Statistics - Unlocking the Power of Data 4th edition Lock Solutions Manual

SKU: 9781394336623-SOLUTIONS

Section 1.1 Solutions

1.1 (a) The cases are the people who are asked the question.

(b) The variable is whether each person supports the law or not. It is categorical.

1.2 (a) The cases are the 100 stocks.

(b) The variable is the percentage change, which is a numerical quantity, for each of the stocks. It is quantitative.

1.3 (a) The cases are the teenagers in the sample.

(b) The variable is the result (yes or no) indicating whether each teenager eats at least five servings a day of fruits and vegetables. It is categorical.

1.4 (a) The cases are the bunches of bananas in the sample.

(b) The variable is the number of days until the bananas go bad. It is quantitative.

1.5 (a) The 10 beams that were tested.

(b) The force at which each beam broke. It is quantitative.

1.6 (a) The cases are countries of the world.

(b) The variable is whether or not the literacy rate is over 75%. It is categorical.

1.7 Since we expect the number of years smoking cigarettes to impact lung capacity, we think of the number of years smoking as the explanatory variable and the lung capacity as the response variable.

1.8 Since we expect the amount of fertilizer used to impact the yield (and not the other way around), we think of the amount of fertilizer as the explanatory variable and the yield of the crop as the response variable.

1.9 Ingesting more alcoholic drinks will cause the level of alcohol in the blood to increase, so the number of drinks is the explanatory variable and blood alcohol content is the response.

1.10 The world record time will continue to decrease as the years go by so we expect the year to impact marathon record time. We think of the year as the explanatory variable and the record time as the response variable.

1.11 (a) *Year* and *HigherSAT* are categorical. The other six variables are all quantitative, although *Siblings* might be classified as either categorical or quantitative.

(b) There are many possible answers, such as "What proportion of the students are first year students?" or "What is the average weight of these students?"

(c) There are many possible answers, such as "Do seniors seem to weigh more than first year students?" or "Do students with high Verbal SAT scores seem to also have high Math SAT scores?"

1.12 (a) In addition to the identification column, *Country*, there are 24 variables. We see that *Developed* is a categorical variable, while the other 23 variables are all quantitative.

(b) There are many possible answers, such as "What is the average life expectancy for all countries of the world?" or "What proportion of countries are developed?"

(c) There are many possible answers, such as "Do countries with a greater land area have a larger percent rural?" or "Do countries that spend a relatively large amount on the military spend a relatively small amount on health care?" or "Do developed countries have a longer life expectancy than developing countries?"

1.13 (a) The explanatory variable is the type of format in which the story is presented. It is categorical, with three categories (audio, illustrated, and animated).

(b) The response variable is the measure of brain connectivity. It is quantitative.

(c) The cases are the four-year-olds in the study, so there are 27 cases.

1.14 (a) The cases are female gamers in Great Britain.

(b) There are three variables mentioned: Whether the gamers had received obscene messages (categorical), how many hours a week they played (quantitative), and whether they felt there were enough strong female characters in games (categorical).

(c) There are 1151 cases and 3 variables, so the dataset will have 1151 rows and 3 columns.

1.15 (a) The cases are the students in the college physics class.

(b) There are 3 variables: Whether the student was assigned to an active or passive learning class, the measure of how much the student thought they learned, and the score of actual learning on the test. The class assignment is categorical, while the other two are quantitative.

(c) The class assignment (active or passive) is the explanatory variable, while the two measures of learning are response variables.

(d) The students in the physics class are the cases, so there are $154 + 142 = 296$ cases. There are three variables, so the dataset will have 296 rows and 3 columns.

1.16 There are at least two variables. One variable is whether or not the spider engaged in mock-sex. This variable is categorical and the explanatory variable. Another variable is length of time to reach the point of real mating once the spider is fully mature. This variable is quantitative and the response variable.

1.17 The individual cases are the lakes from which water samples were taken. For each lake in the sample, we record the concentration of estrogen in the water and the fertility level of fish. Both are quantitative variables.

1.18 There are two variables. One variable indicates the presence or absence of the gene variant and the second variable indicates which of the three ethnic groups the individual belongs to. Both variables are categorical.

1.19 (a) There are 10 cases, corresponding to the 10 cities. The two variables are population, which is quantitative, and the hemisphere the city is in, which is categorical.

(b) We need two columns, one for each variable. The columns can be in either order. See the table.

Population	Hemisphere
37	Eastern
26	Eastern
23	Eastern
22	Eastern
21	Eastern
21	Eastern
21	Eastern
21	Western
20	Western
19	Western

1.20 (a) There are 7 cases, representing the seven pigeons. There are two variables. One is the sex of the pigeon, which is categorical, and the other is the speed of the pigeon, which is quantitative.

(b) The dataset will have 7 rows and 2 columns. See the table. (The seven cases can be listed in any order.)

Sex	Speed
Hen	1676
Hen	1452
Hen	1449
Cock	1458
Cock	1435
Cock	1418
Cock	1413

1.21 (a) The cases are the homing pigeons, so there are 1412 of them.

(b) There are 4 variables. Two are categorical (loft, sex) and two are quantitative (distance, speed).

(c) The dataset will have 1412 rows (one for each pigeon) and 4 columns (one for each variable).

1.22 One variable is whether each male was fed a high-fat diet or a normal diet. This is the explanatory variable and it is categorical. The response variable is whether or not the daughters developed metabolic syndrome, which is also categorical.

1.23 One variable is whether the young female mice lived in an enriched environment or not. This is the explanatory variable and it is categorical. The response variable is how fast the offspring learned to navigate mazes and is quantitative.

1.24 (a) The mode of transportation (e.g., bus, car, walk) gives a categorical value for each student.

(b) The answer about allergies (yes or no) gives a categorical value for each student.

(c) The proportion of students in the sample who are vegetarians is a single number, based on all of the students, and does not give a value for each individual student. It is not a variable for the dataset.

(d) The number of hours worked at a paid job gives a quantitative value for each student.

(e) The difference in hours of sleep on school nights and non-school nights gives a quantitative value for each student.

(f) The maximum time to get to school is a single number, based on all of the students, and does not give a value for each individual student. It is not a variable for the dataset.

(g) The desired super power chosen from a list gives a categorical value for each student.

1.25 (a) The total tuition and fees is a quantitative value for each school.

(b) The number of schools in the Northeast is a single value, based on all of the schools, and does not give a value for each individual school. It is not a variable for the dataset.

(c) The type of school (public, private, for profit) is a categorical value for each school.

(d) The number of undergraduates is a quantitative value for each school.

(e) The percentage of part-time students is a quantitative value for each school.

