Homework 14: Do 5.43 and 5.44 Plus (2) more

SECTION 5.2 Exercises

For Exercises 5.32 and 5.33, see pages 313–314; for Exercises 5.34 and 5.35, see page 315; for Exercises 5.36 and 5.37, see page 319; for Exercise 5.38, see page 322; for Exercise 5.39, see page 326; and for Exercise 5.40, see page 330.

Most binomial probability calculations required in these exercises can be done by using Table C or the Normal approximation. Your instructor may request that you use the binomial probability formula or software. In exercises requiring the Normal approximation, you should use the continuity correction if you studied that topic.

5.41 What is wrong? Explain what is wrong in each of the following scenarios.

(a) If you toss a fair coin three times and heads appears each time, then the next toss is more likely to be a tail than a head.

(b) If you toss a fair coin three times and heads appears each time, then the next toss is more likely to be a head than a tail.

(c) $\hat{p}$ is one of the parameters for a binomial distribution.

5.42 What is wrong? Explain what is wrong in each of the following scenarios.

(a) In the binomial setting, X is a proportion.

(b) The variance for a binomial count is $\sqrt{p(1-p)/n}$.

(c) The Normal approximation to the binomial distribution is always accurate when n is greater than 1000.

5.43 Should you use the binomial distribution? In each of the following situations, is it reasonable to use a binomial distribution for the random variable X? Give reasons for your answer in each case. If a binomial distribution applies, give the values of n and p.

(a) A poll of 200 college students asks whether or not you are usually irritable in the morning. X is the number who reply that they are usually irritable in the morning.

(b) You toss a fair coin until a head appears. X is the count of the number of tosses that you make.

(c) Most calls made at random by sample surveys don't succeed in talking with a live person. Of calls to New York City, only one-twelfth succeed. A survey calls 500 randomly selected numbers in New York City. X is the number of times that a live person is reached.

(d) You deal 10 cards from a shuffled deck and count the number X of black cards.

5.44 Should you use the binomial distribution? In each of the following situations, is it reasonable to use a binomial distribution for the random variable X? Give reasons for your answer in each case.

(a) In a random sample of students in a fitness study, X is the mean systolic blood pressure of the sample.

(b) A manufacturer of running shoes picks a random sample of the production of shoes each day for a detailed inspection. Today's sample of 20 pairs of shoes includes one pair with a defect.

(c) A nutrition study chooses an SRS of college students. They are asked whether or not they usually eat at least five servings of fruits or vegetables per day. X is the number who say that they do.

5.45 Typographic errors. Typographic errors in a text are either nonword errors (as when "the" is typed as "teh") or word errors that result in a real but incorrect word. Spell-checking software will catch nonword errors but not word errors. Human proofreaders catch 70% of word errors. You ask a fellow student to proofread an essay in which you have deliberately made 10 word errors.

(a) If the student matches the usual 70% rate, what is the distribution of the number of errors caught? What is the distribution of the number of errors missed?

(b) Missing 4 or more out of 10 errors seems a poor performance. What is the probability that a proofreader who catches 70% of word errors misses 4 or more out of 10?

5.46 Streaming online music. A recent survey of 1000 United Kingdom music fans aged 14 to 64, revealed that roughly 30% of the teenage music fans are listening to streamed music on their computer everyday.[11] You decide to interview a random sample of 20 U.S. teenage music fans. For now assume they behave similarly to U.K. teenagers.

(a) What is the distribution of the number who listen to streamed music daily? Explain your answer.

(b) What is the probability that at least 8 of the 20 listen to streamed music daily?

5.47 Typographic errors. Return to the proofreading setting of Exercise 5.45.

(a) What is the mean number of errors caught? What is the mean number of errors missed? You see that these two means must add to 10, the total number of errors.

(b) What is the standard deviation σ of the number of errors caught?

(c) Suppose that a proofreader catches 90% of word errors, so that $p = 0.9$. What is σ in this case? What is σ if $p = 0.99$? What happens to the standard deviation of a binomial distribution as the probability of a success gets close to 1?

5.48 Streaming online music, continued. Recall Exercise 5.46. Suppose that only 25% of the U.S. teenage music fans listen to streamed music daily.

(a) If you interview 20 at random, what is the mean of the count X who listen to streamed music daily? What is the mean of the proportion $\hat{p}$ in your sample who listen to streamed music daily?

(b) Repeat the calculations in part (a) for samples of size 200 and 2000. What happens to the mean count of successes as the sample size increases? What happens to the mean proportion of successes?

5.49 Typographic errors. In the proofreading setting of Exercise 5.45, what is the smallest number of misses m with $P(X \geq m)$ no larger than 0.05? You might consider m or more misses as evidence that a proofreader actually catches fewer than 70% of word errors.

5.50 Attitudes toward drinking and behavior studies. Some of the methods in this section are approximations rather than exact probability results. We have given rules of thumb for safe use of these approximations.

(a) You are interested in attitudes toward drinking among the 75 members of a fraternity. You choose 30 members at random to interview. One question is, "Have you had five or more drinks at one time during the last week?" Suppose that in fact 30% of the 75 members would say "Yes." Explain why you *cannot* safely use the $B(30, 0.3)$ distribution for the count X in your sample who say "Yes."

