers while they are driving.
Step 3: Interview or observe the drivers in the sample to determine whether or not they are using a cell phone at the time of the observation.
Step 4: Use statistical techniques to infer the likely results for the entire population of drivers.
Step 5: Based on the likely results for the population, draw conclusions about the percentage of drivers who are using cell phones while driving at any given time.
Second scenario:
Step 1: Goal: Determine the percentage of all drivers who use cell phones while they are driving.
Step 2: Choose a sample of drivers.
Step 3: Interview the drivers in the sample to determine whether or not they admit to using a cell phone at any time while driving.
Step 4: Use statistical techniques to infer the likely results for the entire population of drivers.
Step 5: Based on the likely results for the population, draw conclusions about the percentage of drivers who use cell phones while driving.

28 Step 1: Goal: Determine the mean FICO score of all adult consumers in the U.S.
Step 2: Choose a sample of adult consumers.
Step 3: Obtain the FICO scores of the selected consumers and calculate the mean FICO score for those consumers in the sample.
Step 4: Use statistical techniques to infer the likely results for the entire population adult consumers in the U.S..
Step 5: Based on the likely results for the population, draw conclusions about the mean FICO score of all adult consumers in the U.S.

29 Step 1: Goal: Determine the mean weight of airline passengers.
Step 2: Choose a sample of airline passengers.
Step 3: Weigh each selected passenger and calculate the mean weight of those in the sample.
Step 4: Use statistical techniques to infer the likely results for the entire population of airline passengers.
Step 5: Based on the likely results for the population, draw conclusions about the average weight of all airline passengers.

30 Step 1: Goal: Determine the mean time to failure of all pacemaker batteries.
Step 2: Choose a sample of pacemaker =batteries.
Step 3: Record the length of time that each battery in the sample lasts until failure and then calculate the mean time to failure for the batteries in the sample.
Step 4: Use statistical techniques to infer the likely mean time to failure for the entire population of pacemaker batteries.
Step 5: Based on the likely results for the population, form a conclusion about the mean time to failure for all pacemaker batteries.

Section 1.2
Statistical Literacy and Critical Thinking

1 A *census* is the collection of data from every member of the population. A sample is the collection of data from some, but not all, members of the population. For a given population, a sample will contain less data than will a census.

2 Yes. If the goal is to obtain information useful for predicting the outcome of the election, the sample consisting only of registered Democrats is

certainly biased and of no use in predicting the election.

3 Cluster sampling involves randomly selecting subgroups of a population and then selecting all members of the population in each subgroup. For example, one might randomly select some city blocks and then interview all people living on those blocks. Stratified sampling involves randomly selecting members from each of different subgroups (or strata) of the population. For example, one could randomly select some men and randomly select some women, keeping the results separate for each of the two gender subgroups.

4 If the professor obtained information only from the members of his classes, the sample was a convenience sample. It is not likely that the sample was biased since there is probably nothing right-handedness that would cause the proportion of right-handed students in a particular college class to be different from the proportion of right-handed students in the entire college population.

5 This statement does not make sense. A census would mean getting age data for person who earns a bachelor's degree in the country (or world), which is clearly not practical.

6 This statement makes sense. A convenience sample is often prone to bias, but there may be cases in which it works just fine.

7 This statement makes sense. It's quite apparent that most Americans are not more than 6 feet tall, so a study that comes to a ridiculous conclusion must have suffered from some form of bias.

8 This statement makes sense. This procedure does result in a simple random sample and it is a commonly used technique.

9 Since the number of players on the LA Lakers is small, a census is practical, and it is easy to obtain their heights (for example, from a Laker website).

Concepts and Applications

10 A census is not practical since it would require obtaining the height of every high school basketball player in the country. The number of players is much too large to obtain all of the information.

11 A census is not practical since it would require obtaining the IQ of every statistics instructor in the U.S. The number of statistics instructors is very large and it would be difficult to get them all to take an IQ test.

12 A census is practical. The number of statistics instructors at the university of Colorado is relatively small. Given the interest of statistics instructors in things statistical, it would probably not be difficult to get their ages through a survey that promised anonymity.

