

Chapter 4

Section 4.1

Check Your Understanding, page 211:

1. The company inspector is using a convenience sample. This could lead him to think that the oranges are of better quality than they really are, if the farmer puts the best oranges on top.
2. *Nightline* was using a voluntary response sample. Only those who feel particularly strongly about the issue are likely to respond. In this case, those who are happy that the United Nations has its headquarters in the US already have what they want and so are less likely to worry about responding to the question.

Check Your Understanding, page 219:

1. It might be difficult to give a survey to an SRS of 200 fans because you would have to identify 200 different seats, go to those seats in the arena and find the people who are sitting there. This means going to 200 different locations throughout the arena, which would take time. There is also the problem that people are not always in their seats throughout the game and not all seats will necessarily be occupied in any given game.
2. For a stratified sample, it would be better to take the lettered rows as the strata because each lettered row is the same distance from the court and so would contain only seats with the same (or nearly the same) ticket price. This means that all people in any given stratum would have paid roughly the same amount for their tickets.
3. For a cluster sample, it would be better to take the numbered sections as the clusters because they include all different seat prices. Each section contains seats with many different ticket prices so the people in a section would mirror the characteristics of the population as a whole.

Check Your Understanding, page 224:

1. (a) Using the telephone directory as a sampling frame is an example of a sampling error. This will result in undercoverage because those who are not listed in the phone book (those who do not have a phone or have only a cell phone) do not have the opportunity to be chosen. (b) If the person cannot be contacted, this is an example of a nonsampling error. This did not occur because of the way the sample was chosen, but rather was an effect of the way the survey was administered. (c) If you choose to interview people walking past you on a sidewalk, this is a sampling error. Who you find will depend on where (in what neighborhood, etc.) you are standing. This has to do with how you choose your sample.
2. This question will result in fewer people suggesting that we should ban disposable diapers by making it sound like they are not a problem in the landfill. The author of the question highlights several other items that take up more space in the landfill, which makes it look like disposable diapers are really not a problem.

Exercises, page 226:

- 4.1 The population is (all) local businesses. The sample is the 73 businesses that return the Questionnaire.
- 4.2 The population is all the artifacts discovered at the dig. The sample is those artifacts (2% of the population) that are chosen for inspection.
- 4.3 The population is the 1000 envelopes stuffed during a given hour. The sample is the 40 envelopes selected.

4.4 The population is all 45,000 people who made credit card purchases. The sample is the 137 people who returned the survey form.

4.5 Only persons with a strong opinion on the subject, strong enough that they are willing to spend the time and money, will respond to this advertisement.

4.6 Letters to legislators are an example of a voluntary response sample—the proportion of letters opposed to the insurance should not be assumed to be a fair representation of the attitudes of the congresswoman's constituents. Only those who have very strong opinions will write in.

4.7 This is a voluntary response sample. It is biased in favor of those who feel most strongly about the issue being surveyed.

4.8 (a) A voluntary response sample. (b) It is biased toward readers who feel most strongly about the issue. 85% is probably higher than the true percent because it is likely that readers who feel most strongly about this issue have in some way been involved in an accident caused by cell phone use while driving.

4.9 (a) A convenience sample. (b) It is unlikely that the first 100 students to arrive at school are representative of the student population in general. 7.2 hours is probably higher since you might expect that the students who arrive first are those who got a good night's sleep the night before. Students who received less sleep the night before are probably more likely to run late the next morning.

4.10 This is a convenience sample. It is easy to find lots of people in a mall. However, it is likely to give a higher percentage for the unemployment figure because the unemployed have more time to be at the mall than those who are employed.

