Chapter 11 – Experiments and Observational Studies

Section 11.1

1. Steroids.

This is an observational study because the sports writer is not randomly assigning players to take steroids or not take steroids; the writer is merely observing home run totals between two eras. It would be unwise to conclude steroids caused any increases in home runs because we need to consider other factors besides steroids—factors possibly leading to more home runs include better equipment, players training more in the offseason, smaller ballparks, better scouting techniques, etc.

2. E-commerce.

This is an observational study because the student is not randomly assigning companies to use or not use the Internet for business transactions. If profitability did increase in the 2000s, it could have been due to a number of factors, not specifically the Internet as a means for conducting business.

Section 11.2

3. Tips.

Each of the 40 deliveries is an experimental unit. He has randomized the experiment by flipping a coin to decide whether or not to phone.

4. Tomatoes.

Each tomato plant is an experimental unit. The tastiness and juiciness of the tomatoes will be the response variables.

5. Tips II.

The factor is calling, and the levels are whether or not he calls the customer. The response variable is the tip percentage for each delivery.

6. Tomatoes II.

The factor is the fertilizer, applied at three levels 0, half, and full dose. To measure tastiness and juiciness, we'll need trained tasters.

Section 11.3

7. Tips again.

By calling some customers but not others during the same run, the driver has controlled many variables, such as day of the week, season, and weather. The experiment was randomized because he flipped a coin to determine whether or not to phone and it was replicated because he did this for 40 deliveries.

8. Tomatoes again.

Tomato plants should be grown in the same field, near each other so differences in soil, sun, and rain can be controlled. The experiment is randomized because plants are assigned at random to treatment levels. It is replicated because 6 plants are assigned to each level.

Section 11.4

9. More tips.

Because customers don't know about the experiment, those that are called don't know that others are not, and vice versa. Thus, the customers are blind. That would make this a single-blind study. It can't be double-blind because the delivery driver must know whether or not he phones.

10. More tomatoes.

If the tomato taster is blind, then this is a single-blind study. To make it double-blind, everyone who cares for the tomato plants must be blind to their treatment. This might be done, for example, by treating all plants with solutions that look the same, but applying a "placebo" fertilizer to the plants assigned to receive none.

Section 11.5

11. Block that tip.

Yes. Driver is now a block. The experiment is randomized within each block. This is a good idea because some drivers might generally get higher tips than others, but the goal of the experiment is to study the effect of phone calls. Blocking on driver eliminates the variability in tips inherent to the driver.

12. Blocking tomatoes.

Yes. Garden center is the blocking factor. It is important to randomize the assignment of plants to treatments within each block so that any differences between garden centers won't affect the results.

Section 11.6

13. Confounded tips.

Answers may vary. The cost or size of a delivery may confound his results. Larger orders may generally tip a higher or lower percentage of the bill.

14. Tomatoes finis.

Answers may vary. Confounding factors could include variations in soil fertility, sunlight availability, or rainfall. Some plants might become infested with pests.

Chapter Exercises

15. Standardized test scores.

a) No, this is not an experiment. There are no imposed treatments. This is a retrospective observational study.

b) We cannot conclude that the differences in score are caused by differences in parental income. There may be lurking variables that are associated with both SAT score and parental income.

16. Heart attacks and height.

a) No, this is not an experiment. There are no imposed treatments. This is a retrospective observational study.

b) We cannot conclude that shorter men are at higher risk of dying from a heart attack. There may be lurking variables that are associated with both height and risk of heart attack.

17. MS and vitamin D.

a) This is a retrospective observational study.

b) This is an appropriate choice, since MS is a relatively rare disease.

c) The subjects were U.S. military personnel, some of whom had developed MS.

d) The variables were the vitamin D blood levels and whether or not the subject developed MS.

18. Super Bowl commercials.

a) This is a stratified sample. The question was about population values, namely the proportions of men and women who look forward to more commercials. No treatment was applied, so this is not an experiment.

b) Yes, the design was appropriate.

19. Menopause.

a) This was a randomized, comparative, placebo-controlled experiment.

b) Yes, such an experiment is the right way to determine whether black cohosh is an effective treatment for hot flashes.

c) The subjects were 351 women, aged 45 to 55 who reported at least two hot flashes a day.

d) The treatments were black cohosh, a multi-herb supplement, plus advice to consume more soy foods, estrogen, and a placebo. The response was the women's self-reported symptoms, presumably the frequency of hot flashes.