(b) The National AIDS Behavioral Surveys found that 0.2% (that's 0.002 as a decimal fraction) of adult heterosexuals had both received a blood transfusion and had a sexual partner from a group at high risk of AIDS. Suppose that this national proportion holds for your region. Explain why you *cannot* safely use the Normal approximation for the sample proportion who fall in this group when you interview an SRS of 1000 adults.

5.51 Random digits. Each entry in a table of random digits like Table B has probability 0.1 of being a 0, and digits are independent of each other.

(a) What is the probability that a group of five digits from the table will contain at least one 5?

(b) What is the mean number of 5s in lines 40 digits long?

5.52 Use the *Probability* applet. The *Probability* applet simulates tosses of a coin. You can choose the number of tosses n and the probability p of a head. You can therefore use the applet to simulate binomial random variables.

The count of misclassified sales records in Example 5.15 (page 316) has the binomial distribution with $n = 15$ and $p = 0.08$. Set these values for the number of tosses and probability of heads in the applet. Table C shows that the probability of getting a sample with exactly 0 misclassified records is 0.2863. This is the long-run proportion of samples with no bad records. Click "Toss" and "Reset" repeatedly to simulate 25 samples. Record the number of bad records (the count of heads) in each of the 25 samples. What proportion of the 25 samples had exactly 0 bad records? Remember that probability tells us only what happens in the long run.

5.53 Inheritance of blood types. Children inherit their blood type from their parents, with probabilities that reflect the parents' genetic makeup. Children of Juan and Maria each have probability 1/4 of having blood type A and inherit independently of each other. Juan and Maria plan to have 4 children; let X be the number who have blood type A.

(a) What are n and p in the binomial distribution of X?

(b) Find the probability of each possible value of X, and draw a probability histogram for this distribution.

(c) Find the mean number of children with type A blood, and mark the location of the mean on your probability histogram.

5.54 The ideal number of children. "What do you think is the ideal number of children for a family to have?" A Gallup Poll asked this question of 1007 randomly chosen adults. Over half (52%) thought two children was ideal.[12] Suppose that $p = 0.52$ is exactly true for the population of all adults. Gallup announced a margin of error of ± 3 percentage points for this poll. What is the probability that the sample proportion $\hat{p}$ for an SRS of size $n = 1007$ falls between 0.49 and 0.55? You see that it is likely, but not certain, that polls like this give results that are correct within their margin of error. We will say more about margins of error in Chapter 6.

5.55 Visiting a casino and betting on college sports. A Gallup Poll finds that 24% of adults visited a casino in the past 12 months, and that 4% bet on college sports.[13] These results come from a random sample of 1027 adults. For an SRS of size $n = 1027$:

(a) What is the probability that the sample proportion $\hat{p}$ is between 0.22 and 0.26 if the population proportion is $p = 0.24$?

(b) What is the probability that the sample proportion $\hat{p}$ is between 0.02 and 0.06 if the population proportion is $p = 0.04$?

(c) Using the results from parts (a) and (b), how does the probability that $\hat{p}$ falls within ± 0.02 of the true p change as p gets closer to 0?

5.56 How do the results depend on the sample size? Return to the Gallup Poll setting of Exercise 5.54. We are supposing that the proportion of all adults who think that two children is ideal is $p = 0.52$. What is the probability that a sample proportion $\hat{p}$ falls between 0.49 and 0.55 (that is, within ± 3 percentage points of the true p) if the sample is an SRS of size $n = 300$? Of size $n = 5000$? Combine these results with your work in Exercise 5.54 to make a general statement about the effect of larger samples in a sample survey.

5.57 Shooting free throws. Since the mid-1960s, the overall free throw percent at all college levels, for both men and women, has remained pretty consistent. For men, players have shot roughly 69% with the season percent never falling below 67% or larger than 70%.[14] Assume that 300,000 free throws will be attempted in the upcoming season.

(a) What is the mean and standard deviation of $\hat{p}$ if the population proportion is $p = 0.69$?

(b) Using the 68–95–99.7 rule, we'd expect $\hat{p}$ to fall between what two percents about 95% of the time?

(c) Given the width of this interval in part (b) and the range of season percents, do you feel it reasonable to assume the population proportion has been the same over the last 50 seasons? Explain your answer.

5.58 Online learning. Recently the U.S. Department of Education released a report on online learning stating that blended instruction, a combination of conventional face-to-face and online instruction, appears more effective in terms of student performance than conventional teaching.[15] You decide to poll the incoming students at your institution to see if they prefer courses that blend face-to-face instruction with online components. In an SRS of 400 incoming students, you find 294 prefer this type of course.

(a) What is the sample proportion who prefer this type of blended instruction?

(b) Suppose the population proportion for all students nationwide is 80%. What is the standard deviation of $\hat{p}$?

(c) Using the 68–95–99.7 rule, if you had drawn an SRS from the United States, you would expect $\hat{p}$ to fall between what two percents about 95% of the time?

(d) Based on your result in part (a), do you feel the incoming students at your institution prefer this type of instruction more, less, or about the same as students nationally? Explain your answer.

