5

Producing Data

5.1 The population is employed adult women, the sample is the 48 club members who returned the survey.

5.2 (a) An individual is a person; the population is all adult U.S. residents. (b) An individual is a household; the population is all U.S. households. (c) An individual is a voltage regulator; the population is all the regulators in the last shipment.

5.3 This is an experiment: A treatment is imposed. The explanatory variable is the teaching method (computer assisted or standard), and the response variable is the increase in reading ability based on the pre- and posttests.

5.4 We can never know how much of the change in attitudes was due to the explanatory variable (reading propaganda) and how much to the historical events of that time. The data give no information about the effect of reading propaganda.

5.5 Observational. The researcher did not attempt to change the amount that people drank. The explanatory variable is alcohol consumption. The response variable is survival after 4 years.

5.6 (a) The data was collected after the anesthesia was administered. (b) Type of surgery, patient allergy to certain anesthetics, how healthy the patient was before the surgery.

5.7 Only persons with a strong opinion on the subject—strong enough that they are willing to spend the time, and 50 cents—will respond to this advertisement.

5.8 Letters to legislators are an example of a voluntary response sample—the proportion of letters opposed to the insurance should not be assumed to be a fair representation of the attitudes of the congresswoman's constituents.

5.9 Labeling from 001 to 440, we select 400, 077, 172, 417, 350, 131, 211, 273, 208, and 074.

5.10 Starting with 01 and numbering down the columns, one chooses 04-Bonds, 10-Fleming, 17-Liao, 19-Naber, 12-Goel, and 13-Gomez.

5.11 Assign 01 to 30 to the students (in alphabetical order). The exact selection will depend on the starting line chosen in Table B; starting on line 123 gives 08-Ghosh, 15-Jones, 07-Fisher, and 27-Shaw. Assigning 0–9 to the faculty members gives (from line 109) 3-Gupta and 6-Moore. (We could also number faculty from 01 to 10, but this requires looking up 2-digit numbers.)

5.12 Label the 500 midsize accounts from 001 to 500, and the 4400 small accounts from 0001 to 4400. We first encounter numbers 417, 494, 322, 247, and 097 for the midsize group, then 3698, 1452, 2605, 2480, and 3716 for the small group.

5.13 (a) Households without telephones, or with unlisted numbers. Such households would likely be made up of poor individuals (who cannot afford a phone), those who choose not to have phones, and those who do not wish to have their phone number published.

(b) Those with unlisted numbers would be included in the sampling frame when a random-digit dialer is used.

5.14 The higher no-answer was probably the second period—more families are likely to be gone for vacations, etc. Nonresponse of this type might underrepresent those who are more affluent (and are able to travel).

5.15 The first wording would pull respondents toward a tax cut because the second wording mentions several popular alternative uses for tax money.

5.16 *Variable:* Approval of president's job performance. *Population:* Adult citizens of the U.S., or perhaps just registered voters. *Sample:* The 1210 adults interviewed. *Possible sources of bias:* Only adults with phones were contacted. Alaska and Hawaii were omitted.

5.17 (a) 13,147 + 15,182 + 1448 = 29,777. (b) There's nothing to prevent a person from answering several times. Also, the respondents were only those who went to that Web site and took the time to respond. We cannot define "nonresponse" in this situation. (c) The results are slanted toward the opinions of men, who might be less likely to feel that female athletes should earn as much as men.

5.18 (a) The wording is clear. The question is somewhat slanted in favor of warning labels. (b) The question is clear, but it is clearly slanted in favor of national health insurance by asserting it would reduce administrative costs. (c) The question could be clearer by using simpler language. It is slanted in favor of incentives by starting out discussing environmental degradation.

5.19 (a) The adults in the country. (b) All the wood sent by the supplier. (c) All households in the U.S.

5.20 The call-in poll is faulty in part because it is a voluntary sample. Furthermore, even a small charge like 50 cents can discourage some people from calling in—especially poor people. Reagan's Republican policies appealed to upper-class voters, who would be less concerned about a 50 cent charge than lower-class voters who might favor Carter.

5.21 Number the bottles across the rows from 01 to 25, then select 12-B0986, 04-A1101, and 11-A2220. (If numbering is done down columns instead, the sample will be A1117, B1102, and A1098.)

5.22 In order to increase the accuracy of its poll results. Larger samples give more accurate results than smaller samples.

5.23 The blocks are already marked; select three-digit numbers and ignore those that do not appear on the map. This gives 214, 313, 409, 306, and 511.

5.24 (a) False—if it were true, then after looking at 39 digits, we would know whether or not the 40th digit was a 0, contrary to property 2. (b) True—there are 100 pairs of digits 00 through 99, and all are equally likely. (c) False—0000 is just as likely as any other string of four digits.

5.25 It is *not* an SRS, because some samples of size 250 have no chance of being selected (e.g., a sample containing 250 women).

5.26 (a) This question will likely elicit more responses against gun control (that is, more people will choose 2). The two options presented are too extreme; no middle position on gun control is allowed.

(b) The phrasing of this question will tend to make people respond in favor of a nuclear freeze. Only one side of the issue is presented.

(c) The wording is too technical for many people to understand—and for those that *do* understand it, it is slanted because it suggests reasons why one should support recycling. It could be rewritten to something like: "Do you support economic incentives to promote recycling?"

5.27 A smaller sample gives less information about the population. "Men" constituted only about one-third of our sample, so we know less about that group than we know about all adults.