13 The sample consists of the service times of the four selected Senators. The population consists of the service times of all 100 Senators. The sampling method is simple random sampling. Since the sample is so small, there is a good chance that it is not representative of the entire population.

14 The sample is the 5108 selected households. The population is the complete set of all households. The sample is selected using simple random sampling. Because the sample size is quite large and sampling was done by a well-established and reputable firm, the sample is likely to be representative of the population.

15 The sample consists of the 1059 randomly selected adults. The population consists of all adults. Simple random sampling was used. Because the sample size is quite large and sampling was done by a well-established and reputable firm, the sample is likely to be representative of the population.

16 The sample consists of the 65 responses she received. The population consists of the responses of all American adults (if they had been asked). The sampling method is convenience sampling since the adults to whom she sent the survey were people she already knew. The final sample is also the result of self-selection since those who received letters decided themselves

whether or not to respond. Since the survey was about communication and mailing was required to respond, those who preferred to use email may not have chosen to respond in writing, while those who preferred to use snail mail may have been more likely to respond by mail. As a result, the sample is not likely to representative of the entire population.

17 The most representative sample is likely to be Sample 3 because the list will contain people from all over Florida and there is no reason to suspect that the people with the first 1000 numbers would differ in any particular way from the other people. [This assumes that the list is alphabetical, not in order by phone number, in which case the first three digits of the phone are likely to be the same and the entire sample would come from one area, possibly one city, of Florida.] Sample 1 is biased because it involves owners of expensive vehicles. Such owners may be able to pay off their credit cards monthly or may be people with greater credit limits on their credit cards. Sample 2 is biased because is includes only people from the Fort Lauderdale area. Sample 4 is biased because it includes only people who are self-selected and may have strong feelings about the issue of credit card debt.

18 The most representative sample is likely to be Sample 4, which is a good use of systematic sampling. Samples 1 and 2 are likely to be unrepresentative because they each involve people from one geographic region of the state. Sample 3 is likely to be biased because it is a self-selected sample and it is further limited to people who have internet access and who receive the AOL survey.

19 The critic may be under real or imagined pressure to give a favorable review to the film since she works for the same company that produced the film.

20 There are no sources of bias in this situation. Because *Consumer Reports* does not accept any advertising and it does not accept free products, it is not influenced by the manufacturers of the cars that it reviews.

21 The university scientists are receiving funding from Monsanto, which might make them eager to please Monsanto in hopes of getting additional funding opportunities in the future. Thus, there is a potential for bias toward giving Monsanto the results it wants, even though they do not work for Monsanto.

22 Yes. Because some of the physicians who wrote the article receive funding from the pharmaceutical company, they might be more inclined to provide more favorable results so that they can get additional funding in the future. The Journal of the American Medical Association now requires that all such physician authors disclose any funding, and those disclosures are included in the articles.

23 This sample is a simple random sample that is likely to be representative because there is no bias in the selection process.

24 This is an example of systematic sampling, and it is likely to be representative because there is no bias in the selection process.

25 This is an example of cluster sampling. It is likely to be unbiased as long as there are enough polling stations selected for the sample so that the entire sample has a chance to be representative on a national level. Since the actual results of the election are usually known within a few hours of exit poll results and the exit polls are unlikely to influence the voting of any other voters, poor sampling techniques that have a good chance of resulting in embarrassment for the news media are likely to be avoided.

26 This is a stratified sample. However, even if the participants are randomly selected in each of the strata, the sample is likely to be biased because strata representing other sports are not being used, and because the number of people who participate in various sports do not do so in equal numbers for every sport, let alone for golfing, swimming, and tennis.

27 This is an example of convenience sampling. It is likely to be biased because members of a family are likely to be more similar in their physical

characteristics and strength than would a sample taken from the population as a whole.

28 This sample is a cluster sample. Waiters and waitresses who cheat on their taxes are unlikely to give truthful answers, biasing the study. Also, the small number of restaurants chosen could easily result in a sample that is not representative of all waitresses and waiters.