20. Honesty.

a) This is an experiment. The picture is the controlled factor. Randomization may have been used to decide which days each picture appeared.

b) The treatment was the picture behind the coffee station. The response variable was the average contribution.

c) The differences in money contributed were larger than could be reasonably attributed to usual day-to-day variation.

21. What's the design?

a) This is an experiment, since treatments were imposed.

b) The subjects studied were 30 patients with bipolar disorder.

c) The experiment has 1 factor (omega-3 fats from fish oil), at 2 levels (high dose of omega-3 fats from fish oil and no omega-3 fats from fish oil).

d) 1 factor, at 2 levels gives a total of 2 treatments.

e) The response variable is "improvement", but there is no indication of how the response variable was measured.

f) There is no information about the design of the experiment.

g) The experiment is blinded, since the use of a placebo keeps the patients from knowing whether or not they received the omega-3 fats from fish oils. It is not stated whether or not the evaluators of the "improvement" were blind to the treatment, which would make the experiment double-blind.

h) Although it needs to be replicated, the experiment can determine whether or not omega-3 fats from fish oils cause improvements in patients with bipolar disorder, at least over the short term. The experiment design would be stronger if it were double-blind.

22. What's the design?

a) This is an observational study. The researchers are simply studying traits that already exist in the subjects, not imposing new treatments.

b) This is a prospective study. The subjects were identified first, then traits were observed.

c) The subjects were disabled women aged 65 and older, with and without a vitamin B-12 deficiency. The selection process is not stated.

d) The parameter of interest is the percentage of women in each group who suffered severe depression.

e) There is no random assignment, so a cause-and-effect relationship between B-12 deficiency and depression cannot be established. The most that can be determined is an association, if this is supported by the data.

23. What's the design?

a) This is an observational study. The researchers are simply studying traits that already exist in the subjects, not imposing new treatments.

b) This is a prospective study. The subjects were identified first, then traits were observed.

c) The subjects are roughly 200 men and women with moderately high blood pressure and normal blood pressure. There is no information about the selection method.

d) The parameters of interest are difference in memory and reaction time scores between those with normal blood pressure and moderately high blood pressure.

23. (continued)

e) An observational study has no random assignment, so there is no way to know that high blood pressure caused subjects to do worse on memory and reaction time tests. A lurking variable, such as age or overall health, might have been the cause. The most we can say is that there was an association between blood pressure and scores on memory and reaction time tests in this group, and recommend a controlled experiment to attempt to determine whether or not there is a cause-and-effect relationship.

24. What's the design?

a) This is an experiment, since treatments were imposed on randomly assigned groups.

b) The subjects were 40 volunteers suffering from insomnia.

c) There are 2 factors in this experiment (dessert and exercise). The dessert factor has 2 levels (no dessert and normal dessert). The exercise factor has 2 levels (no exercise and an exercise program).

d) 2 factors, with 2 levels each, results in 4 treatments.

e) The response variable is improvement in ability to sleep.

f) This experiment is probably completely randomized.

g) This experiment does not use blinding.

h) This experiment indicates that insomniacs who exercise and refrain from desserts will experience improved ability to sleep.

25. What's the design?

a) This is an experiment, since treatments were imposed on randomly assigned groups.

b) 24 post-menopausal women were the subjects in this experiment.

c) There is 1 factor (type of drink), at 2 levels (alcoholic and non-alcoholic). (Supplemental estrogen is not a factor in the experiment, but rather a blocking variable. The subjects were not given estrogen supplements as part of the experiment.)

d) 1 factor, with 2 levels, is 2 treatments.

e) The response variable is an increase in estrogen level.

f) This experiment utilizes a blocked design. The subjects were blocked by whether or not they used supplemental estrogen. This design reduces variability in the response variable of estrogen level that may be associated with the use of supplemental estrogen.

g) This experiment does not use blinding.

h) This experiment indicates that drinking alcohol leads to increased estrogen level among those taking estrogen supplements.

