Chapter 4

Discussion Question Solutions

D1. a. The population consists of the new models produced by the manufacturer. Census is used. Each new car is inspected. This procedure is used because every customer expects his or her new car to be perfect. Also, if a safety-related problem gets by in even one car, the cost is high.

b. The population is all chocolate chip cookies produced. Sampling is used. Counting the chocolate chips in a cookie is destructive, and counting all the chips in every cookie is time-consuming.

c. The population is the set of voters in a presidential election in the United States. Sampling is used. Theoretically, elections in the United States are a census of voters, but because not everyone who is eligible actually votes, elections are in practice a nonrandom and probably nonrepresentative sample.

d. The population consists of individual movie theater attendance figures. Sampling is used. Not every movie theater owner reports ticket sales every weekend.

e. The population is the set of all teachers in Los Angeles. Sampling is used. With thousands of teachers in the Los Angeles area, it would be too time-consuming and expensive to do in-depth interviews with all of them.

D2. Suppose you have 32 students. Assign a two-digit number to each student: 01, 02, 03, . . . , 32. Then, go through the random digit table one pair of digits at a time until you find six of these assigned numbers. Ignore 00, any pairs that give a number larger than 32, and any duplicate pairs. The sample should be representative on some variables and unrepresentative on others. Individual results will vary.

D3. a. The population is the set of all possible test tube samples from the reservoir. One sampling procedure would be to divide the surface of the reservoir into 20 equal size areas and fill a test tube from each area. If the depth at which the sample is taken is important, the surface could be divided into 10 areas with a two samples being taken in each area at two different depths.

b. The population is all recreational boats on the lake. If a list exists of all the boats, that is, if the owners are required to register their boats say, then a random digit method such as that outlined in D2 could be used to select a sample.

c. The population in part a. is extremely large, much larger than in part b., which could mean a sample of size 20 is nonrepresentative. It seems likely that a representative sample of boats on the lake is possible. A census may be possible in part b. but not in part a.

D4. Number the characters in the book. To be concrete, suppose there are exactly one million characters. Use a computer or random number table to choose 10,000 random numbers between

1 and 1,000,000. Each of these 10,000 numbers is the number of a character in your list—that character goes in the sample. Finding the characters in the SRS would take forever, so this is not a very practical way to sample! Treating each page as a cluster and taking a two-stage cluster sample would be much more efficient.

D5. It is reasonable to stratify students by class year if there is good reason to expect that the time spent traveling to school varies by class year. This would be the case if, for example, older students tend to drive themselves to school while younger students tend to walk. However, a better method probably would be to stratify students according to how far they live from school. That variable probably has the biggest effect on travel time and so would make the strata as different from each other as possible.

D6. Cities are typically divided into neighborhoods. Further, neighborhoods often correspond to economic status: lower, middle, and upper class. One stratified random sampling plan would be to use the neighborhoods as the strata and the choose the relative sizes proportionally to the size of the neighborhoods. Given a HDTV is a luxury item, one would expect little variability in the responses in the upper class neighborhoods and greater variability in the middle and lower class neighborhoods.

D7. Not all possible groups have a chance of being in the sample. For example, two friends who are standing in line together at the time of the sample cannot both be in the sample.

D8. a. This is closest to three-stage cluster sampling. First, one lot was chosen. (This is like choosing one cluster—a lot—from all possible lots.) Then three containers were chosen from that lot—again like cluster sampling. At the final stage, however, not all pills in the bottle were in the sample, only ten were. We do not know whether the sampling was random at any of the three stages.

b. It might be better to use more lots, more containers, and possibly fewer pills per container. The key is to sample more where there is more variability. If there is great variability among lots, then more than one lot of each product should be sampled. If, on the other hand, all the pills in any one container are nearly identical, then ten pills may be more than you need in the sample.

D9. One suggestion is to sample from auto registrations. Another is to take samples from parking lots (clusters of cars) and other open areas where cars are visible. The latter method might introduce bias because it is conceivable that owners of one type of car (foreign or domestic) might be more likely to keep the car out of sight than owners of the other type of car.

D10. In stratified random sampling, the population is divided into groups (strata) that are as different as possible but the individuals within a group are as alike as possible. Samples are then selected randomly and separately from each group (strata). Stratification is useful because you can expect the sample mean to be closer to the population mean than it would be with simple random sampling. Cluster sampling tends to be used when people are already in groups that cannot easily be separated, such as classrooms of students or cartons of food; everyone in the sampled cluster is in the sample in single-stage cluster sampling. Cluster samples are often easier to obtain in practice. If you have the option of constructing your own clusters, ideally each

cluster should contain as much variation as is in the population as a whole, just the opposite of the optimal rule for constructing strata.

D11. Agree on an ordering of the students in your class, perhaps by rows of seats or by names on the class roster, and number them. Divide the number of students by 5 to see how many possible systematic samples there are. For example, if there are 33 students in the class, a random start between numbers 1 and 6 would give you six possible systematic samples. But the 3 students at the end of the list have no chance of being sampled under this plan. There is no good solution. Suppose you start at random between 1 and 7. If the random start is student number 6, then the sampled students are 6, 13, 20, 27, and 34. Because there is no number 34, start over again and turn 34 into number 1. But now numbers 1 and 2 have twice the probability of being selected as other students.

D12. a. The landline survey indicates neither Obama nor McCain was favored at the time of the survey. On the other hand, the cell phone respondents clearly put Obama in the lead. One explanation could be that a greater percentage of cell phone users are Democrats or that they are more willing to respond to a marketing research survey on their cell phones.

b. The survey results for those under age 30 displays a much higher percentage of Democrats among both the cell phone and landline respondents. The percentage of the Other preference basically remains constant across phone types in both the under 30 class and also in the general registered voter survey. The high percentage of Democrats in both the landline and cell phone respondents under 30 may indicate that younger people and newly registered voters typically are registered Democrats. Note this may explain the difference observed in part a. If people under 30 are more likely to use their cell phones as their primary or only telephone, then the cell phone category in part a. may have included a much larger percentage of voters under 30.

c. The data in a. and b. can be used to argue that cell phone users form a unique demographic with different voting patterns. Further, if there is a strong relationship between voters under 30 and cell phone users, then the omission of cell phone respondents from election polls could omit a significant part of the population resulting in bias in a preference poll.

d. Any telephone survey will contact people who will hang up or simply not answer the phone. In the case of a cell phone users who pay by the minute, not answering would likely be more common. This could lead to bias as discussed in part c. Over time this could change as unlimited calling plans become more common and affordable. Since the TCPA specifically bans automated calls to cell phones, again the cell phone demographic could be excluded resulting in bias.

