- Know how to compare the responses in different treatment groups to assess whether the differences are larger than could be reasonably expected from ordinary sampling variability.
- Know how to report the results of an experiment. Tell who the subjects are and how their assignment to treatments was determined. Report on how the response variable was measured and in what measurement units.
- Understand that your description of an experiment should be sufficient for another researcher to replicate the study with the same methods.
- Be able to report on the statistical significance of the result in terms of whether the observed group-to-group differences are larger than could be expected from ordinary sampling variation.

Experiments on the Computer

Most experiments are analyzed with a statistics package. You should almost always display the results of a comparative experiment with side-by-side boxplots. You may also want to display the means and standard deviations of the treatment groups in a table.

The analyses offered by statistics packages for comparative randomized experiments fall under the general heading of Analysis of Variance, usually abbreviated ANOVA. These analyses are beyond the scope of this chapter. You'll find a discussion of ANOVA in Chapter 28 on the CD-ROM that comes with this book.

EXERCISES

1–20. What's the design? Read each brief report of statistical research, and identify:

a) whether it was an observational study or an experiment.

If it was an observational study, identify (if possible)

b) whether it was retrospective or prospective.
c) the subjects studied, and how they were selected.
d) the parameter of interest.
e) the nature and scope of the conclusion the study can reach.

If it was an experiment, identify (if possible)

b) the subjects studied.
c) the factor(s) in the experiment, and the number of levels for each.
d) the number of treatments.
e) the response variable measured.
f) the design (completely randomized, blocked, or matched).
g) whether it was blind (or double-blind).
h) the nature and scope of the conclusion the experiment can reach.

1. Over a 4-month period, among 30 people with bipolar disorder, patients who were given a high dose (10 g/day) of omega-3 fats from fish oil improved more than those given a placebo. (*Archives of General Psychiatry* 56 [1999]: 407)

2. The leg muscles of men aged 60 to 75 were 50% to 80% stronger after they participated in a 16-week, high-intensity resistance-training program twice a week. (*Journal of Gerontology* 55A [2000]: B336)

3. In a test of roughly 200 men and women, those with moderately high blood pressure (averaging 164/89 mm Hg) did worse on tests of memory and reaction time than those with normal blood pressure. (*Hypertension* 36 [2000]: 1079)

4. Among a group of disabled women aged 65 and older who were tracked for several years, those who had a vitamin B_{12} deficiency were twice as likely to suffer severe depression as those who did not. (*American Journal of Psychology* 157 [2000]: 715)

5. An examination of the medical records of more than 360,000 Swedish men showed that those who were overweight or who had high blood pressure had a higher risk of kidney cancer. (*New England Journal of Medicine* 3434 [2000]: 1305)

6. To research the effects of "dietary patterns" on blood pressure in 459 subjects, subjects were randomly assigned to three groups and had their meals prepared by dieticians. Those who were fed a diet low in fat and cholesterol and high in fruits, vegetables, and low-fat dairy foods (known as the DASH diet) lowered their systolic blood pressure by an average of 6.7 points when compared with subjects fed a control diet.

7. After menopause many women take supplemental estrogen. There is some concern that if these women also drink alcohol, their estrogen levels will rise too high. Twelve volunteers who were receiving supplemental estrogen were randomly divided into two groups, as were 12 other volunteers not on estrogen. In each case, one group drank an alcoholic beverage, the other a nonalcoholic beverage. An hour later everyone's estrogen level was checked. Only those on supplemental estrogen who drank alcohol showed a marked increase.

8. Is diet or exercise effective in combating insomnia? Some believe that cutting out desserts can help alleviate the problem, while others recommend exercise. Forty volunteers suffering from insomnia agreed to participate in a month-long test. Half were randomly assigned to a special no-desserts diet; the others continued desserts as usual. Half of the people in each of these groups were randomly assigned to an exercise program, while the others did not exercise. Those who ate no desserts and engaged in exercise showed the most improvement.

9. Some gardeners prefer to use nonchemical methods to control insect pests in their gardens. Researchers have designed two kinds of traps, and want to know which design will be more effective. They randomly choose 10 locations in a large garden and place one of each kind of trap at each location. After a week they count the number of bugs in each trap.

10. Researchers have linked an increase in the incidence of breast cancer in Italy to dioxin released by an industrial accident in 1976. The study identified 981 women who lived near the site of the industrial explosion and were under age 40 at the time. Fifteen of the women had developed breast cancer at an unusually young average age of 45. Medical records showed that these women had heightened concentrations of dioxin in their blood, and that each 10-fold increase in dioxin level was associated with a doubling of the risk of breast cancer. (*Science News*, Aug. 3, 2002)

11. In 2002 the journal *Science* reported that a study of women in Finland indicated that having sons shortened the lifespans of mothers by about 34 weeks per son, but that daughters helped to lengthen the mothers' lives. The data came from church records from the period 1640 to 1870.