5.59 A college alcohol study. The Harvard College Alcohol Study finds that 67% of college students support efforts to "crack down on underage drinking." The study took a sample of almost 15,000 students, so the population proportion who support a crackdown is very close to $p = 0.67$.[16] The administration of your college surveys an SRS of 200 students and finds that 140 support a crackdown on underage drinking.

(a) What is the sample proportion who support a crackdown on underage drinking?

(b) If in fact the proportion of all students on your campus who support a crackdown is the same as the national 67%, what is the probability that the proportion in an SRS of 200 students is as large or larger than the result of the administration's sample?

(c) A writer in the student paper says that support for a crackdown is higher on your campus than nationally. Write a short letter to the editor explaining why the survey does not support this conclusion.

5.60 How large a sample is needed? The changing probabilities you found in Exercises 5.54 and 5.56 are due to the fact that the standard deviation of the sample proportion $\hat{p}$ gets smaller as the sample size n increases. If the population proportion is $p = 0.52$, how large a sample is needed to reduce the standard deviation of $\hat{p}$ to $\sigma_{\hat{p}} = 0.004$? (The 68–95–99.7 rule then says that about 95% of all samples will have $\hat{p}$ within 0.01 of the true p.)

5.61 A test for ESP. In a test for ESP (extrasensory perception), the experimenter looks at cards that are hidden from the subject. Each card contains either a star, a circle, a wave, or a square. As the experimenter looks at each of 20 cards in turn, the subject names the shape on the card.

(a) If a subject simply guesses the shape on each card, what is the probability of a successful guess on a single card? Because the cards are independent, the count of successes in 20 cards has a binomial distribution.

(b) What is the probability that a subject correctly guesses at least 10 of the 20 shapes?

(c) In many repetitions of this experiment with a subject who is guessing, how many cards will the subject guess correctly on the average? What is the standard deviation of the number of correct guesses?

(d) A standard ESP deck actually contains 25 cards. There are five different shapes, each of which appears on 5 cards. The subject knows that the deck has this makeup. Is a binomial model still appropriate for the count of correct guesses in one pass through this deck? If so, what are n and p? If not, why not?

5.62 Admitting students to college. A selective college would like to have an entering class of 950 students. Because not all students who are offered admission accept, the college admits more than 950 students. Past experience shows that about 75% of the students admitted will accept. The college decides to admit 1200 students. Assuming that students make their decisions independently, the number who accept has the $B(1200, 0.75)$ distribution. If this number is less than 950, the college will admit students from its waiting list.

(a) What are the mean and the standard deviation of the number X of students who accept?

(b) Use the Normal approximation to find the probability that at least 800 students accept.

(c) The college does not want more than 950 students. What is the probability that more than 950 will accept?

(d) If the college decides to increase the number of admission offers to 1300, what is the probability that more than 950 will accept?

5.63 Is the ESP result better than guessing? When the ESP study of Exercise 5.61 discovers a subject whose performance appears to be better than guessing, the study continues at greater length. The experimenter looks at many cards bearing one of five shapes (star, square, circle, wave, and cross) in an order determined by random numbers. The subject cannot see the experimenter as he looks at each card in turn, in order to avoid any possible nonverbal clues. The answers of a subject who does not have ESP should be independent observations, each with probability 1/5 of success. We record 900 attempts.

(a) What are the mean and the standard deviation of the count of successes?

(b) What are the mean and standard deviation of the proportion of successes among the 900 attempts?

(c) What is the probability that a subject without ESP will be successful in at least 24% of 900 attempts?

(d) The researcher considers evidence of ESP to be a proportion of successes so large that there is only probability 0.01 that a subject could do this well or better by guessing. What proportion of successes must a subject have to meet this standard? (Example 1.41 shows how to do an inverse calculation for the Normal distribution that is similar to the type required here.)

5.64 Scuba-diving trips. The mailing list of an agency that markets scuba-diving trips to the Florida Keys contains 65% males and 35% females. The agency calls 30 people chosen at random from its list.

(a) What is the probability that 20 of the 30 are men? (Use the binomial probability formula.)

(b) What is the probability that the first woman is reached on the fourth call? (That is, the first 4 calls give MMMF.)

5.65 Checking for problems with a sample survey. One way of checking the effect of undercoverage, nonresponse, and other sources of error in a sample survey is to compare the sample with known demographic facts about the population. The 2000 census found that 23,772,494 of the 209,128,094 adults (aged 18 and over) in the United States called themselves "Black or African American."

(a) What is the population proportion p of American adults who called themselves black or African American?

(b) An opinion poll chooses 1200 adults at random. What is the mean number of individuals who call themselves black of African American in such samples? (Explain the reasoning behind your calculation.)

(c) Use a Normal approximation to find the probability that such a sample will contain 100 or fewer individuals who call themselves black or African American. Be sure to check that you can safely use the approximation.

5.66 Show that these facts are true. Use the definition of binomial coefficients to show that each of the following facts is true. Then restate each fact in words in terms of the number of ways that k successes can be distributed among n observations.

(a) $\binom{n}{n} = 1$ for any whole number $n \geq 1$.

(b) $\binom{n}{n-1} = n$ for any whole number $n \geq 1$.

(c) $\binom{n}{k} = \binom{n}{n-k}$ for any n and k with $k \leq n$.