5.28 The chance of being interviewed is 3/30 for students over age 21 and 2/20 for students under age 21. This is 1/10 in both cases. It is not an SRS because not all combinations of students have an equal chance of being interviewed. For instance, groups of 5 students all over age 21 have no chance of being interviewed.

5.29 Answers will vary, of course. One possible approach: Obtain a list of schools, stratified by size or location (rural, suburban, urban). Choose SRSs (not necessarily all the same size) of schools from each strata. Then choose SRSs (again, not necessarily the same size) of students from the selected schools.

5.30 (a) Split the 200 addresses into 5 groups of 40 each. Looking for 2-digit numbers from 01 to 40, we find 35, and so take 35, 75, 115, 155, and 195. (b) Every address has a 1-in-40 chance of being selected, *but* not every subset has an equal chance of being picked—for example, 01, 02, 03, 04, and 05 cannot be selected by this method.

5.31 Units are the individual trees. Factor is the amount of light. Treatments are full light and reduced light. Response variable is the weight of the trees.

5.32 The liners are the experimental units. The heat applied to the liners is the factor; the levels are 250°F, 275°F, 300°F, and 325°F. The force required to open the package is the response variable.

5.33 The units are the individuals who were called. One factor is what information is offered. Treatments are (1) giving name, (2) identifying university, (3) both of these. Second factor is offering to send a copy of the results. The treatments are either offering or not offering. The response is whether the interview was completed.

5.34 Subjects: 300 sickle cell patients. Factor: drug given. Treatments: hydroxyurea and placebo. Response variable: number of pain episodes.

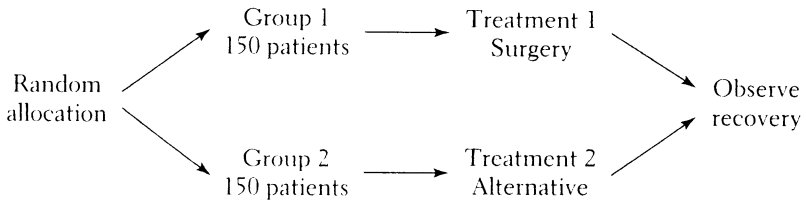
5.35 (a) This is an experiment, since the teacher imposes treatments (instruction methods). (b) The explanatory variable is the method used (computer software or standard curriculum), and the response is the change in reading ability.

5.36 (a) The experimental units are the batches of the product; the yield of each batch is the response variable. (b) There are two factors: temperature (with 2 levels) and stirring rates (with 3 levels), for a total of 6 treatments. (c) Since two experimental units will be used for each treatment, we need 12.

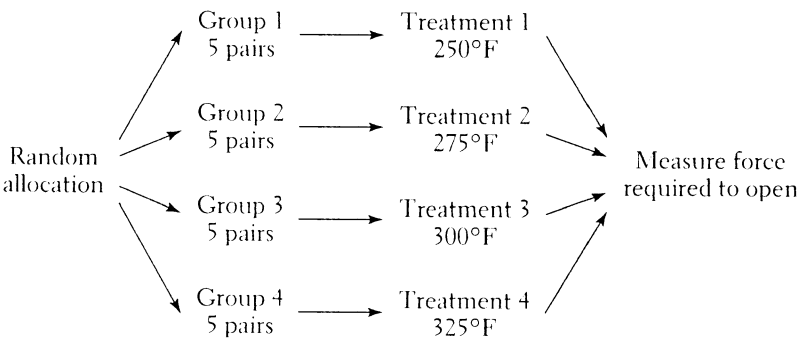
		Factor B: Stirring rates		
		60 rpm	90 rpm	120 rpm
Factor A: Temperature	50°C	1	2	3
	60°C	4	5	6

5.37 (a) In a serious case, when the patient has little chance of surviving, a doctor might choose not to recommend surgery; it might be seen as an unnecessary measure, bringing expense and a hospital stay with little benefit to the patient.

(b)



5.38 (a)



(b) Number the liners from 01 to 20, then take Group 1 to be 16, 04, 19, 07, and 10; Group 2 is 13, 15, 05, 09, and 08; Group 3 is 18, 03, 01, 06, and 11. The others are in Group 4.

5.39 (a) Randomly select 20 women for Group 1, which will see the "childcare" version of Company B's brochure, and assign the other 20 women to Group 2 (the "no childcare" group). Allow all women to examine the appropriate brochures, and observe which company they choose. Compare the number from Group 1 who choose Company B with the corresponding number from Group 2.

(b) Numbering from 01 to 40, Group 1 is 05-Cansico, 32-Roberts, 19-Hwang, 04-Brown, 25-Lippman, 29-Ng, 20-Iselin, 16-Gupta, 37-Turing, 39-Williams, 31-Rivera, 18-Howard, 07-Cortez, 13-Garcia, 33-Rosen, 02-Adamson, 36-Travers, 23-Kim, 27-McNeill, and 35-Thompson.

5.40 If this year is considerably different in some way from last year, we cannot compare electricity consumption over the two years. For example, if this summer is warmer, the customers may run their air conditioners more often. The possible differences between the two years would confound the effects of the treatments.

5.41 The second design is an experiment—a treatment is imposed on the subjects. The first is a study; it may be confounded by the types of men in each group. In spite of the researcher's attempt to match "similar" men from each group, those in the first group (who exercise) could be somehow different from men in the non-exercising group.

