Chapter 7
Data for Decisions

Solutions

Exercises:

1. Population: U.S. residents aged 18 and older. Sample: The 1,002 who responded.

3. Students passing the student center may not fairly represent all students. For example, they may underrepresent commuters or students whose classes are far from the center. In addition, a woman student may be reluctant to stop men, thus underrepresenting male students.

5. (a) This isn't clear: possibly its readers, possibly all adults in its circulation area.

(b) Larger; People with strong opinions, especially negative opinions, are more likely to respond. This is bias due to voluntary response.

7. If we assign labels 01 to 33 to the complexes in alphabetical order and start at line 117 in Table 7.1, our sample is 16 = Fairington, 32 = Waterford Court, and 18 = Fowler.

9. (a) 001 to 371.

(b) Area codes labeled 214, 235, 119, 033, 199.

11. If you always start at the same point in the table, your sample is predictable in advance. Repeated samples of the same size from the same population will always be the same - that's not random.

13. (a) Because $\frac{200}{5} = 40$ we divide the list into 5 groups of 40. (By the way, if the list has 204 rooms, we divide it into 5 groups of 40 and final group of 4. A sample contains a room from the final group only when the first room chosen is among the first 4 in the list.) Label the first 40 rooms 01 to 40. Line 120 chooses room 35. The sample is rooms 35, 75, 115, 155, and 195.

(b) Each of the first 40 rooms has chance 1 in 40 to be chosen. Each later room is chosen exactly when the corresponding room in the first 40 is chosen. So every room has equal chance, 1 in 40. The only possible samples consist of 5 rooms spaced 40 apart in the list. An SRS gives *all* samples of 5 rooms an equal chance to be chosen.

15. (a) All people aged 18 and over living in the United States.

(b) Of the 1800 called, 669 did not respond. The rate is $\frac{669}{1800} \approx 0.37 = 37\%$.

(c) It is hard to remember exactly how many movies you saw in exactly the past 12 months.

17. Answers will vary.
People are more reluctant to "change" the Constitution than to "add to" it. So the wording "adding to" will produce a higher percent in favor.

19. No treatment was imposed on the subjects. This observational study collected unusually detailed information about the subjects, but made no attempt to influence them.

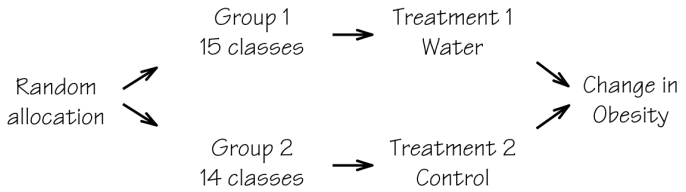
21. Answers will vary.

(a) It is an observational study that gathers information (e.g., through interviews) without imposing any treatment.

(b) "Significant" means "unlikely to be due simply to chance."

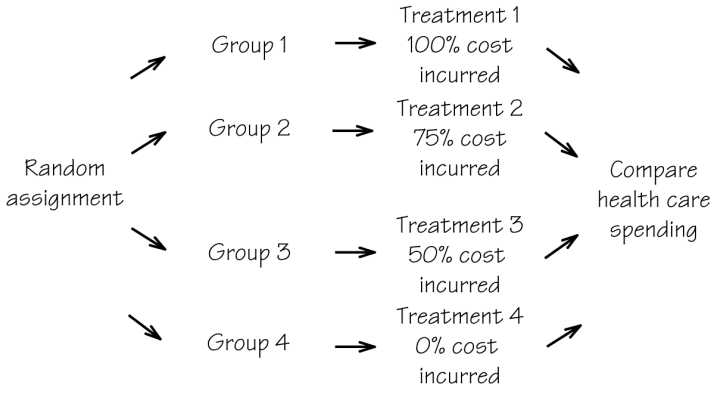
(c) Nondrinkers might be more elderly or in poorer health than moderate drinkers.

23. The design resembles Figure 7.3. Be sure to show randomization, two groups and their treatments, and the response variable (change in obesity).



If we label the 29 classes 01 to 29 and choose 15 for the treatment group, this group contains classes 17, 09, 22, 13, 07, 02, 27, 01, 18, 25, 29, 19, 14, 15, 08. We used lines 103 to 106 of Table 7.1, skipping many duplicate pairs of digits. The remaining 14 classes make up the control group.

25. This is a randomized comparative experiment with four branches, similar to Figure 7.4 with one more branch. The "flow chart" outline must show random assignment of subjects to groups, the four treatments, and the response variable (health care spending).