D13. a. This is a convenience sample. The estimate could be too high if, for example, the class is an AP Statistics class where students are typically more motivated than the general population of students.

b. This is a volunteer sample, and the estimate would likely be too high. Graduates who feel they have not lived up to the expectations of old classmates might be less likely to respond. In addition, some people may exaggerate a bit in their response.

c. This is a random sample of people with dementia alive on a given day. However, size bias will affect the estimate wanted.

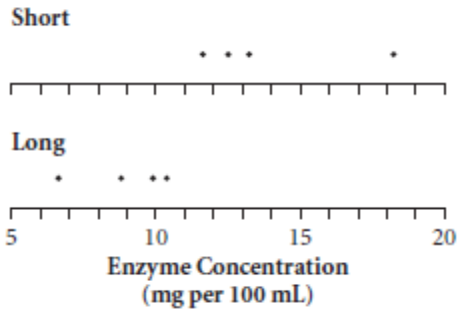
D14. Answers will vary, but students should realize that the frame does not represent the population well. Two groups that belong to the population but not to the frame are those likely to vote who have unlisted telephone numbers and those likely to vote who do not have a residential telephone. Those too young to vote or those who are not registered can be in the phone book but not in the population. Because a substantial number of eligible voters do not go to the polls, the frame (phone book) will list many nonvoters and unlikely voters. These people are not a part of the target population. With such a poor fit between frame and population, it is hard to say with confidence what sort of bias there might be.

D15. Those graduates who have been successful are more likely to respond to surveys about the high quality of their education. They will probably be happier with their education than those who do not respond.

D16. The second question: 62% of Americans agreed with the statement that it is not the government's job to financially support television programming. On the other hand, when Americans were asked if they would be disappointed if Congress cut its funding for public television, 54% said yes.

D17. Nonresponse bias occurs when some of the sampled units, perhaps in a well-designed random sample, could not be measured for one reason or another. Voluntary response bias comes from poor design of the sampling plan. It is often possible to measure nonresponse bias by working a little harder to get some responses from those who initially formed the nonresponse group. It is usually impossible to measure the bias in a voluntary response survey.

D18. This stacked dot plot shows the concentrations of the enzyme (Na^+ K^+ ATP-ase) in the brains of Kelly's hamsters. The difference in the responses for the two treatments supports Kelly's hypothesis that there tends to be more of the enzyme in the brains of hamsters raised in short days.



D19. Almost. If we are convinced that the two groups were treated exactly alike except for the length of their day, only two possible explanations can account for the difference: (1) all the hamsters with the higher concentrations of this enzyme happened to get assigned to the short-day group or (2) the day length caused the enzyme concentrations to change. Because she randomly

assigned the hamsters to the two treatments, it is unlikely that all the hamsters with high enzyme concentration were assigned to the short-day group. (In fact, the probability of getting the four highest values in the *short days* treatment group is only $\frac{1}{70}$ and the probability of getting the four highest values in the same treatment group is only $\frac{1}{35}$.)

D20. a. No. It may be the other way around. The respiratory problems may have caused the thymus to enlarge.

b. For an experiment, randomly divide the children with respiratory problems into two groups: one that gets surgery and one that does not. Otherwise, treat them exactly alike. Record how many survived under each treatment.

D21. For example, older people are more likely to get cancer than younger people. If older people are also more likely to have sedentary jobs than younger people, then the researchers wouldn't know whether the higher cancer rate of those with sedentary jobs was because of the sedentary jobs or the older ages. Amount of exercise and age would be confounded. So, the researchers certainly must have matched a person with a sedentary job with someone of the same age with a non-sedentary job. Another example of a variable that might be confounded with amount of exercise is smoking. Smokers have a higher cancer rate than nonsmokers. If smokers tend to be more sedentary, then researchers wouldn't be able to tell whether the higher cancer rate in those with less exercise was because of the amount of exercise or because there were more smokers in that group. Other variables that might be confounded with amount of exercise include diet and, for some cancers, obesity, ethnicity, exposure to certain chemicals in the workplace, and sex.

D22. In an experiment, treatments are assigned to units. In an observational study, the conditions you want to compare come already attached to the units.

D23. Experiments should be blind because of the placebo effect and because people are eager to please. People do better, for example, when they know they are getting a drug rather than the placebo. People who participate in an experiment want to help it have a favorable outcome and have been known to try harder when getting a new treatment. The experiment should be double-blind for the same reasons. The person who evaluates the effects of the treatment should not know which treatment the person received so that subconscious beliefs about which treatment is better do not shape his or her judgment and analysis, and so that the person's expectations won't be unintentionally conveyed to the subjects.

D24. Assign subjects who suffer from depression at random to three groups, one to receive St. John's worth, one to receive an antidepressant drug, and one to receive a placebo. Treat the subjects exactly the same otherwise, including having the pills look exactly alike. After at least eight weeks, have a doctor evaluate the extent of the depression. The doctor should not know which treatment the person received. Students may suggest the good strategy of evaluating the extent of the depression at both the beginning and the end of the study and recording the degree to which it improved.

D25. Assign volunteers randomly to one of two groups, one to receive nothing and one to receive a placebo. Those receiving the placebo must think they are getting a real treatment. This experiment cannot be completely blind because the point is for the subjects to know whether they are getting a treatment or not. Have subjects report whether their pain has improved, and compare the results of the two groups.

D26. The effectiveness of a method for teaching reading highly depends on the teacher and the classroom atmosphere. The reading scores of the children in the classroom are not independent. Typically, an entire classroom will do well, or poorly. Besides, it would be difficult to randomly assign different teaching methods to be used on different students within the same classroom. Remember: *The things that receive the randomization are your experimental units.* For the experimental unit to be the school, the researcher would need several schools and each school could be randomly assigned one of the two teaching methods.

D27. The things that receive the randomization are your experimental units. Since the hospitals were randomly assigned the treatments, the hospitals are the experimental units.

D28. For the same reason that you have more faith in a poll with a large sample size—as the sample size gets larger, the percentage of successes in the sample tends to approach that percentage of successes in the population. In an experiment, a large number of subjects tend to ensure that the ones in each treatment group are more representative of the whole group designated for study in the experiment. Also, as you will see in later chapters, larger numbers of measurements allow for more powerful inferences because they provide better estimates of variability.

D29. Yes. The completely randomized design resulted in so much within-treatment variability that any difference between treatments was obscured. The dot plots in Display 4.19 do not show any noticeable difference between pulse rates from the different treatments other than a much greater spread for standing. Both of the randomized paired comparison designs showed that the standing pulse rate was generally greater than the sitting pulse rate, since most differences $(stand - sit)$ were greater than 0.