(f) The school with the highest average faculty salary is a single value, based on all of the schools, and does not give a value for each individual school. It is not a variable for the dataset.

1.26 In the first study, the cases are the students. The only variable is whether or not the student has smoked a hookah. This is a categorical variable.
In the second study, the cases are the people in a hookah bar. The variables are the length of the session, the frequency of puffing, and the depth of inhalation. All are quantitative.
In the third study, the cases are the smoke samples, and the variables are the amount of tar, nicotine, and heavy metals. All three variables are quantitative.

1.27 (a) This description of the study mentions six variables: age, nose volume, nose surface area, nose height, nose width, and gender.

(b) One of the variables (gender) is categorical, and the other five are quantitative.

(c) There are six variables so the dataset will have six columns. The 859 participants are the cases, so the dataset will have 859 rows.

1.28 (a) The cases are the 47 participants.

(b) The description of the study includes three different variables: the score on the no-distractions test, the score on the test while texting, and whether or not the student considered him or herself to be good at multitasking. The two test score variables are quantitative and the multitasking variable is categorical.

(c) The dataset would have 47 rows (one for each participant) and three columns (one for each of the three variables).

1.29 (a) The cases are the 40 people with insomnia who were included in the study.

(b) There are two variables. One is which group the person is assigned to, either therapy or not, and the other is whether or not the person reported sleep improvements. Both are categorical.

(c) The dataset would have two columns, one for each of the two variables, and 40 rows, one for each of the people in the study.

1.30 If we simply record age in years and income in dollars, the variables are quantitative. Often, however, in a survey, we don't ask for the exact age but rather what age category the participant falls in ($20 - 29$, $30 - 39$, etc.). Similarly, we often don't ask for exact income but for an income category (less than \$10,000, between \$10,000 and \$25,000, etc.). If we ask participants what category they are in for each variable, then the variables are categorical.

1.31 We could sample people eligible to vote and ask them each their political party and whether they voted in the last election. The cases would be people eligible to vote that we collect data from. The variables would be political party and whether or not the person voted in the last election. Alternatively, we could ask whether each person plans to vote in an upcoming election.

1.32 We could survey a sample of people and ask their household income and measure happiness in some way, such as asking how happy they are on a scale of 1–10. The cases would be the people we collect data from. The variables in this case would be household income and happiness rating, although any two variables measuring wealth and happiness are possible.

1.33 Answers will vary.

Section 1.2 Solutions

1.34 This is a sample, because only a subset of fish are measured.

1.35 This is a population, because all customers are accounted for.

1.36 This is a population, because all registered vehicles are accounted for.

1.37 This is a sample, because only a subset of college students were sent the questionnaire.

1.38 The sample is the 120 people interviewed. The population might be all people in that town or all people that go to the mall in that town or a variety of other groups larger than and containing the 120 people in the sample.

1.39 The sample is the five hundred Canadian adults that were asked the question; the population is all Canadian adults.

1.40 The sample is the 100 customers surveyed; the population is all customers of the cell phone carrier.

1.41 The sample is the 1000 households which have databoxes attached to the televisions. The population is all US households with televisions.

1.42 (a) The sample is the 100 college students who were asked the question.

(b) The population we are interested in is all Americans.

(c) A population we can generalize to, given our sample, is college students.

1.43 (a) The sample is the 10 selected twitter accounts.

(b) The target population is all twitter accounts.

(c) The population we can generalize to, given the sample, is only twitter accounts of this author's followers, since this is the population from which the sample was drawn.

1.44 (a) The sample is the 1500 people who were contacted.

(b) The population we are interested in is all residents of the US.

(c) A population we can generalize to, given our sample, is residents of Minnesota.

1.45 (a) The sample is the girls who are on the selected basketball teams.

(b) The population we are interested in is all female high school students.

(c) A population we can generalize to, given our sample, is female high school students who are on a basketball team.

1.46 Yes, this is a random sample from the population.

1.47 Yes, this is random sample from the population.

1.48 No, this is not a random sample, because some employees may be more likely than others to actually complete the survey.

1.49 No, this is not a random sample, because certain segments of the population (e.g., those not attending college) cannot be selected.

1.50 No, this is not a random sample. We might think we can pick out a "representative sample", but we probably can't. We need to let a random number generator do it for us.

1.51 No, this is not a random sample, this is a volunteer sample, since the only people in the sample are those that self-select to respond to the online poll.

1.52 This sample is definitely biased because only students who are at the library on a Friday night can be selected. The random sample should be from all students.

1.53 This is biased because the way the question is worded is not at all objective. Although the sample is a random sample, the wording bias may distort the results.

1.54 This sample is biased because taking 10 apples off the top is not a random sample. The apples on the bottom of the truckload are probably more likely to be bruised.

1.55 From the description, it appears that this method of data collection is not biased.

1.56 This sample is biased because it is a volunteer survey in which people choose to participate or not. Most likely, the people taking the time to respond to the email will have stronger opinions than the rest of the student body.

1.57 Because this was a random sample of parents in Kansas City, the result can be generalized to all parents in Kansas City.

1.58 (a) No, the sample is almost certainly not representative, since it is a volunteer sample and only includes people who visit that website and who chose to participate in the poll.

(b) No, it is not appropriate to generalize since the sample is not representative.

1.59 (a) Yes, the sample is likely to be representative since it is a random sample.

(b) Yes, since the sample is a random sample, we can generalize to the population of all Canadian consumers.

1.60 (a) The sample is the $1,000$ US adults that were contacted. The intended population is all US adults.

(b) Yes, it is reasonable to generalize since the sample was selected randomly.

1.61 (a) The sample is the 800 people who participated in the survey. The intended population is all US smartphone users.

(b) The cases are the people who participated in the survey. There are two variables: Whether or not a food delivery app was used in the last month, and which app (if any) was used. Both are categorical.

1.62 (a) The 457 students in the sample is not a larger population. We would know the exact answers for this group so no need to generalize.

(b) The sample was randomly selected from among all Pennsylvania high school seniors who participated in the Census at School project, so that would be a reasonable population to generalize results to.

(c) There might be something special about schools (or students) who participate in the Census at School project, so it might not be reasonable to generalize the results from that sample to all Pennsylvania high school seniors.

(d) It would not be reasonable to generalize from a sample of only Pennsylvania students to students from all states.

1.63 (a) There is no sampling bias here, since we are told that a random sample was used. The results are being manipulated based only on the order of the options (the wording of the question) so this is an example of wording bias.

(b) We see that people are more likely to select the first option given, so you should ask your friend to choose between "Option W and Option Q,“ with Option W presented first.