26. What's the design?

a) This is an observational study.

b) The study is retrospective. Results were obtained from pre-existing medical records.

c) The subjects in this study were 981 women who lived near the site of dioxin release.

d) The parameter of interest is the incidence of breast cancer.

e) As there is no random assignment, there is no way to know that the dioxin levels caused the increase in breast cancer. There may have been lurking variables that were not identified.

27. What's the design?

a) This is an observational study.

b) The study is retrospective. Results were obtained from pre-existing church records.

c) The subjects of the study are women in Finland. The data were collected from church records dating 1640 to 1870, but the selection process is unknown.

d) The parameter of interest is difference in average lifespan between mothers of sons and daughters.

e) For this group, having sons was associated with a decrease in lifespan of an average of 34 weeks per son, while having daughters was associated with an unspecified increase in lifespan. As there is no random assignment, there is no way to know that having sons caused a decrease in lifespan.

28. What's the design?

a) This is an experiment, since treatments were imposed on randomly assigned groups.

b) The subjects were volunteers exposed to a cold virus.

c) There is 1 factor (herbal compound), at 2 levels (herbal compound and sugar solution).

d) 1 factor, at 2 levels, results in 2 treatments.

e) The response variable is the severity of cold symptoms.

f) There is no mention of any randomness is the design. Hopefully, subjects were randomly assigned to treatment groups.

g) The experiment uses blinding. The use of a sugar solution as a placebo kept the subjects from knowing whether or not they had received the herbal compound. If the doctors responsible for assessing the severity of the patients' colds were also unaware of the treatment group assignments, then the experiment incorporates double blinding.

h) There is no evidence to suggest that the herbal treatment is effective.

29. What's the design?

a) This is an observational study. (Although some might say that the sad movie was "imposed" on the subjects, this was merely a stimulus used to trigger a reaction, not a treatment designed to attempt to influence some response variable. Researchers merely wanted to observe the behavior of two different groups when each was presented with the stimulus.)

b) The study is prospective. Researchers identified subjects, and then observed them after the sad movie.

c) The subjects in this study were people with and without depression. The selection process is not stated.

d) The parameter of interest is the difference in crying response between depressed and nondepressed people exposed to sad situations.

e) There is no apparent difference in crying response to sad movies for the depressed and nondepressed groups.

30. What's the design?

a) This is an experiment.

b) The subjects were racing greyhounds.

c) There is 1 factor (level of vitamin C in diet). The 3 levels of diet were not specified.

d) One factor, at 3 levels, results in 3 treatments.

e) The response variable is speed.

f) The experiment uses a matched design. Each greyhound was given each of the 3 levels of diet, in random order. The matched design reduces variation due to the racing ability of each greyhound.

30. (continued)

g) There is no mention of blinding.

h) Greyhounds with diets high in vitamin C run more slowly than greyhounds with diets lower in vitamin C.

31. What's the design?

a) This is an experiment. Subjects were randomly assigned to treatments.

b) The subjects were people experiencing migraines.

c) There are 2 factors (pain reliever and water temperature). The pain reliever factor has 2 levels (pain reliever or placebo), and the water temperature factor has 2 levels (ice water and regular water).

d) 2 factors, at 2 levels each, results in 4 treatments.

e) The response variable is the level of pain relief.

f) The experiment is completely randomized.

g) The subjects are blinded to the pain reliever factor through the use of a placebo. The subjects are not blinded to the water factor. They will know whether they are drinking ice water or regular water.

h) The experiment may indicate whether pain reliever alone or in combination with ice water give pain relief, but patients are not blinded to ice water, so the placebo effect may also be the cause of any relief seen due to ice water.

32. What's the design?

a) This is an experiment. Hopefully, dogs are randomly assigned to different treatment groups.

b) The subjects are inactive dogs.

c) There is 1 factor (type of dog food), at 2 levels (low-calorie and standard). One possible difficulty with this experiment is that some owners might feed their dogs more food than others. We will assume that the dog food company has given the owners specific instructions about the quantity of food required, based on the size of each dog.

d) 1 factor, at 2 levels, results in 2 treatments.

e) The response variable is the weight of the dogs.

f) The experiment uses blocking by size of breed. Blocking by size reduces variation in weight that may be due to overall size of the dog.

g) Assuming that the dog owners do not know which type of dog food their dog is receiving, the experiment is blinded.

h) Assuming the dog owners followed the prescribed feeding levels, there could be a conclusion as to whether or not the dog food helped the dogs maintain a healthy weight.