5.67 Multiple-choice tests. Here is a simple probability model for multiple-choice tests. Suppose that each student has probability p of correctly answering a question chosen at random from a universe of possible questions. (A strong student has a higher p

than a weak student.) The correctness of an answer to a question is independent of the correctness of answers to other questions. Jodi is a good student for whom $p = 0.88$.

(a) Use the Normal approximation to find the probability that Jodi scores 85% or lower on a 100-question test.

(b) If the test contains 250 questions, what is the probability that Jodi will score 85% or lower?

(c) How many questions must the test contain in order to reduce the standard deviation of Jodi's proportion of correct answers to half its value for a 100-item test?

(d) Laura is a weaker student for whom $p = 0.72$. Does the answer you gave in part (c) for the standard deviation of Jodi's score apply to Laura's standard deviation also?

5.68 Tossing a die. You are tossing a balanced die that has probability 1/6 of coming up 1 on each toss. Tosses are independent. We are interested in how long we must wait to get the first 1.

(a) The probability of a 1 on the first toss is 1/6. What is the probability that the first toss is not a 1 and the second toss is a 1?

(b) What is the probability that the first two tosses are not 1's and the third toss is a 1? This is the probability that the first 1 occurs on the third toss.

(c) Now you see the pattern. What is the probability that the first 1 occurs on the fourth toss? On the fifth toss?

5.69 The geometric distribution. Generalize your work in Exercise 5.68. You have independent trials, each resulting in a success or a failure. The probability of a success is p on each trial. The binomial distribution describes the count of successes in a fixed number of trials. Now the number of trials is not fixed; instead, continue until you get a success. The random variable Y is the number of the trial on which the first success occurs. What are the possible values of Y? What is the probability $P(Y = k)$ for any of these values? (*Comment:* The distribution of the number of trials to the first success is called a **geometric distribution**.)

CHAPTER 5 Exercises

5.70 The cost of Internet access. In Canada, households with Internet spent an average of \$430 annually for access.[17] This distribution is non-Normal: while roughly 61% of the households have high-speed access, many of the remaining households have limited dial-up access and pay much less. Assume the standard deviation is \$140. If you ask an SRS of 500 households with Internet access how much they pay, what is the probability that the average amount will exceed \$440?

5.71 Dust in coal mines. A laboratory weighs filters from a coal mine to measure the amount of dust in the mine atmosphere. Repeated measurements of the weight of dust on the same filter vary Normally with standard deviation $\sigma = 0.08$ milligram (mg) because the weighing is not perfectly precise. The dust on a particular filter actually weighs 123 mg.

(a) The laboratory reports the mean of 3 weighings of this filter. What is the distribution of this mean?

(b) What is the probability that the laboratory reports a weight of 124 mg or higher for this filter?

5.72 The effect of sample size on the standard deviation. Assume that the standard deviation in a very large population is 100.

(a) Calculate the standard deviation for the sample mean for samples of size 1, 4, 25, 100, 250, 500, 1000, and 5000.

(b) Graph your results with the sample size on the x-axis and the standard deviation on the y-axis.

(c) Summarize the relationship between the sample size and the standard deviation that you showed in your graph.

5.73 Carpooling. Although cities encourage carpooling to reduce traffic congestion, most vehicles carry only one person. For example, nationally 75.5% of the people drive to work alone.[18]

(a) If you choose 12 vehicles driving to work at random, what is the probability that more than half (that is, 7 or more) carry just one person?

(b) If you choose 80 vehicles at random, what is the probability that more than half (that is, 41 or more) carry just one person?

5.74 Common last names. The Census Bureau says that the 10 most common names in the United States are (in order) Smith, Johnson, Williams, Jones, Brown, Davis, Miller, Wilson, Moore, and Taylor.[19] These names account for 5.6% of all U.S. residents. Out of curiosity, you look at

5.41. **(a)** Separate flips are independent (coins have no "memory," so they do not try to compensate for a lack of tails). **(b)** Separate flips are independent (coins have no "memory," so they do not get on a "streak" of heads). **(c)** $\hat{p}$ can vary from one set of observed data to another; it is not a parameter.

5.42. **(a)** X is a *count;* $\hat{p}$ is a proportion. **(b)** The given formula is the *standard deviation* for a binomial *proportion*. The variance for a binomial count is $np(1-p)$. **(c)** The rule of thumb in the text is that np and $n(1-p)$ should both be at least 10. If p is close to 0 (or close to 1), $n = 1000$ might not satisfy this rule of thumb. (See also the solution to Exercise 5.22.)

5.43. **(a)** A $B(200, p)$ distribution seems reasonable for this setting (even though we do not know what p is). **(b)** This setting is not binomial; there is no fixed value of n. **(c)** A $B(500, 1/12)$ distribution seems appropriate for this setting. **(d)** This is not binomial, because separate cards are not independent.

5.44. **(a)** This is not binomial; X is not a count of successes. **(b)** A $B(20, p)$ distribution seems reasonable, where p (unknown) is the probability of a defective pair. **(c)** This should be (at least approximately) the $B(n, p)$ distribution, where n is the number of students in our sample, and p is the probability that a randomly-chosen student eats at least five servings of fruits and vegetables.

