Other Random Sampling Procedures

A commonly used method for selecting a sample of 100 from the list of 20,000 is to pick a single random start from the positions 1 to 200 and then to select every 200th subject going down the frame. This is called *systematic random sampling*. Another method is to group the frame into 200 blocks of 100 consecutive subjects each and to select one of the blocks at random. This is called *random cluster sampling*. In sampling units such as lakes of different sizes, it is sometimes useful to allow larger units to have higher probabilities of being sampled than smaller units. This constitutes *variable probability sampling*. These and other random sampling schemes can be useful, but they differ from simple random sampling in fundamental ways.

1.5.5 On Being Representative

Random samples and randomized experiments are representative in the same sense that flipping a coin to see who takes out the garbage is fair. Flipping is always fair before the coin hits the table, but the outcome is always unfair to the loser. In the same way, close examination of the results of randomization or random sampling can usually expose ways in which the chosen sample is not representative. The key, however, is not to abandon the procedure when its result is suspect. Uncertainty about representativeness will be incorporated into the statistical analysis itself. If randomization were abandoned, there would be no way to express uncertainty accurately.

1.6 SUMMARY

Cause-and-effect relationships can be inferred from randomized experiments but not from observational studies. The problem with observational studies is that confounding variables—identifiable or not—may be responsible for observed differences. Randomized experiments eliminate this problem by ensuring that differences between groups (other than those of the assigned treatments) are due to chance alone. Statistical measures of uncertainty account for this chance.

Statistically, the statements that generalize sample results to more general contexts are based on a probability model. When the model corresponds to the planned use of randomization or random sampling, it provides a firm basis for drawing inferences. The probability model may also be a fiction, created for the purposes of assessing uncertainty. Results from fictitious probability models must be viewed with skepticism.

1.7 EXERCISES

Conceptual Exercises

1. Creativity Study. In the motivation and creativity experiment, the poems were given to the judges in random order. Why was that important?

2. Sex Discrimination Study. Explain why it is difficult to prove sex discrimination (that males in a company receive higher starting salaries because they are males) even if it has occurred.

3. A study found that individuals who lived in houses with more than two bathrooms tended to have higher blood pressure than individuals who lived in houses with two or fewer bathrooms. (a) Can cause and effect be inferred from this? (b) What confounding variables may be responsible for the difference?

4. A researcher performed a comparative experiment on laboratory rats. Rats were assigned to group 1 haphazardly by pulling them out of the cage without thinking about which one to select. Should others question the claim that this was as good as a randomized experiment?

5. In 1930 an experiment was conducted on 20,000 school children in England. Teachers were responsible for randomly assigning their students to a treatment group—to receive $\frac{3}{4}$ pint of milk each day—or to a control group—to receive no milk supplement. Weights and heights were measured before and after the four-month experiment. The study found that children receiving milk gained more weight during the study period. On further investigation, it was also found that the controls were heavier and taller than the treatment group *before* the milk treatment began (more so than could be attributed to chance). What is the likely explanation and the implication concerning the validity of the experiment?

6. Ten marijuana users, aged 14 to 16, were drawn from patients enrolled in a drug abuse program and compared to nine drug-free volunteers of the same age group. Neuropsychological tests for short-term memory were given, and the marijuana group average was found to be significantly lower than the control group average. The marijuana group was held drug-free for the next six weeks, at which time a similar test was given with essentially the same result. The researchers concluded that marijuana use caused adolescents to have short-term memory deficits that continue for at least six weeks after the last use of marijuana. (a) Can a genuine causal relationship be established from this study? (b) Can the results be generalized to other 14- to 16-year-olds? (c) What are some potential confounding factors?

7. Suppose that random samples of Caucasian-American and Chinese-American individuals are obtained from patient records of doctors participating in a study to compare blood pressures of the two populations. Suppose that the individuals selected are asked whether they want to participate and that some decline. The study is conducted only on those that volunteer to participate, and a comparison of the distributions of blood pressures is conducted. Where does this study fit in Display 1.5? What assumption would be necessary to allow inferences to be made to the sampled populations?

8. More people get colds during cold weather than during warm weather. Does that prove that cold temperatures cause people to get colds? What is a potential confounding factor?

9. A study showed that children who watch more than two hours of television each day tend to have higher cholesterol levels than children who watch less than two hours of television each day. Can you think of any use for the result of this study?

10. What is the difference between a randomized experiment and a random sample?

11. A number of volunteers were randomly assigned to one of two groups, one of which received daily doses of vitamin C and one of which received daily placebos (without any active ingredient). It was found that the rate of colds was lower in the vitamin C group than in the placebo group. It became evident, however, that many of the subjects in the vitamin C group correctly guessed that they were receiving vitamin C rather than placebo, because of the taste. Can it still be said that the difference in treatments is what caused the difference in cold rates?

12. Fish Oil and Blood Pressure. Researchers used 7 red and 7 black playing cards to randomly assign 14 volunteer males with high blood pressure to one of two diets for four weeks: a fish oil diet and a standard oil diet. The reductions in diastolic blood pressure are shown in Display 1.14.