33. What's the design?

a) This is an experiment. Athletes were randomly assigned to one of two exercise programs.

b) The subjects are athletes suffering hamstring injuries.

c) There is one factor (type of exercise), at 2 levels (static stretching, and agility and trunk stabilization).

d) 1 factor, at 2 levels, results in 2 treatments.

e) The response variable is the time before the athletes were able to return to sports.

f) The experiment is completely randomized.

g) The experiment employs no blinding. The subjects know what kind of exercise they do.

h) Assuming that the athletes actually followed the exercise program, this experiment can help determine which of the two exercise programs is more effective at rehabilitating hamstring injuries.

34. What's the design?

a) This is an observational study. The researchers are simply studying traits that already exist in the subjects, not imposing new treatments.

b) This is a prospective study. Researchers identified two groups, and studied their traits.

c) The subjects are members of the general public, chosen in two random samples.

d) The purpose of this study was to identify variables on which there was a difference, so no response variable(s) could be identified at the start of the study.

e) This study will allow researchers to identify differences between people who can be reached by ordinary 5-day polling methods and those who cannot be reached.

35. Omega-3.

The experimenters need to compare omega-3 results to something. Perhaps bipolarity is seasonal and would have improved during the experiment anyway.

36. Insomnia.

The experimenters need a basis for comparison. Perhaps insomnia is related to the amount of daylight, and that changed during the time when the experiment was conducted.

37. Omega-3 revisited.

a) Subjects' responses might be related to other factors, like diet, exercise, or genetics. Randomization should equalize the two groups with respect to unknown factors.

b) More subjects would minimize the impact of individual variability in the responses, but the experiment would become more costly and time-consuming.

38. Insomnia again.

a) Subjects responses might be related to many other factors, such as diet, medications, or genetics. Randomization should equalize the two groups with respect to unknown factors.

b) More subjects would minimize the impact of individual variability in the responses, but the experiment would become more costly and time-consuming.

39. Omega-3 finis.

The researchers believe that people who engage in regular exercise might respond differently to the omega-3. This additional variability could obscure the effectiveness of the treatment.

40. Insomnia, at last.

The researchers believe that people who are overweight might respond differently to exercise. This additional variability could obscure the effectiveness of the treatment.

41. Injuries.

Answers may vary. Generate random values of 0 or 1 from an Internet site. Assign participants with a 1 to one program and those with a 0 to the other. Participants must follow the program to which they are assigned; they can't choose to switch programs.

42. Tomatoes II.

Answers may vary. Number the tomatoes plants 1 to 24. Use a random number generator to randomly select 24 numbers from 1 to 24 without replication. Assign the tomato plants matching the first 8 numbers to the first group, the second 8 numbers to the second group, and the third group of 8 numbers to the third group.

43. Shoes.

a) First, the manufacturers are using athletes who have a vested interest in the success of the shoe by virtue of their sponsorship. They should try to find some volunteers that aren't employed by the company! Second, they should randomize the order of the runs, not run all the races with the new shoes second. They should blind the athletes by disguising the shoes, if possible, so they don't know which is which. The experiment could be double blinded, as well, by making sure that the timers don't know which shoes are being tested at any given time. Finally, they should replicate several times since times will vary under both shoe conditions.

b) First of all, the problems identified in part (a) would have to be remedied before *any* conclusions can be reached. Even if this is the case, the results cannot be generalized to all runners. This experiment compares effects of the shoes on speed for Olympic class runners, not runners in general.

44. Swimsuits.

The "control" in this experiment is not the same for all swimmers. We don't know what "their old swim suit" means. They should compare their new swim suit to the same suit design. The order in which the swims are performed should be randomized. There may be a systematic difference from one swim to the next. For instance, swimmers may be tired after the first swim (or more warmed up). Finally, there is no way to blind this test. The swimmer will know which kind of suit they have on, and this may bias their performance.