5.42 "Significantly higher" means that these returns are higher than would be expected due to random chance. "Not significantly different from zero" means that, although the sample may not be exactly zero, the difference from zero is within the range that can occur simply due to chance.

5.43 Because the experimenter knew which subjects had learned the meditation techniques, he (or she) may have had some expectations about the outcome of the experiment: if the experimenter believed that meditation was beneficial, he may subconsciously rate that group as being less anxious.

5.44 (a) If only the new drug is administered, and the subjects are then interviewed, their responses will not be useful, because there will be nothing to compare them to: How much "pain relief" does one expect to experience?

(b) Randomly assign 20 patients to each of three groups: Group 1, the placebo group; Group 2, the aspirin group; and Group 3, which will receive the new medication. After treating the patients, ask them how much pain relief they feel, and then compare the average pain relief experienced by each group.

(c) The subjects should certainly not know what drug they are getting—a patient told that she is receiving a placebo, for example, will probably not expect any pain relief.

(d) Yes—presumably, the researchers would like to conclude that the new medication is better than aspirin. If it is not double-blind, the interviewers may subtly influence the subjects into giving responses that support that conclusion.

5.45 (a) Ordered by increasing weight, the five blocks are (1) Williams-22, Deng-24, Hernandez-25, and Moses-25; (2) Santiago-27, Kendall-28, Mann-28, and Smith-29; (3) Brunk-30, Obrach-30, Rodriguez-30, and Loren-32; (4) Jackson-33, Stall-33, Brown-34, and Cruz-34; (5) Birnbaum-35, Tran-35, Nevesky-39, and Wilansky-42.

(b) The exact randomization will vary with the starting line in Table B. Different methods are possible; perhaps the simplest is to number from 1 to 4 within each block, then assign the members of block 1 to a weight-loss treatment, then assign block 2, etc. For example, starting on line 133, we assign 4-Moses to treatment A, 1-Williams to B, and 3-Hernandez to C (so that 2-Deng gets treatment D), then carry on for block 2, etc. (either continuing on the same line, or starting over somewhere else).

5.46 (a) Assume that the 6 circular areas are given in advance. Number them in any order. Use Table B to select 3 for the treatment. We used line 104. The first 4 digits are: 5 2 7 1. We cannot use the 7 because it is more than 6. Therefore, we would treat areas 5, 2 and 1.

(b) If the pairs are not given in advance, divide the 6 areas into 3 pairs so that the elements of each pair are close to each other and therefore of similar fertility. For each pair, we randomly pick one of the two to receive the treatment. Label the two areas in each pair A and B. If the random number from Table B is even, then apply the treatment to area A. Otherwise, apply the treatment to Area B. Alternatively, we could go along the table looking for either a 0 or a 1, ignoring the other digits. If we find a 0 before a 1, then treat area A. Otherwise, treat B.

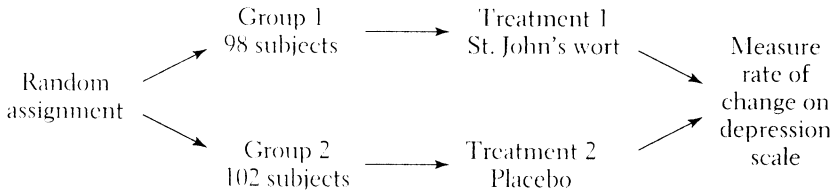
5.47 (a) Assign 10 subjects to Group 1 (the 70° group) and the other 10 to Group 2 (which will perform the task in the 90° condition). Record the number of correct insertions in each group.

(b) All subjects will perform the task twice—once in each temperature condition. Randomly choose which temperature each subject works in first, either by flipping a coin, or by placing 10 subjects in Group 1 (70°, then 90°) and the other 10 in Group 2.

5.48 The randomization will vary with the starting line in Table B. *Completely randomized design:* Randomly assign 10 students to Group 1 (which has the trend-highlighting software) and the other 10 to Group 2 (which does not). Compare the performance of Group 1 with that of Group 2. *Matched pairs design:* Each student does the activity twice, once with the software and once without. Randomly decide (for each student) whether they have the software the first or second time. Compare performance with the software and without it. (This randomization can be done by flipping a coin 20 times, or by picking 20 digits from Table B, and using the software first if the digit is even, etc.) *Alternate matched pairs design:* Again, all students do the activity twice. Randomly assign 10 students to Group 1 and 10 to Group 2. Group 1 uses the software the first time; Group 2 uses the software the second time.

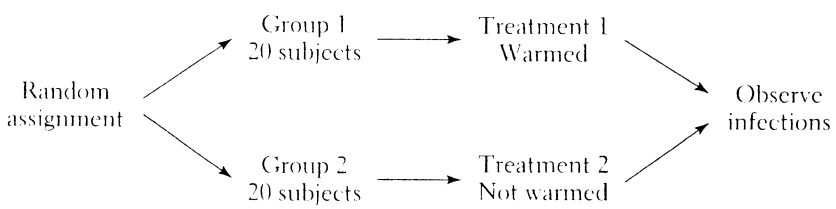
5.49 (a) "Randomized" means that patients were randomly assigned either St. John's wort or the placebo. "Placebo controlled" means that we will compare the results for the group using St. John's wort to the group that received the placebo.