5.45. **(a)** C, the number caught, is $B(10, 0.7)$. M, the number missed, is $B(10, 0.3)$. **(b)** Referring to Table C, we find $P(M \geq 4) = 0.2001 + 0.1029 + 0.0368 + 0.0090 + 0.0014 + 0.0001 = 0.3503$ (software: 0.3504).

5.46. **(a)** The $B(20, 0.3)$ distribution (at least approximately). **(b)** $P(X \geq 8) = 0.2277$.

5.47. **(a)** The mean of C is $(10)(0.7) = 7$ errors caught; for M the mean is $(10)(0.3) = 3$ errors missed. **(b)** The standard deviation of C (or M) is $\sigma = \sqrt{(10)(0.7)(0.3)} \doteq 1.4491$ errors. **(c)** With $p = 0.9$, $\sigma = \sqrt{(10)(0.9)(0.1)} \doteq 0.9487$ errors; with $p = 0.99$, $\sigma \doteq 0.3146$ errors. σ decreases toward 0 as p approaches 1.

5.48. X, the number who listen to streamed music daily, has the $B(20, 0.25)$ distribution. **(a)** $\mu_X = np = 5$, and $\mu_{\hat{p}} = 0.25$. **(b)** With $n = 200$, $\mu_X = 50$ and $\mu_{\hat{p}} = 0.25$. With $n = 2000$, $\mu_X = 500$ and $\mu_{\hat{p}} = 0.25$. μ_X increases with n, while $\mu_{\hat{p}}$ does not depend on n.

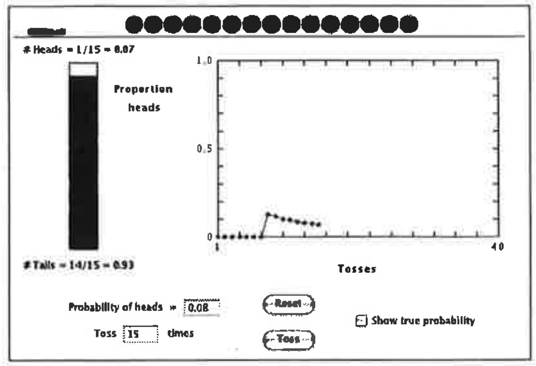
5.49. $m = 6$: $P(X \geq 6) = 0.0473$ and $P(X \geq 5) = 0.1503$.

5.50. **(a)** The population (the 75 members of the fraternity) is only 2.5 times the size of the sample. Our rule of thumb says that this ratio should be at least 20. **(b)** Our rule of thumb for the Normal approximation calls for np and $n(1-p)$ to be at least 10; we have $np = (1000)(0.002) = 2$.

5.51. The count of 5s among n random digits has a binomial distribution with $p = 0.1$. **(a)** $P(\text{at least one 5}) = 1 - P(\text{no 5}) = 1 - (0.9)^5 \doteq 0.4095$. (Or take 0.5905 from Table C and subtract from 1.) **(b)** $\mu = (40)(0.1) = 4$.

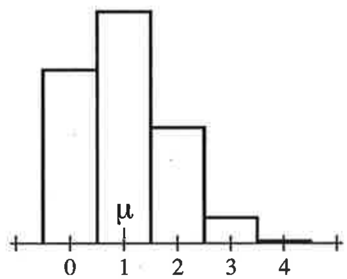
5.52. One sample of 15 flips is shown on the right. Results will vary quite a bit; Table C shows that 99.5% of the time, there will be 4 or fewer bad records in a sample of 15.

Out of 25 samples, most students should see 2 to 12 samples with no bad records. That is, N, the number of samples with no bad records, has the $B(25, 0.2683)$ distribution, and $P(2 \le N \le 12) = 0.9894$.



5.53. **(a)** $n = 4$ and $p = 1/4 = 0.25$. **(b)** The distribution is below; the histogram is on the right. **(c)** $\mu = np = 1$.

x	0	1	2	3	4
$P(X = x)$	.3164	.4219	.2109	.0469	.0039



5.54. For $\hat{p}$, $\mu = 0.52$ and $\sigma = \sqrt{p(1-p)/n} \doteq 0.01574$. As $\hat{p}$ is approximately Normally distributed with this mean and standard deviation, we find:

$$P(0.49 < \hat{p} < 0.55) \doteq P(-1.91 < Z < 1.91) \doteq 0.9438$$

(Software computation of the Normal probability gives 0.9433. Using a binomial distribution, we can also find $P(493 \le X \le 554) \doteq 0.9495$.)

5.55. Recall that $\hat{p}$ is approximately Normally distributed with mean $\mu = p$ and standard deviation $\sqrt{p(1-p)/n}$. **(a)** With $p = 0.24$, $\sigma \doteq 0.01333$, so $P(0.22 < \hat{p} < 0.26) = P(-1.50 < Z < 1.50) \doteq 0.8664$. (Software computation of the Normal probability gives 0.8666. Using a binomial distribution, we can also find $P(226 \le X \le 267) \doteq 0.8752$.) **(b)** With $p = 0.04$, $\sigma \doteq 0.00611$, so $P(0.02 < \hat{p} < 0.06) = P(-3.27 < Z < 3.27) = 0.9990$. (Using a binomial distribution, we can also find $P(21 \le X \le 62) \doteq 0.9992$.) **(c)** $P(-0.02 < \hat{p} - p < 0.02)$ increases to 1 as p gets closer to 0. (This is because σ also gets close to 0, so that $0.02/\sigma$ grows.)