D30. With the repeated measures design, only one difference was less than 0, compared with four differences less than 0 for the matched pairs design. The repeated measures design shows stronger evidence of a difference between treatments with a mound-shaped distribution centered at about 5 with SD about 4, whereas the matched pairs design is not quite as mound-shaped, with center at about 5 but with SD about 10.

D31. In a completely randomized design, treatments are assigned randomly to the subjects. In a randomized block design, the subjects first are divided into blocks of subjects that are expected to be somewhat homogeneous with respect to the characteristic being measured (the response) and then the treatments are assigned randomly *within each block*.

D32. Blocking is desirable because you can then look, within each block of like subjects, at the differences between the treatments. This method reduces the variability due to individual differences in the subjects. Students' examples of where you would want to block should involve

situations where individual differences would be large compared with the difference between treatments. In general, you would not want to block in situations where it is impossible or inconvenient. As an example where the completely randomized design is best, suppose you are a theater owner and you want to determine the effect of giving movie-goers a flier advertising popcorn when they buy their tickets. Your response variable will be whether the ticket purchaser goes directly to buy popcorn or not. If you expect, say, 200 people at the movie where you are doing this experiment, you would have to use a completely randomized design, assigning each treatment randomly to the people as they enter. You can't use a matched pairs design because you can't examine the people beforehand in order to match them. You cannot use repeated measures because once a movie-goer gets one method, you cannot use the other method on the person.

Practice Problem Solutions

P1. a. The population is the collection of all the households in your community; the units are the individual households.
b. Sampling is quicker, cheaper, and easier than conducting a census.

P2. This is a two-stage cluster sample, not a simple random sample. Any sample with more than 10 students from a single dorm, for example, cannot occur by this process.

P3. a. No
b. No; two students with last names starting with different letters cannot both be in the sample.
c. No; two students sitting in different rows cannot both be in the sample.
d. Yes
e. No; a group of six girls cannot all be in the sample.

P4. Design c is not valid because the groups proposed as strata are not mutually exclusive.

P5. a. Compute the respective percentages of 1200: 516; 384; 300
b. Age since 40 seems to be an established cutoff point for likelihood of getting receding gums.

P6. $84(0.65) + 69(0.35) \approx 0.79$

P7. a. Choose a random start between persons 1 and 20, and take every 20th person.
b. Choose a random start between persons 1 and 5, and take every 5th person.

P8. Choose a random start between 1 and 17, and take every 17th person.

P9. a. Pick a random sample of zip codes. The sample consists of all households on the carrier routes along in each zip code in the sample.

b. Start as in part b, then take a random sample of households within each "zip-code cluster" – maybe certain streets within the carrier route within each zip code.

P10. a. Use pages as clusters. You can concentrate on only a small number of pages in the book, but some sections of the text may contains more capitalized words than other parts (for instance, an author index). A sample containing such a non-representative page would through off the results.

b. Take an SRS of pages. Then take an SRS of lines from each of those pages.

c. Take an SRS of characters from each line selected in part b.

P11. a. Al. The shots are closely spaced but off center.
b. Cal. The shots are widely spaced and off center.
c. Bal. The shots are closely spaced and cluster around the center.
d. Dal. The shots are widely spaced but cluster around the center.
e. Al. His shots are consistent. You just need to get him to adjust his direction. That would be easier than getting one of the others to be more consistent.

P12. a. Size bias. Larger farms are more likely to be selected than smaller farms.

b. This is a judgment sample. In his attempt to get a diverse sample, the professor may have inadvertently selected too few valedictorians from the most typical high schools. He may also have missed some important groups. The students may also see this as a convenience sample because the professor only used high schools in Illinois.

c. Voluntary response bias. Teachers whose students did well are more likely to report. Overall, the teachers from the AP discussion list reported that 426 out of 535, or 80%, passed (got a 3 or better). When the official results were released, the pass rate was 62.1%.

d. Size bias. Longer strings are more likely to be selected.

e. Voluntary response bias. Responders tend to have stronger opinions than nonresponders. Even though Ann Landers received a very large number of responses, a large sample offers no guarantees about bias. Voluntary response samples are so likely to be biased that you should not trust them.

P13. The estimate will be too high. The sample will catch almost all those who came to Boston to see the museum, but only a tiny fraction of those who came to Boston for other reasons because those in the latter group are not likely to go to the museum at all. This is an example of a convenience sample.

P14. Question I is more likely to draw "yes" responses, which would show support for some control of gun ownership. Question II is more likely to draw "no" responses, which would indicate a lack of support for controls.

P15. This is incorrect response bias where people gave an answer that they thought made them look knowledgeable. They probably considered Princeton's overall reputation and assumed that any program there must be strong.

P16. a. The subjects are the children; the response variable is having/not having leukemia; and the treatments are living near to or not living near to a major power line.

b. No, it's merely observational. Samples of children are taken from a collection of areas, some near to a major power line and some not, and they are assessed as to whether or not they have leukemia.

c. It is very difficult to prove a causal relationship. They can list many other factors that could contribute to the children developing leukemia.

P17. a. Number the twelve test sites and randomly select four of them for each type of glass. Collect the measures for the amount of energy lost at each site and compare the results from each of the three types of glass.

b. The subjects, or experimental units, are the twelve site-generator combinations. The treatments are the three types of glass. Each treatment is replicated four times.

c. If each site had six generators, each treatment would be randomly assigned two generators at each site. The experimental units would then be the 72 site-generator combinations, but you are "blocking" on sites so that each treatment gets studied fairly on each site. You will learn about blocking in the next section.

P18. a. The lurking variable is the person's age. Older people are more likely to do both.

b. I causes II. Experiments have shown that drinking more milk, which is high in calcium, results in stronger bones. Of course, there are other causes of strong bones, such as heredity and exercise.

c. II causes I. People who go to school longer tend to earn more money. However, this statement is based on observational studies, not on randomized experiments, so occasionally the cause and effect relationship is questioned. Some people believe that family background is a lurking variable: People from more well-off families tend to go to school longer and to earn more money, but it isn't the schooling that caused the higher income, it was their family's support in getting established. However, most people believe that although family background certainly contributes, the number of years of schooling is another cause of higher income.

P19.

		Motivation	
		High	**Low**
Took Course?	**Yes**	Higher SAT	No evidence
	No	No evidence	Lower SAT

P20. The variable "time spent on the activity" could be confounded by the use of activities other than those assigned, in the study or outside of it.

P21. a. The factors are type of lighting (brightness of the room) and type of music. For the type of lighting, the levels are low, medium, and high. For the type of music, the levels are pop, classical, and jazz.

b. Answers will vary. Possibilities are heart rate, blood pressure, or self-description of anxiety level.