12. In 2001 a report in the *Journal of the American Cancer Institute* indicated that women who work nights have a 60% greater risk of developing breast cancer. Researchers based these findings on the work histories of 763 women with breast cancer and 741 women without the disease.

13. The May 4, 2000, issue of *Science News* reported that, contrary to popular belief, depressed individuals cry no more often in response to sad situations than nondepressed people. Researchers studied 23 men and 48 women with major depression, and 9 men and 24 women with no depression. They showed the subjects a sad film about a boy whose father has died, noting whether or not the subjects cried. Women cried more often than men, but there were no significant differences between the depressed and nondepressed groups.

14. Scientists at a major pharmaceutical firm investigated the effectiveness of an herbal compound to treat the common cold. They exposed each subject to a cold virus, then gave him or her either the herbal compound or a sugar solution known to have no effect on colds. Several days later they assessed the patient's condition, using a cold severity scale ranging 0 to 5. They found no evidence of benefits associated with the compound.

15. Scientists examined the glycogen content of rats' brains at the rats' normal bedtimes and after they had been kept awake for an extra 6, 12, or 24 hours. The scientists found that glycogen was 38% lower among rats that had been sleep-deprived for 12 hours or more, and that the levels recovered during subsequent sleep. These researchers speculate that we may need to sleep in order to restore the brain's energy fuel. (*Science News*, July 20, 2002)

16. Some people who race greyhounds give the dogs large doses of vitamin C in the belief that the dogs will run faster. Investigators at the University of Florida tried three different diets in random order on each of five racing greyhounds. They were surprised to find that when the dogs ate high amounts of vitamin C they ran more slowly. (*Science News*, July 20, 2002)

17. Some people claim they can get relief from migraine headache pain by drinking a large glass of ice water. Researchers plan to enlist several people who suffer from migraines in a test. When a participant experiences a migraine headache, he or she will take a pill that may be a standard pain reliever or a placebo. Half of each group will also drink ice water. Participants will then report the level of pain relief they experience.

18. Weight is an issue for both humans and their pets. A dog food company wants to compare a new lower-calorie

food with their standard dog food to see if it's effective in helping inactive dogs maintain a healthy weight. They have found several dog owners willing to participate in the trial. The dogs have been classified as small, medium, or large breeds, and the company will supply some owners of each size of dog with one of the two foods. The owners have agreed not to feed their dogs anything else for a period of 6 months, after which the dogs' weights will be checked.

19. Athletes who had suffered hamstring injuries were randomly assigned to one of two exercise programs. Those who engaged in static stretching returned to sports activity in a mean of 37.4 days (SD = 27.6 days). Those assigned to a program of agility and trunk stabilization exercises returned to sports in a mean of 22.2 days (SD = 8.3 days). (*Journal of Orthopaedic & Sports Physical Therapy* 34 [March 2004]: 3)

20. Pew Research compared respondents to an ordinary 5-day telephone survey with respondents to a 4-month-long rigorous survey designed to generate the highest possible response rate. They were especially interested in identifying any variables for which those who responded to the ordinary survey were different from those who could be reached only by the rigorous survey.

21. **Tomatoes.** Describe a strategy to randomly split the 24 tomatoes into the three groups for the chapter's completely randomized single factor test of OptiGro fertilizer.

22. **Tomatoes II.** The chapter also described a completely randomized two-factor experiment testing OptiGro fertilizer in conjunction with two different routines for watering the plants. Describe a strategy to randomly assign the 24 tomato plants to the six treatments.

23. **Shoes.** A running shoe manufacturer wants to test the speed of its new sprinting shoe on the 100-meter dash times. The company sponsors 5 athletes who are running the 100-meter dash in the 2004 Summer Olympic games. To test the shoe, they have all 5 runners run the 100-meter dash with a competitor's shoe and then again with their new shoe. They use the difference in times as the response variable.

a) Suggest some improvements to the design.

b) Why might the shoe manufacturer not be able to generalize the results they find to all runners?

24. **Swimsuits.** A swimsuit manufacturer wants to test the speed of its newly designed suit. They design an experiment by having 6 randomly selected Olympic swimmers swim as fast as they can with their old swim suit first and then swim the same event again with the new, expensive swim suit. They'll use the difference in times as the response variable. Criticize the experiment and point out some of the problems with generalizing the results.

25. **Hamstrings.** Exercise 19 discussed an experiment to see if the time it took athletes with hamstring injuries to be able to return to sports was different depending on which of two exercise programs they engaged in.

a) Explain why it was important to assign the athletes to the two different treatments randomly.

b) There was no control group of athletes who did not participate in a special exercise program. Explain the advantage of including such a group in this experiment.

c) How might blinding have been used in this experiment?

a) One group returned to sports activity in a mean of 37.4 days (SD = 27.6 days), and the other in a mean of 22.2 days (SD = 8.3 days). Do you think this difference is statistically significant? Explain.

