29. 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.

30. 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 form 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.

31. 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.

32. 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.

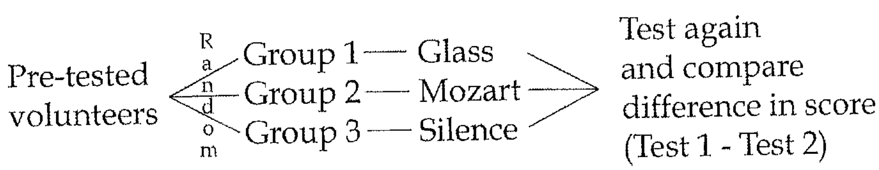
c) The researchers can compare the change in blood pressure observed in the DASH group to the control group. They need to rule out the possibility that external variables (like the season, news events, etc.) affected everyone's blood pressure.

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.

33. 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.

34. Full moon.

Answers may vary. Use a retrospective observational study. For example, collect records from a random selection of police and emergency room logs for the past 3 years. Find the number of cases for the days when there is a full moon, when there is a waxing moon, a waning moon, and when the moon is nearly dark. Compare the numbers for each group.

35. 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.

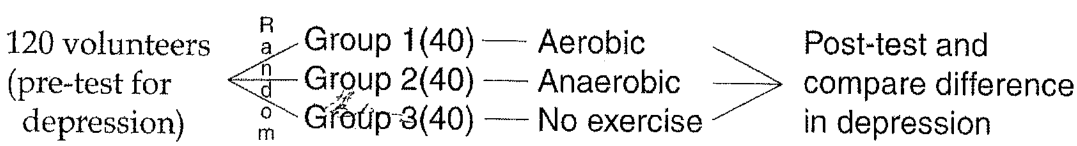
36. Swimming.

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.

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.



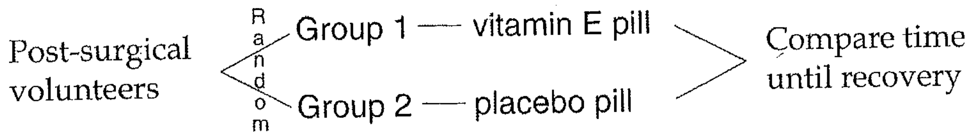
37. Dowsing.

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.

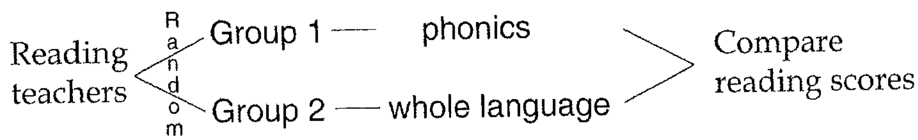
c) Answers may vary. A high level of success would need to be observed. 90% to 100% success (18 to 20 correct identifications) would be convincing.

38. Healing.



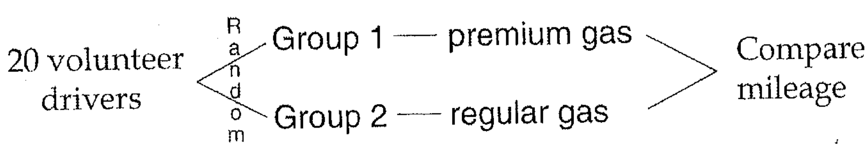
Answers will vary. This double-blind experiment has 1 factor (vitamin E), at 2 levels (vitamin E and no vitamin E), resulting in 2 treatments. The response variable measured is the time it takes the patient to recover from the surgery. Randomly select half of the patients who agree to the study to get 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.

39. 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.

40. 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.

41. 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.

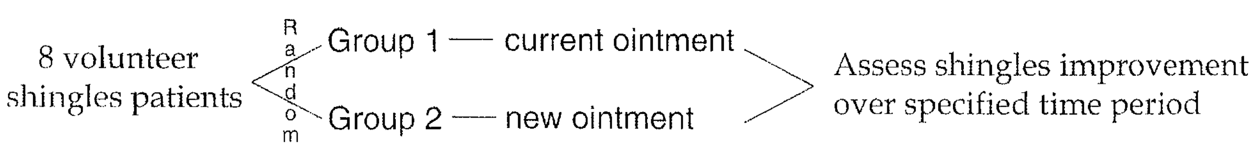
b) This was a prospective observational study. The researchers identified hospitals in Ontario, Canada, and tracked admissions to the emergency rooms. This certainly cannot be an experiment. People can't be assigned to become injured on a specific day of the week!

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.

42. 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.



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.

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

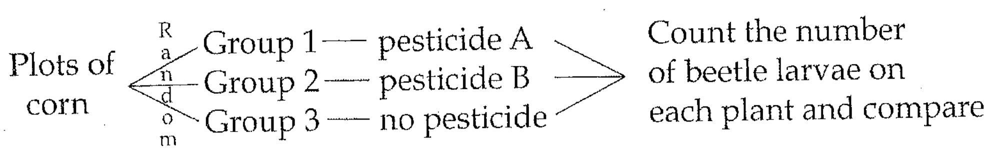
Group 1 (current ointment): D, A, H, C
Group 2 (new ointment): B, G, E, F

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.

43. Beetles.

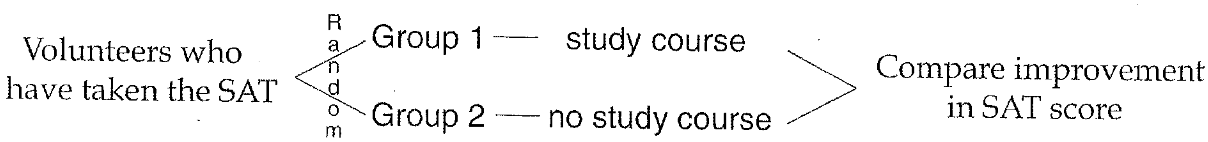
Answers may vary. This experiment has 1 factor (pesticide), at 3 levels (pesticide A, pesticide B, no pesticide), resulting in 3 treatments. The response variable is the number of beetle larvae found on each plant. Randomly select a third of the plots to be sprayed with pesticide A, a third with pesticide 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.



44. 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.

45. 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.

46. 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.