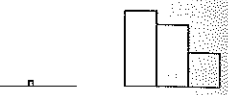
ment from the box
ogram for the num-
sum. And one is the
Why?

(iii)



o simulate tossing a
o one million tosses.
e programmer looks

o, wait. Compare it to
f a million is two out
a percent. Very small.

ke

2

give or take 5 or so.

ke 5 or so.

ent from a box with
." True or false, and

in the range 0 to 2.

10. Ten thousand draws are made at random with replacement from a box with ninety-nine tickets marked "0" and one ticket marked "1." True or false, and explain:
    (a) The sum will be around 100, give or take 10 or so.
    (b) There is about a 68% chance that the sum will be in the range 90 to 110.

11. One hundred draws are made at random with replacement from the box |1 2 2 5|. The draws come out as follows: 17 1's, 54 2's, 29 5's. Fill in the blanks, using the options below; show work.
    (a) For the _____, the observed value is 0.8 SEs above the expected value.
    (b) For the _____, the observed value is 1.33 SEs above the expected value.

    Options (one will be left over):

    sum of the draws number of 1's number of 2's

12. A box contains ten tickets, four marked with a positive number and six with a negative number. All the numbers are between −10 and 10. One thousand draws will be made at random with replacement from the box. You are asked to estimate the chance that the sum will be positive.
    (a) Can you do it on the basis of the information already given?
    (b) Can you do it if you are also told the average and SD of the numbers in the box, but are not told the numbers themselves?

    Explain briefly.

13. Repeat exercise 12, if you are asked to estimate the chance of getting 100 or more 3's.

14. Repeat exercise 12, if you are asked to estimate the chance of getting 425 or more positive numbers.

15. A box contained 1,500 marbles; 600 were red and the others, blue. The following procedure was repeated many times.

    > One hundred draws were made at random with replacement from the box; the number of red marbles among the draws was counted.

    The first 10 counts were 38, 35, 37, 31, 36, 39, 36, 33, 30, 34. Is anything fishy? Answer yes or no, and explain.

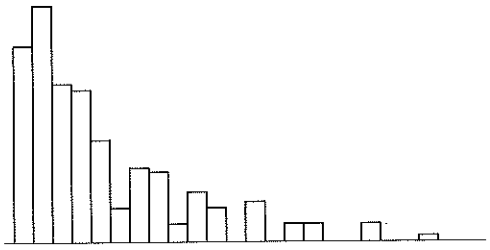
8. SUMMARY

1. If the chance process for getting a sum is repeated many times, the empirical histogram for the observed values converges to the *probability histogram*.

6. "Ecstasy" was a popular drug in the 1990s. It produced a sense of euphoria derisively called the "yuppie high." One investigator made a careful sample survey to estimate the prevalence of drug use at Stanford University. Two assistants were stationed on the main campus plaza and instructed to interview all students who passed through at specified times. As it turned out, 39% of 369 students interviewed said they had used Ecstasy at least once.[30] Does the investigator's procedure give a probability sample of Stanford students? Answer yes or no, and explain.

7. A coin is tossed 1,000 times. There are two options:
   (i) To win \$1 if the number of heads is between 490 and 510.
   (ii) To win \$1 if the percentage of heads is between 48% and 52%.
   Which option is better? Or are they the same? Explain.

8. Can you tell whether the figure below is a probability histogram or a histogram for data? If so, which is it and why? If you can't tell, why not?



9. One hospital has 218 live births during the month of January.[31] Another has 536. Which is likelier to have 55% or more male births? Or is it equally likely? Explain. (There is about a 52% chance for a live-born infant to be male.)

10. A coin will be tossed 100 times. You get to pick 11 numbers. If the number of heads turns out to equal one of your 11 numbers, you win a dollar. Which 11 numbers should you pick, and what is your chance (approximately) of winning? Explain.

11. A sorcerer has hidden a Porsche in one of an infinite row of boxes

The sorcerer will let you drive away with the car if you can find it. But you are only allowed to look in 11 boxes. He agrees to give you a hint, by tossing a coin 100 times and counting the number of heads. He will not tell you this number, or the number of the box in which he hid the car. But he will tell you the sum of the two numbers.