(b)



5.50 (a) The subjects are the 210 children. (b) The factor is the "choice set"; there are three levels (2 milk/2 fruit drink, 4 milk/2 fruit drink, and 2 milk/4 fruit drink). (c) The response variable is the choice made by each child.

5.51 (a) Outline:



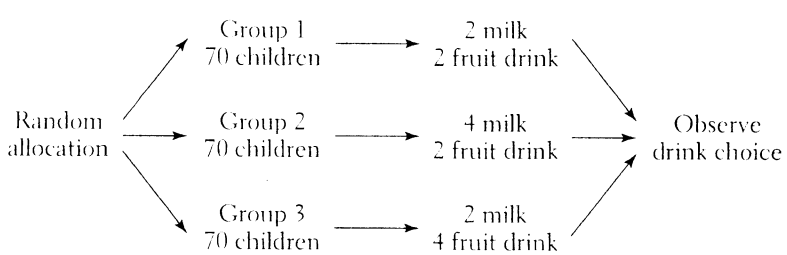
(b) Number the subjects from 01 to 40. Divide digits into groups of 2. Omit groups that are over 40. First 20 pairs will be Group 1. The rest will be Group 2.

(c) Assign each subject a group, 01 to 40, by alphabetical order. Starting at line 121 in Table B, the first twenty different groups we see are: 29-Ng, 07-Cordoba, 34-Sugiwara, 22-Kaplan, 10-Devlin, 25-Lucero, 13-Garcia, 38-Ullmann, 15-Green, 05-Cansico, 09-Decker, 08-Curzakis, 27-McNeill, 23-Kim, 30-Quinones, 28-Morse, 18-Howard, 03-Afifi, 01-Abbott, 36-Travers. These subjects will be assigned to Treatment Group 1; the remaining subjects go into Group 2.

(d) We want the treatment groups to be as alike as possible. If the same operating team was used to operate on "warmed" and "unwarmed" patients, then the effect of the "warming" on the occurrence of infection might be confounded with the effect of the surgical team (e.g., how skillful the team was in performing the necessary preventive measures).

(e) Double-blindness. We would prefer a double-blind experiment here to ensure that the patients would not be treated differently with regard to preventing and monitoring infections due to prior knowledge of how they were treated.

5.52



5.53 (a) Randomly assign 20 men to each of two groups. Record each subject's blood pressure, then apply the treatments: a calcium supplement for Group 1, and a placebo for Group 2. After sufficient time has passed, measure blood pressure again and observe any change.

(b) Number from 01 to 40 down the columns. Group 1 is 18-Howard, 20-Imrani, 26-Maldonado, 35-Tompkins, 39-Willis, 16-Guillen, 04-Bikalis, 21-James, 19-Hruska, 37-Tullock, 29-O'Brian, 07-Cranston, 34-Solomon, 22-Kaplan, 10-Durr, 25-Liang, 13-Fratianna, 38-Underwood, 15-Green, and 05-Chen.

(c) We prefer large treatment groups because differences in responses arising from such groups are more likely due to the treatments themselves, rather than chance variation.

5.54 Label the children from 001 to 210, then consider three digits at a time. The first five children in Group 1 are numbers 119, 033, 199, 192, and 148.

5.55 Responding to a placebo does not imply that the complaint was not "real"—38% of the placebo group in the gastric freezing experiment improved, and those patients really had ulcers. The placebo effect is a *psychological* response, but it may make an actual *physical* improvement in the patient's health.

5.56 (a) The explanatory variable is the vitamin(s) taken each day; the response variable is whether colon cancer develops.

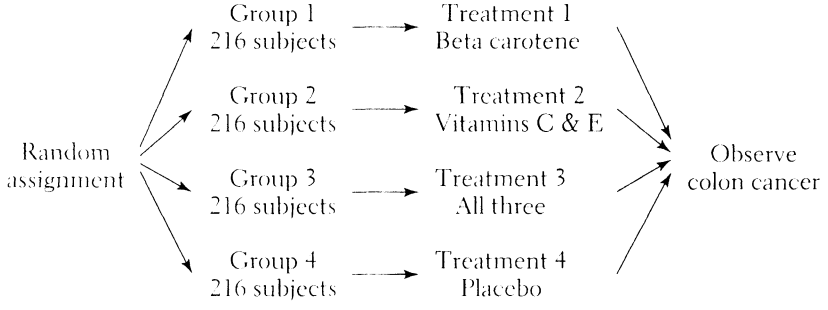
(b) Diagram below; equal group sizes are convenient but not necessary.

(c) Using labels 001 through 864 (or 000 through 863), we choose 731, 253, 304, 470, and 296.

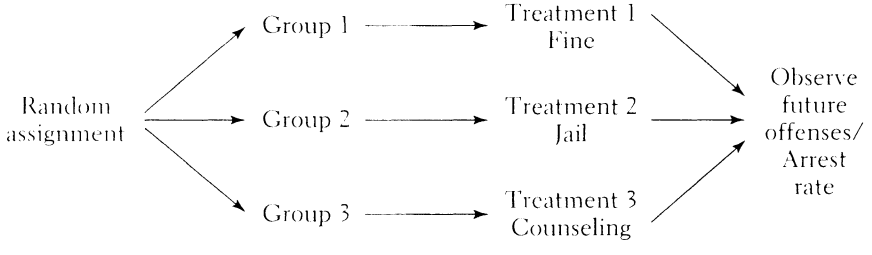
(d) "Double-blind" means that both the subjects and those who work with the subjects do not know who is getting what treatment. This prevents the expectations of those involved from affecting the way in which the subjects' conditions are diagnosed.