1.64 (a) Yes, there is probably sampling bias since the group has not shared the method. No, we cannot know if it is appropriate to generalize, and it is probably not appropriate to do so. The survey might have been asked of children, for example. This would certainly create sampling bias.

(b) Yes, the way the question was worded almost certainly biased the results. A better way might be to ask an open-ended question such as describe how chocolate milk is made,

1.65 (a) The individual cases are the over 6000 restroom patrons who were observed. The description makes it clear that at least three variables are recorded. One is whether or not the person washed their hands, another is the gender of the individual, and a third is the location of the observation. All three are categorical.

(b) In a phone survey, people are likely to represent themselves in the best light and not always give completely honest answers. That is why it is important to also find other ways of collecting data, such as this method of observing people's actual habits in the restroom.

1.66 (a) The sample is the survey participants, the population is all professors at the University of Nebraska.

(b) No, we cannot conclude that the sample of survey responders is not representative of professors at the University of Nebraska since we are not given enough information to decide one way or the other.

(c) No, the 94% is based on self descriptions, which can be (and in this case, probably are) biased.

1.67 No. This is a volunteer sample, and there is reason to believe the participants are not representative of the population. For example, some may choose to participate because they LIKE alcohol and/or marijuana, and those in the sample may tend to have more experience with these substances than the overall population. In addition, the advertisements for the study were aired on rock radio stations in Sydney, so only those people who listen to rock radio stations in Sydney would hear about the option to participate.

1.68 Yes! The sample is a random sample so we can be quite confident that it is probably a representative sample.

1.69 (a) This is not a simple random sample from the population, since only those who saw and wanted to click and complete the survey were included.

(b) These results could also have been biased by how the survey was constructed. The wording of the questions might also introduce bias.

1.70 The study given found a relationship in a sample of rats. This relationship may not generalize to the human population.

1.71 The sample of planes that return from bombing missions was biased. More bullet holes were found in the wings and tail because planes that were shot in other regions were more likely to crash and not return.

1.72 (a) The population in the CPS is all US residents. (Also acceptable: US citizens, US households...)

(b) The population in the CES survey is all non-farm businesses and government agencies in the US.

(c) i. The CES survey would be more relevant, because the question pertains to companies.
   ii. The CPS would be more relevant, because the question pertains to American people.
   iii. The CPS would be more relevant, because the question pertains to people, not businesses.

1.73 (a) Since the NHANES sample is drawn from all people in the US, that is the population we can generalize to.

(b) Since the NHAMCS sample is drawn from patients in emergency rooms in the US, we can generalize the results to all emergency room patients in the US.

(c) i. NHANES: The question about an association between being overweight and developing diabetes applies to all people in the US, not just those who visit an emergency room.
   ii. NHAMCS: This question asks specifically about the type of injury for people who go to an emergency room.
   iii. NHAMCS: This question of average waiting time only applies to emergency room patients.
   iv. NHANES: This question is asking about all US residents. Note that the proportion would be equal to one for the people sampled in NAMCS since they only get into the sample if they visit an emergency room!

1.74 Answers will vary. See the technology notes to see how to use specific technology to select a random sample.

1.75 Answers will vary. See the technology notes to see how to use specific technology to select a random sample.

Section 1.3 Solutions

1.76 The use of "improves" implies this is a causal association.

1.77 Since "no link is found" there is neither association nor causation.

1.78 The phrase "leads to deaths" indicates a causal association.

1.79 The phrase "more likely" indicates an association, but there is no claim that wealth *causes* people to lie, cheat or steal.

1.80 The phrase "tend to be more educated" indicates an association, but there is no claim that owning a cat causes more education (or that better education causes people to prefer cats).

1.81 The statements imply that eating more fiber will cause people to lose weight, so this is a causal association.

1.82 One possible confounding variable is temperature (or season). More people eat ice cream, and go swimming, in warm weather. Other answers are possible. Remember that a confounding variable should be associated with both of the variables of interest.

1.83 One possible confounding variable is population. Increasing population in the world over time may mean more beef and more pork is consumed. Other answers are possible. Remember that a confounding variable should be associated with both of the variables of interest.

1.84 One possible confounding variable is wealth. People who own a yacht are likely wealthy and can afford a sports car. Other answers are possible. Remember that a confounding variable should be associated with both of the variables of interest.

1.85 One possible confounding variable is snow in the winter. When there is more snow, sales of both toboggans and mittens will be higher. Remember that a confounding variable should be associated with both of the variables of interest.

1.86 One possible confounding variable is number of cars (and also number of people). If there are lots of cars, there will be more pavement and more air pollution. Remember that a confounding variable should be associated with both of the variables of interest.

1.87 One possible confounding variable is gender. Males usually have shorter hair and are taller. Other answers are possible. Remember that a confounding variable should be associated with both of the variables of interest.

1.88 We are not manipulating any variables in this study, we are only collecting information (rice preference and metabolism) as they exist. This is an observational study.

1.89 We are actively manipulating the explanatory variable (playing music or not), so this is an experiment.

1.90 We are actively manipulating the explanatory variable (planting trees or not), so this is an experiment.

1.91 We are not manipulating any variables in this study, we are only measuring things (omega-3 oils and water acidity) as they exist. This is an observational study.

1.92 Data were collected after the fact from sprinters, marathon runners, and non-athletes. No genes were manipulated. These data came from an observational study.

1.93 The penguins in this study were randomly assigned to get either a metal or an electronic tag so this is an experiment.

1.94 All three studies are experiments since the scientists actively control the treatment (tears or salt solution).

1.95 A possible confounding variable is amount of snow and ice on the roads. When more snow and ice has fallen, more salt will be needed *and* more people will have accidents. Notice that the confounding variable has an association with *both* the variables of interest.

1.96 Age or grade level! Certainly, students in sixth grade can read substantially better than students in first grade and they are also substantially taller. Grade level is influencing both of the variables and it is a confounding variable. If we look at individual grades one at a time, the association could easily disappear.

1.97 (a) The cases are people. The two variables are whether or not the person is a golfer and how long the person lives.

(b) No variables were manipulated, so this result comes from an observational study.

(c) A confounding variable is a variable that is associated with the other variables and may help explain the association. Wealth or income is one obvious confounding variable, since golfing is a relatively expensive sport and therefore associated with wealth and wealth is also associated with living longer. Other confounding variables are also possible.

1.98 (a) The cases are the $1,773$ participants.

(b) The explanatory variable is the group to which the participant is assigned. It is categorical.

(c) This is an experiment since the group is randomly assigned by the researchers.

(d) There are $1,773$ cases and four variables (group assignment and all three attitude measures), so the dataset will have 1773 rows and 4 columns.