We can't show the group sizes because we don't know how many people or households are available to participate.

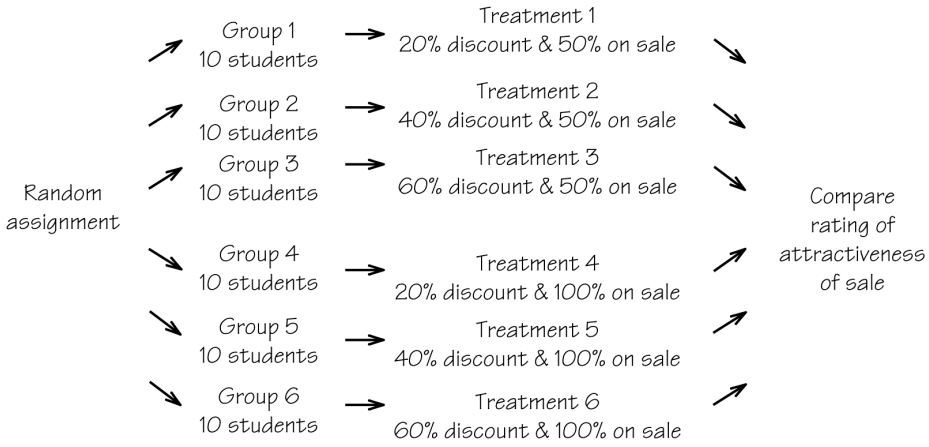
27. (a) There are 6 treatments, each combination of a level of discount and fraction on sale. In table form, the treatments are

	Discount level		
	20%	40%	60%
50% on sale	1	2	3
100% on sale	4	5	6

Continued on next page

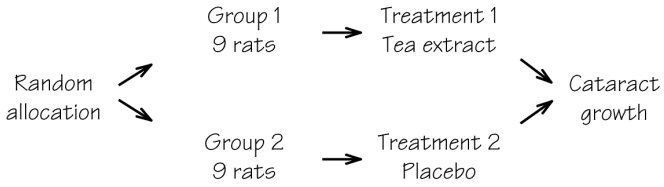
27. continued

(b) The outline randomly assigns 10 students to each of the 6 treatment groups, then compares the attractiveness ratings. It resembles Figure 7.4, but with 6 branches.



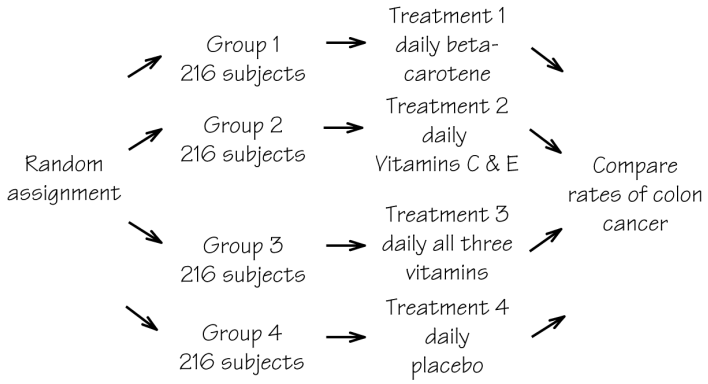
Label the subjects 01 to 60 and read line 123 of Table 7.1. The first group contains subjects labeled 54, 58, 08, 15, 07, 27, 10, 25, 60, 55.

29. The design resembles Figure 7.3:



(b) Label the rats 01 to 18. The tea group contains 15, 06, 13, 09, 18, 03, 05, 16, 17.

31. (a) This is a randomized comparative experiment with four branches, similar to Figure 7.4 with one more branch. The "flow chart" outline must show random assignment of subjects to groups, the group sizes and treatments, and the response variable (colon cancer). It is best to use groups of equal size, 216 people in each group.



(b) With labels 001 to 864, the first five chosen are 731, 253, 304, 470, and 296.

(c) Those working with the subjects did not know the contents of the pill each subject took daily.

(d) The differences in colon cancer cases in the four groups were so small that they could easily be due to the chance assignment of subjects to groups.

(e) People who eats lots of fruits and vegetables may eat less meat or more cereals than other people. They may drink less alcohol or exercise more.

33. During the experiment, only the experimental cars had center brake lights, so they attracted attention. Once most cars had them, they were less noticed and so did a poorer job of preventing collisions. This is an example of an experiment that could not be completely realistic.

35. (a) What percent of college students say that being very wealthy is one of their goals in life? Consider themselves conservatives? Have had five drinks in one sitting in the past week?