4.11 (a) Number the 40 students from 01 to 40 ~~alphabetically~~. Go to the random number table and pick a starting point. Record two-digit numbers, skipping any that aren't between 01 and 40 or are repeats, until you have 5 unique numbers between 01 and 40. (b) Starting at line 107 we read off the following numbers: 82 (ignore) 73 (ignore) 95 (ignore) 78 (ignore) 90 (ignore) **20** 80(ignore) 74 (ignore) 75 (ignore) **11** 81 (ignore) 67 (ignore) 65 (ignore) 53 (ignore) 00 (ignore) 94 (ignore) **38 31** 48 (ignore) 93 (ignore) 60(ignore) 94 (ignore) **07**. So we have picked: **Johnson (20), Drasin (11), Washburn (38), Rider (31), and Calloway (07).**

4.12 (a) Number the 33 complexes from 01 to 33 alphabetically. Go to the random number table and pick a starting point. Record two-digit numbers, skipping any that aren't between 01 and 33 or are repeats, until you have 3 unique numbers between 01 and 33. (b) Starting at line 117 we read off the following numbers: 38 (ignore) **16** 79 (ignore) 85 (ignore) **32** 62 (ignore) **18**. So we have picked: Fairington (16), Waterford Court (32) and Fowler (18).

4.13 (a) Number the plots from 0001 to 1410. Go to the random number table and pick a starting point. Record four-digit numbers, skipping any that aren't between 0001 and 1410 or are repeats, until you have 141 unique numbers between 0001 and 1410. (b) Starting at line 131 we read off the following numbers: **0500** 7166 3281 1941 4873 **0419** 7855 7645 1959 6565 6873 2552 5984 2920 8796 4316 5937 3931 6859 7150 4574 **0418** (ignore all numbers not in bold). So the first three plots in our sample are plots 0500, 0419 and 0418.

4.14 (a) Number the gravestones from 00001 to 55914. Go to the random number table and pick a starting point. Record 5-digit numbers, skipping any that aren't between 00001 and 55914 or are repeats, until you have 395 unique numbers between 00001 and 55914. (b) Starting at line 127 we read off the

following numbers: **43909** 99477 **25330** 64359 **40085** (ignore all numbers not in bold). So the first three gravestones picked are those numbered 43909, 25330 and 40085.

4.15 If you always begin at the same place, then the results would not be random. You would end up using the same sample in every case.

4.16 (a) False—if it were true, then after looking at 39 digits, we would know whether or not the 40th digit was a 0. (b) True—there are 100 pairs of digits 00 through 99, and all are equally likely. (c) False—0000 is just as likely as any other string of four digits.

4.17 (a) Assuming none of the phones can be shipped until after the inspection, inspecting a random sample of 20 phones could hold up the shipping process. Additionally, in order to obtain a random sample, the phones must be numbered in some way. Keeping track of the ordering of 1000 phones may be difficult. (b) It is possible that the quality of the phones produced changes over the course of the day so that the last phones manufactured are not representative of the day's production. (c) This is not an SRS because each sample of 20 phones does not have the same probability of being selected. In fact, the 20 phones that are sampled will be the 50th, 100th, ..., 1000th, the others have no chance of being sampled.

4.18 (a) To obtain an SRS, every tree would need to have an equal chance of being included in the sample. It is not practical to even identify every tree in the park. (b) This sampling method is biased because these trees are unlikely to be representative of the population. Trees along the main road are more likely to be damaged by cars and people and may be more susceptible to infestation. (c) The scientists can be confident that the actual percentage of pine trees in the area that are infected by the pine beetle is near 35% although there is always some error associated with using sampling to estimate population parameters.

4.19 Assign 01 to 30 to the students (~~in alphabetical order~~). Starting on line 123 gives 08-Ghosh, 15-Jones, 07-Fisher, and 27-Shaw. Assigning 0-9 to the faculty members gives 1-Besicovitch and 0-Andrews.

4.20 Label the 500 midsize accounts from 001 to 500, and the 4400 small accounts from 0001 to 4400. Starting at line 115, the first five accounts in each strata are 417, 494, 322, 247, and 097 for the midsize group, then 3698, 1452, 2605, 2480, and 3716 for the small group.

4.21 (a) Use the three types of seats (sideline, corner and end zone) as the three strata since ticket prices will be similar within each stratum but different between the three strata. (b) It might be easier to obtain a cluster sample because a stratified random sample will still likely choose seats all over the stadium, which would make it very time consuming to get to everyone. A cluster sample would be easier to obtain, because there would be many people sitting all together who would be part of the sample. In this case one would use the section numbers for the clusters.