1.99 Yes, this study provides evidence that louder music causes people to drink more beer, because the explanatory variable (volume of music) was randomly determined by the researchers and an association was found.

1.100 Study 1 is a randomized, controlled experiment, while Study 2 is observational. Therefore, Study 1 provides better evidence of causation, while Study 2 is more prone to confounding variables (e.g., people who eat more nuts may also be more healthy in other ways).

1.101 (a) Yes. Because the study in mice was a randomized experiment, we can conclude causation.

(b) No. Since it appears that the study in humans was an observational study, it is not appropriate to conclude causation. Although the headline may still be true for humans, we cannot make this conclusion based on the study described.

1.102 (a) Since the students weren't assigned to one type of diet or another, this is an observational study.

(b) No, we cannot conclude a cause-and-effect relationship since the result does not come from a randomized experiment.

(c) A confounding variable is a variable that is associated with the other variables and may help explain the association. In this case, there are many possible confounding variables. One is exercise, since students who exercise may also tend to eat well and may also have lower rates of depression.

1.103 (a) The two variables are amount of iron in the soil and amount of potassium in spinach grown in that soil. Both variables are quantitative.

(b) Amount of iron in the soil is the explanatory variable since it appears to affect the amount of potassium in the spinach.

(c) Since there is no indication that the amount of iron in the soil was controlled by the researchers, the data appear to come from an observational study.

(d) No. Since the data do not come from an experiment, we cannot assume causation.

(e) A confounding variable is a variable that might affect both the amount of iron in the soil and the amount of potassium in the spinach leaves. Examples might be the amount of rainfall or the temperature of the soil or other chemicals in the soil.

1.104 (a) The cases are the children being studied and there are 19 of them. There are three variables: Hours spent reading, Hours of screen time, and Connectivity in the brain.

(b) All three variables are quantitative. It appears that hours spent reading and hours of screen time are the explanatory variables and connectivity in the brain is the response variable.

(c) This is an observational study, since no variables were controlled.

(d) No, we cannot conclude causation since this was not an experiment.

(e) A confounding variable affects both variables in an association and might help explain the apparent association. In this case, one possible confounding variable might be the connectivity in the parents' brains, since parents with higher brain connectivity might be more likely to read to their kids and less likely to have them spend time on screens, and children of parents with high brain connectivity might be more likely to inherit high brain connectivity. Other answers are possible.

1.105 (a) This is an observational study since none of the variables was controlled.

(b) There are several possible answers. Wealth is one, since wealth has been shown to be associated with better health and wealthy people might find it easier to take vacations. Job stress might be another since job stress might harm health and job stress might also make it difficult to take vacations.

(c) The results of this study indicate an association but not a causation association. Thus the first two statements are appropriate and the second two are not. (i) Yes (ii) Yes (iii) No (iv) No.

(d) No. "Want to be Healthier? Take a Vacation!" implies cause-and-effect, which is not an appropriate conclusion from this observational study.

1.106 No, this was an observational study and allowed for many confounding variables.

1.107 (a) The cases are the 2,623 schoolchildren.

(b) The explanatory variable is the amount of greenery around the schools.

(c) The response variable is the score on the memory and attention tests.

(d) Yes, the headline implies that more green space causes kids to be smarter.

(e) No variables were manipulated so this is an observational study.

(f) No! Since this is not an experiment, we cannot conclude causation.

(g) The socioeconomic status of the children is a possible confounding variable, since it is likely to effect both the amount of green space and also the test scores. There are other possible answers.

1.108 (a) The cases are the Danish men who were included in the study.

(b) The explanatory variable is whether or not the person has been in the hospital for an infection. Since the answer is either yes or no, this is a categorical variable.

(c) The response variable is the IQ score for the person. We can find an average so this is a quantitative variable.

(d) Yes, the headline implies that infections lower IQ, which is causation.

(e) The explanatory variable was not manipulated, so this is an observational study.

(f) No, since this was not an experiment, it is not appropriate to conclude causation. There are many possible lurking variables that might impact both IQ and the likelihood of having an infection.

1.109 (a) The explanatory variable is amount of leisure time spent sitting and the response variable is whether or not the person gets cancer.

(b) This is an observational study because the explanatory variable was not randomly assigned.

(c) No, we cannot conclude spending more leisure time sitting causes cancer in women because this is an observational study.

(d) No, we also cannot conclude that spending more leisure time sitting does not cause cancer in women; because this was an observational study we can make no conclusions about causality. Sitting may or may not cause cancer.

1.110 (a) The explanatory variable is whether or not the participants were given access to food and drink after 10pm, or just water. The response variables are reaction time and number of attention lapses.

(b) This is a randomized experiment because the explanatory variable was randomly assigned.

(c) Yes, we can conclude that eating late at night worsens reaction time and increases attention lapses for sleep deprived people.

(d) No, there are not likely to be confounding variables, because this was a randomized experiment.

1.111 (a) This is an observational study because we cannot randomly determine when a child learns to talk.

(b) No, we cannot conclude that early language skills reduce preschool tantrums because this is an observational study.

(c) There are many possible confounding variables, such as the intelligence of the child or parental involvement.

1.112 (a) No, because it is an observational study, not a randomized experiment. People were not randomly assigned to buy organic food or not, but rather chose for themselves, so the groups almost certainly differed to begin with.

(b) Answers will vary, but possible answers may include income (organic food costs more money and wealthier people are generally healthier) or other health-conscious behaviors (people who care about eating organic may eat healthier in general).

(c) No, because we do not have evidence against alternative explanation (ii); we cannot determine whether explanation (i) or (ii) (or both) are driving the observed difference.

1.113 (a) Yes, because this was a randomized experiment; it was randomly determined which flies ate organic and which ate conventional food.

(b) Answer to part (a) was not no.

(c) Yes, because we have evidence against both of the competing alternative explanations, we have convincing evidence that eating organic food really does make fruit flies live longer.

1.114 (a) Since all the children are presented with all three formats, this is a matched pairs experiment (although it might more accurately be called a matched triples experiment!)

(b) Yes, we can conclude causation since the results come from an experiment.

(c) The formats (and the stories used) should be presented to the children in random order. (This is how the experiment was conducted.)

1.115 (a) The explanatory variable is whether the person just had a full night of sleep or 24 hours of being awake. The response variable is ability to recognize facial expressions.

(b) This is a randomized experiment, a matched pairs experiment because each person received both treatments.

(c) Yes, we can conclude that missing a night of sleep hinders the ability to recognize facial expressions, because the explanatory variable was randomly assigned.