29 This is a stratified sample with the strata being the various age groups. As described, the sample is likely to be biased because it contains equal numbers of people in each of the each groups whereas the population is not equally distributed among these age groups. There are two ways to remedy this problem. The results from the age strata could be combined by "weighting" the results from each stratum to reflect the sizes of the strata in the population as a whole. A second way is use proportionate sampling in which each stratum in the sample has a number of members that is proportional to its presence in the population as a whole.

30 This sample is a convenience sample. The sample is likely to be biased because all of the students are attending the same college. They are not likely to be representative of all college students.

31 This sample is a systematic sample. It is unlikely to be biased because there is nothing about an alphabetical list that is likely to produce a biased sample when testing a telemarketing technique.

32 This is a simple random sample. Because the sample size is fairly large and the sample is random, it is unlikely to be a biased sample.

33 This is a stratified sample. It is likely to be a biased sample because population does not consist of employed, unemployed, and employed part time in equal numbers. It is possible to correct this bias by "weighting" the strata results to reflect the strata sizes in the population.

34 This is a cluster sample. It could easily be biased, but that may depend on what types of classes were selected.

35 This is a convenience sample, and it is one that is unlikely to be representative because people with strong feelings are more likely to return the survey. The magazine probably chose this sampling method because it was easy; the magazine might even be interested in the opinions of those with the strongest feelings.

36 This is a simple random sample. It is likely to be representative for that reason. There are situations in which a sample size of 50 is regarded as large, but 50 would be considered small in other situations. Whether or not the sample has a good chance of being representative depends to some extent on what characteristic of the patients is being measured.

37 This is a simple random sample and is therefore likely to be representative. The sample size is not specified, but the larger the sample size, the better the chance that the sample is representative.

38 This is a systemic sample. It is likely to be representative unless there is something systematic in the manufacturing process that produces defects. For example, if every 50^{th} seat belt produced is defective, then every 500^{th} seat belt is also defective. If the sampling plan is to select seat belts 3, 503, 1003, 1503, ..., and the defective seat belts are 17, 67, 107, 167, ..., then all of the defective seat belts will be missed and the proportion of defectives will be thought to be lower than it actually is. On the other hand, if the defective seat belts are 3, 53, 103, 153, ..., then every seat belt tested will be found to be defective and the proportion of defectives will be thought to be higher than it actually is.

39 Simple random sampling should be adequate for a student election if the sample is large enough.

40 Simple random sampling should be adequate. However, stratified sampling in which the strata are different ethnic groups is also a possibility. This would enable one to gather information about the differences in percentages of blood types among the different ethnic groups and would make it possible

to better estimate the overall percentage of people in each of the four blood groups.

41 Since all states have single departments that keep all death records, it should be easy to randomly select some states and then search the computer records to determine the number and percentage of deaths due to heart disease each year. This is an example of cluster sampling with each cluster being a state. [The U.S. Center for Disease Control (CDC) routinely collects these data from all states and they are available on the CDC website.]

42 You will need stratified sampling in which you measure the mercury content of tuna in different markets that represent different sources of tuna fish.

Section 1.3
Statistical Literacy and Critical Thinking

1 A placebo is physically similar to a treatment, but it lacks any active ingredient, so it should not have any effects on the subject. A placebo is important so that results from subjects given a real treatment can be compared to the results from subjects given a placebo.

2 Blinding is a process used in an experiment in which the subjects and/or the experimenters do not know who is in the treatment group and who is in the control group. It is important to use blinding for subjects so that they are not affected by the knowledge that they are receiving (or not receiving) the real treatment. It is important to use blinding for the experimenters so that they can evaluate results objectively without their judgments being affected by knowledge about who is getting the real treatment(s) and who is getting the placebo.

3 Confounding occurs when it is not possible to ascertain what caused the effects that were observed. In this instance, if males were chosen for the real treatment and females were chosen for the placebo group, and if a difference resulted in the effects on the two groups, it would not be possible to tell whether those effects were caused by the treatment or by the gender of the subjects.

4 This is an experiment because subjects were modified by the treatments. The study is morally and criminally wrong because an effective treatment was withheld from men who needlessly continued to suffer.