DISPLAY 1.14 Reductions in diastolic blood pressure (mm of mercury) for 14 men after 4 weeks of a special diet containing fish oil or a regular, nonfish oil

Fish oil diet:	8	12	10	14	2	0	0
Regular oil diet:	−6	0	1	2	−3	−4	2

(Based on a study by H. R. Knapp and G. A. FitzGerald, "The Antihypertensive Effects of Fish Oil," *New England Journal of Medicine* 320 (1989): 1037–43.) Why might the results of this study be important, even though the volunteers do not constitute a random sample from any population?

13. Why does a stem-and-leaf diagram require less space than an ordinary table?

14. What governs the *width* of a box plot?

15. What general features are evident in a box plot of data from a normal distribution? from a skewed distribution? from a short-tailed distribution? from a long-tailed distribution?

Computational Exercises

16. Gross Domestic Product (GDP) per Capita. The data file ex0116 contains the gross domestic product per capita for 228 countries—the data used to construct the box plot in Display 1.11. (a) Make a box plot of the per capita GDPs with a statistical computer program. Include a y-axis label (for example, "Gross Domestic Product per Capita in $U.S."). (b) In what ways, if any, is the display from your software different from Display 1.11? (c) Use a statistical computer program to draw a *histogram* of the per capita GDPs. Include an x-axis label. Use the program's default bin width. Report that bin width. (d) The program's default bin width for histograms is not necessarily the best choice. If it's possible with your statistical computer program, redraw the histogram of part (c) using bin widths of $5,000. (Data from Central Intelligence Agency, "Country Comparison: GDP — per capita (PPP)," *The World Factbook*, June 24, 2011 https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html (June 30, 2011).)

17. Seven students volunteered for a comparison of study guides for an advanced course in mathematics. They were randomly assigned, four to study guide A and three to study guide B. All were instructed to study independently. Following a two-day study period, all students were given an examination about the material covered by the guides, with the following results:

Study Guide A scores: 68, 77, 82, 85
Study Guide B scores: 53, 64, 71

Perform a randomization test by listing all possible ways that these students could have been randomized to two groups. There are 35 ways. For each outcome, calculate the difference between sample averages. Finally, calculate the two-sided p-value for the observed outcome.

18. Using the creativity study data (Section 1.1.1) and a computer, assign a set of random numbers to the 47 subjects. Order the entire data set by increasing values of the random numbers, and divide the subjects into group 1 with the 24 lowest random numbers and group 2 with the 23 highest. Compute the difference in averages. Repeat this process five times, using different sets of random numbers. Did you get any differences larger than the one actually observed (4.14)?

19. Write down the names and ages of 10 people. Using coin flips, divide them into two groups, as if for a randomized experiment. Did one group tend to get many of the older subjects? Was there any way to predict which group would have a higher average age in advance of the coin flips?

20. Repeat Exercise 19 using a randomization mechanism that ensures that each group will end up with exactly five people.

21. Read the methods and design sections of five published studies in your own field of specialization. (a) Categorize each according to Display 1.5. (b) Now read the conclusions sections. Determine whether inferential statements are limited to or go beyond the scope allowed in Display 1.5.

22. Draw back-to-back stem-and-leaf diagrams of the creativity scores for the two motivation groups in Display 1.1 (by hand).

23. Use the computer to draw side-by-side box plots for the creativity scores in Display 1.1.

24. Using the stem-and-leaf diagrams from Exercise 22, compute the median, lower quartile, and upper quartile for each of the motivation groups. Identify these on the box plots from Exercise 23. Using a ruler, measure the length of the box (the IQR) for each group, and make horizontal lines at 1.5 IQRs above and below each box and at 3 IQRs above and below the box. Are there any *extreme points* in either group? Are there any *very extreme points* in either group?

25. The following are zinc concentrations (in mg/ml) in the blood for two groups of rats. Group A received a dietary supplement of calcium, and group B did not. Researchers are interested in variations in zinc level as a side effect of dietary supplementation of calcium.

Group A: 1.31 1.45 1.12 1.16 1.30 1.50 1.20 1.22 1.42 1.14 1.23 1.59 1.11 1.10 1.53 1.52 1.17 1.49 1.62 1.29

Group B: 1.13 1.71 1.39 1.15 1.33 1.00 1.03 1.68 1.76 1.55 1.34 1.47 1.74 1.74 1.19 1.15 1.20 1.59 1.47

(a) Make a back-to-back stem-and-leaf diagram (by hand) for these measurements. (b) Use the computer to draw side-by-side box plots.

Data Problems

26. Environmental Voting of Democrats and Republicans in the U.S. House of Representatives. Each year, the League of Conservation Voters (LCV) identifies legislative votes taken in each house of the U.S. Congress—votes that are highly influential in establishing policy and action on environmental problems. The LCV then publishes whether each member of Congress cast a pro-environment or an anti-environment vote. Display 1.15 shows these votes during the years 2005, 2006, and 2007 for members of the House of Representatives. Evaluate the evidence supporting party differences in the percentage of pro-environment votes. Write a brief report of your conclusion, including a graphical display and summary statistics.