45. Hamstrings.

a) Allowing the athletes to choose their own treatments could confound the results. Other issues such as severity of injury, diet, age, etc., could also affect time to heal, and randomization should equalize the two treatment groups with respect to any such variables.

b) A control group could have revealed whether either exercise program was better (or worse) than just letting the injury heal without exercise.

c) Although the athletes cannot be blinded, the doctors who approve their return to sports should not know which treatment the subject had engaged in.

d) It's difficult to say with any certainty, since we aren't sure if the distributions of return times are unimodal and roughly symmetric, and contain no outliers. Otherwise, the use of mean and standard deviation as measures of center and spread is questionable. Assuming mean and standard deviation are appropriate measures, the subjects who exercised with agility and trunk stabilization had a mean return time of 22.2 days compared to the static stretching group, with a mean return time of 37.4 days. The agility and trunk stabilization group also had a much more consistent distribution of return times, with a standard deviation of 8.3 days, compared to the standard deviation of 27.6 days for the static stretching group. This appears to be a statistically significant difference.

46. Diet and blood pressure.

a) Self-selection could result in groups that are very different at the start of the experiment, making it impossible to attribute differences in the results to the diet alone.

b) The meals were prepared by dieticians to ensure that the diets were followed and that all subjects received comparable treatments.

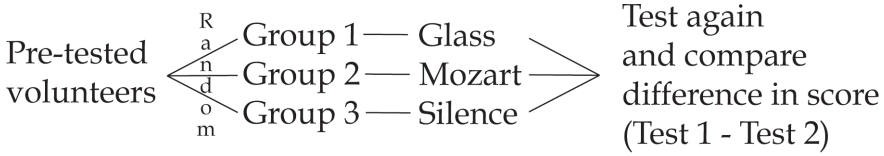
c) The researchers can compare the change in blood pressure observed in the DASH group to the control group. They need to rule out the possibility that external variables (like the season, news events, etc.) affected everyone's blood pressure.

d) We would like to know the standard deviation of the changes, as well. If the standard deviation is very small, then 6.7 points would seem like a significant change. If not, 6.7 points could be due to naturally occurring variability.

47. Mozart.

a) The differences in spatial reasoning scores between the students listening to Mozart and the students sitting quietly were more than would have been expected from ordinary sampling variation.

b)



c) The Mozart group seems to have the smallest median difference in spatial reasoning test score and thus the *least* improvement, but there does not appear to be a significant difference.

d) No, the results do not prove that listening to Mozart is beneficial. If anything, there was generally less improvement. The difference does not seem significant compared with the usual variation one would expect between the three groups. Even if type of music has no effect on test score, we would expect some variation between the groups.

48. Contrast Baths.

a) There is no evidence that the changes in swelling were different among the three treatments. Any differences see between the treatments could be attributed to chance alone. Contrast baths, with or without exercise, did not appear to reduce swelling any more than exercise alone.

b) Patients were assigned to treatments at random. The exercise treatment was the control group.

c) If exercise is the standard treatment, then using it as a control seems appropriate (and would correspond to the Helsinki guidelines.) The use of a placebo could be viewed as unethical, since effective treatment is withheld.

49. Wine.

a) This is a prospective observational study. The researchers followed a group of children born at a Copenhagen hospital between 1959 and 1961.

b) The results of the Danish study report a link between high socioeconomic status, education, and wine drinking. Since people with high levels of education and higher socioeconomic status are also more likely to be healthy, the relation between health and wine consumption might be explained by the confounding variables of socioeconomic status and education.

c) Studies such as these prove none of these. While the variables have a relation, there is no indication of a cause-and-effect relationship. The only way to determine causation is through a controlled, randomized, and replicated experiment.

50. Swimming.

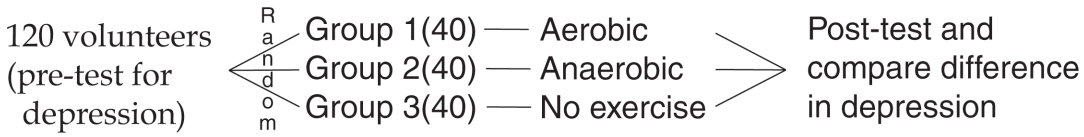
a) The swimmers showed a rate of depression that was lower than would be expected from a sample of that size drawn at random from the population. This rate was so low that it was unlikely to be due to natural sampling variation.

b) This is a retrospective observational study. There was no imposition of treatments. The researchers simply identified a group and evaluated them for depression.