5 It almost always makes sense to use double blinding for an experiment, but it is sometimes impossible or difficult to do. In this case, both subjects and experimenters can see the clothing worn by the subjects. Blinding must therefore be achieved by some other method. The subjects may be blinded by not telling them the purpose of the experiment or even that there is an experiment so that their knowledge of the color of their clothes does not affect the results. The experimenters clearly know the purpose of the experiment, so blinding is not possible for them. It is therefore necessary that data be based on objective measures that are not influenced by any judgments of the experimenters.

6 The lawn does not know what treatment it is getting and therefore its response to the treatment cannot be affected by any knowledge of what treatment is used. It is important that those who evaluate the results be blinded to the treatment so that their judgments are not affected by the knowledge of what sections of lawn received the treatment. Since neither the subjects (lawns) nor the experimenters have knowledge of the treatment, this is a double-blind experiment.

7 The experimenter effect occurs when a researcher somehow influences subjects by such things as tone of voice, facial expressions, or attitude. It can be avoided by using blinding so that those who evaluate the results do not know which subjects are given an actual treatment and which subjects are given a placebo. It might also help if the subjects responded to written, rather than oral, questioning or to a computerized voice that conveys exactly the

same attitude to every subject and does not have different tones of voice or facial expressions associated with it.

8 Such results are possible. For example, if the treatment is totally ineffective, random variation might produce better results for the placebo group than for the treatment group. If the placebo group does get better results than the treatment group, it is easy to conclude that the treatment is ineffective.

Concepts and Applications

9 This is an observational study because the subjects were tested, but they were not given any treatment. The variable of interest is whether the subjects were right or wrong on each trial.

10 This is an observational study because the aspirin tablets were tested, but not treated. The variable of interest is the amount of aspirin in each tablet.

11 This is an experiment. There is a treatment group of subjects that received the magnetic bracelets and a placebo group that received the non-magnetic bracelets. The variable of interest is whether or not the passengers experienced motion sickness. Blinding might not be totally successful since passengers might happen to detect whether their bracelets are magnetic by holding them near something made of iron.

12 This is an experiment because the subjects were given a treatment. The treatment group consists of the 152 couples given the YSORT treatment. The control group consists of others not given any treatment. The variable of interest is whether a baby is a boy.

13 This is an observational study since there was no treatment of the subjects. The mental abilities of twins were observed and the results placed in one of two categories (fraternal or identical twins). The variable of interest is the difference in mental abilities of the two people in each twin pair.

14 This is an observational, retrospective study comparing those who took vitamin B supplements and those who did not. The variable of interest is whether heart disease occurred in each subject.

15 This is a meta-analysis in which the results of several studies are combined. The variable of interest is whether each subject developed prostate cancer.

16 This is an observational study since no one received any treatment. The variable of interest is the response to the survey question.

17 This is an experiment. The treatment group consists of the Bt corn and the control group consists of normal corn. The variable of interest is the measured amount of insecticide that is released through the corn roots.

18 This is an observational study. The variable of interest is the response to the survey question.

19 This is an experiment since the subjects received different treatments. The variable of interest is the measured level of back pain.

20 This is a meta-analysis, combining the results of previous studies. The variable of interest is whether the subject developed cancer.

21 Confounding is likely to occur. If there are differences in tree growth in the two groups, it will be impossible to tell if those differences are due to the treatment (fertilizer or irrigation) or to the type of region (moist or dry). This confounding can be avoided by using blocks of trees fertilized trees in both regions and blocks of irrigated trees in both regions.

22 Confounding is very possible. If there are differences between the two groups, we won't be able to tell whether it was because of the group they were in or because they were already comfortable (or not) with computer and internet usage for shopping. Since all of the subjects are volunteers, the entire study is subject to self-selection bias, and since the volunteers

were allowed to select the group, self-selection is again a factor. Since it probably not possible to erase computer and internet experience, nor is it possible to give quick experience to those who do not have computer and internet experience, this study is replete with problems no matter how it is designed. It is not clear how the purchases will be compared – total spent, types of purchases, etc. Clearly, those who do shop on the internet also buy things in stores as well, so making comparisons is going to be difficult.