26. **Diet and blood pressure.** Exercise 6 reports on an experiment that showed that subjects fed the DASH diet were able to lower their blood pressure by an average of 6.7 points when compared to a group fed a "control diet." All meals were prepared by dieticians.

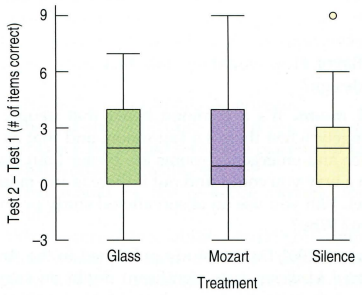
a) Why were the subjects randomly assigned to the diets instead of letting people pick what they wanted to eat?

b) Why were the meals prepared by dieticians?

c) Why did the researchers need the control group? If the DASH diet group's blood pressure was lower at the end of the experiment than at the beginning, wouldn't that prove the effectiveness of that diet?

d) What additional information would you want to know in order to decide whether an average reduction in blood pressure of 6.7 points was statistically significant?

27. **Mozart.** Will listening to a Mozart piano sonata make you smarter? In a 1995 study, Rauscher, Shaw, and Ky reported that when students were given a spatial reasoning section of a standard IQ test, those who listened to Mozart for 10 minutes improved their scores more than those who simply sat quietly.



a) These researchers said the differences were statistically significant. Explain what that means in this context.

b) Steele, Bass, and Crook tried to replicate the original study. The subjects were 125 college students who participated in the experiment for course credit. Subjects first took the test. Then they were assigned to one of three groups: listening to a Mozart piano sonata, listening to music by Philip Glass, and sitting for 10 minutes in silence. Three days after the treatments, they were retested. Draw a diagram displaying the design of this experiment.

c) The boxplots on p. 315 show the differences in score before and after treatment for the three groups. Did the Mozart group show improvement?

d) Do you think the results prove that listening to Mozart is beneficial? Explain.

28. More Mozart. An advertisement selling specially designed music CDs of Mozart's music specifically because they will "strengthen your mind, heal your body, and unlock your creative spirit" claims (we *swear!*) that "In Japan, a brewery reports that their best sake is made when Mozart is played near the yeast." Suppose, just for the sake (as it were) of discussion, you wished to design an experiment to test whether this is true. Assume you have the full cooperation of the sake brewery. Specify how you would design the experiment. Indicate factors and response and how they would be measured, controlled, or randomized.

29. Frumpies. The makers of Frumpies, "the breakfast of rug rats," want to improve their marketing, so they consult you:

a) They first want to know what fraction of children, ages 10 to 13, like their celery-flavored cereal. What kind of study should they perform?

b) They are thinking of introducing a new flavor, maple-marshmallow Frumpies, and want to know whether children will prefer the new flavor to the old one. Design a completely randomized experiment to investigate this question.

c) They suspect that children who regularly watch the Saturday morning cartoon show starring Frump, the flying teenage warrior rabbit who eats Frumpies in every episode, may respond differently to the new flavor. How would you take that into account in your design?

30. Full moon. It's a common belief that people behave strangely when there's a full moon, and that as a result police and emergency rooms are busier than usual. Design a way you could find out if there is any merit to this belief. Will you use an observational study or an experiment? Why?

31. Wine. A 2001 Danish study published in the *Archives of Internal Medicine* casts significant doubt on suggestions that adults who drink wine have higher levels of "good" cholesterol and fewer heart attacks. These researchers followed a group of individuals born at a Copenhagen hospital between 1959 and 1961 for 40 years. Their study found that in this group the adults who drank wine were richer and better educated than those who did not.

a) What kind of study was this?

b) It is generally true that people with high levels of education and high socioeconomic status are healthier than others. How does this call into question the supposed health benefits of wine?

c) Can studies such as these prove causation (that wine helps prevent heart attacks, that drinking wine makes one richer, that being rich helps prevent heart attacks, etc.)? Explain.

32. Swimming. Recently, a group of adults who swim regularly for exercise were evaluated for depression. It turned out that these swimmers were less likely to be depressed than the general population. The researchers said the difference was statistically significant.

a) What does "statistically significant" mean in this context?

b) Is this an experiment or an observational study? Explain.

c) News reports claimed this study proved that swimming can prevent depression. Explain why this conclusion is not justified by the study. Include an example of a possible confounding variable.

d) But perhaps it is true. We wonder if exercise can ward off depression, and whether anaerobic exercise (like weight training) is as effective as aerobic exercise (like swimming). We find 120 volunteers not currently engaged in a regular program of exercise. Design an appropriate experiment.