5.56. When $n = 300$, the distribution of $\hat{p}$ is approximately Normal with mean 0.52 and standard deviation 0.02884 (nearly twice that in Exercise 5.54). When $n = 5000$, the standard deviation drops to 0.00707 (less than half as big as in Exercise 5.54). Therefore:

$$n = 300: \quad P(0.49 < \hat{p} < 0.55) \doteq P(-1.04 < Z < 1.04) \doteq 0.7016$$

$$n = 5000: \quad P(0.49 < \hat{p} < 0.55) \doteq P(-4.25 < Z < 4.25) \doteq 1$$

Larger samples give a better probability that $\hat{p}$ will be close to the true proportion p. (Software computation of the first Normal probability gives 0.7017; using a binomial distribution, we can also find $P(147 \le X \le 165) \doteq 0.7278$. These more accurate answers do not change our conclusion.)

5.57. **(a)** The mean is $\mu = p = 0.69$, and the standard deviation is $\sigma = \sqrt{p(1-p)/n} \doteq 0.0008444$. **(b)** $\mu \pm 2\sigma$ gives the range 68.83% to 69.17%. **(c)** This range is considerably narrower than the historical range. In fact, 67% and 70% correspond to $z = -23.7$ and $z = 11.8$—suggesting that the observed percents do not come from a $N(0.69, 0.0008444)$ distribution; that is, the population proportion has changed over time.

5.58. **(a)** $\hat{p} = \frac{294}{400} = 0.735$. **(b)** With $p = 0.8$, $\sigma_{\hat{p}} = \sqrt{(0.8)(0.2)/400} = 0.02$. **(c)** Still assuming that $p = 0.8$, we would expect that about 95% of the time, $\hat{p}$ should fall between 0.76 and 0.84. **(d)** It appears that these students prefer this type of course less than the national average. (The observed value of $\hat{p}$ is quite a bit lower than we would expect from a $N(0.8, 0.2)$ distribution, which suggests that it came from a distribution with a lower mean.)

5.59. **(a)** $\hat{p} = \frac{140}{200} = 0.7$. **(b)** We want $P(X \geq 140)$ or $P(\hat{p} \geq 0.7)$. The first can be found exactly (using a binomial distribution), or we can compute either using a Normal approximation (with or without the continuity correction). All possible answers are shown on the right. **(c)** The sample results are higher than the national percentage, but the sample was so small that such a difference could arise by chance even if the true campus proportion is the same.

			Continuity correction	
Exact prob.	Table Normal	Software Normal	Table Normal	Software Normal
0.2049	0.1841	0.1835	0.2033	0.2041

5.60. As $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$, we have $0.004^2 = (0.52)(0.48)/n$, so $n = 15{,}600$.

5.61. **(a)** $p = 1/4 = 0.25$. **(b)** $P(X \geq 10) = 0.0139$. **(c)** $\mu = np = 5$ and $\sigma = \sqrt{np(1-p)} = \sqrt{3.75} \doteq 1.9365$ successes. **(d)** No: The trials would not be independent because the subject may alter his/her guessing strategy based on this information.

5.62. **(a)** $\mu = (1200)(0.75) = 900$ and $\sigma = \sqrt{225} = 15$ students. **(b)** $P(X \geq 800) \doteq P(Z \geq -6.67) = 1$ (essentially). **(c)** $P(X \geq 951) \doteq P(Z \geq 3.4) = 0.0003$. **(d)** With $n = 1300$, $P(X \geq 951) \doteq P(Z \geq -1.54) = 0.9382$. Other answers are shown in the table on the right.

		Continuity correction	
Table Normal	Software Normal	Table Normal	Software Normal
0.9382	0.9379	0.9418	0.9417

5.63. **(a)** X, the count of successes, has the $B(900, 1/5)$ distribution, with mean $\mu_X = np = (900)(1/5) = 180$ and $\sigma_X = \sqrt{(900)(0.2)(0.8)} = 12$ successes. **(b)** For $\hat{p}$, the mean is $\mu_{\hat{p}} = p = 0.2$ and $\sigma_{\hat{p}} = \sqrt{(0.2)(0.8)/900} \doteq 0.01333$. **(c)** $P(\hat{p} > 0.24) \doteq P(Z > 3) = 0.0013$. **(d)** From a standard Normal distribution, $P(Z > 2.326) = 0.01$, so the subject must score 2.326 standard deviations above the mean: $\mu_{\hat{p}} + 2.326\sigma_{\hat{p}} = 0.2310$. This corresponds to 208 or more successes.

5.64. **(a)** M has the $B(30, 0.65)$ distribution, so $P(M = 20) = \binom{30}{20}(0.65)^{20}(0.35)^{10} \doteq 0.1502$. **(b)** $P(\text{1st woman is the 4th call}) = (0.65)^3(0.35) = 0.0961$.