23 Confounding is likely. If there ios a difference in the amount of gasoline consumed between the two groups, it will not be possible to tell whether the difference is due to the type of vehicles in the two groups or to the octane rating of the gasoline used. Confounding can be avoided by using 87 octane gasoline in half of the vehicles in each group and 91 octane gasoline in the other half.

24 Confounding is possible if either the subjects or the experimenters know who is getting the aspirin and who is getting the placebo. The experiment should be double-blind to avoid a placebo effect or an experimenter effect. Furthermore, the sample sizes are much too small for this kind of study. No meaningful results could be obtained with sample sizes of 3 and 7.

25 Confounding is possible. If there is a difference between the two groups, it will not be possible to determine whether it is because some subjects received the drug and some received the placebo or whether it is because some subjects **were told** that they had received the drug or placebo. The confounding can be avoided by using a double-blind experiment.

26 Confounding is possible, depending on how evaluations of the treatments are done. If a difference between the results for the two groups is found, you want to ensure that it is a result of the difference in effectiveness of the treatment and the placebo, not a result of the physician's evaluation of the effectiveness of the treatment. It would be better to use double-blinding so that the physicians do not know who is given the treatment and who is given the placebo.

27 Confounding is possible. If a difference is found in the effects on blood pressure from lifting weights or tennis balls, you want to ensure that the difference is a result of the two treatments, not from some subjects' apprehension over having their blood pressure measured or from an experimenter's judgment of the effect on blood pressure. The experimenter effect can be avoided by using technology and trained personnel to measure the blood pressure without any interaction with the experimenter. The placebo effect can be reduced or eliminated by having the same subjects use the heavy weights and tennis balls at different times, with the order mixed. Any apprehension over the measurement process should be the same for both sets of weights and will therefore be cancelled out.

28 Confounding is possible if the researchers have a bias toward either of the mixtures. There is no effect due to the subjects' reactions since the painted objects have no way of reacting to the treatment. However, if the evaluation of the mixtures requires judgments on the part of the experimenters, then the experimenters should be blinded so that they evaluate the results without knowing which batch each mixture came from.

29 The control group consists of those who do not listen to Beethoven, and the treatment group consists of those who do listen to Beethoven. Blinding of the subjects is automatic since the infants won't know they are part of an experiment. By coding the subjects, blinding could be used so that those who measure intelligence are not influenced by any knowledge of which group the subjects were in. There is an additional problem that could arise in interpreting the data from the experiment. If a difference between the two groups is found, is it a result of listening to Beethoven, or is it a result of just listening to some kind of music? As designed, this experiment will not be able to determine the answer. If the real interest is Beethoven's

music, the experiment must be expanded to include more groups with other kinds of music.

30 This should be a double-blind experiment with a control group consisting of subjects given a placebo and a treatment group with consisting of those treated with Lipitor. Subjects should be randomly assigned to the two groups.

31 The control group consists of a group of cars using gasoline without the ethanol additive. The treatment data should be obtained by using the same cars with gasoline containing the ethanol additive, mixing the order in which the two gasoline blends are used for the different cars. In this way, it is possible to ensure that any observed difference in mileage is the result of the difference in gasoline blend. If two different groups of cars were used, a difference in mileage between the two groups might have been caused by differences in the cars themselves, even if they were all of the same brand and model. There is no need to blind the cars, and since the mileage will be determined without any judgments on the part of the experimenters, there is no need to blind the experimenters.

32 The control group consists of houses with wood siding, and the treatment group consists of houses with aluminum siding. Blinding is not necessary for the houses, and it is unnecessary for the researchers if the longevity is measured with objective tools. Blinding would be difficult to implement for the evaluators anyway because anyone could tell whether a home has wood siding or aluminum siding.

Section 1.4
Statistical Literacy and Critical Thinking

1 Peer review is a process by which experts in a field evaluate a research report before the report is published. It is useful for lending credibility to the research because it implies that other experts agree that it was carried out properly.

2 Selection bias occurs when researchers select their sample in a biased way, and participation bias occurs when the participants themselves decide to be included in the study.