(a) If the sum is 65, which 11 boxes would you look in?
(b) As in (a), except replace 65 by 95.
(c) What is the general rule?
(d) Following this rule, how likely are you to get the Porsche?

12. The *San Francisco Chronicle* reported on a survey of top high-school students in the U.S. According to the survey,

> Cheating is pervasive. Nearly 80 percent admitted some dishonesty, such as copying someone's homework or cheating on an exam. The survey was sent last spring to 5,000 of the nearly 700,000 high achievers included in the 1993 edition of *Who's Who Among American High School Students*. The results were based on the 1,957 completed surveys that were returned. "The survey does not pretend to be representative of all teenagers," said *Who's Who* spokesman Andrew Weinstein. "Students are listed in *Who's Who* if they are nominated by their teachers or guidance counselors. Ninety-eight percent of them go on to college."

(a) Why isn't the survey "representative of all teenagers"?
(b) Is the survey representative "of the nearly 700,000 high achievers included in the 1993 edition of *Who's Who Among American High School Students*"? Answer yes or no, and explain briefly.

10. SUMMARY

1. A *sample* is part of a *population*.

2. A *parameter* is a numerical fact about a population. Usually a parameter cannot be determined exactly, but can only be estimated.

3. A *statistic* can be computed from a sample, and used to estimate a parameter. A statistic is what the investigator knows. A parameter is what the investigator wants to know.

4. When estimating a parameter, one major issue is accuracy: how close is the estimate going to be?

5. Some methods for choosing samples are likely to produce accurate estimates. Others are spoiled by *selection bias* or *non-response bias*. When thinking about a sample survey, ask yourself:

- What is the population? the parameter?
- How was the sample chosen?
- What was the response rate?

6. Large samples offer no protection against bias.

7. In *quota sampling*, the sample is hand picked by the interviewers to resemble the population in some key ways. This method seems logical, but often

Poll: the likely size of the chance error in sample percentages depends mainly on the absolute size of the sample, and hardly at all on the size of the population. The huge number of eligible voters makes it hard work to draw the sample, but does not affect the standard error.

Is 2,500 a big enough sample? The square root law provides a benchmark. For example, with 2,500 tosses of a coin, the standard error for the percentage of heads is only 1%. Similarly, with a sample of 2,500 voters, the likely size of the chance error is only a percentage point or so. That is good enough unless the election is very close, like Bush versus Gore in 2000. The Electoral College would be a major complication: the Gallup Poll only predicts the popular vote.

6. REVIEW EXERCISES

Review exercises may also cover material from previous chapters.

1. Complete the following table for the coin-tossing game.

	Number of heads		*Percent of heads*	
Number of tosses	*Expected value*	*SE*	*Expected value*	*SE*
100	50	5	50%	5%
2,500				1%
10,000				
1,000,000				

2. A die is rolled one thousand times. The percentage of aces ([•]) should be around _____, give or take _____ or so.
   (a) The first step in solving this problem is
      (i) computing the SD of the box.
      (ii) computing the average of the box.
      (iii) setting up the box model.

   Choose one option and explain.
   (b) Now solve the problem.

3. A group of 50,000 tax forms has an average gross income of $37,000, with an SD of $20,000. Furthermore, 20% of the forms have a gross income over $50,000. A group of 900 forms is chosen at random for audit. To estimate the chance that between 19% and 21% of the forms chosen for audit have gross incomes over $50,000, a box model is needed.
   (a) Should the number of tickets in the box be 900 or 50,000?
   (b) Each ticket in the box shows

   a zero or a one a gross income

   (c) True or false: the SD of the box is $20,000.
   (d) True or false: the number of draws is 900.
   (e) Find the chance (approximately) that between 19% and 21% of the forms chosen for audit have gross incomes over $50,000.

(f) With the information given, can you find the chance (approximately) that between 9% and 11% of the forms chosen for audit have gross incomes over $75,000? Either find the chance, or explain why you need more information.

4. As in exercise 3, except it is desired to find the chance (approximately) that the total gross income of the audited forms is over $33,000,000. Work parts (a) through (d); then find the chance or explain why you need more information.

5. (Hypothetical.) On the average, hotel guests who take elevators weigh about 150 pounds with an SD of about 35 pounds. An engineer is designing a large elevator for a convention hotel, to lift 50 such people. If she designs it to lift 4 tons, the chance it will be overloaded by a random group of 50 people is about _____. Explain briefly.