50. (continued)

c) The news reports made a claim of a cause-and-effect relationship. Causation can only be determined through the use of a controlled, randomized, and replicated experiment, not an observational study. The difference in depression rates might be explained by lurking variables. For example, swimmers might tend to have higher incomes than the general population. Swimmers need to have access to a pool, either by having their own, or paying for a membership to a health club. Perhaps it is their financial situation that makes them happier, not the swimming. Another possible explanation is a reversal of the direction of the relationship implied by the news reports. Perhaps depression makes people not want to swim.

d) Answers may vary. Give the subjects a test to measure depression. Then randomly assign the 120 subjects to one of three groups: the control group (no exercise program), the anaerobic exercise group, and the aerobic exercise group. Monitor subjects' exercise (have them report to a particular gym or pool). At the end of 12 weeks, administer the depression test again. Compare the post-exercise and pre-exercise depression scores.



51. Dowsing.

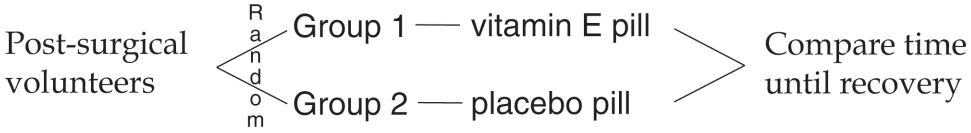
a) Arrange the 20 containers in 20 separate locations. Number the containers 01 – 20, and use a random number generator to identify the 10 containers that should be filled with water.

b) We would expect the dowser to be correct about 50% of the time, just by guessing. A record of 60% (12 out of 20) does not appear to be significantly different than the 10 out of 20 expected.

c) Answers may vary. A high level of success would need to be observed. 90% to 100% success (18 to 20 correct identifications) would be convincing.

52. Healing.

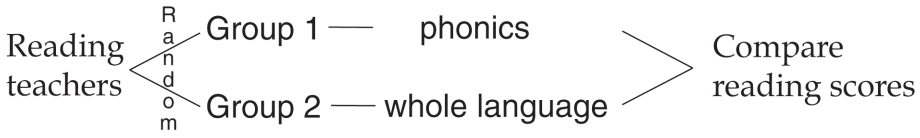
Answers may vary. This double-blind experiment has 1 factor (vitamin E), at 2 levels (vitamin E and no vitamin E), resulting in 2 treatments. The response variable measured is the time it takes the patient to recover from the surgery. Randomly select half of the patients who agree to the study to get large doses of vitamin E after surgery. Give the other patients in the study a similar looking placebo pill. Monitor their progress, recording the time until they have reached an easily agreed upon level of healing. Have the evaluating doctor blinded to whether the patient received the vitamin E or the placebo. Compare the number of days until recovery of the two groups.



53. Reading.

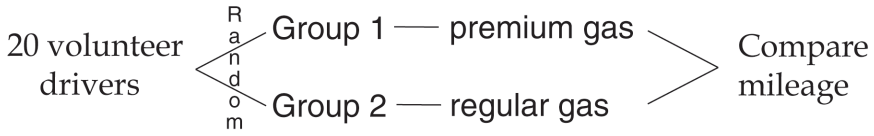
Answers may vary. This experiment has 1 factor (reading program), at 2 levels (phonics and whole language), resulting in 2 treatments. The response variable is reading score on an appropriate reading test after a year in the program. After randomly assigning students to teachers, randomly assign half the reading teachers in the district to use each method. There may be variation in reading score based on school within the district, as well as by grade. Blocking by both school and grade will reduce this variation.

53. (continued)



54. Gas mileage.

Answers may vary. This experiment has 1 factor (type of gasoline), at 2 levels (premium and regular), resulting in two treatments. The response variable is gas mileage. An experiment diagram for a simple design appears above. Randomly assign each of the 20 volunteers to the premium or regular groups. Ask them to keep driving logs (the number of miles driven and the gallons of gasoline) for one month. Compare the differences in the fuel economy for the two groups.