3 When participants select themselves for a survey, those with strong opinions about the topic being surveyed are more likely to participate, and this group is typically not representative of the general population.

4 Confounding variables are those that affect results in such a way that we cannot determine the effects of the specific variables being studied. Another way of saying this is that whatever effects were observed could have been the result of differences in the variables studied, but they could also have been unintended results of (confounding) variables that were not under study.

5 This answer does not make sense. The phone-in responses were self-selected and may have involved mostly people with strong opinions. There is a good chance that the sample was not representative of the population, even if it happened to be large. The poll taken by professional pollsters would have been taken in such a way that the people surveyed were randomly selected, making it more likely that the sample was representative of the population.

6 This makes sense. By handing out the survey at Catholic churches, it was more likely to reach Catholics than people of other religious faiths or beliefs.

7 This does not make sense. Often we don't even know if there are confounding variables, let alone how many, so we can't know for certain that they have all been taken into account.

8 This does not make sense. While it's possible that studies could show these results, there is no point to the studies. Most joggers jog for exercise, not for competition, so a 1% increase in speed has no practical significance

for them.

9 Since the researchers are from the public relations department of the company, there is a good chance that they may be biased, so Guideline 2 *(Consider the source..) is the most relevant.

10 Guidelines 7 and 8 appear to be the most pertinent. Although "most" of the 44 babies were girls (23 of them), this is only one more than one would expect on the average if there were no treatment at all. The concluding statement is not justified by the results. One could also say that the conclusions do not make sense or that the results have no practical significance, making Guideline 8 relevant as well.

11 Guideline 4 is most relevant since the phrase "good ethics" is not well defined and it is probably not possible to measure "good ethics."

12 Guideline 5 is the most relevant since weather and soil conditions are different in Arizona and California, making it impossible to determine whether differences are due to the irrigation system or to the weather and soil conditions.

13 Guideline 5 is the most relevant since the sample is self-selected, resulting in possible participation bias.

14 Guideline 8 is the most relevant since the conclusion in the headline is not consistent with the results of the poll. Many people consider a landslide victory to occur when a candidate receives 60% or more of the vote.

15 Guideline 6 is the most relevant since the wording of the question is biased and intended to elicit negative responses.

16 Guideline 4 is the most relevant since it is very difficult to measure the value of counterfeit goods (in any year).

17 Because companies involved in the chocolate business provided much of the funding for the research, the researchers may have been more inclined to provide favorable results, to search for only positive aspects of eating chocolate, or to report only results that would be deemed positive by the companies. The bias could have been avoided if the researchers were not paid by the chocolate manufacturers. If that was the only way to fund the research, then the researchers should institute procedures to ensure that they submit all results for publication, including any negative ones.

18 The sample is self-selected and the replies represent only a small proportion of the questionnaires sent out, so the responses were more likely to come from those with strong feelings about the issues. A better sampling procedure, such as interviews with a random sample of women, would have been better.

19 The sample is self-selected from among visitors to the Newsweek Web site. As such, they are likely to be people who have strong feelings about the issue. A better sampling method, such as the simple random sampling used by most polling companies, is needed.

20 The list of property owners is clearly biased toward those who can afford to own property. All of those who live in rented housing units would be excluded from participating in the survey. In addition, the responses come from a group of people who are self-selected. A better method of sampling, such as the simple random sampling used by most polling companies, was needed.

21 The results are not necessarily contradictory, but might appear to be so. The word "wrong" in the first question could be misleading or confusing. Some people might believe that abortion is wrong, but still favor choice. Such people would respond "yes" to the first question and "no" top the second. The second question could also be confusing, as some people might think that "advice of her doctor" means that the woman's life is in danger, which could alter their opinion about abortion. Groups opposed to abortion would be likely to cite the results of the first question, while groups favoring choice would be more likely to cite the results of the second question.

22 The first question refers only to "government programs," which many people tend to think of as being generally wasteful. The second question lists specific programs that are very popular among the general public. Groups favoring tax cuts would be likely to cite the results of the first question, while groups opposing the tax cuts would be likely to cite the second question.