(d) No, we cannot conclude that better quality of REM sleep improves ability to recognize facial expressions, because the explanatory variable in this case (quality of REM sleep) was not randomly assigned.

1.116 (a) No, we cannot conclude that drinking diet soda causes weight gain because the study was observational and prone to confounding variables. For example, it is possible that seniors who drink more diet soda do so because they know they are prone to gaining weight.

(b) Neither study is perfect. The study on senior citizens is observational and the study on rats may not generalize to humans. However, together these studies provide a more convincing case that diet soda can cause weight gain. Opinions can vary on how strong this evidence for causation is.

1.117 (a) It is an observational study since no one assigned some people to live in a city and some to live in the country.

(b) No, since we can never conclude from an observational study that there is a causal association.

(c) The 2011 study is also an observational study, since, again, no one assigned some people to live in a city and some to live in the country.

(d) The explanatory variable is whether or not the participant lives in the city or the country, which is a categorical variable. The response variable is level of activity in stress centers of the brain, which is quantitative.

(e) No! The results come from an observational study, so we cannot conclude a causal relationship.

1.118 The explanatory variables are the type of *payment* and *sex*. Only the type of payment can be randomly assigned. The number of *items* ordered and *cost* are response variables.

1.119 (a) The explanatory variable is whether or not the person had a good night's sleep or is sleep-deprived. The response variable is attractiveness rating.

(b) Since the explanatory variable was actively manipulated, this is an experiment. The two treatments are well-rested and sleep-deprived. Since all 23 subjects were photographed with both treatments, this is a matched pairs experiment.

(c) Yes, we can conclude that sleep-deprivation causes people to look less attractive, because this is an experiment.

1.120 (a) We randomly divide the participants into two groups of 25 each. Half will be given fluoxetine and half will get a placebo.

(b) The placebo pills will look exactly like the fluoxetine pills and will be taken the same way, but they will not have any active ingredients.

(c) The patients won't know who is getting which type of pill (the fluoxetine or the placebo) and the people treating the patients and administering the questionnaire won't know who is in which group.

1.121 (a) The explanatory variable is amount of sleep and the response variable is growth in height.

(b) We would take a sample of children and randomly divide them into two groups. One group would get lots of sleep and the other would be deprived of sleep. Then after some time passed, we would compare the amount of height increase for the children in the two groups.

(c) An experiment is necessary in order to verify a cause and effect relationship, but it would definitely not be appropriate to randomly assign some of the kids to be sleep-deprived for long periods of time just for the purposes of the experiment!

1.122 (a) Randomly assign 25 people to carbo-load and 25 people to not carbo-load and then measure each person's athletic performance the following day.

(b) We would have each person carbo-load and not carbo-load, on different days (preferably different weeks). The order would be randomly determined, so some people would carbo-load first and other people would carbo-load second. In both cases athletic performance would be measured the following day and we would look at the difference in performance for each person between the two treatments.

(c) The matched pairs experiment is probably better because we are able to compare the different effects for the same person. It is more precise comparing one person's athletic performance under two different treatments, rather than different people's athletic performance under two different treatments.

1.123 (a) Randomly divide the students into two groups of 20 students each. One group gets alcohol and the other gets water. Measure reaction time for students in both groups.

(b) Measure reaction time for all 40 students both ways: after drinking alcohol and after drinking water. Do the tests on separate days and randomize the order in which the students are given the different treatments. Measure the difference in reaction time for each student.

1.124 Answers will vary. Example: The total amount of pizza consumed and the total amount of cheese consumed, per year, over the last century. Eating more pizza causes people to eat more cheese, but the overall rise in population is also a confounding variable.

Chapter 1

Collecting Data

Chapter Overview

We begin this chapter by introducing the basic structure of datasets, with cases and variables. In the first section, we also include motivating examples to get students thinking about the kinds of questions that can be answered using data. The goal is also to build enthusiasm for how powerful and fun statistics can be.

The rest of the chapter is about collecting data. The second section is about sampling and the third about experiments and observational studies. In these sections, we have chosen *not* to include a long list of many of the standard definitions. (We don't, for example, formally define non-response bias or stratified sampling.) The focus, instead, is on helping students think critically about how data are collected, and on helping students understand how the collection method influences the possible scope of inference. We revisit these topics regularly throughout the rest of the book to continue to reinforce these ideas.

1.1 The Structure of Data

Key Concepts

- **Datasets have cases and variables**
- **Variables can be classified as either quantitative or categorical**
- **We can use data to investigate interesting questions**
- **In investigating relationships between variables, we can sometimes identify an explanatory variable and a response variable**

Timing

One class or less. If your students have weaker backgrounds, you might delay talking about explanatory and response variables until a later section. This section is designed to be covered in the same class in which you go over your syllabus, and the material can shrink or expand to fill the allotted time.

Class Notes

Datasets have cases and variables. A great way to set the tone for the course is to begin the first class by collecting data directly from the students. This dataset can then provide examples to use in class all semester long. It can also be a perfect jumping off point to talk about cases and variables. See more details in the class activity discussion below. You will also want to have some other examples ready to go, where you can ask the students to identify the cases and the variables. Suggestions are given below.

Variables can be classified as either quantitative or categorical. For each of the examples used to identify cases and variables, have the students identify each of the variables as categorical or quantitative. Again, suggested examples are given below. Make sure students know that all of the analysis that follows in the entire book depends on first identifying variables as categorical or quantitative. It will pay off later to spend time on this now!

We can use data to investigate interesting questions. Try to find a few examples that your students will be interested in, to begin to generate interest in the subject. Some ideas are given below, or use examples from the section. Another idea is to pull some recent headlines[1] and have the students talk about how to use data to investigate the question of interest. Get them to identify cases and variables, and to begin to recognize how powerful data analysis can be!

In investigating relationships between variables, we can sometimes identify an explanatory variable and a response variable. This concept is less important than the others (at this point) and can be delayed until a later section if desired. The concept is helpful, though, in getting the students thinking about relationships and can fit in well with the earlier examples. It also leads nicely into the discussion of experiments that will come in Section 1.3.

[1] A good source for links to recent news articles of interest for teaching statistics is the Chance News project at http://test.causeweb.org/wiki/chance/index.php/Main_Page

Class Activity

Have students answer questions on a student survey. We give two options below, but feel free to modify them to work for your students. Be sure to have at least two quantitative and at least two categorical variables. The more variables, the better!

Tell your students that statistics is all about data and we need to collect data! Give them time during class to fill out the survey (or have them do it online[2] outside of class). After collecting the data, have them go through the variables and identify each as categorical or quantitative. Also, have them come up with questions they might like to answer, using these variables. (If you ask the "Pulse" question, be sure to have a way to time 60 seconds while they count their pulse.) If you have multiple sections of your Intro Stats course, it is great to combine the data from all the sections into one large file, since that increases anonymity.