Stronger designs would control for several variables that may have an effect on fuel economy, such as size of engine, type of driving (for example, city or highway), and driving style (for example, if the is driver aggressive, or if the driver exceeds the speed limit). With only 20 volunteers, it would be difficult to block for all of these variables, but a matched design would work well. Have each volunteer use regular gasoline for a specified time period and record the mileage, and also use premium for a specified time period. Randomize which type of gasoline is used first.

55. Weekend deaths.

a) The difference between death rate on the weekend and death rate during the week is greater than would be expected due to natural sampling variation.

b) This was a prospective observational study. The researchers identified hospitals in Ontario, Canada, and tracked admissions to the emergency rooms. This certainly cannot be an experiment. People can't be assigned to become injured on a specific day of the week!

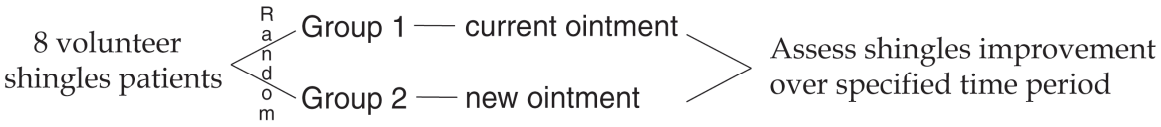
c) Waiting until Monday, if you were ill on Saturday, would be foolish. There are likely to be confounding variables that account for the higher death rate on the weekends. For example, people might be more likely to engage in risky behavior on the weekend.

d) Alcohol use might have something to do with the higher death rate on the weekends. Perhaps more people drink alcohol on weekends, which may lead to more traffic accidents, and higher rates of violence during these days of the week.

56. Shingles.

a) Answers may vary. This experiment has 1 factor (ointment), at 2 levels (current and new), resulting in 2 treatments. The response variables are the improvements in severity of the case of shingles and the improvements in the pain levels of the patients. Randomly assign the eight patients to either the current ointment or to the new ointment. Before beginning treatment, have doctors assess the severity of the case of shingles for each patient, and ask patients to rate their pain levels. Administer the ointments for a prescribed time, and then have doctors reassess the severity of the case of shingles, and ask patients to once again rate their pain levels. If the neither the patients nor the doctors are told which treatment is given to each patient, the experiment will be double-blind. Compare the improvement levels for each group.

56. (continued)



b) Answers may vary. Let numbers 1 through 8 correspond to letter A through H, respectively. Ignore digits 0 and 9, and ignore repeats. The first row contains the random digits, the second row shows the corresponding patient (X indicates an ignored or repeated digit), and the third row shows the resulting group assignment, alternating between Group 1 and Group 2.

Group 1 (current ointment): D, A, H, C
Group 2 (new ointment): B, G, E, F

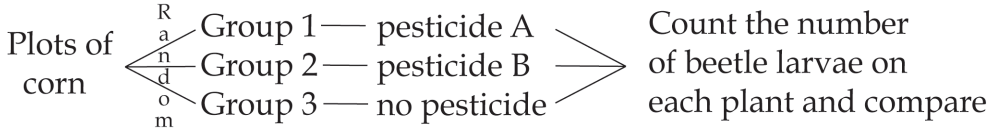
```
41098 18329 78458 31685 55259
DAXXH XXCBX GXXEX XXF
11  1   12  2  2    2
```

c) Assuming that the ointments looked alike, it would be possible to blind the experiment for the patient and the evaluating doctor. If both the subject and the evaluator are blinded, the experiment is double-blind.

d) Before randomly assigning patients to treatments, identify them as male or female. Having blocks for males and females will eliminate variation in improvement due to gender.