33. Dowsing. A water dowser claims to be able to sense the presence of water using a forked stick. Suppose we wish to set up an experiment to test his ability. We get 20 identical containers, fill some with water, and ask the dowser to tell which ones are full and which empty.

a) How will we randomize this procedure?

b) The dowser correctly identifies the contents of 12 out of 20 containers. Do you think this level of success is statistically significant? Explain.

c) How many correct identifications (out of 20) would the dowser have to make to convince you that the forked stick trick works? Explain.

34. Healing. A medical researcher suspects that giving postsurgical patients large doses of vitamin E will speed their recovery times by helping their incisions heal more quickly. Design an experiment to test this conjecture. Be sure to identify the factors, levels, treatments, response variable, and the role of randomization.

35. Reading. Some schools teach reading using phonics (the sounds made by letters) and others using whole language (word recognition). Suppose a school district

wants to know which method works better. Suggest a design for an appropriate experiment.

36. Gas mileage. Do cars get better gas mileage with premium instead of regular unleaded gasoline? While it might be possible to test some engines in a laboratory setting, we'd rather use real cars and real drivers in real day-to-day driving, so we get 20 volunteers. Design the experiment.

37. Weekend deaths. A study published in the *New England Journal of Medicine* (Aug. 2001) suggests that it's dangerous to enter a hospital on a weekend. During a 10-year period, researchers tracked over 4 million emergency admissions to hospitals in Ontario, Canada. Their findings revealed that patients admitted on weekends had a much higher risk of death than those who went to the emergency room on weekdays.

a) The researchers said the difference in death rates was "statistically significant." Explain in this context what that means.
b) What kind of study was this? Explain.
c) If you think you're quite ill on a Saturday, should you wait until Monday to seek medical help? Explain.
d) Suggest some possible explanations for this troubling finding.

38. Shingles. A research doctor has discovered a new ointment that she believes will be more effective than the current medication in the treatment of shingles (a painful skin rash). Eight patients have volunteered to participate in the initial trials of this ointment. You are the statistician hired as a consultant to help design a completely randomized experiment.

a) Describe how you will conduct this experiment.
b) Suppose the eight patients' last names start with the letters A to H. Using the random numbers listed below, show which patients you will assign to each treatment. Explain your randomization procedure clearly.

41098 18329 78458 31685 55259

c) Can you make this experiment double-blind? If so, explain how.
d) The initial experiment revealed that males and females may respond differently to the ointment. Further testing of the drug's effectiveness is now planned, and many patients have volunteered. What changes in your first design, if any, would you make for this second stage of testing?

39. Beetles. Hoping to learn how to control crop damage by a certain species of beetle, a researcher plans to test two different pesticides in small plots of corn. A few days after application of the chemicals, he'll check the number of beetle larvae found on each plant. The researcher wants to know if either pesticide works, and whether there is a significant difference in effectiveness between them. Design an appropriate experiment.

40. SAT Prep. Can special study courses actually help raise SAT scores? One organization says that the 30 students they tutored achieved an average gain of 60 points when they retook the test.

a) Explain why this does not necessarily prove that the special course caused the scores to go up.
b) Propose a design for an experiment that could test the effectiveness of the tutorial course.
c) Suppose you suspect that the tutorial course might be more helpful for students whose initial scores were particularly low. How would this affect your proposed design?

41. Safety switch. An industrial machine requires an emergency shutoff switch that must be designed so that it can be easily operated with either hand. Design an experiment to find out whether workers will be able to deactivate the machine as quickly with their left hands as with their right hands. Be sure to explain the role of randomization in your design.

42. Washing clothes. A consumer group wants to test the effectiveness of a new "organic" laundry detergent and make recommendations to customers about how to best use the product. They intentionally get grass stains on 30 white T-shirts in order to see how well the detergent will clean them. They want to try the detergent in cold water and in hot water on both the "regular" and "delicates" wash cycles. Design an appropriate experiment, indicating the number of factors, levels, and treatments. Explain the role of randomization in your experiment.

43. Skydiving, anyone? A humor piece published in the *British Medical Journal* notes ("Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomized control trials," Gordon, Smith, and Pell, *BMJ*, 2003:327) that we can't tell for sure whether parachutes are safe and effective because there has never been a properly randomized, double-blind, placebo-controlled study of parachute effectiveness in sky diving. (Yes, this is the sort of thing statisticians find funny . . . .) Suppose you were designing such a study:

a) What is the factor in this experiment?
b) What experimental units would you propose?[6]
c) Explain what would serve as a placebo for this study.
d) What would the treatments be?
e) What would be the response variable for such a study?
f) What sources of variability would you control?
g) How would you randomize this "experiment"?
h) How would you make the experiment double-blind?

[6] Don't include your Statistics instructor!