Once you have the data, you can use it for class examples throughout the semester. Using data about YOUR students makes it more real to them. For privacy reasons, though, don't share the full dataset with the students - just share summary statistics and graphs as needed.

- *Option 1:* Have a student survey ready to go to collect data in the first class. One option is given as the Section 1.1 Activity in the Instructor Resources. You can use the questions there or modify them in any way that works for you. The key is to have enough variables that you can use the data for interesting analysis later in the course.
- *Option 2:* Have the students themselves create the student survey! During the first class, ask them what they would like to ask their peers and have them decide the questions to put on the survey. Use these suggestions to create a student survey to bring to the second class (or set it up online) and have them fill out. This idea is also given in more detail in the Activity handout.

Class Examples

In addition to the class examples that arise from discussing the student survey, here are a couple other suggestions. See also the PowerPoint slides and the Class Examples handout. Also, check the text exercises for additional ideas. There are more exercises than needed and many are quite interesting and would make great class examples. The datasets that come with the book also provide a good source for getting your students comfortable thinking about cases and variables.

Example *Hollywood Movies 2011*

Here is a small part of the dataset that includes information on all 136 movies to come out of Hollywood in 2011. (The full dataset is available in **HollywoodMovies2011**.) *Rating* is audience rating on a 100-point scale, *Budget* is budget in millions of dollars, and *Opening* is opening weekend gross, in millions of dollars.

Title	LeadStudio	Rating	Genre	Budget	Opening
Insidious	Sony	65	Horror	1.5	13.27
Harry Potter and Deathly Hallows P2	Warner Bros	92	Fantasy	125.0	169.19
Bridesmaids	Relativity	77	Comedy	32.5	26.25
The Help	DreamWorks	91	Drama	25.0	26.04
Horrible Bosses	Warner Bros	72	Comedy	35.0	28.30
Transformers: Dark of the Moon	DreamWorks	67	Action	195.0	197.85

[2]For example, you can set up a survey relatively easily at www.surveymonkey.com.

(a) What are the cases in this dataset?

(b) What are the variables? Identify each as quantitative or categorical:

(c) Name a question we might ask about this dataset that is about a single variable. Name a question we might ask that is about a relationship between two of the variables.

Solution

(a) The movies are the cases.

(b) There are five variables. (Note that "Title" is just an identifier column and is not a variable.) "LeadStudio" and "Genre" are both categorical while "Rating", "Budget", and "Opening" are all quantitative.

(c) There are many possible answers. For a question about a single variable: What proportion are comedies or what is an average audience rating or which movie had the largest budget? Let the students suggest ideas! For a relationship between two of the variables: Which genre of movie has the highest budget? Is audience rating related to opening weekend gross? Is one of the studios more likely to make horror films? Do movies with a higher budget tend to get higher audience ratings? You also might use some of these questions (or student suggestions) to ask about which would be the explanatory variable and which the response variable.

◇

Example *Yogurt and Weight Loss*

Can eating a yogurt a day cause people to lose weight (as compared to not eating yogurt)? In a study investigating this:

(a) What are the cases?

(b) What are the variables and is each quantitative or categorical?

(c) What might a dataset look like that might help us answer this question? Make up some data to show the first couple of rows of such a dataset.

(d) Which variable is the explanatory variable? Which is the response variable?

Solution

(a) The cases are the people involved in the study.

(b) One variable is whether or not the person eats a yogurt a day, which is categorical. Another variable is the amount of weight loss, which is quantitative. (Other reasonable student responses might be: Did the person lose weight (yes/no), which is categorical or record weights as two quantitative variables before and after the study.)

(c) Students have a surprisingly hard time with this, and asking questions such as this really seems to help them internalize the idea of cases and variables. We have:

Person	Yogurt?	Weight loss
1	Yes	10
2	No	2
3	No	−3

(d) Since we are interested in the effect of yogurt on weight loss, the explanatory variable is whether or not the person eats a yogurt a day and the response variable is the weight loss.

◇

Exercise Notes

Exercises 1 - 6, 11, 12, 13, 14, 15, 16, 17, 18, 19, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33 ask students to identify cases and categorical/quantitative variables.
Exercises 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33 ask students to think about how data is used to answer a question.
Exercises 7 - 10, 13, 14, 15, 16, 22, 23 ask students to identify explanatory and response variables.

Exercises 30, 31, 32, and 33 ask students to think more deeply about how data is recorded. Exercises 30, 31, 32, and 33 are open-ended.

There are no exercises requiring technology, although students might choose to look at the online datasets for Exercises 11 and 12.

Suggested Exercises:

Skill Builders: 1 - 10 odd (or even); Exercises: 12, 16, 22, 27, 30

1.2 Sampling from a Population

Key Concepts

- **Samples come from populations**
- **Samples can be biased**
- **If a sample is representative of a population, we can use information from a sample to infer information about a population**
- **Random sampling is the most effective way to get a representative sample**

Timing

One class.

Section Overview

The main goal of this section is to get students to recognize and think about possible sources of bias in sampling, and to recognize how hard it is to avoid bias. We do *not* give a laundry list of terms (such as systematic sampling, stratified sampling, non-response bias, etc.) but rather just encourage students to think critically about how data are collected. This approach may be different than that often taken in other books. At the end of the section, students should know:

- Random sampling is very good!
- Random does not mean "haphazard", and people can't do random without using an objective randomization method.
- Volunteer samples are generally not representative samples.
- There are many forms of bias.
- The statistical inference we talk about in the rest of the book depends on having good sampling methods.

Class Notes

Samples come from populations. This is an easy concept for the students and an important one for them to internalize. The activity below involving Lincoln's Gettysburg Address provides a great way to get students thinking about sampling from a population and is an excellent way to start the class. You will also want a couple more quick examples of samples from populations. Suggestions are given below.

Samples can be biased. If you use the Lincoln activity, that will probably provide an immediate example of biased sampling since students are likely to over-sample the large words. Give some more very obvious examples of biased samples (such as sampling students in the library on a Friday night) and then some more subtle ways in which a sample might give biased results. The example in the book asking whether parents would choose to have children again is a great one. Include at least one example of a volunteer sample and at least one example of wording bias. See examples below.

If a sample is representative of a population, we can use information from a sample to infer information about a population. We talk about inference here in a very informal way, just to help prepare students for what is to come. The only goal is to get them to think generally about using information from a sample to estimate information for a population. Examples are below or use any recent survey to talk about generalizing the results to the broader population. If the survey is from a random sample (as are all the surveys put out by the Pew Research Center on pewresearch.org), all the better.