57. Beetles.

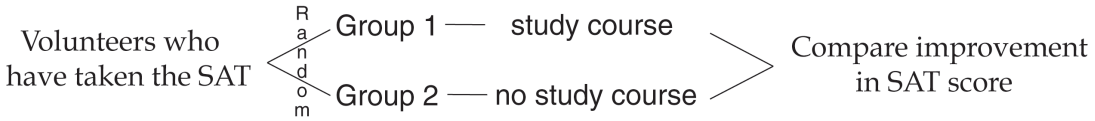
Answers may vary. This experiment has 1 factor (pesticide), at 3 levels (pesticide A, pesticide B, no pesticide), resulting in 3 treatments. The response variable is the number of beetle larvae found on each plant. Randomly select a third of the plots to be sprayed with pesticide A, a third with pesticide B, and a third to be sprayed with no pesticide (since the researcher also wants to know whether the pesticides even work at all). To control the experiment, the plots of land should be as similar as possible, with regard to amount of sunlight, water, proximity to other plants, etc. If not, plots with similar characteristics should be blocked together. If possible, use some inert substance as a placebo pesticide on the control group, and do not tell the counters of the beetle larvae which plants have been treated with pesticides. After a given period of time, count the number of beetle larvae on each plant and compare the results.



58. SAT prep.

a) The students were not randomly assigned to the special study course. Those who signed up for the course may be a special group whose scores would have improved anyway, due to motivation, intelligence, parental involvement, or other reasons.

b) Answers may vary. This experiment has 1 factor (study course), at 2 levels (study course, no study course), resulting in 2 treatments. The response variable is improvement in SAT score on the second test. Find a group of volunteers who are willing to participate. Have all volunteers take the SAT exam. Randomly assign the subjects to the study course or no study course groups. After giving the study course to the appropriate group, have both groups take the SAT again. Check to see if the group given the study course had a significant improvement in scores when compared with the group receiving no study course.



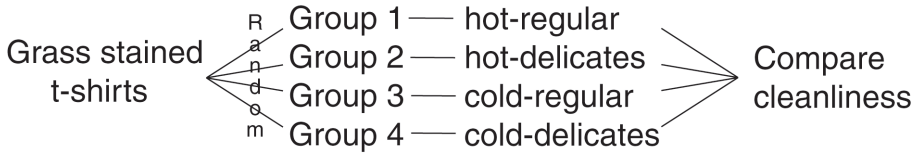
c) After the volunteers have taken the first SAT exam, block the volunteers by Low, Average, and High SAT exam score performance. For each block, replicate the experiment design described in part b.

59. Safety switch.

Answers may vary. This experiment has 1 factor (hand), at 2 levels (right, left), resulting in 2 treatments. The response variable is the difference in deactivation time between left and right hand. Find a group of volunteers. Using a matched design, we will require each volunteer to deactivate the machine with his or her left hand, as well as with his or her right hand. Randomly assign the left or right hand to be used first. Hopefully, this will equalize any variability in time that may result from experience gained after deactivating the machine the first time. Complete the first attempt for the whole group. Now repeat the experiment with the alternate hand. Check the differences in time for the left and right hands. Since the response variable is difference in times for each hand, workers should be blocked into groups based on their dominant hand. Another way to account for this difference would be to use the absolute value of the difference as the response variable. We are interested in whether or not the difference is significantly different from the zero difference we would expect if the machine were just as easy to operate with either hand.

60. Washing clothes.

Answers may vary. This experiment has two factors (water temperature, wash cycle). The factor water temperature has 2 levels (cold, hot), and the factor wash cycle has 2 levels (regular, delicates). 2 factors, at 2 levels each, results in 4 treatments (hot-regular, hot-delicates, cold-regular, cold-delicates). The response variable is the level of cleaning of the grass stains. It would be nice to have 32 shirts with which to experiment, so that we could randomly assign 8 shirts to each treatment group, but equal numbers of shirts in each group are not necessary. After washing, have "laundry experts" rate the cleanliness of each shirt. Compare the level of cleanliness in each group.



61. Skydiving, anyone?

a) There is 1 factor, jumping, with 2 levels, with and without a working parachute.

b) You would need some (dim-witted) volunteers skydivers as the subjects.

c) A parachute that looked real, but didn't open, would serve as the placebo.

d) 1 factor at 2 levels is 2 treatments, a good parachute and a placebo parachute.

e) The response variable is whether the skydiver survives the jump (or the extent of injuries).

f) All skydivers should jump from the same altitude, in similar weather conditions, and land on similar surfaces.

g) Make sure that you randomly assign the skydivers to the parachutes.

h) The skydivers (and the distributers of the parachutes) shouldn't know who got a working chute. Additionally, the people evaluating the subjects after the jumps should not be told who had a real chute, either.