(e) The observed differences were no more than what might reasonably occur by chance even if there is no effect due to the treatments.

(f) Fruits and vegetables contain fiber; this could account for the benefits of those foods. Also, people who eat lots of fruit and vegetables may have healthier diets overall (e.g., less red meat).



5.57 Three possible treatments are (1) fine, (2) jail time, and (3) attending counseling classes. The response variable would be the rate at which people in the three groups are rearrested.



5.58 (a) Each subject takes both tests; the order in which the tests are taken is randomly chosen. (b) Take 22 digits from Table B. If the first digit is even, subject 1 takes the BI first; if it is odd, he or she takes the ARSMA first. (Or, administer the BI first if the first digit is 0–4, and the ARSMA first if it is 5–9.)

5.59 (a) Flip the coin twice. Let HH ⇔ failure, and let the other three outcomes, HT, TH, TT ⇔ success.

(b) Let 1,2,3 ⇔ success, and let 4 ⇔ failure. If 5 or 6 come up, ignore them and roll again.

(c) Peel off two consecutive digits from the table; let 01 through 75 ⇔ success, and let 76 through 99 and 00 ⇔ failure.

(d) Let diamond, spade, club ⇔ success, and let heart ⇔ failure.

5.60 Flip both nickels at the same time. Let HH ⇔ success (the occurrence of the phenomenon of interest) and HT, TH, TT ⇔ failure (the nonoccurrence of the phenomenon).

5.61 (a) Obtain an alphabetical list of the student body, and assign consecutive numbers to the students on the list. Use a random process (table or random digit generator) to select 10 students from this list.

(b) Let the two-digit groups 00 to 83 represent a "Yes" to the question of whether or not to abolish evening exams and the groups 84 to 99 represent a "No."

(c) Starting at line 129 in Table B ("Yes" in **boldface**) and moving across rows:

Repetition 1:	**36, 75,** 95, 89, 84, **68, 28, 82, 29, 13**	# "Yes": 7
Repetition 2:	**18, 63,** 85, **43, 03, 00, 79, 50,** 87, **27**	# "Yes": 8
Repetition 3:	**69, 05, 16, 48, 17,** 87, **17, 40,** 95, **17**	# "Yes": 8
Repetition 4:	84, **53, 40, 64,** 89, 87, **20, 19, 72, 45**	# "Yes": 7
Repetition 5:	**05, 00, 71, 66, 32, 81, 19, 41, 48, 73**	# "Yes": 10

(Theoretically, we should achieve 10 "Yes" results approximately 10.7% of the time.)

5.62 (a) A single random digit simulates one shot, with 0 to 6 a hit and 7, 8, or 9 a miss. Then 5 consecutive digits simulate 5 independent shots.

(b) Let 0–6 ⇔ "hit" and 7, 8, 9 ⇔ "miss." Starting with line 125, the first four repetitions are:

9 6 7 4 6	1 2 1 4 9	3 7 8 2 3	7 1 8 6 8
(3)	(4)	(3)	(2)

Each block of 5 digits in the table represents one repetition of the 5 attempted free throws. The underlined digits represent hits. We perform 46 more repetitions for a total of 50, and calculate the proportion of times the player makes 2 or fewer shots. Here are the number of hits for the 50 repetitions.

3 4 3 2 4	4 5 4 2 3	3 3 2 2 3	4 3 1 5 4
4 4 3 4 3	4 5 4 5 3	2 2 3 2 2	4 3 5 3 2
4 3 2 3 4	3 3 3 4 5		

The frequency counts are

X	0	1	2	3	4	5
Freq.	0	1	10	18	15	6

The relative frequency of 2 or fewer hits in 5 attempts is 11/50 = .22.

Note: It will be shown in Chapter 8 that the theoretical probability of missing 3 or more shots (i.e., making 2 or fewer shots) is 0.1631, or about one time in six.

5.63 The choice of digits in these simulations may of course vary from that made here. In (a)–(c), a single digit simulates the response; for (d), two digits simulate the response of a single voter.

(a) Odd digits — voter would vote Democratic
Even digits — voter would vote Republican

(b) 0, 1, 2, 3, 4, 5 — Democratic
6, 7, 8, 9 — Republican

(c) 0, 1, 2, 3 — Democratic
4, 5, 6, 7 — Republican
8, 9 — Undecided

(d) 00, 01, ..., 52 — Democratic
53, 54, ..., 99 — Republican

5.64 For the choices made in the solution to Exercise 5.63,

(a) D, R, R, R, R, R, R, D, R, D — 3 Democrats, 7 Republicans

(b) R, D, D, R, R, R, R, D, R, R — 3 Democrats, 7 Republicans

(c) R, U, R, D, R, U, U, U, D, R — 2 Democrats, 4 Republicans, 4 undecided

(d) R, R, R, D, D, D, D, D, D, R — 6 Democrats, 4 Republicans

5.65 Let 1 = girl and 0 = boy. The command `randInt(0, 1)` produces a 0 or 1 with equal likelihood. Continue to press ENTER. In 50 repetitions, we got a girl 47 times, and all 4 boys three times. Our simulation produced a girl 94% of the time, vs. a theoretical probability of 0.938.