5.65. **(a)** $p = \frac{23,772,494}{209,128,094} \doteq 0.1137$. **(b)** If B is the number of blacks, then B has (approximately) the $B(1200, 0.1137)$ distribution, so the mean is $np \doteq 136.4$ blacks. **(c)** $P(B \le 100) \doteq P(Z < -3.31) = 0.0005$.

Note: *In (b), the population is at least 20 times as large as the sample, so our "rule of thumb" for using a binomial distribution is satisfied. In fact, the mean would be the same even if we could not use a binomial distribution, but we need to have a binomial distribution for part (c), so that we can approximate it with a Normal distribution—which we can safely do, because both np and n(1 − p) are much greater than 10.*

5.66. **(a)** $\binom{n}{n} = \frac{n!}{n!\,0!} = 1$. The only way to distribute n successes among n observations is for all observations to be successes. **(b)** $\binom{n}{n-1} = \frac{n!}{(n-1)!\,1!} = \frac{n \cdot (n-1)!}{(n-1)!} = n$. To distribute $n-1$ successes among n observations, the one failure must be either observation $1, 2, 3, \ldots, n-1$, or n. **(c)** $\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n!}{(n-k)!\,[n-(n-k)]!} = \binom{n}{n-k}$. Distributing k successes is equivalent to distributing $n-k$ failures.

5.67. Jodi's number of correct answers will have the $B(n, 0.88)$ distribution. **(a)** $P(\hat{p} \le 0.85) = P(X \le 85)$ is on line 1. **(b)** $P(\hat{p} \le 0.85) = P(X \le 212)$ is on line 2. **(c)** For a test with 400 questions, the standard deviation of $\hat{p}$ would be half as big as the standard deviation of $\hat{p}$ for a test with 100 questions: With $n = 100$, $\sigma = \sqrt{(0.88)(0.12)/100} \doteq 0.03250$; and with $n = 400$, $\sigma = \sqrt{(0.88)(0.12)/400} \doteq 0.01625$. **(d)** Yes: Regardless of p, n must be quadrupled to cut the standard deviation in half.

			Continuity correction	
Exact prob.	Table Normal	Software Normal	Table Normal	Software Normal
0.2160	0.1788	0.1780	0.2206	0.2209
0.0755	0.0594	0.0597	0.0721	0.0722

5.68. **(a)** P(first ⚀ appears on toss 2) $= \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \frac{5}{36}$.

(b) P(first ⚀ appears on toss 3) $= \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \frac{25}{216}$.

(c) P(first ⚀ appears on toss 4) $= \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)$.

P(first ⚀ appears on toss 5) $= \left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)$.

5.69. Y has possible values $1, 2, 3, \ldots$. P(first ⚀ appears on toss k) $= \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right)$.

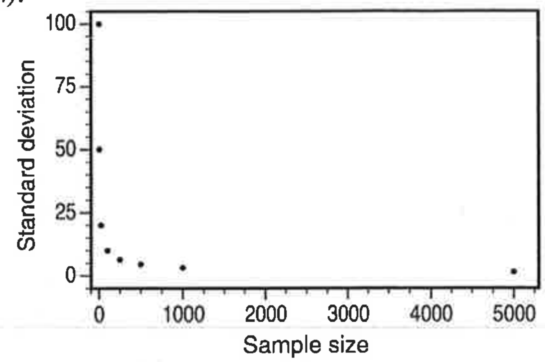
5.70. With $\mu = \$430$, $\sigma = \$140$, and $n = 500$, the distribution of $\bar{x}$ is approximately Normal with mean \$430 and $\sigma_{\bar{x}} = 140/\sqrt{500} \doteq \$6.2610$, so $P(\bar{x} > 440) \doteq P(Z > 1.60) = 0.0548$ (software: 0.0551).

5.71. **(a)** With $\sigma_{\bar{x}} = 0.08/\sqrt{3} \doteq 0.04619$, $\bar{x}$ has (approximately) a $N(123 \text{ mg}, 0.04619 \text{ mg})$ distribution. **(b)** $P(\bar{x} \ge 124) = P(Z \ge 21.65)$, which is essentially 0.

5.72. **(a)** The table of standard deviations is given below. **(b)** The graph is below on the right; it is shown as a scatterplot, but in this situation it would be reasonable to "connect the dots" because the relationship between standard deviation and sample size holds for all n. **(c)** As

n increases, the standard deviation decreases—at first quite rapidly, then more slowly (a demonstration of the law of diminishing returns).

n	$\sigma/\sqrt{n}$
1	100
4	50
25	20
100	10
250	6.32
500	4.47
1000	3.16
5000	1.41



5.73. **(a)** Out of 12 independent vehicles, the number X with one person has the $B(12, 0.755)$ distribution, so $P(X \geq 7) = 0.9503$ (using software or a calculator). **(b)** Y (the number of one-person cars in a sample of 80) has the $B(80, 0.755)$ distribution. Regardless of the approach used—Normal approximation, or exact computation using software or a calculator—$P(Y \geq 41) \doteq 1$.

5.74. This would not be surprising: Assuming that all the authors are independent (for example, none were written by siblings or married couples), we can view the 12 names as being a random sample so that the number N of occurrences of the ten most common names would have a binomial distribution with $n = 12$ and $p = 0.056$. Then $P(N = 0) = (1 - 0.056)^{12} \doteq 0.5008$.