4.22 (a) Using a stratified random sample would assure the manager that he got opinions from each type of room. Use each type of view as the strata and randomly pick 60 guests from each stratum. (b) A cluster sample would be a simpler option because someone could just slip the forms under a specific pattern of doors. All rooms ending in a specific number would be the clusters. For example, all rooms ending in 16. Presumably these are all stacked above each other on the 30 floors. The manager should just pick two random numbers that represent rooms on the water side and two random numbers that represent rooms on the golf course side and then survey all 30 rooms (one per floor) that end in that number.

4.23 It is not an SRS, because some samples of size 250 have no chance of being selected (e.g., a sample containing 250 women).

4.24 The chance of being interviewed is $3/30$ for students over age 21 and $2/20$ for students under age 21. This is $1/10$ in both cases. It is not an SRS because **not all combinations of students have an equal chance of being interviewed**. For instance, groups of 5 students all over age 21 have no chance of being interviewed.

4.25 (a) This is cluster sampling. (b) Answers will vary. Label each block from 01 through 65. Beginning at line 142, record two-digit numbers, skipping any that aren't between 01 and 65 or are repeats. The 5 identified blocks are 02, 32, 26, 34, and 08. The statistical applet selected blocks 10, 20, 45, 36, and 60.

4.26 (a) Split the 200 addresses into 5 groups of 40 each. Looking for 2-digit numbers from 01 to 40, the table gives 35 so the systematic random sample consists of 35, 75, 115, 155, and 195. (b) Every address has a 1-in-40 chance of being selected, but not every subset has an equal chance of being picked—for example, 01, 02, 03, 04, and 05 cannot be selected by this method.

4.27 Households without telephones or with unlisted numbers are omitted from this frame. Such households would likely be made up of poor individuals (who cannot afford a phone), those who choose not to have phones, and those who do not wish to have their phone number published. If the variable being measured tends to have different values for those excluded by this sampling method, then our sample result would be off in a particular direction from the truth about the population of households.

4.28 This will miss those who do not have telephones. This means that we will be likely underrepresenting the poor in our sample.

4.29 (a) You are sampling only from the lower priced ticket holders. (b) This is a sampling error. The sampling frame differs from the population of interest (undercoverage).

4.30 (a) **Nonsampling error**. People may lie in response to questions about past drug use. It is not an error due to the act of taking a sample, rather it is a response error. (b) **Nonsampling error**. This is an example of a processing error. (c) **Sampling error**. This will suffer from the same forms of bias as any voluntary response survey.

4.31 (a) The response rate was $\frac{5,029}{45,956} = 0.1094$ so the nonresponse rate is $1 - 0.1094 = 0.8906$ or

89.1%. (b) It is likely that the high amount of nonresponse gave the researchers a lower mean number of miles driven because those who drive more are at home less to answer the phone.

4.32 The higher no-answer rate was probably the second period—when families are likely to be vacationing or spending time outdoors. A high rate of nonresponse makes sample results less reliable because you don't know how these individuals would have responded. It is very risky to assume that they would have responded exactly the same way as those individuals who did respond.

4.33 More than 171 respondents have run red lights. We would not expect very many people to claim they *have* run red lights when they have not, but some people will deny running red lights when they have.

E - Essentially P - Partially I - Incorrect

2013 AP® STATISTICS FREE-RESPONSE QUESTIONS

2. An administrator at a large university wants to conduct a survey to estimate the proportion of students who are satisfied with the appearance of the university buildings and grounds. The administrator is considering three methods of obtaining a sample of 500 students from the 70,000 students at the university.
- (a) Because of financial constraints, the first method the administrator is considering consists of taking a convenience sample to keep the expenses low. A very large number of students will attend the first football game of the season, and the first 500 students who enter the football stadium could be used as a sample. Why might such a sampling method be biased in producing an estimate of the proportion of students who are satisfied with the appearance of the buildings and grounds?
- (b) Because of the large number of students at the university, the second method the administrator is considering consists of using a computer with a random number generator to select a simple random sample of 500 students from a list of 70,000 student names. Describe how to implement such a method.
- (c) Because stratification can often provide a more precise estimate than a simple random sample, the third method the administrator is considering consists of selecting a stratified random sample of 500 students. The university has two campuses with male and female students at each campus. Under what circumstance(s) would stratification by campus provide a more precise estimate of the proportion of students who are satisfied with the appearance of the university buildings and grounds than stratification by gender?