23 The first question requires a study of Internet dates generally, while the second examines people who are married to see whether their first date was an Internet date. The first question is a more difficult one to study since, at the time of the study, some Internet dates will not yet have led to marriage, but may eventually. In addition, there is no good way to determine who is in the population of Internet daters.

24 The first question requires a study of introductory classes while the second requires a study of full-time faculty members. The results could be very different in their percentage terms. For example, it may be that every full-time faculty member teaches at least one introductory class along with one or more advanced classes. On the other hand, there may also be many additional introductory classes that are taught by part-time faculty. It is therefore possible to get 100% as a result for the first question while getting only 50%, say, for the second.

25 The first question involves a study of college students in general, and the second question involves a study of those who do binge drinking. The first question might be addressed by surveying college students. The second question would be addressed by surveying binge drinkers, and it would be much more difficult to survey or even identify this group.

26 The first question involves a study of college graduates and the second involves a study of people who have taken one or more statistics courses. The second group includes college graduates, college students, high school students, people who take statistics courses at their workplace, and people who take the courses for self-improvement purposes. The first question involves a group that is much easier to identify, locate, and survey.

27 The headline says "drugs" whereas the story says "drug use, drinking, or smoking." Because "drugs" is usually taken to mean drugs other than smoking or alcohol, the headline is very misleading.

28 The story does not include the margin of error for the survey. Since with 500 people surveyed, this is likely to be about 4%, the likely range for a satisfying sex life is 78% to 86% while the likely range for job satisfaction is 75% to 83%. Since these ranges overlap, it is quite possible that the conclusion in the headline, "Sex more important than jobs," is incorrect for the population as a whole.

29 No information is given about what the "confidence" refers to. For example, does it mean that the public is confident about the military leaders only in military situations, or in other situations (such as business or politics) as well? The sample size and margin of error are also missing in the report, but even if they were present, we still don't know what "confidence" refers to.

30 The report seems to be making an implication of restaurant quality in New York (the "Big Apple"), but there is nothing unusual about the case of New York City. With only nine scores of 29, most large cities will not have a restaurant with a score of 29. In addition, data are missing. What about restaurants receiving scores of 30, or 28, or 27? What criteria were used for the ratings? Without much more information, it would be difficult to act on these data.

31 No information is given to justify the statement that "more" companies try to bet on weather forecasting. If only the four cited companies are new, the increase is certainly insignificant.

32 The article suggests that China is thrown off balance by this improbable (under normal circumstances) ratio of boys to girls among newborns. This

suggests that some change is having this dramatic effect or that this imbalance in births is somehow having a dramatic effect on something else, but no information is given about any such changes. How is China thrown off-balance?

Chapter 1 Review Exercises

1 a) The range of values likely to contain the proportion of all households with guns is from 38% - 3% to 38% + 3% or from 35% to 41%.

b) The population consists of all adults in the U.S.

c) This an observational study because the subjects were not treated or modified in any way. The variable of interest is *gun in home*, which for this study, can take on either of two values, *yes* or *no*.

d) The 38% is a sample statistic based on the sample of 1012 adults, not the population of all adults.

e) No. In this case, the sample would be self-selected with a likely participant bias.

f) A perfect simple random sample of all adults in the U.S. is probably not possible since some have no phones and there are some with no addresses. You not only need a list from which to choose participants, but also a way to contact them. However, the percentage with phones or addresses is very high, so a sample taken from the population of those with phones will likely yield a sample that is very representative of the population. Therefore, we can use a computer to randomly generate telephone numbers, call those numbers until a desired sample size of adults has been contacted.

g) We could stratify the sample by state, taking a simple random sample of households in each state.

h) Select all of the households in each of a number of random selected voting precincts.

i) Systematic sampling would be difficult for the entire U.S. since it would require an ordered list of all adults in the U.S. However, one could select every 10^{th} address on each street in a city. Such a sample would be systematic, but it would not be very representative of the population of U.S. adults as a whole.

j) Select the households of your classmates. Again, this type of sample will not be representative of all U.S. households.