Random sampling is the most effective way to get a representative sample. It is important for the students to recognize that they can't do random on their own and that random does not mean "haphazard." A random sample is found only by pulling names out of a hat or using a random number generator or table or some equivalent method. You may or may not choose to show your students how to use a random number generator or table – this is an optional topic. Either way, though, the students should recognize the importance of selecting a random sample. *StatKey* can be used to demonstrate random samples, using the Sampling Distributions options. If time, return to the Lincoln activity from the start of class and repeat the activity using true random samples. Compare the results to those obtained with their biased samples.

Class Activity

Give students a copy of Lincoln's Gettysburg Address and ask them each to select a sample of 10 words to estimate the average number of letters in all words in the speech. (This is available on the Activity handout for this section.) Have them then record the number of letters in each of their 10 words and compute the average length for their sample.

This is a good opportunity to point out the population (all 268 words in Lincoln's Gettysburg Address) and the different samples that students selected from that population. Have the students all put their averages on a dotplot on the board, or you can enter the averages into a dataset as the students go around the room and announce theirs. Compare the resulting distribution to the correct average number of letters for all 268 words, which is 4.3.

It is likely that almost all the student averages are too high. Talk about the fact that they are more likely to sample large words while mostly ignoring the very small words. Discuss the fact that it is virtually impossible for humans to take a "random" sample, even when we know about the problem and are actively trying to compensate. We need to use an objective method, such as a random number generator.

If you have time, it is great to repeat this exercise at the end of class, using a random number generator or table to select random samples of size 10 from the list of all 268 words. The distribution students get using this method should be centered near the population mean of 4.3. This really drives the point home of how hard it is for humans to do "random".

Class Examples

In addition to the class examples that arise from the activity, here are a couple other suggestions. See also the PowerPoint slides and the Class Examples handout. Also, check the text exercises for additional ideas. There are more exercises than needed and many are quite interesting and would make great class examples. Websites with recent public opinion polls, such as pewresearch.org or gallup.com or rasmussenreports.com, also can provide great current examples to illustrate sampling.

Example *Number of Cell Phone Calls per Day*

Princeton Survey Research reports on a survey of 1,917 cell phone users in the US, conducted in May 2010, asking "On an average day, about how many phone calls do you make and receive on your cell phone?"

(a) What is the sample in this study?
(b) What is the population?
(c) Is the variable in this study quantitative or categorical?

Solution

(a) The 1,917 people who were asked the question.
(b) All cell phone users in the US
(c) Quantitative

◇

Example *Driving with a Pet on your Lap*

Over 30,000 people participated in an online poll on cnn.com conducted in April 2012 asking "Have you ever driven with a pet on your lap?" We see that 34% of the participants answered yes and 66% answered no.

(a) What is the sample?
(b) Can we conclude that 34% of all drivers have driven with a pet on their lap?
(c) Explain why it is not appropriate to generalize these results to all drivers, or even to all drivers who visit cnn.com. What is the problem with this method of data collection?
(d) How might we select a sample of people that would give us results that we can generalize to a broader population?
(e) Is the variable in this study quantitative or categorical?

Solution

(a) The 30,000 people who participated in the poll.
(b) No! This is a volunteer sample, and volunteer samples are usually biased. It is likely that people who have driven with a pet on their lap are more likely to respond to the poll.
(c) Volunteer samples are generally biased since people who feel strongly about the subject are much more likely to participate.
(d) Use a RANDOM SAMPLE!
(e) Categorical

◇

Example *Sampling Bias*

You wish to estimate the number of hours a week that students at your university spend studying. For each sampling method below, indicate whether the results are likely to be representative of the entire student body or whether the method is likely to give biased results.

(a) Ask students you find at the library on a Friday night.
(b) Ask students who are at a party on a Tuesday night.
(c) Ask a random sample of students.
(d) Ask the students in Organic Chemistry.
(e) Ask all the students on one of the school's athletic teams.

(f) Send an email out to all students and use all the responses you get.
(g) Ask friends that you know, trying to get a reasonably representative group.

Solution

(a) Biased
(b) Biased
(c) Representative
(d) Biased
(e) Biased
(f) Biased, since this is a volunteer sample
(g) Biased, since bias can occur when people try pick what they think is representative

◇

Exercise Notes

Exercises 34 - 41, 42 - 45, 57, 66, 72, 73 ask students to identify samples and populations.
Exercises 42 - 45, 52 - 56, 58, 65, 66, 67, 69 ask students to identify when a sample is likely to be biased.
Exercises 46 - 51, 59, 60, 58, 74, 75 discuss random sampling.

Exercises 74 and 75 require students to use a random number generator or a random number table. Exercises 70 and 71 ask students to think more creatively about when we can generalize from a sample to a population.

Suggested Exercises:

Skill Builders: 34 - 56 odd (or even); Exercises: 58, 59, 67, 68

1.3 Experiments and Observational Studies

Key Concepts

- **Association does not imply causation!**
- **Confounding variables can explain an apparent association between two other variables**
- **Distinguishing between experiments and observational studies**
- **Randomized experiments provide the only method that allows us to infer causality**
- **Aspects of a well-designed experiment**

Timing

One or two classes.

Section Overview

This material will permeate the rest of the course and students especially need to recognize when they can and cannot conclude that there is a causal relationship. However, the more specific material at the end of the section (placebos, blinding, matched pairs) can be covered either very quickly or in detail, depending on the interests of the instructor.

Class Notes

Association does not imply causation! Start the class by discussing the difference between an association and a *causal* association. Students should be able to recognize when a statement implies a causal relationship. The distinction can be subtle sometimes, so include several examples. Use the examples below or pull some recent headlines and ask which appear to imply causation. You might want to share with the students this quote from Stephen Jay Gould: "The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning."

Confounding variables can explain an apparent association between two other variables. Define a confounding variable as a third variable that is associated with both the explanatory variable and the response variable in a relationship. (In this text, we don't differentiate between confounding variables, lurking variables, confounding factors, etc.) You want the students to recognize that confounding variables are often present in observational studies and that a confounding variable can be the real reason behind an apparent association. Give several associations (such as those suggested below in the class examples) and ask the students to suggest confounding variables. They have a surprisingly hard time with this.

Distinguishing between experiments and observational studies. As long as students can remember that, in an experiment, the explanatory variable is actively manipulated and, in an observational study, it isn't, they probably won't have a hard time with this. Again, a couple short examples will do the trick.