DISPLAY 1.15 Number of pro- and anti-environment votes in 2005, 2006, and 2007, according to the League of Conservation Voters, of Republican (R) and Democratic (D) members of the U.S. House of Representatives, and the total percentage of their votes that were deemed pro-environment; first 5 of 492 rows

State	Representative	Party	Pro05	Anti05	Pro06	Anti06	Pro07	Anti07	PctPro
Alabama	Bonner	R	2	16	3	9	2	18	14.0
Alabama	Everett	R	0	18	1	11	2	18	6.0
Alabama	Rogers	R	1	17	2	10	3	17	12.0
Alabama	Aderholt	R	0	18	0	12	2	18	4.0
Alabama	Cramer	D	5	13	4	7	14	6	46.9

27. Environmental Voting of Democrats and Republicans in the U.S. Senate. Display 1.16 shows the first five rows of a data set with pro- and anti-environment votes (according to the League of Conservation Voters; see Exercise 26) during 2005, 2006, and 2007, cast by U.S. senators. Evaluate the evidence supporting party differences in the percentage of pro-environment votes. Write a brief report of your conclusion, and include a graphical display and summary statistics.

DISPLAY 1.16 Number of pro- and anti-environment votes in 2005, 2006, and 2007, according to the League of Conservation Voters, of Republican (R) and Democratic (D) members of the U.S. Senate; and the total percentage of their votes that were deemed pro-environment; first 5 of 112 rows

State	Senator	Party	Pro05	Anti05	Pro06	Anti06	Pro07	Anti07	PctPro
Alabama	Session	R	1	18	0	7	2	13	7.3
Alabama	Shelby	R	1	19	0	7	1	14	4.8
Alaska	Murkowski	R	2	17	1	6	6	9	22.0
Alaska	Stevens	R	1	19	1	6	4	11	14.3
Arizona	Kyle	R	1	19	2	5	2	13	11.9

Answers to Conceptual Exercises

1. If one of the two treatment groups had their poems judged first, then the effect of motivation treatment would be confounded with time effects in the judges' marking of creativity scores. Judges are influenced by memories of previous cases. They may also discern a change in average quality, which could alter their expectations.

2. Statistically, the distributions of male and female starting salaries may be compared after adjusting for possible confounding variables. Since the data are necessarily observational, a difference in distribution cannot be linked to a specific cause. Once more, for emphasis: The best possible statistical analysis using the best possible data cannot establish causation in an observational study, but observational data are the only data likely to be available for discrimination cases. If, therefore, courts required plaintiffs to produce scientifically defensible proof of discrimination in order to prevail, defendants would win all discrimination cases *by definition*. As a result, some courts that wish to give weight to statistical information in discrimination cases adopt rules of evidence that allow proof to be established negatively—by the lack of an adequate rebuttal.

3. (a) No. (b) Wealth and richness of diet.

4. Yes. First of all, there is the possibility that the rats that were easier to get out of the cage are different from the others—bigger, less mobile, perhaps. Second, even if there is no obvious reason why the rats in the two groups might be different, that does not ensure they are not. Researchers must maintain skepticism about the conclusions in light of the uncertainty. With proper randomization, which is easy to carry out, there would be no doubt.

5. Teachers apparently gave the milk to the students they thought would most benefit from it. As a consequence, the results of the experiment were not valid.

6. (a) No. In this observational study, it is possible that the drug users were different from the nonusers in ways other than drug use. (b) No, because the samples are volunteers, not random samples. (c) Happiness of the child, stability of family, success of child in school.

7. It is an observational study. The population that is randomly sampled is the population of consenting subjects. To draw inference to all subjects requires the assumption that the response is unrelated to the reasons for consent.

8. No. The amount of time people spend indoors.

9. It may be possible for doctors to identify children who are likely to have high cholesterol by asking about their television watching habits. This requires no causative link. It is a minor use, however, and there is apparently no other.

10. In a randomized experiment a random mechanism is used to allocate the available subjects to treatment groups. In a random sample a random mechanism is used to select subjects from the populations of interest.

11. Yes, sort of. The treatment difference caused the different responses, but the actual "treatment" received in the vitamin C group was both a daily dose of vitamin C and knowledge that it was vitamin C. It's possible that the second aspect of this treatment is what was responsible for the difference. Researchers must make sure that the two groups are treated as similarly as possible in all respects, except for the specific agent under comparison.

12. The conclusion that the fish oil diet causes a reduction in blood pressure for these volunteers is a strong and useful one, even if it formally applies only to these particular individuals.

13. The leading digit or digits (the stems) are listed only once.

14. The width of the box is chosen to make the overall picture pleasing to the eye. It does not represent anything. For side-by-side box plots, the widths of the two boxes should be equal.

15. Normal: Median line in the middle of the box; equal whiskers; few if any extreme points; no very extreme points. Skewed: extreme points in one direction only; very extreme points possible, but few; long whisker on side of extreme points and short whisker (if any) on other side; median line closer to short whisker than to long one. Short-tailed: Like normal with no extreme points and very short whiskers. Long-tailed: Like normal (roughly symmetric) with extreme points strung out in both tails; some very extreme points possible.