a) Plausible explanation why sample not rep

Link

Direction ✓

b) Assign numbers

Generate 500 distinct #s

Select students who correspond...

c) More variability based on campus than by gender

4-EEE

2-EEI

PPP

EPI

EPP

3-EEP

1-EII

PPI

0-III

4.34 People likely claim to wear their seat belts because they know they should; they are embarrassed or ashamed to say that they do not always wear seat belts. Such bias is likely in most surveys about seat belt use (and similar topics).

4.35 (a) The wording is clear. The question is slanted in favor of warning labels. (b) The question is clear, but it is clearly slanted in favor of national health insurance by asserting it would reduce administrative costs. (c) The wording is too technical for many people to understand—and for those who do understand the question, it is slanted because it suggests reasons why one should support recycling. It could be rewritten to something like: “Do you support economic incentives to promote recycling?”

4.36 (a) The question is clear, but the two options presented are too extreme; no middle position on gun control is allowed. Many students may suggest that this question is likely to elicit more responses against gun control (that is, more people will choose 2). (b) The question is so complicated that it isn't clear. It is also slanted; the phrasing of this question will tend to make people respond in favor of a nuclear freeze. Only one side of the issue is presented.

4.37 c

4.38 d

4.39 d

4.40 c

4.41 e

4.42 c

4.43 The predicted sleep debt for a 5-day school week, based on the least-squares regression equation, is $2.23 + 3.17(5) = 18.08$ hours, a little more than 3 hours greater than what was found in the research study. Based on their collected data, the students have reason to be skeptical of the research study's reported results.

4.44 (a) The 95th percentile is the amount of bandwidth below which 95 percent of all 5 minute measurements fall. (b) The method using the 98th percentile would cost the company more because it would suggest a higher usage of bandwidth by the company.

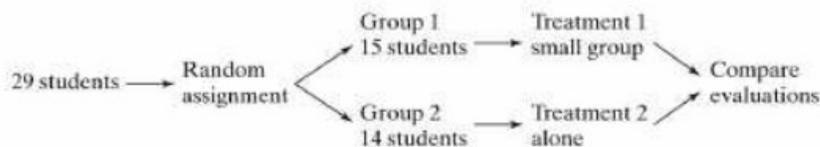
Section 4.2

Check Your Understanding, page 233:

1. This was an experiment because a treatment (brightness of screen) was imposed on the laptops.
2. This was an observational study. Students were not assigned to a particular number of meals to eat with their family per week.
3. The explanatory variable is the number of meals per week eaten with their family and the response variable is probably their GPA (or some other measure of their grades).
4. This is an observational study and there may well be lurking variables that are actually influencing the response variable. For instance, families that eat more meals together may also be families where the parents show more interest in their childrens' education and therefore help them to do better in school.

Check Your Understanding, page 240:

1.



2. Using an alphabetical list of the students, assign each student a number between 01 and 29. Pick a line of the random number table and read off two digit numbers until you have 15 numbers between 01 and 29. These students belong in the treatment group where students will meet in small groups. The other students will view the videos alone.

3. The purpose of the control group is to have a group to compare to. Presumably the students have been evaluating their own performances by themselves before. If you incorporate such a group into your experiment, you can evaluate if the group work is actually better.

Check Your Understanding, page 244:

1. No, this experiment did not take the placebo effect into account. It is possible that women who “thought” they were getting an ultrasound would have different reactions to pregnancy than those who knew that they hadn’t received an ultrasound.