Randomized experiments provide the only method that allows us to infer causality. This is the most important concept of the section. You will want to make the point strongly in this class, and then reinforce it throughout the rest of the semester. It is worth asking, whenever we find an association in the remainder of the course, whether we can conclude that it is a causal association. Emphasize the importance of assigning the explanatory variable randomly. The class activity recommended below will help to make this point.

Aspects of a well-designed experiment. This part of the section can take another full class or can be covered quickly, depending on the course. It is worth talking at least quickly about the use of placebos and blinding, since the students will see those words in many of the examples and exercises throughout the rest of the book. Also, students find the placebo effect very interesting! There are some fun examples and information in the PowerPoint slides. It is fine to delay discussion of a matched pairs design until later in the course if desired.

Class Activity

This activity, on conducting a randomized experiment, should come right after you have talked about randomized experiments in the class (and before talking about placebos, matched pairs, etc.) We conduct an experiment to investigate the question:
Is just 10 seconds of exercise enough to increase pulse rates?
The treatment groups are exercise vs sitting and the response variable is pulse rate.

Randomly divide the class into two groups, using some legitimate randomization method (such as a random number generator or random number table, or putting all student names in a hat and drawing out half, or flipping a coin for each student to determine the group.) It is valuable for the students to see the actual randomization happen.

Have the students in one group exercise while you count out 10 seconds. (Let each person choose how to exercise: run in place, jumping jacks, or whatever. This is part of the fun!) Have the other group remain sitting. As soon as the 10 seconds is over, have all students find their pulse and then give them the start and stop signal as you time one minute for them as they count their pulse. Collect the data from them (using the class activity handout if desired) for later use, or enter it right away into a dataset that will allow you to compare means or to look at side-by-side dotplots.

You can also use this activity to connect to other class concepts. Ask the students to identify the cases and the variables, and to indicate whether variables are categorical or quantitative. Have students identify the explanatory variable and the response variable. Discuss possible confounding variables if we had *not* assigned groups randomly but instead had let students pick which treatment group they wished to join. Finally, ask students why a matched pairs design might be an even more effective way to collect data to address the question of interest.

Class Examples

Here are a couple suggestions for class examples. See also the PowerPoint slides and the Class Examples handout. Also, check the test exercises, including Review Exercises, for additional ideas. There are more exercises than needed and many are quite interesting and would make great class examples.

Example *Association or Causation?*

For each of the following statements, indicate whether the statement implies causation or just indicates association without causation.

(a) If you study more, your grade will improve in this course.

(b) Want to lose weight? Drink tea!
(c) Aging of the brain tends to be delayed in people with a college education.

Solution

(a) Causation. "Will improve" implies causation.
(b) Causation. The implication is that drinking tea will cause you to lose weight.
(c) Association without causation. The statement is not claiming that a college education will cause a delay in aging of the brain. In fact, there are many confounding factors.

◇

Example *Television Sets and Life Expectancy*

There is a very strong positive association between television sets per person in a country and the life expectancy of the people in that country, which means people in countries with lots of TVs tend to live longer.

(a) Does this mean we can improve the life expectancy of people in developing countries by sending them lots of television sets?
(b) Describe a possible confounding variable in this association.

Solution

(a) No! This strong association is not causation.
(b) The wealth of the countries is a confounding variable. Wealth will affect both the number of television sets and the life expectancy.

◇

Example *Does Kindergarten Lead to Crime?*

A state legislator in New Hampshire claimed in July 2012 that kindergarten programs lead to more crime. He noted as evidence that the largest of all the communities in his county was the only community with a kindergarten program and also had the most crime.

(a) What common mistake is the legislator making?
(b) Describe a possible confounding variable in this association.

Solution

(a) The legislator is assuming association implies causation. This a very common mistake! Just because there is a strong association does NOT mean that one is causing the other.
(b) The population of the communities. Larger communities are more likely to have a kindergarten program and are also more likely to have more crime.

◇

Example *Exercise and Depression*

In each case below, we describe a study to determine whether exercise helps increase certain mood enhancing chemicals in the brain. Indicate whether each describes an experiment or an observational study.

(a) We contact a random sample of 100 people and record how much each person exercises and also measure the chemicals in the brain for each person.
(b) Using a random sample of 100 people, we randomly assign half of them to participate in a regular exercise program for a six-week period while the other half makes no changes. At the end of the time period, we measure the brain chemicals.

Solution (a) Observational study. No variables were manipulated.
(b) Experiment. The experimenters actively imposed the explanatory variable (exercise or not).

◇

Example *Caffeine and Learning*
Does ingesting caffeine help mice learn the way through a maze faster? Describe a well-designed randomized experiment to test this. Give all details, including the use of placebos and blinding. We have 20 mice to work with.

Solution We randomly divide the mice into two groups of 10 mice each. We give half the mice caffeine in their food while the other group gets identical food without caffeine (placebo). We make the study double-blind by not telling the mice or anyone interacting with the mice (for example, whoever is timing the maze runs) which mice are getting the caffeinated food and which are not. We measure the time it takes for the mice to learn the maze and compare the results between the two groups. ◇

Example *Is the Dominant Hand Stronger?*
Describe a Matched Pairs experiment to see if right hand strength is greater than left-hand strength in right-handed people. You have a hand-grip to measure strength and 30 right-handed people to use in your study. Be sure to include how randomization is used in your design.

Solution We are comparing right-hand strength to left-hand strength and in a matched pairs experiment, all participants are measured for both, so we will measure the strength in both the left and right hand for all participants. We randomize the order in which the participants have hand strength measured: some randomly assigned to measure the strength in the left hand first and some randomly assigned to measure strength in the right hand first. Then we compare the right and left hand strength for each individual. ◇

Exercise Notes

Exercises 76 - 81, 99, 100, 101, 106, 107, 108, 109, 110, 111, 115, 116, 117, 119, 124 require an understanding of association versus causation.
Exercises 82 - 87, 95, 96, 107, 110, 111, 91 require an understanding of confounding variables.
Exercises 88 - 91, 92 - 94, 99, 100, 101, 106, 107, 108, 109, 110, 111, 115, 116, 117, 119 ask to distinguish between experiments and observational studies.
Exercises 115, 118, 120, 121, 122, 123 ask students to think about aspects of well-designed experiments.

Exercises 115, 119, 122, 123 are the only exercises about matched pairs design. Exercises 116, 120, 121, 122, 123 are open-ended problems asking students to describe an well-designed experiment. Exercise 124 is harder and requires students to have a deeper understanding of the ideas.
None of the exercises in this section require technology.

Suggested Exercises:

Skill Builders: 76 - 91 odd (or even); Exercises: 96, 99, 101, 110, 111, 117, 120, 122