6. The Census Bureau is planning to take a sample amounting to 1/10 of 1% of the population in each state in order to estimate the percentage of the population in that state earning over $100,000 a year. Other things being equal:

   (i) The accuracy to be expected in California (population 35 million) is about the same as the accuracy to be expected in Nevada (population 2 million).
   (ii) The accuracy to be expected in California is quite a bit higher than in Nevada.
   (iii) The accuracy to be expected in California is quite a bit lower than in Nevada.

   Explain.

7. Five hundred draws are made at random from the box

60,000 [0]'s 20,000 [1]'s

   True or false, and explain:

   (a) The expected value for the percentage of 1's among the draws is exactly 25%.
   (b) The expected value for the percentage of 1's among the draws is around 25%, give or take 2% or so.
   (c) The percentage of 1's among the draws will be around 25%, give or take 2% or so.
   (d) The percentage of 1's among the draws will be exactly 25%.
   (e) The percentage of 1's in the box is exactly 25%.
   (f) The percentage of 1's in the box is around 25%, give or take 2% or so.

8. In a certain town, there are 30,000 registered voters, of whom 12,000 are Democrats. A survey organization is about to take a simple random sample of 1,000 registered voters. There is about a 50–50 chance that the percentage of Democrats in the sample will be bigger than _____. Fill in the blank, and explain.

mal approximation has been used. If the normal approximation does not apply, neither do the methods of this chapter. There is no hard-and-fast rule for deciding. The best way to proceed is to imagine that the population has the same percentage composition as the sample. Then try to decide whether the normal approximation would work for the sum of the draws from the box. For instance, a sample percentage near 0% or 100% suggests that the box is lopsided, so a large number of draws will be needed before the normal approximation takes over (section 5 of chapter 18). On the other hand, if the sample percentage is near 50%, the normal approximation should be satisfactory when there are only a hundred draws or so.

Exercise Set B

1. Fill in the blanks, and explain.
   (a) In example 2 on p. 382, the 917 is the _____ value for the number of Democrats in the sample. Options: (i) expected (ii) observed
   (b) The SD of the box is _____ $\sqrt{0.573 \times 0.427}$. Options:
   (i) exactly equal to (ii) estimated from the data as
   (c) The SE for the number of Democrats in the sample is _____ 20. Options:
   (i) exactly equal to (ii) estimated from the data as

2. Refer back to exercise 2 on p. 379.
   (a) Find a 95%-confidence interval for the percentage of persons age 18 to 24 in the city who are currently enrolled in college.
   (b) Repeat, for a confidence level of 99.7%.
   (c) Repeat, for a confidence level of 99.7%, supposing the size of the sample was 2,000, of whom 776 were currently enrolled in college.

3. A box contains 1 red marble and 99 blues; 100 marbles are drawn at random with replacement.
   (a) Find the expected number of red marbles among the draws, and the SE.
   (b) What is the chance of drawing fewer than 0 red marbles?
   (c) Use the normal curve to estimate this chance.
   (d) Does the probability histogram for the number of red marbles among the draws look like the normal curve?

4. A box contains 10,000 marbles, of which some are red and the others blue. To estimate the percentage of red marbles in the box, 100 are drawn at random without replacement. Among the draws, 1 turns out to be red. The percentage of red marbles in the box is estimated as 1%, with an SE of 1%. True or false: a 95%-confidence interval for the percentage of red marbles in the box is $1\% \pm 2\%$. Explain.

The answers to these exercises are on pp. A82–83.

3. INTERPRETING A CONFIDENCE INTERVAL

In example 1 on p. 378, a simple random sample was taken to estimate the percentage of students registered at a university in fall 2005 who were living at home. An approximate 95%-confidence interval for this percentage ran from 75%

3. The following article appeared on the *New York Times* Op Ed page of August 27, 1988, headlined MAYBE BUSH HAS ALREADY WON.