2. This experiment was not double-blind. While the people weighing the babies at birth may not have known whether that particular mother had an ultrasound or not, the mothers did know whether they had had an ultrasound or not. This means that the mothers may have affected the outcome since they knew whether they had received the treatment or not.

3. An improved design would have been one in which all mothers were treated as if they had an ultrasound, but for some mothers the ultrasound machine just wasn’t turned on (but this fact would not be obvious to the woman). This means that the ultrasound would have to have been done in such a way so that the woman could not see the screen.

Exercises, page 253:

4.45 (a) This was an observational study because no treatment was imposed on the mothers. The researchers simply asked them to report both their chocolate consumption and their babies’ temperament. (b) The explanatory variable is the mother’s chocolate consumption and the response variable is the baby’s temperament. (c) No, this study is an observational study so we cannot make a conclusion of cause and effect. There could be a lurking variable that is actually causing the difference in temperament.

4.46 (a) This was an observational study because no treatment was imposed on the children. The researchers simply followed them through their 6th year in school, asking adults to rate their behavior at several times along the way. (b) The explanatory variable was the amount of time in child care from birth to age four-and-a-half. The response variable was the adult ratings of their behavior. (c) No, this study is an observational study so we cannot make a conclusion of cause and effect. There could be a lurking variable that is actually causing the difference in adult ratings of their behavior.

4.47 (a) This was an experiment because students were randomly assigned to the different teaching methods. (b) Since this was an experiment with proper randomization, the teacher can conclude that using the computer animation appears to result in higher increases in test scores.

4.48 (a) This is an example of an observational study. The researchers did not assign people to either use or not use cell phones. (b) No, this study is an observational study so we cannot make a conclusion of cause and effect.

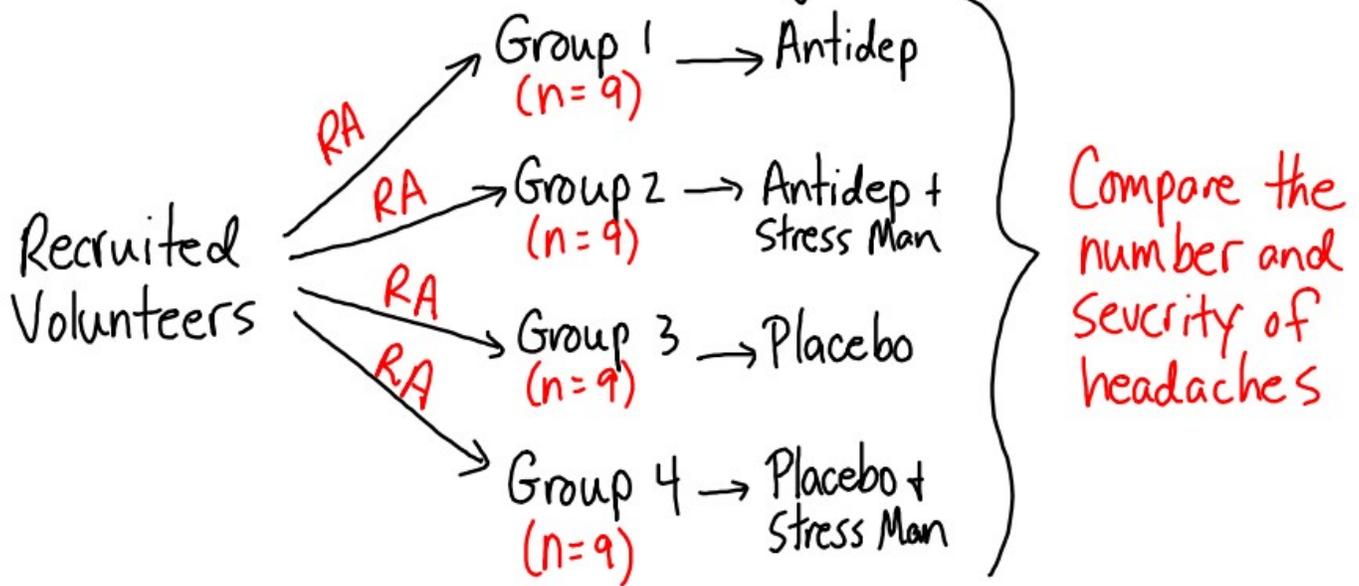
49) Class Size → Academic Achievement

50) Binge Drinking → Low GPA

- Adjustment to "college life"