P22. a. Students should not believe this. Their explanations will vary but the actual explanation is in P23. The pipe- and cigar-smokers are a bit older on average.

b. Observational study

c. The factor is smoking behavior. The levels are nonsmoking, cigarette smoking, and pipe or cigar smoking. The response variable is the number of deaths per 1000 men per year.

P23. Older men have a higher death rate; the pipe- and cigar-smokers are older than the others. Here are the age-adjusted death rates per 1000 men: nonsmokers 20.3, cigarette-smokers 28.3, and pipe- and cigar-smokers 21.2. The new factor is the age. The discussion loosely follows Paul R. Rosenbaum's *Observational Studies* (New York: Springer-Verlag, 1995, page 60). (These data come from William G. Cochran, "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," *Biometrics* 24 (1968): 205–213. The original study was by Best and Walker.)

P24. a. This study is observational.

b. The factor is the legal age for driving. The levels are the age groups. The response variable is the highway death rate by state.

c. The "treatments," legal driving age, may be confounded with driver education, as states with higher age limits may generally be more restrictive about getting a driver's license and so also require driver's education courses. Also, if it is the case that the legal driving age tends to be lower in western states, the geography of the states would be confounded with age because western states with many miles of high-speed highways tend to have higher death rates than congested eastern states.

P25. a. The response variable is death or survival; the explanatory variable is low birth weight.

b. This is an observational because the data had already been obtained prior to the study and are simply being analyzed.

c. Low birth weight appears to be strongly associated with neonatal death.
d. The following could be confounded with birth weight: other diseases or physical

impairments of the baby; health of the mother; difficulty of delivery.

e. Within each group, the percentages are nearly equal and most of the deaths come from the premature births; thus, it is unclear if the deaths are mainly due to low birth weight, or premature birth (or some other cause).

P26. For his control group, Dr. Mayo used adults who died of nonrespiratory causes. He was right to use a control group, but he failed to isolate the effect of interest because when it comes to thymus size, adults and infants are not comparable. If Mayo had used infants who died of nonrespiratory causes as his control group, he would have seen that they all had large thymus glands and would not have concluded that the thymus was responsible for the breathing problems. The placebo effect could have been a factor because Dr. Mayo was a trusted physician. It was not possible for this study to be blind or double-blind because it was obvious to everyone which infants had had surgery.

P27. a. There are quite a few problems with this study. It is possible that students were pointing home not from some magnetic sense, but because they could feel how much "turning" they had done or because they knew the roads. It is also possible that they could tell direction because they knew the direction of the sun even though they were blindfolded.

b. The group that has the magnets.

c. Whether the students were assigned randomly to the treatments. Students may also mention that the study was not double-blind.

d. Answers will vary, but the second design is certainly better than the first.

e. Because the subjects could tell whether or not they had the magnet because of the magnet's attraction to the walls of the van, this study was not blind. It was not double-blind either because the experimenter who evaluated how well the students pointed home was the same person who chose which kind of bar they wore.

P28. <u>treatments</u>: the two text books <u>units</u>: the 10 classes
<u>sample size</u>: 10, with 5 for each treatment.

P29. The experimental units are carnation plants, and the ones used for the experiment should be as nearly alike as possible. Randomly assign the new product to about half of a number of plants, and leave the others growing under standard conditions. There must be at least two plants in each group, and preferably many more. The plants receiving the standard treatment can be used as a control; the goal is to see whether the new product is better than the standard treatment at producing large blooms. Putting the plants in their places before randomly assigning treatments helps ensure that the two treatments aren't confounded with variables such as the amounts of heat, light, water, and temperature the various plants receive.

P30. a. The first factor is the brand of paper towel, with levels Brand A and Brand B. The second factor is wetness, with levels dry and wet. The response variable is the number of pennies it can hold before breaking.

b. Randomize the assignment of wet or dry to each set of ten towels (five from each brand to each treatment) and the order of testing. There are four treatments: Brand A wet, Brand A dry, Brand B wet, Brand B dry. Write each treatment name on five slips of paper, then draw to determine the order. The experimental units are the 20 time-towel combinations in which the tests will be performed.
c. This is an experiment, although the results may not generalize beyond the towels tested unless it can be established that they are essentially the same as all towels of these brands.

P31. a. Students should list reasons why dorms that get the same soap still vary in the number of visits to the infirmary. Some possible reasons include the following: Some dorms have a less healthy atmosphere than others (e.g., they are too stuffy or contain contaminated drinking fountains), some dorms may be for athletes or others that tend to be healthier on average, or some dorms may have students who just happen to be sicker more often than usual.

b. To equalize variation between the treatments.

P32. a. randomized paired comparison with repeated measures

b. randomized paired comparison with matched pairs

c. The choice to use randomized paired comparison with matched pairs was likely made because there may be a residual effect of one or both of the drugs—that is, it does not clear out of the bloodstream in the time allowed between treatments.

P33. The most striking difference would likely occur in the factor "whether the person can roller skate." One should block on the answer to this question so that prior experience with skating doesn't confound the results.

P34. a. There is a lot of variability in how well students memorize, so some sort of blocking seems desirable. Your students may suggest a randomized paired comparison design with repeated measures. First, find the seniors in the school who are willing to participate. Then randomly select half to go into a room that has the radio playing and the other half to go into a room that is quiet. The students would be given a familiar task to study, such as a list of new vocabulary words. Afterward, they would be tested on the meaning of the words. The number that they remember is the response variable. Then, the students would switch rooms and be given another new list of vocabulary words. If your students feel that this would take too much time, they might suggest matching students by GPA or, if that is impossible, a completely randomized design.

b. For this experiment, the treatments are soup with MSG and soup without MSG. The

response variable is the amount of soup eaten by a customer who orders soup. Even though there is a lot of variation in how much soup people will eat, blocking seems difficult to do. (However, it might be possible to block by estimating a person's weight.) In this case a completely randomized design seems best. For each customer who orders soup that night, a treatment is randomly selected, perhaps by flipping a coin each time.

P35. a. The question is difficult to answer from the plots. The plots show that the number of eyelid flicks per minute cover about the same range with both treatments, and there is no way to tell which dots represent measures from the same patient.
b. Here it is clear that the average number of flicks per minute is higher for Drug A, indicating that Drug A works better than Drug C for reducing drowsiness. If they were equal, the points would lie on the line.

c. You could subtract the number of flickers per minute for Drug C from that for Drug A. If the differences are generally greater than zero, Drug A is more effective at reducing drowsiness.

d. This is a randomized paired comparison with repeated measures. By using the same unit for each treatment, the within-treatment variability is reduced, allowing the between treatment variability to become apparent.