5.66 (a) Let the digits 0, 1, 2, 3, 4, 5 ⇔ the American League team winning a Series game and 6, 7, 8, 9 ⇔ the National League team winning. We choose single digits until one team has won four games, with a minimum of four digits and a maximum of seven digits being chosen. On the TI-83, you can use the command `randInt (0, 9, 1)` repeatedly to generate the digits. Here are several sample simulations:

0, 3, 9, 2, 7, 9, 2	AL, AL, NL, AL, NL, NL, AL	# games = 7
3, 0, 9, 1, 0	AL, AL, NL, AL, AL	# games = 5

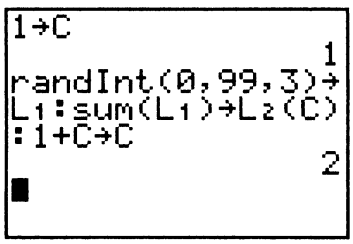
The long-term average of many simulations will give the approximate number of games one would expect the Series to last.

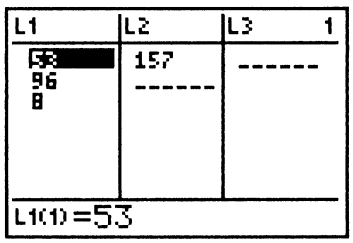
(b) Other factors might include: the starting pitchers, the weather conditions, the injury status of key players.

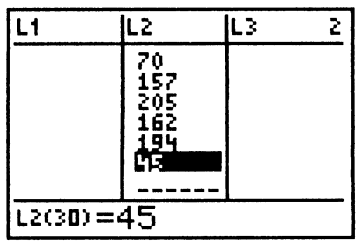
5.67 Let 01 to 15 ⇔ breaking a racquet, and let 16 to 99 and 00 ⇔ not breaking a racquet. Starting with line 141 in the random digit table, we peel two digits off at a time and record the results: 96 76 73 59 64 23 82 29 60 12. In the first repetition, Brian played 10 matches until he broke a racquet. Addition repetitions produced these results: 3, 11, 6, 37, 5, 3, 4, 11, 1. The average for these 10 repetitions is 9.1. We will learn in Chapter 8 that the expected number of matches until a break is about 6.67. More repetitions should improve our estimate.

5.68 (a) Read two random digits at a time from Table B. Let 01 to 13 represent a Heart, let 14 to 52 represent another suit, and ignore the other two-digit numbers. (b) You should beat Slim about 44% of the time.

5.69 On the TI-83, we started a counter (C), and then executed the command shown, pressing the ENTER key 30 times for 30 repetitions.







For five sets of 30 repetitions, we observed 5, 3, 3, 8, and 4 numbers that were multiples of 5. The mean number of multiples of 5 in 30 repetitions was 3.6, so 3.6/30 = 12% is our estimate for the proportion of times a person wins the game.

5.70 The command randInt (1, 365, 23) → L_1 : SortA (L_1) randomly selects 23 birthdays and assigns them to L_1. Then it sorts the day in increasing order. Scroll through the list to see duplicate birthdays. Repeat many times. For a large number of repetitions, there should be duplicate birthdays about half the time. To simulate 41 people, change 23 to 41 in the command and repeat many times. We assume that there are 365 days for birthdays, and that all days are equally likely to be a birthday.

5.71 (a) Let 000 to 999 ⇔ a bats, 000 to 319 ⇔ hits, and 320 to 999 ⇔ no hits.

(b) We entered 1 → c ENTER to set a counter. Then enter randInt (0, 999, 20) → L_1 : sum (L_1 ≥ 0 and L_1 ≤ 319) → L_2 (C) : C + 1 → C and press ENTER repeatedly. The count (number of the repetition) is displayed on the screen to help you see when to stop. The results for the 20 repetitions are stored in list L_2. We obtained the following frequencies:

Number of hits in 20 at bats	4	5	6	7	8	9
Frequency	3	5	4	3	2	3

(c) The mean number of hits in 20 at bats was $\bar{x}$ = 6.25. And 6.25/20 = .3125, compared with the player's batting average of .320. Notice that even though there was considerable variability in the 20 repetitions, ranging from a low of 3 hits to a high of 9 hits, the results of our simulation were very close to the player's batting average.

5.72 (a) One digit simulates system A's response: 0 to 8 shut down the reactor, and 9 fails to shut it down.

(b) One digit simulates system B's response: 0 to 7 shut down the reactor, and 8 or 9 fail.

(c) A pair of consecutive digits simulates the response of both systems, the first giving A's response as in (a), and the second B's response as in (b). If a single digit were used to simulate both systems, the reactions of A and B would be dependent—for example, if A fails, then B must also fail.

(d) The true probability that the reactor will shut down is 1 - (0.2)(0.1) = 0.98.

5.73 This simulation is fun for students, but the record-keeping can be challenging! Here is one method. First number the (real or imaginary) participants 1–25. Write the numbers 1–25 on the board so that you can strike through them as they hear the rumor. We used randInt (1, 25) to randomly select a person to begin spreading the rumor, and then pressed ENTER repeatedly to randomly select additional people to hear the rumor. We made a table to record the round (time increment), those who knew the rumor and were spreading it, those randomly selected to hear the rumor, and those who stopped spreading it because the person randomly selected to hear it had already heard it. Here is the beginning of our simulation, to illustrate our scheme:

Time incr	Knows		Tells	Stopped
1	16	→	2	
2	2	→	25	
	16	→	3	
3	2		19	
	3		6	
	16		15	
	25		1	
4	1		21	
	2		5	
	3		23	
	6		13	
	15		25	15
	16		9	
	19		16	19
	25		15	25
5				

Eventually we crossed off all but 7, 12, 14, and 24, so 4 out of 25 or 4/25 = 16% never heard the rumor. It can be shown that with a sufficiently large population, approximately 20% of the population will not hear the rumor.