(b) Will offering nightly tutoring sessions improve performance in a math course? Will showing videos of drunken students reduce binge drinking?

37. Both are statistics because both describe the sample (the subjects who took part in the study).

39. (a) The mean is $p = 0.14$; the standard deviation is as follows.

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.14(1-0.14)}{500}} = \sqrt{\frac{0.14(0.86)}{500}} \approx 0.0155$$

(b) $0.14 \pm (2)(0.0155) = 0.14 \pm 0.031$

$$0.14 - 0.031 = \underline{0.109} \text{ to } 0.14 + 0.031 = \underline{0.171}$$

41. (a) Each digit in the table has one chance in 10 to be any of the ten possible digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. So in the long run, 60% of the digits we encounter will be 0, 1, 2, 3, 4, or 5 and 40% will be 6, 7, 8, or 9.

(b) Line 101 contains 29 digits 0 to 5. This stands for a sample with $\frac{29}{40} = 0.725 = 72.5\%$ "yes" responses. If we use lines 101 to 110 to simulate ten samples, the counts of "yes" responses are 29, 24, 23, 23, 20, 24, 23, 19, 24, and 18. Thus, three samples are exactly correct $\left(\frac{24}{40} = 0.60 = 60\%\right)$, one overestimates, and six underestimate.

43. The sample proportion who claim to have attended is $\hat{p} = \frac{750}{1785} \approx 0.420$. The approximate 95% confidence interval is calculated as follows.

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.420 \pm 2\sqrt{\frac{0.420(1-0.420)}{1785}} = 0.420 \pm 2\sqrt{\frac{0.420(0.580)}{1785}} \approx 0.420 \pm 0.023$$

$$0.420 - 0.023 = 0.397 \text{ to } 0.420 + 0.023 = 0.443$$

45. (a) The sample proportion who admit running a red light is $\hat{p} = \frac{171}{880} \approx 0.194$. The approximate 95% confidence interval is calculated as follows.

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.194 \pm 2\sqrt{\frac{0.194(1-0.194)}{880}} = 0.194 \pm 2\sqrt{\frac{0.194(0.806)}{880}} \approx 0.194 \pm 0.027$$

$$0.194 - 0.027 = 0.167 \text{ to } 0.194 + 0.027 = 0.221$$

(b) It is likely that more than 171 ran a red light, because some people are reluctant to admit illegal or antisocial acts.

47. If Harris takes a very large number of poll samples using the same methods, the poll result will be within $\pm 3\%$ percentage points of the truth about the population in 95% of the samples. The usual language for this is "95% confidence."

49. (a) $\frac{1468}{13{,}000} \approx 0.113 = 11.3\%$.

(b) The response rate is so low that it is likely that those who responded differ from the population as a whole. That is, there is a bias that the margin of error does not include.

51. (a) No. The number of e-filed returns in all states is much larger than the sample size. When this is true, the margin of error depends only on the size of the sample, not on the size of the population.

(b) The sample sizes vary from 970 to 49,000, so the margins of error will also vary.

53. The margin of error for 90% confidence comes from the central 90% of a normal sampling distribution. We need not go as far out to cover 90% of the distribution as to cover 95%. So the margin of error for 90% confidence is smaller than for 95% confidence.

55. The sample proportion of successes is $\hat{p} = \frac{7}{97} \approx 0.072$. That is, there were 7.2% successes in the sample. The approximate 95% confidence interval is calculated as follows.

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} = 0.072 \pm 2\sqrt{\frac{0.072\left(1-0.072\right)}{97}} = 0.072 \pm 2\sqrt{\frac{0.072\left(0.928\right)}{97}} \approx 0.072 \pm 0.052$$

$$0.072 - 0.052 = 0.020 \text{ to } 0.072 + 0.052 = 0.124$$

We are 95% confident that the true proportion of articles that discuss the success of blinding is between 0.020 and 0.124 (that is, 2.0% to 12.4%).

57. The distribution of the sample proportion $\hat{p}$ is approximately normal with mean $p = 0.1$ (that is, 10%) and standard deviation

$$\sqrt{\frac{p\left(1-p\right)}{n}} = \sqrt{\frac{0.1\left(1-0.1\right)}{97}} = \sqrt{\frac{0.1\left(0.9\right)}{97}} \approx 0.030$$

or 3%. Notice that 7% is one standard deviation below the mean. By the 68 part of the 68-95-99.7 rule, 68% of all samples will have between 7% and 13% that discuss blinding. Half the remaining 32% lie on either side. So 16% of samples will have fewer than 7% articles that discuss blinding. That is, the probability is about 0.16.