P36. a. Green bears went farther. The display below shows the average distances for each team for both red bears and green bears, using the data in the student text. On the surface, it looks as though color has a real effect on launch distance.

Team	Red	Green	Difference
1	26.2	32.4	6.2
2	54.4	89.8	35.4
3	31.8	60.2	28.4
4	30.0	34.8	4.8
5	31.2	41.2	10.0
6	102.6	115.4	12.8

b. The greater distances traveled by the green bears is due to confounding of launch order with bear color. As students got more practice, they were able to launch the bear farther. This table shows how the variables are confounded.

		Practice Doing Launches	
		Less	More
Color of Bear	Red	Shorter distances	No evidence
	Green	No evidence	Longer distances

c. The plots show a clear time trend: Later launches tend to travel farther, because most people get better with practice. (If you do this experiment and your class gets results that do not show a time trend, it is worth spending a few minutes trying to figure out why. For

example, several practice launches before starting to record data could minimize confounding of the launch order with bear color.)

d. Randomize the order in which the bears of different colors are launched. For example, put five bears of each color in a box and mix them up. For each launch, draw out a bear without looking. Don't return that bear to the box. A more risky strategy is to use practice launches until the learning effect appears to go away.

Exercise Solutions

E1. a. This is a stratified random sample with strata of grocery store owners and restaurant owners.

b. Perhaps a systematic sample of customers coming into the businesses would work.

E2. a. Cookies are more typically packaged together in grocery stores rather than being sold separately. So, it is more convenient and practical to use cluster sampling, where each bag is considered a cluster.

b. Choose, at random, entire bags of cookies to obtain a cluster sample.

c. Start as in (b), but then randomly choose a predetermined number of cookies from each bag to obtain the final sample.

E3. Stratification by gender is likely to be the best strategy. This could potentially avoid cost bias in the sense that females are more likely to spend a higher amount on hair cuts than males.

E4. Choose every 20^{th} page to form such a sample of 30 pages.

E5. Consider the farms as the five strata, and take a random sample of, say, 10 acres from each farm.

E6. Stratified random sampling by age (adult, teen, child) and then choose at random (0.64)(50)= 32 from the adult group, (0.30)(50) = 15 from the teen group, and (0.06)(50) = 3 from the child group to get a representative sample.

E7. Systematic sampling with random start in both cases; in the second case this leads to stratified sampling.

E8. (1) is more likely since, all things being equal, you would assume an equal number of women and men to choose from, and choosing 25 people at random is highly unlikely to yield 25 of the same gender.

E9. a. Too high. Your polltakers will never get in touch with people who are away from home between 9:00 a.m. and 5:00 p.m., so these people will eventually be dropped from the sample. Because these are the people who are less likely to have children under the age of 5 at home (those with small children are more likely to stay home to take care of them), the sample will contain an artificially high percentage of those who do have children under the age of 5 at home.

b. At dinnertime, adults are more likely to be at home.

c. This is a case of sampling bias. The problem is with the design of the survey: calling between 9 a.m. and 5 p.m. The people not at home did not get the opportunity to respond.

E10. a. The population is all customers of the wholesale food distributor. The sample is the group of customers randomly selected by the sampling firm.

b. This is a sampling bias based on convenience.

c. Busy restaurants would be underrepresented, as would the owner-operated restaurants.

E11. a. A lower percentage of abstainers responded to the mailed survey, perhaps due to lack of interest; among the drinkers, a higher percentage were found to be frequent excessive drinkers in the follow-up, perhaps more accurately brought out by the interview.

b. Yes. This is not surprising because excessive drinkers may be reticent to respond to such a survey for fear of ramifications if the results were to ever be non-anonymous.

E12. a. For the "active and collaborative learning" and "student-faculty interaction" questions, the N group exhibited a higher level of engagement than the R group, while the opposite was true for the "enriching educational experience" items.

b. Nonresponse effectively lowered the means of the items of the first two categories and raised the means in the third category. This throws off the observed levels of engagement.

E13. a. Among adults aged 18 to 24, 39.1% are not at risk.

b. The young and the old have the higher percentages at risk, due perhaps to risky life styles among the young and illness among the old.

c. It is self-reported data on sensitive issues.

E14. a. Among adults aged 18 to 24, 46.9% of males are at risk for obesity and 38.3% of females are at risk.

b. The pattern for both males and females is an increasing change of being at risk of obesity, up to age 65, then a slight drop in risk beyond 65 years. Males, in general, are at higher risk for obesity than females for all age groups.
c. The "at risk" numbers seem a bit high, so perhaps people who were obese were more vested in the study and hence more likely to reply.

E15. This is a convenience sample, resulting in sample selection bias. Students who take a statistics course are more likely to like math than students in general, unless it is a required class. The estimated response will tend to be too high in the first case and about right in the second case.

E16. This convenience sample probably will be too optimistic about how well their candidate will do.

E17. From respondents 40 years old or older you would expect to get an estimate that is too high. 40-year-olds are older than average. As people get older, they tend to visit more and more states. (They can't visit fewer!) From the residents of Rhode Island you might get an estimate that is too high as well, as Rhode Island is a small state close to many other small states. Compare this result to what you might expect from asking people who live in Texas or Montana.

E18. There are tens of millions of adult males in the United States. It would be impossible to contact them all. Questionnaire bias is unlikely for a factual question like this. Assuming that a nonbiased sampling method is used, there still might be an incorrect response bias because some people might be reluctant to admit to smoking.

E19. a. A convenience sample seems likely here.

b. No. The results of the survey are a bit suspect because they were obtained using self-reported data in a survey on guitars conducted by a guitar manufacturer.

c. Less than 67% because those interested in learning to play the guitar may have no interest at all in video games. In fact, it is arguable that those serious about playing an instrument would not tend toward the video game version; rather, those who are less talented may be more apt to pursue the video game version as a compromise measure.

E20. a. This is a volunteer sample.

b. No. With a volunteer sample, people with stronger feelings one way or the other are more likely to respond.
The Consumer Union is well aware of this, as evidenced by their disclaimer: "These polls are not scientific. It reflects the opinions of only those Internet users who have chosen to participate. The results cannot be assumed to represent the opinions of Internet

users in general, nor the public as a whole. Consumers Union is not responsible for content, functionality or the opinions expressed therein."

c. Answers will vary. It is difficult to say whether people who drive less would be overrepresented or underrepresented in the voluntary sample.

E21. The estimate will be too high. There are two reasons. First, families with no children have no chance of being in the sample. Second, families with many children will be over- represented. In Activity 4.1a, a person who stays 5 days is five times as likely to be chosen as a person who stays 1 day. Similarly, families with 5 children would be five times more likely to be represented in the sample than families with only 1 child.