5.74 (a) The population is Ontario residents; the sample is the 61,239 people interviewed. (b) The sample size is very large, so if there were large numbers of both sexes in the sample—this is a safe assumption since we are told this is a "random sample"—these two numbers should be fairly accurate reflections of the values for the whole population.

5.75 (a) Explanatory variable: treatment method; response: survival times. (b) No treatment is actively imposed; the women (or their doctors) chose which treatment to use. (c) Doctors may make the decision of which treatment to recommend based in part on how advanced the case is. Some might be more likely to recommend the older treatment for advanced cases, in which case the chance of recovery is lower. Other doctors might view the older treatment as not being worth the effort, and recommend the newer method as a way of providing *some* hope for recovery while minimizing the trauma and expense of major surgery.

5.76 (a) Sample survey. (b) Experiment. The treatment would be classroom or online course. (c) Observational study.

5.77 Divide the players into two groups—one to receive oxygen and the other without oxygen during the rest period. Match the players so each player receiving oxygen has a corresponding player of similar speed who does not receive oxygen.

5.78 A stratified random sample would be useful here; one could select 50 faculty members from each level. Alternatively, select 25 (or 50) institutions of each size, then choose 2 (or 1) faculty members at each institution.

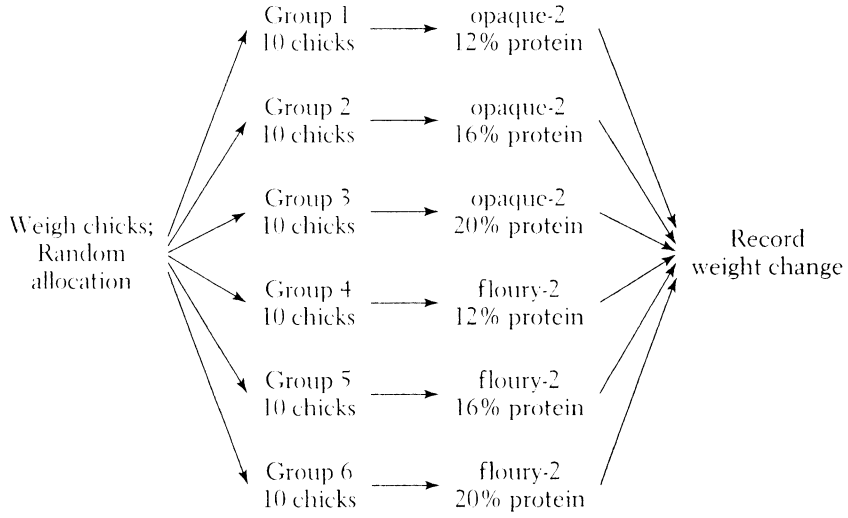
If a large proportion of faculty in your state work at a particular class of institution, it may be useful to stratify unevenly. If, for example, about 50% teach at Class I institutions, you may want half your sample to come from Class I institutions.

5.79 (a) The chicks are the experimental units; weight gain is the response variable.

(b) There are two factors: corn variety (2 levels) and percent of protein (3 levels). This makes 6 treatments, so 60 chicks are required.

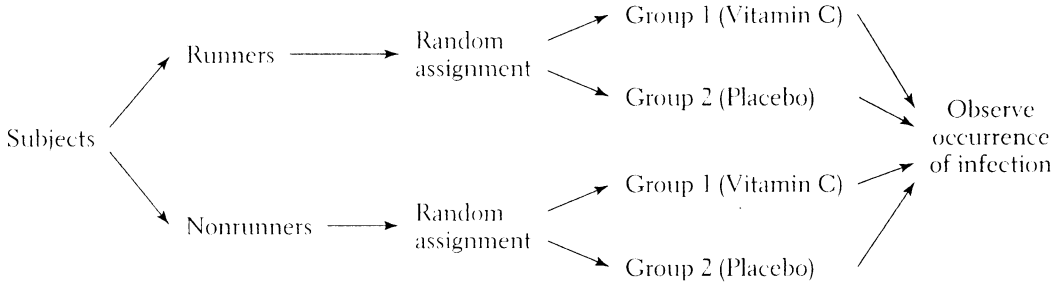
		Factor B: Protein level 12%	16%	20%
Factor A: Corn variety	opaque-2	1	2	3
	floury-2	4	5	6

(c)



5.80 (a) This is a randomized block design. The blocks here are "runners" and "nonrunners."

(b)



(c) A difference in rate of infection may have been due to the effects of the treatments, or it may simply have been due to random chance. Saying that the placebo rate of 68% is "significantly more" than the Vitamin C rate of 33% means that the observed difference is too large to have occurred by chance alone. In other words, Vitamin C appears to have played a role in lowering the infection rate of runners.