- Relationship problems

6(a) Between Groups Design:



6(b) Draw names out of a hat

Number volunteers → Random Digit Table
↳ Calculator

72) Double Blind

Neither subjects nor experimenter know who received a placebo vs supplement

Randomized

Subjects were put into groups randomly

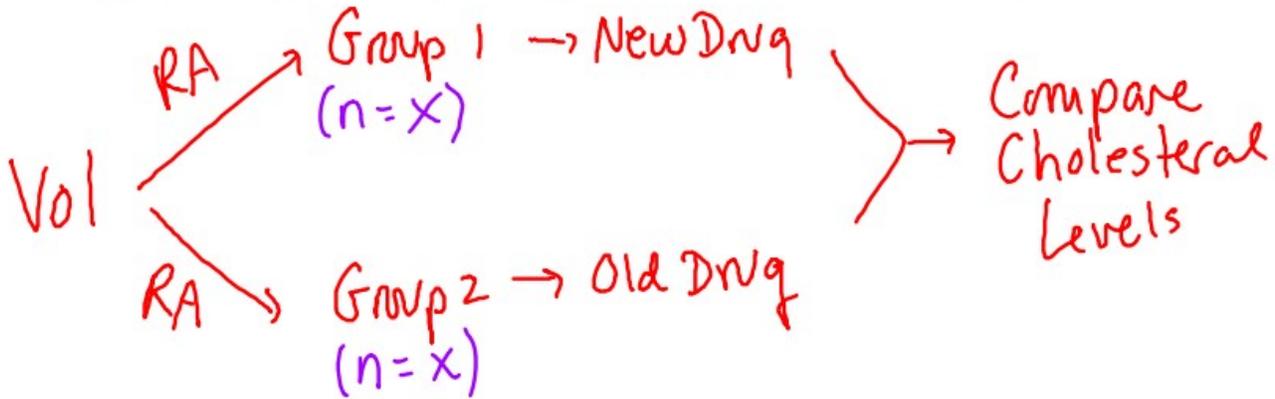
Placebo Controlled

One group received an inactive pill

2000 AP® STATISTICS FREE-RESPONSE QUESTIONS

5. High cholesterol level in people can be reduced by exercise or by drug treatment. A pharmaceutical company has developed a new cholesterol-reducing drug. Researchers would like to compare its effects to the effects of the cholesterol-reducing drug that is currently available on the market. Volunteers who have a history of high cholesterol and who are currently not on medication will be recruited to participate in a study.

(a) Explain how you would carry out a completely randomized experiment for the study.



(b) Describe an experimental design that would improve the design in (a) by incorporating blocking.

- Race
- Gender
- Age
- ★ Physical Activity
- Diet/Weight
- Medical Conditions
- Cholesterol Levels

(c) Can the experimental design in (b) be carried out in a double blind manner? Explain.

Yes - Experimenter / subject don't know...

2001 AP® STATISTICS FREE-RESPONSE QUESTIONS

4. Students are designing an experiment to compare the productivity of two varieties of dwarf fruit trees. The site for the experiment is a field that is bordered by a densely forested area on the west (left) side. The field has been divided into eight plots of approximately the same area. The students have decided that the test plots should be blocked. Four trees, two of each of the two varieties, will be assigned at random to the four plots within each block, with one tree planted in each plot.

The two blocking schemes shown below are under consideration. For each scheme, one block is indicated by the white region and the other block is indicated by the gray region in the figures.



- (a) Which of the blocking schemes, A or B, is better for this experiment? Explain your answer.
- (b) Even though the students have decided to block, they must randomly assign the varieties of trees to the plots within each block. What is the purpose of this randomization in the context of this experiment?

- Soil conditions
- Moisture
- Drainage
- Light Conditions
- Wildlife / Insects