E22. Sampling Method II will give you an unbiased estimate of the average class size. The unit is the class, and you are taking a random sample from the classes. Sampling Method I is a size-biased method, and the estimate will be too high. Because there are more of them, students who are in larger classes will be more likely to be selected than students in smaller classes. In addition, more than one student might be selected from the same large class, thereby inflating the influence of that class size.

E23. From the Gallup poll, 36% were in favor of U.S. air strikes. From the ABC poll, 65% were in favor. Note that the ABC News poll mentions the involvement of allies of the United States and also makes it sound as though the air strikes will hit only objects and not soldiers. Students might reasonably think that the Gallup poll will produce a higher favorable response, because it mentions why the United States would want to conduct air strikes, i.e., it justifies air strikes. It can be hard to predict in advance which direction the psychology of questionnaire bias will go.

E24. Here's a possible way to word the questions, in this case about the issue capital punishment:

I. Should society have the right to protect itself by executing convicted murderers who are known sociopaths and likely to kill again if given the opportunity?

II. Given the established fact that innocent people have been falsely convicted and executed, do you favor life imprisonment without possibility of parole as an alter native to the death penalty?

E25. There is no right answer here, but some answers are better than others. A good way to potentially balance the characteristics is to take a random sample. This method may not perfectly balance the characteristics, but has a better chance of producing something close to a balanced sample than most other methods.

E26. Examples will vary.

E27. Observational study. Grade level is the 'treatment' and this cannot be assigned randomly to students.

E28. Observational study

E29. **a.** Randomization: yes, five plants are selected at random from the ten.

Replication: yes, each treatment is applied to five plants
Control: yes, five plants are replanted in their normal environment.

b. All ten needed to be dug up so that the variables *shade* and *being dug up* are not confounded.

E30. **a.** The factors are sintering temperatures and type of powder. The levels for temperature are the three different temperatures and the levels for type of powder are the two types. There are six treatments: each combination of a level of temperature and a level of powder type.

b. Since only one oven is available, only one temperature can be considered at a time. The experimenter should place equal numbers of each type of powder at randomly selected locations in the oven for each run. Then, the order at which the temperatures are run should be randomized across the time slots available.

c. Blinding would require a second person to assist. The assistant could set and record the temperature and the type of powder and the metallurgist would then prepare the disks.

E31. **a.** dormitories **b.** 20 **c.** experimental

E32 **a.** the dishes of insects **b.** 8 **c.** experimental

E33. **a.** Factors are feeding method, education program, and hospital stay, each at two levels; all possible combinations lead to 8 treatments

b. Researchers may have randomized treatments as follows: List the n mothers who volunteer to participate in the study; form a box of n chips of 8 different colors (one for each treatment) with the numbers of chips of each color being as equal as possible; draw a chip at random for each mother.

c. No; one cannot assign birth weight to subjects.

E34. **a.** A possible response variable could be the percentage of hatchings from each enclosure. An explanatory variable could be the type of cover used for the enclosure.

b. A factor is the type of cover. The levels are (i) cover that blocked UV rays, (ii) cover that did not block UV rays, and (iii) no cover.
There are six treatments (% birth and % non-birth for each of the three levels of cover).

c. The experimental units are the 12 enclosures housing the eggs.

d. To be able to determine if a cover without UV protection had any effect, in contrast to no cover at all, on % birthrate.

E35. A reason for using death *rate* instead of total number of deaths is that the sizes of the populations are different.

There is a possibility that climate is confounded with age. Florida has a much higher proportion of older people than does Alaska because Florida is a state where a great number of people move after retiring.
This is an observational study.

E36. a. Graduation rate could be the proportion of enrolled seniors who graduate from college, or the proportion of the freshman class four years prior that graduate within four years. (Which of these is likely to produce higher graduation rates?)

b. The main issue is the question you are trying to answer. If you want to simply see what graduation rates are at different schools you would do a census or a sample survey. Cost or time would be the main issue to determine which of these you want to do. If you want to determine something about the cause of graduation rates you would want to do an experiment or observational study. You could do an experiment if there is some way to randomly assign treatments to the units. If there is not, you would need to do an observational study.

c. No, since we are simply looking at data for the schools and not randomly assigning treatments, we cannot determine cause.

E37. No, not unless the subjects are in random order to begin with. Suppose the subjects in poorest health are all at the end of the list. Then, all of those in poorest health probably are going to be assigned to the same treatment group, a serious confounding effect.

E38. No. Consider the first two subjects to arrive. No matter how the coin comes up, these two will be in different groups. Thus, not all possible groupings are possible, and they certainly are not equally likely. (If the subjects had been paired, this would be a good way of splitting pairs in a matched pairs design, to be considered in the next section.)

E39. The modified SAT study is a completely randomized experiment, where students are randomly assigned to two groups, one group for each treatment.
Experimental Units: high school students who want to take the special course
Factor and levels: the only factor is course-taking behavior, and the levels are the same as the treatments
Response: SAT score
Blocks: none
Treatments: course, or no course (control)

E40. a. This is a completely randomized design where schools are randomly assigned to two groups, one group for each treatment.

Treatments: extensive calculator use or limited calculator use.
***Experimental Unit*:** each school
There are no blocks.

b. This is a randomized block (paired comparison) design.
***Treatments*:** extensive calculator use or limited calculator use
***Experimental Unit**:* each school
***Block*:** the pairs of schools with similar gain scores from the previous year.

E41. This is a randomized paired comparison (repeated measures) design.

Block: patient (more precisely, two weeks of a patient's time)
Treatments: low phenylalanine diet or regular diet
Experimental Unit: one week of a patient's time
Design: the design is a randomized paired comparison design with repeated measures
Response: dopamine level

E42. The treatments of interest to manufacturer are the glass-panel combinations, of which there are six (three types of glass within each of two panel designs). Think of the four locations around the world as blocks, as they offer relatively homogeneous conditions for experiments within but they may differ considerable between. Now, each block offers twelve test sites; each of the treatments should be assigned at random to two sites within each block (location).

Treatment: each type of glass-panel combination
Experimental Unit: Each test site
Block: The four locations: Mojave Desert, mountains of Tennessee, the desert of northern Africa, and central Europe.

E43. Treatments were assigned to units using a block design. In particular, the design was a randomized block with repeated measures on the same subject.

Response: rate of finger tapping
Unit: one day of a subject's time
Treatments: caffeine, theobromine, or placebo
Block: subject (more precisely, three days of a subject's time)

E44. The treatments were assigned completely at random in this completely randomized design.
Response: time (months) until baby walks unaided
Unit: infant
Treatments: four kinds of exercise/follow-up program
Blocks: none

E45. The treatments were assigned completely at random in this completely randomized design.