5.75. The probability that the first digit is 1, 2, or 3 is $0.301 + 0.176 + 0.125 = 0.602$, so the number of invoices for amounts beginning with these digits should have a binomial distribution with $n = 1000$ and $p = 0.602$. More usefully, the proportion $\hat{p}$ of such invoices should have approximately a Normal distribution with mean $p = 0.602$ and standard deviation $\sqrt{p(1-p)/1000} \doteq 0.01548$, so $P(\hat{p} \leq \frac{560}{1000}) \doteq P(Z \leq -2.71) = 0.0034$. Alternate answers shown on the right.

		Continuity correction	
Table Normal	Software Normal	Table Normal	Software Normal
0.0034	0.0033	0.0037	0.0037

5.76. **(a)** If R is the number of red-blossomed plants out of a sample of 12, then $P(R = 9) = 0.2581$, using a binomial distribution with $n = 12$ and $p = 0.75$. (For Table C, use $p = 0.25$ and find $P(X = 3)$, where $X = 12 - R$ is the number of flowers with nonred blossoms.) **(b)** With $n = 120$, the mean number of red-blossomed plants is $np = 90$. **(c)** If R_2 is the number of red-blossomed plants out of a sample of 120, then $P(R_2 \geq 80) \doteq P(Z \geq -2.11) = 0.9826$. (Other possible answers are given in the table on the right.)

			Continuity correction	
Exact prob.	Table Normal	Software Normal	Table Normal	Software Normal
0.9845	0.9826	0.9825	0.9864	0.9866

5.77. If $\bar{x}$ is the average weight of 12 eggs, then $\bar{x}$ has a $N(65 \text{ g}, 5/\sqrt{12} \text{ g}) = N(65 \text{ g}, 1.4434 \text{ g})$ distribution, and $P(\frac{755}{12} < \bar{x} < \frac{830}{12}) \doteq P(-1.44 < Z < 2.89) = 0.9231$ (software: 0.9236).

5.78. **(a)** The machine that makes the caps and the machine that applies the torque are not the same. **(b)** T (torque) is $N(7.0, 0.9)$ and S (cap strength) is $N(10.1, 1.2)$, so $T - S$ is $N(7 - 10.1, \sqrt{0.9^2 + 1.2^2}) = N(-3.1 \text{ inch} \cdot \text{lb}, 1.5 \text{ inch} \cdot \text{lb})$. The probability that the cap breaks is $P(T > S) = P(T - S > 0) = P(Z > 2.07) = 0.0192$ (software: 0.0194).

5.79. The center line is $\mu_{\bar{x}} = \mu = 4.25$ and the control limits are $\mu \pm 3\sigma/\sqrt{5} = 4.0689$ to 4.4311.

5.80. **(a)** $\bar{x}$ has a $N(32, 6/\sqrt{25}) = N(32, 1.2)$ distribution, and $\bar{y}$ has a $N(29, 5/\sqrt{25}) = N(29, 1)$ distribution. **(b)** $\bar{y} - \bar{x}$ has a $N(29 - 32, \sqrt{5^2/25 + 6^2/25}) \doteq N(-3, 1.5620)$ distribution. **(c)** $P(\bar{y} \geq \bar{x}) = P(\bar{y} - \bar{x} \geq 0) = P(Z \geq 1.92) = 0.0274$.

5.81. **(a)** $\hat{p}_F$ is approximately $N(0.82, 0.01921)$ and $\hat{p}_M$ is approximately $N(0.88, 0.01625)$. **(b)** When we subtract two independent Normal random variables, the difference is Normal. The new mean is the difference of the two means ($0.88 - 0.82 = 0.06$), and the new variance is the sum of the variances ($0.01921^2 + 0.01625^2 = 0.000633$), so $\hat{p}_M - \hat{p}_F$ is approximately $N(0.06, 0.02516)$. **(c)** $P(\hat{p}_F > \hat{p}_M) = P(\hat{p}_M - \hat{p}_F < 0) \doteq P(Z < -2.38) = 0.0087$ (software: 0.0085).

5.82. **(a)** Yes; this rule works for any random variables X and Y. **(b)** No; this rule requires that X and Y be independent. The incomes of two married people are certainly not independent, as they are likely to be similar in many characteristics that affect income (for example, educational background).

5.83. For each step of the random walk, the mean is $\mu = (1)(0.6) + (-1)(0.4) = 0.2$, the variance is $\sigma^2 = (1 - 0.2)^2(0.6) + (-1 - 0.2)^2(0.4) = 0.96$, and the standard deviation is $\sigma = \sqrt{0.96} \doteq 0.9798$. Therefore, $Y/500$ has approximately a $N(0.2, 0.04382)$ distribution, and $P(Y \geq 200) = P(\frac{Y}{500} \geq 0.4) \doteq P(Z \geq 4.56) \doteq 0$.

Note: *The number R of right-steps has a binomial distribution with $n = 500$ and $p = 0.6$. $Y \geq 200$ is equivalent to taking at least 350 right-steps, so we can also compute this probability as $P(R \geq 350)$, for which software gives the exact value 0.00000215....*