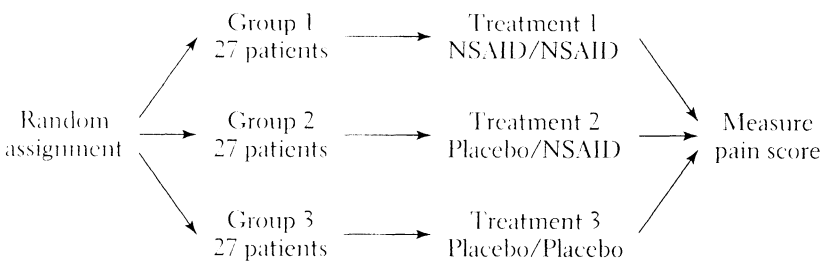
5.81 The factors are whether or not the letter has a ZIP code (2 levels: yes or no), and the time of day the letter is mailed. The number of levels for the second factor may vary.

To deal with lurking variables, all letters should be the same size and should be sent to the same city, and the day on which a letter is sent should be randomly selected. Because most post

offices have shorter hours on Saturdays, one may wish to give that day some sort of "special treatment" (it might even be a good idea to have the day of the week be a *third* factor in this experiment).

5.82 Each subject should taste both kinds of cheeseburger, in a randomly selected order, and then be asked about their preference. Both burgers should have the same "fixings" (ketchup, mustard, etc.). Since some subjects might be able to identify the cheeseburgers by appearance, one might need to take additional steps (such as blindfolding, or serving only the center part of the burger) in order to make this a truly "blind" experiment.

5.83 (a)



The two extra patients can be randomly assigned to two of the three groups.

(b) No one involved in administering treatments or assessing their effectiveness knew which subjects were in which group.

(c) The pain scores in Group A were so much lower than the scores in Groups B and C that they would not often happen by chance if NSAIDs were not effective. We can conclude that NSAIDs provide real pain relief.

5.84 (a) A single run: spin the 1–10 spinner twice; see if the larger of the two numbers is larger than 5. The player wins if either number is 6, 7, 8, 9, or 10.

(b) If using the random digit table, let 0 represent 10, and let the digits 1–9 represent themselves.

(c) `randInt (1, 10, 2)`.

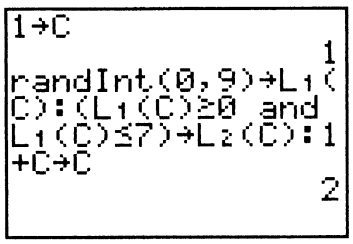
(d) In our simulation of 20 repetitions, we observed 13 wins for a 65% win rate. Using the methods of the next chapter, it can be shown that there is a 75% probability of winning this game.

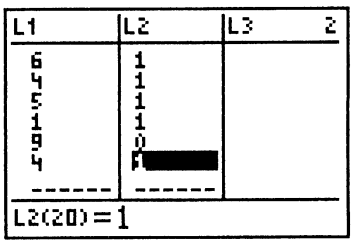
5.85 (a) Let 01 to 05 represent demand for 0 cheesecakes
Let 06 to 20 represent demand for 1 cheesecake
Let 21 to 45 represent demand for 2 cheesecakes
Let 46 to 70 represent demand for 3 cheesecakes
Let 71 to 90 represent demand for 4 cheesecakes
Let 91 to 99 and 00 represent demand for 5 cheesecakes

(b) Our results suggest that the baker should make 2 cheesecakes each day to maximize his profits.

5.86 (a) Since Carla makes 80% of her free throws, let a single digit represent a free throw, and let 0–7 $\Leftrightarrow$ "hit" and 8, 9 $\Leftrightarrow$ "miss."

(b) We instructed the calculator to simulate a free throw, and store the result in L_1. Then we instructed the calculator to see if the attempt was a hit (1) or a miss (0), and record that fact in L_2. Continue to press ENTER until there are 20 simulated free throws.





Scroll through L_2 and determine the longest string of 1's (consecutive baskets). In our first set of 20 repetitions, we observed 8 consecutive baskets. Additional sets of 20 repetitions produced: 11, 9, 9, and 8. The average longest run was 9 consecutive baskets in 20 attempts. The five number summary: Min = 8, Q_1 = 8, Med = 9, Q_3 = 10, and Max = 11. There was surprisingly little variation in our five groups of 20 repetitions.

5.87 (a) A single digit simulates one try, with 0 or 1 a pass and 2 to 9 a failure. Three independent tries are simulated by three successive random digits.

(b) With the convention of (a), 50 tries beginning in line 120 gives 25 successes, so the probability of success is estimated as 25/50 = 1/2. [In doing the simulation, remember that you can end a repetition after 1 or 2 tries if the student passes, so that some repetitions do not use three digits. Though this is a proper simulation of the student's behavior, the probability of at least one pass is the same if three digits are examined in every repetition. The true probability is $1 - (0.8)^3 = 0.488$, so this particular simulation was quite accurate.]

(c) No—learning usually occurs in taking an exam, so the probability of passing probably increases on each trial.

5.88 (a) The three tries are simulated by three consecutive random digits (stopping after a pass): 0 and 1 are a pass on the first try; 0, 1, and 2 a pass on the second try; and 0, 1, 2, and 3 a pass on the third try.

(b) The correct probability is $1 - (0.8)(0.7)(0.6) = 0.664$. Taking groups of three digits at a time (not quitting early after a pass) gives 36 passes, so the estimated probability is 0.72. If one does quit early after a pass, there are 32 passes, so the estimated probability is 0.64.