Response: quantity eaten (relative to body weight)
Unit: hornworm
Treatments: diet of regular food or diet of 80% cellulose
Blocks: none

E46. a. This is a completely randomized design. Since there are no obvious variables on which to block, a CRD seems reasonable.

b. The experimental units are the individual rats. The response measurement is the total length of stomach lesions.

c. Yes. While the rats are unlikely to understand the treatment, being forced to take medication (however it is administered) could be upsetting. The added stress could affect the stomach lesions.

E47. a. Sample. The measurement process is destructive, and the size of the population makes a census too expensive.

b. Census. The population size is small, the information is easy to get, and the measurement process is not destructive.

c. Sample. The population is so large that a census would be too costly and time-consuming.

E48. The frame misses all cases of heart disease that are not yet diagnosed and all those that are diagnosed but did not result in hospitalization in the last five years. If people who smoke are more likely to be hospitalized than others with heart disease, the estimate will be too big. If they are less likely—probably not the case because smokers tend to have multiple health problems—it will be too small.

E49. a. The class is likely to be reasonably representative.

b. The class will not be representative: Its average age will be much lower than that of the population.

c. The class is likely to be reasonably representative.

d. Students who choose to take a statistics course as an elective are more likely to prefer science to English than students who do not take statistics. So in this case, it would not be reasonably representative. If your classmates took statistics because it is required for a variety of majors, they could be reasonably representative.

e. Blood pressure increases with age. The average blood pressure for the class will almost surely be much lower than for the population.

E50. a. Decrease, because take-home pay is less than gross earnings.

b. The worker might overstate his or her earnings, misclassify his or her occupation/title, or include overtime, for example.

E51. Yes; actors and actresses listed in the almanac are better known, and so better paid than those who are not listed.

E52. Method I is size-biased and will tend to overestimate the true average. (States with fewer representatives are less likely to end up in the sample than are states with more representatives. In fact, a state's chance of being chosen is proportional to the number of representatives it has.) Method II is unbiased.

E53. The frame will not include most owners who bought their 2008 models used from either a private party or from dealers of other makes. It will include people who have sold their 2008 model. These owners may be the ones with the higher repair bills.

E54. This method is size-biased. Authors with more quotations are more likely to be chosen than authors with fewer quotations. (Authors with longer quotations also are more likely to be chosen.)

E55. Your net allows tiny fish to escape. Dragging it behind a boat will miss fish that are far below the surface.

E56. Maps will vary. The larger trees are more likely to be chosen. For example, suppose the lot is subdivided into six sections, with five big and five small trees on six plots. Each of the five big trees is on its own section, whereas the five small trees are all clustered on one section. If a section is selected at random, a big tree is more likely to be chosen.

E57. *New York Times* readers tend to have higher incomes and more years of education in comparison with the average New Yorker.

E58. For example, any survey that takes a lot of time to answer or that asks a sensitive question is likely to get nonresponse bias. In the case of the long survey, people with more time on their hands and people who just want to talk to someone are more likely to respond. In the case of the sensitive question, people who can give a socially acceptable response are more likely to respond. But the major cause of nonresponse these days is probably that people do not want to be bothered because of the plethora of polls, both good and bad, that are trying to get information from the public.

E59. Stage 1: Choose an SRS of states.
Stage 2: Choose an SRS of congressional districts from each state chosen in stage 1.
Stage 3: Choose an SRS of precincts from each congressional district chosen in stage 2.
Stage 4: Choose an SRS of voters from each precinct chosen in stage 3.

E60. Stage 1: Choose an SRS of shelves.
Stage 2: Choose an SRS of books from each shelf chosen in stage 1.
Stage 3: Choose an SRS of pages from each book chosen in stage 2.
Stage 4: Choose an SRS of lines from each page chosen in stage 3.

E61. a. This is an observational study. Although subjects were selected at random, they weren't randomly assigned to a diet. Thus, this isn't an experiment.

b. Almost all of the people who follow a Mediterranean diet will be Greek and almost all of the people who follow a typical American diet will be American. Thus it will be impossible for the researcher to tell whether it was the diet or some other difference in American and Greek lifestyles that accounts for the difference in percentage of heart attacks. If it is some other difference in lifestyle, then a young American who switches to a Mediterranean Diet may not decrease his or her chance of getting a heart attack.

c. Lifestyle differences between the two countries may include more physical activity and less stress in Greece and so these factors are confounded with diet.
The American Heart Association page on "Mediterranean Diet" says

> "The incidence of heart disease in Mediterranean countries is lower than in the United States. Death rates are lower, too. But this may not be entirely due to the diet. Lifestyle factors (such as more physical activity and extended social support systems) may also play a part. Before advising people to follow a Mediterranean diet, we need more studies to find out whether the diet itself or other lifestyle factors account for the lower deaths from heart disease."

E62. a. The mother has done an observational study. She has not assigned the two treatments of cola and no cola to her daughter at random at different times.

b. The cola is confounded with the fast food and birthday party food. It also is confounded with the fact that the daughter is eating out rather than eating at home and may be excited about that. A third variable that is confounded with the cola is that the daughter knows her mother doesn't approve of cola and this fact may change the daughter's behavior. Thus, instead of the cola causing the hyperactivity, it may be caused by something in the food, or by the excitement of being away from home, or by the daughter wanting the mother to notice she is doing something "rebellious."

c. The daughter could be assigned cola or not cola at random over a sufficient number of afternoons and her behavior evaluated. The person who evaluates the level of activity must be blinded to whether the daughter had been given cola or not. However, the daughter would know whether she had had cola or not and she already understands how she is expected to behave when she has cola. If the trigger for any possible hyperactivity could be isolated as, say the sugar or caffeine in the cola, it might be possible to remove those from the cola so that the daughter couldn't tell the difference.

E63. a. This is an experiment as subjects were randomly assigned to treatments, there are a sufficient number of subjects, and there are four different treatments. The factors are

presence of mother, with levels present and absent, and presence of siblings, with levels present and absent. The response is the difference in the mouse's weight from the initial weight to the weight 90 days later. The number of units is the total number of baby mice that were divided into the four groups.

b. This is a survey, with the population being all of the different possible 10 ft. by 10 ft. squares on the hill. The explanatory variables are the amounts of the ten different nutrients. The response variable is the number of species of plants. The sample size is 4.

c. This is an observational study as there is no random selection of types of fruit or of particular fruit within a type. The factors are the different types of fruit. (There are no levels.) The response variable is float or not. The observational unit is a fruit and there is one observed unit for each type of fruit.