2 a) A simple random sample is one chosen in such a way that every sample of the same size has the same chance of being selected.

b) No. Not every sample of 2007 people has the same chance of being selected. For example, it would be impossible to select all 2007 people from the same primary sampling unit. In fact, it is impossible to select any sample that has two people from the same unit.

c) Randomly select a primary unit and then randomly select one of its members. For the second person, randomly select a primary unit (which might be the same as the first unit selected) and then randomly select one of its members. Continue doing this until the desired sample size has been obtained. If, at any time, a person is selected who has been previously selected, ignore any such selection.

3 a) No. There is no information about the occurrence of headaches among people who do not use Zocor.

b) Because the headache rate is lower among Zocor users than among the placebo group, it appears that headaches are not an adverse reaction to Zocor use.

c) With blinding, neither the participants nor the researchers know who is receiving Zocor and who is getting the placebo.

d) This clinical trial is an experiment because some of the participants are given a treatment and some are given a placebo.

e) An experimenter effect occurs if the experimenter somehow influences subjects through such factors as facial expression, tone of voice, or attitude. It can be avoided through the use of blinding.

4 a) The second question should be used because the word "welfare" has negative connotations.

b) Use the first question because it is more likely to elicit negative responses.

c) Some professional pollsters are opposed to all such questions that are deliberately biased and strive to make questions as neutral as possible. Others believe that such questions can be used. In any case, survey questions can modify how people think, and it is important that such modification should not occur with their awareness or agreement.

Chapter 1 Quiz

1 A. The sample is a subset of the population. Since the 1000 people were drawn from the registered voters in Texas, the population of interest is the set of all registered voters in Texas.

2 C. Those who contributed to the Governor's campaign are more likely as a group to approve of the job their governor is doing than is the population of Texas as a whole. This might be because he is doing things that they, as supporters of his campaign, expected him to do. It could also be that they don't want to admit to themselves that they supported a person that they no longer approve of.

3 A. A large sample is not necessarily representative (e.g., a convenience sample). Also, even if the sample was chosen in the best possible way, there is no guarantee that it will turn out to reflect the entire population.

4 B. Those receiving the financial reward comprise the treatment group. There is no such thing as an observation group (unless one calls the researchers the observation group).

5 C. The experiment is not blind since one group is told about the incentive for perfect attendance, while the other group is told that they are part of an experiment, but clearly they will find out that they are the control group.

6 A. A placebo is used that all participants receive some kind of treatment. This keeps them from knowing whether they are in the experimental group or the control group.

7 C. A placebo is not supposed to have any effect at all. If some people in the control group experience a result that is supposed to happen only in the experimental group, that is called a placebo effect. Placebos are not supposed to cure warts, so if some people in the control group have warts that are cured even though they haven't received a treatment designed to cure warts, then we have a placebo effect.

8 C. A single blind experiment is one in which the subjects do not know if they are in the control group or the treatment group.

9 B. We could be 95% certain from Poll X that Powell will receiver between 46$ and 52% of the vote, while we could be 95% certain from Poll Y that she will receive between 50% and 56% of the vote. Both polls will be correct if she receives 50-% and 52% of the vote, so the results of the polls are not inconsistent with one another.

10 B. The confidence interval extends from 12% - 4% to 12% + 4% or 8% to 16%.

11 B. The conclusion may be valid even if the study was biased. Answer C doesn't say anything because we don't know what "it" is.

12 B. (A) is not the answer because we don't even know if <u>most</u> Americans

even watch the show, let alone care who wins. (C) is not the answer, in part because the subjects (voters) are not subjected to any treatment. (B) is the correct answer, not only because the voters are self-selected, but also because some of them may vote a number of times.

13 B. If you are measuring the weights of 6-year-olds, the variable of interest is the weight of a 6-year-old.

14 C. People are seldom in the sun don't sunscreen. Those are often in the sun are more likely to use sunscreen. In addition, some of the people using sunscreen are doing so because they previously got sunburned and they don't want that to happen again.

15 B. Whenever we do a statistical study using a sample from a population, there is always a small chance, even when every thing is done correctly to try to ensure that the sample is representative of the population, that the conclusions drawn about the population based on the sample results are not correct.