> The presidential campaign, only now formally set to begin, is in fact virtually finished. Despite the Niagara of news stories about how the candidates are touting their running mates, haggling over debates and sniping at each other, the die is just about cast.
>
> A significant indicator is the Gallup Poll, which this week shows Vice President Bush ahead of Gov. Michael S. Dukakis by 4 percentage points. In the half century since George Gallup began his electoral opinion surveys in Presidential years, his "trial heats" in the last week or so of September have foretold with notable accuracy the outcome on election day.
>
> The late James A. Farley, the Democrats' peerless tactician of 50 years ago, always argued that voters made up their minds by Labor Day.... It is now established, moreover, that when traditional nonvoters—the object of get-out-the-vote efforts—are persuaded to vote, they too cast their ballots in the same proportion as the rest of the electorate.... Significant changes in the percentages from September to November are due only to altered voter enthusiasm....

   (a) How does the article explain differences in voter opinion between September and November?
   (b) What else could explain a difference between Gallup Poll results in late September and election results in early November?
   (c) A difference of several percentage points between Gallup Poll results in late September and election results in early November is: very unlikely, unlikely but possible, quite possible. Choose one option, and explain.

The answers to these exercises are on p. A84.

6. REVIEW EXERCISES

Review exercises may cover material from previous chapters.

1. A survey organization draws a simple random sample of 1,000 registered voters in a certain town. In the sample, 32% approve of the Mayor. The organization estimates that 32% of all 50,000 registered voters in the town approve of the Mayor. How to figure the SE? The organization realizes that the number in the sample who approve _____ 1,000 draws _____ box _____. Fill in each blank (33 words or less). Then work out the SE.

2. The Residential Energy Consumption Survey found in 2001 that 47% of American households had internet access.[10] A market survey organization repeated this study in a certain town with 25,000 households, using a simple random sample of 500 households: 239 of the sample households had internet access.

   (a) The percentage of households in the town with internet access is estimated as _____; this estimate is likely to be off by _____ or so.
   (b) If possible, find a 95%-confidence interval for the percentage of all 25,000 households with internet access. If this is not possible, explain why not.

(a) Is 0.8 of 1% the right SE? Answer yes or no, and explain.
(b) 71% ± 1.6% is a _____ for the _____. Fill in the blanks and explain.

9. (Hypothetical.) A bank wants to estimate the amount of change people carry. They take a simple random sample of 100 people, and find that on the average, people in the sample carry 73¢ in change. They figure the standard error is 4¢, because

$$\sqrt{100} \times \sqrt{0.73 \times 0.27} \approx 4, \quad 4/100 = .04$$

Are they right? Answer yes or no, and explain.

10. In Keno, there are 80 balls numbered from 1 to 80, and 20 are drawn at random. If you play a double-number, you win if both numbers are chosen. This bet pays 11 to 1, and you have very close to a 6% chance of winning.[13] If you play 100 times and stake $1 on a double-number each time, your net gain will be around _____, give or take _____ or so.

11. One hundred draws will be made at random without replacement from a large box of numbered tickets. There are two options:

(i) To win $1 if the sum of the draws is bigger than 710.
(ii) To win $1 if the average of the draws is bigger than 7.1.

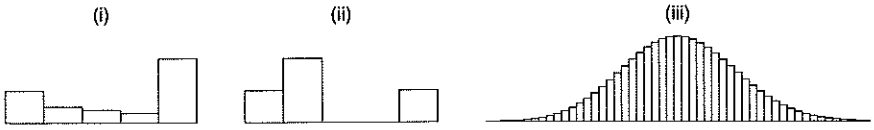
Which is better? Or are they the same? Explain.

12. A monthly opinion survey is based on a sample of 1,500 persons, "scientifically chosen as a representative cross section of the American public." The press release warns that the estimates are subject to chance error, but guarantees that they are "reliable to within two percentage points." The word "reliable" is ambiguous. According to statistical theory, the guarantee should be interpreted as follows:

(i) In virtually all these surveys, the estimates will be within two percentage points of the parameters.
(ii) In most such surveys, the estimates will be within two percentage points of the parameters, but in some definite percentage of the time larger errors are expected.

Explain.

13. One hundred draws are made at random with replacement from the box $\boxed{1}\boxed{2}\boxed{2}\boxed{5}$. One of the graphs below is a histogram for the numbers drawn. Another is the probability histogram for the sum. And the third is irrelevant. Which is which? Why?



14. A coin is tossed 1,000 times.

(a) Suppose it lands heads 529 times. Find the expected value for the number of heads, the chance error, and the standard error.