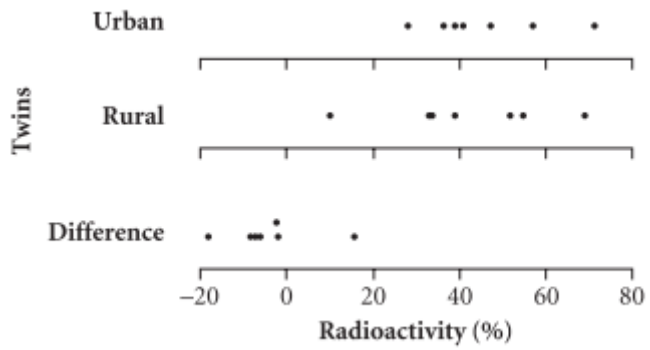
E64. Because of the variability in the amount of dandruff and the extent to which each subject's current shampoo gets rid of it, a "cross-over" experiment is best. Randomly divide the subjects who volunteer for this experiment into two groups. The first group will use the new shampoo first and, after a month or so, switch to the best-selling shampoo. The second group will use the best-selling shampoo first, then, after the same period of time, switch to the new shampoo. A person who does not know which shampoo the person has been using should evaluate the degree of dandruff at the end of each period of treatment. The response variable would be the difference: *amount of dandruff with new shampoo – amount of dandruff with best-selling shampoo*.

The problem remains that the person can tell which shampoo is the best-selling brand and so his or her behavior may differ depending on which shampoo he or she is using. For example, the distinctive smell might cause the person to rinse longer, so you would be unable to tell whether it was the shampoo or the longer rinsing that caused any difference in dandruff. In that case, the type of shampoo would be confounded with how the person shampoos. To prevent this confounding, you would have to have the person or a professional do the shampooing according to a specific protocol. You would have to give them instructions about how often to shampoo, how much shampoo to use, how long to scrub, and how long to rinse.

E65. a. The factor is location, with level being urban or rural. A block is one pair of twins.

b. This is an observational study. The treatment is already built into the units.

c. As the dotplots below show, the variation in response is greatest in rural twins and least in the differences. This study does demonstrate that the lungs of the urban twins took longer to clear than did the lungs of the rural twins.



d. Twins were used because they are genetically identical. This will reduce any variability caused by genetic differences.

E66. One possible experiment is to have each person be a block. Each person is then presented the three kinds of cookies in a random order.

E67. You should include the following points in their design: a description of the treatments, a rationale for any blocking, a description of the blocking scheme, a description of the random assignment of treatments to units *within* blocks, the protocol for the experiment such as making it double-blind if possible, and the response variable to be measured and compared. (Have the students keep their notes so they can refer back to this example after Chapter 9.)

Sample response: The two treatments will be the two types of exercise bike. I will assign four bikes of each type to the eight units or spaces on the floor of the gym. Bikes nearer the door and bikes in front of the televisions may be more (or less) likely to be used than the other bikes. Thus, I will form four blocks of spaces with similar locations in the gym: 1 and 4, 2 and 3, 5 and 8, 6 and 7. Then *for each block,* I will flip a coin, and if it is heads, a bike of the first type goes into the even-numbered space and a bike of the second type into the odd-numbered space. If it is tails, a bike of the second type goes into the even-numbered space and a bike of the first type goes into the odd-numbered space. There is no reason to make this experiment blind or double-blind because we want the customers to observe the different choices and the number of hours of use is recorded automatically for each bike.

E68. Answers will vary

Concept Review Solutions

C1. C. A family with two children in high school is twice as likely to be selected by this method as is a family with one child in high school, so the computed mean is likely to be too high. When finding the mean number of children per family, it is important to randomly select families rather than select children. The sampling method in choice B will have a sampling distribution with a lot of variation. For that reason this method is not a very good sampling plan, but on average it is a method that should give the correct mean.

C2. E. Listeners in the sample are the people who choose to call in.

C3. B

C4. A. Choice A is not a good reason for choosing this plan because stratification reduces variability not confounding. Choice D is a good reason because stratifying tends to reduce variability when the strata have different averages.

C5. A. The two treatments are expected to result in different responses from men and women, so form a block of men and a block of women and then randomly assign the two treatments within these blocks.

C6. A

C7. A. A control group is not necessary as long as there are two or more treatments that can be compared.

C8. D

C9. A design that uses no blocks would be a completely randomized design. Randomly divide the available subjects into four treatments groups (of equal size, if possible) and then randomly assign one of the background colors (treatments) to each group. Each person should then be independently timed in threading a needle with white thread. All needles and thread should be of the same size and type. The main weakness of this design is that there will probably be considerable variation from subject to subject within groups, which will serve to mask any differences in times between the treatments (colors of background).

A block design requires somewhat homogeneous units within each block with regard to variables related to the response (time to thread a needle). These could involve steadiness of nerves and perception of color, among others. One way to decide on blocks is to do a preliminary run of the experiment on every subject, perhaps using a neutral color that is not one of the treatments, and grouping subjects into blocks such as slow, medium, fast based on the trial run. Blocks should be of equal size, a multiple of four, so that each treatment is run the same number of times in each block. Randomly assign treatments to subjects within the blocks in a manner similar to that used in the completely randomized design. The block design should reduce the subject-to-subject (within-block) variation so that any observed treatment differences are more easily detected.

A repeated measures design is a block design that uses each subject as a block. In this design each subject would be assigned each of the four treatments (background colors) in random order and timed in threading the needle with the white thread. This design is the ultimate in reducing subject-to-subject variability, but it has the major drawback of the possibility of learning taking place along the way. Thus, the colors presented later may produce lower times simply because of the experience of the subject. This "carryover" effect can seriously bias the results of the study. (Of course, the carryover effect could be

one of fatigue, which would make the later times slower than the earlier ones.)
C10. Because the mailed portion of the census depends on the goodwill of the citizens to return the questionnaires, many people are left uncounted. Making an intensive effort to find those not counted and estimating the undercount for a sample of regions seems like a good way to improve the accuracy of the census count, at least for that segment of the population in the sample. The adjustment to the census count through sampling can correct many deficiencies of the mail campaign if the Census Bureau pays careful attention to the kinds of errors discussed below. There are three large weaknesses in the sampling plan for adjustment. First, many of those not found by the mail campaign are also difficult to find in a follow-up survey of geographic regions, so the sampling will still underestimate some segments of the population. Second, the matching of records is subject to error. Many of the census forms are incomplete or filled out incorrectly by the respondents, sometimes intentionally, so that deciding whether or not a sampled household has already filled out a form is not always easy. In addition, a sampled block may show completed census forms for persons who are no longer in that block or may not even exist. (It is possible for a block to be over-counted.) Third, applying adjustments from a sampled region to other similar regions for which no sample data are taken is also error prone.