Homework 13
Do at least 4.

SECTION 5.1 Summary

The **sample mean** $\bar{x}$ of an SRS of size n drawn from a large population with mean μ and standard deviation σ has a sampling distribution with mean and standard deviation

$$\mu_{\bar{x}} = \mu$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The sample mean $\bar{x}$ is therefore an unbiased estimator of the population mean μ and is less variable than a single observation.

Linear combinations of independent Normal random variables have Normal distributions. In particular, if the population has a Normal distribution, so does $\bar{x}$.

The **central limit theorem** states that for large n the sampling distribution of $\bar{x}$ is approximately $N(\mu, \sigma/\sqrt{n})$ for any population with mean μ and finite standard deviation σ.

SECTION 5.1 Exercises

For Exercise 5.1, see page 299; for Exercises 5.2 and 5.3, see pag 302; for Exercise 5.4, see page 303; for Exercise 5.5, see page 304; and for Exercise 5.6, see page 306.

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

(a) If the variance of a population is 10, then the variance of the mean for an SRS of 30 observations from this population will be $10/\sqrt{30}$.

(b) When taking SRS's from a population, larger sample sizes will result in larger standard deviations of the sample mean.

(c) The mean of a sampling distribution of $\bar{x}$ changes when the sample size changes.

5.8 What is wrong? Explain what is wrong in each of the following statements.

(a) For large n, the distribution of observed values will be approximately Normal.

(b) The 68–95–99.7 rule says that $\bar{x}$ should be within $\mu \pm 2\sigma$ about 95% of the time.

(c) The central limit theorem states that for large n, μ is approximately Normal.

5.9 Generating a sampling distribution. Let's illustrate the idea of a sampling distribution in the case of a very small sample from a very small population. The population is the 10 scholarship players currently on your men's basketball team. For convenience, the 10 players have been labeled with the integers 0 to 9. For each player, the total amount of time spent (in minutes) on Facebook during the last month is recorded in the table below.

Player	0	1	2	3	4	5	6	7	8	9
Total Time (min)	370	290	358	366	323	319	358	309	327	368

The parameter of interest is the average amount of time on Facebook. The sample is an SRS of size $n = 3$ drawn from this population of players. Because the players are labeled 0 to 9, a single random digit from Table B chooses one player for the sample.

(a) Find the mean of the 10 players in the population. This is the population mean μ.

(b) Use Table B to draw an SRS of size 3 from this population (Note: you may sample the same player's time more than once). Write down the three times in your sample and calculate the sample mean $\bar{x}$. This statistic is an estimate of μ.

(c) Repeat this process 10 times using different parts of Table B. Make a histogram of the 10 values of $\bar{x}$. You are constructing the sampling distribution of $\bar{x}$.

(d) Is the center of your histogram close to μ? Would it get closer to μ the more times you repeated this sampling process? Explain.

5.10 Total sleep time of college students. In Example 5.1, the total sleep time per night among college students was approximately Normally distributed with mean $\mu = 7.02$ hours and standard deviation $\sigma = 1.15$ hours. Suppose you plan to take an SRS of size $n = 200$ and compute the average total sleep time.

(a) What is the standard deviation for the average time?

(b) Use the 95 part of the 68–95–99.7 rule to describe the variability of this sample mean.

(c) What is the probability that your average will be below 6.9 hours?

5.11 Determining sample size. Recall the previous exercise. Suppose you want to use a sample size such that about 95% of the averages fall within ±5 minutes of the true mean $\mu = 7.02$.

(a) Based on your answer to part (b) in Exercise 5.8, should the sample size be larger or smaller than 200? Explain.

(b) What standard deviation of the average do you need such that about 95% of all samples will have a mean within 5 minutes of μ?

(c) Using the standard deviation calculated in part (b), determine the number of students you need to sample.

5.12 Songs on an iPod. An iPod has about 10,000 songs. The distribution of the play time for these songs is highly skewed. Assume that the standard deviation for the population is 280 seconds.

(a) What is the standard deviation of the average time when you take an SRS of 10 songs from this population?

(b) How many songs would you need to sample if you wanted the standard deviation of $\bar{x}$ to be 15 seconds?

5.13 Bottling an energy drink. A bottling company uses a filling machine to fill cans with an energy drink. The cans are supposed to contain 250 milliliters (ml). The machine, however, has some variability, so the standard deviation of the size is $\sigma = 3$ ml. A sample of 6 cans is inspected each hour for process control purposes, and records are kept of the sample mean volume. If the process mean is exactly equal to the target value, what will be the mean and standard deviation of the numbers recorded?

5.14 Play times for songs on an iPod. Averages of several measurements are less variable than individual measurements. Suppose the true mean duration of the play time for the songs in the iPod of Exercise 5.12 is 350 seconds.

(a) Assuming the play times to be Normally distributed, sketch on the same graph the two Normal curves, one for sampling a single song and one for the mean of 10 songs.

(b) What is the probability that the sample mean differs from the population mean by more than 19 seconds when only 1 song is sampled?

(c) How does the probability that you calculated in part (b) change for the mean of an SRS of 10 songs?

5.15 Can volumes. Averages are less variable than individual observations. Suppose that the can volumes in Exercise 5.13 vary according to a Normal distribution. In that case, the mean $\bar{x}$ of an SRS of cans also has a Normal distribution.

(a) Make a sketch of the Normal curve for a single can. Add the Normal curve for the mean of an SRS of 6 cans on the same sketch.

(b) What is the probability that the volume of a single randomly chosen can differs from the target value by 1 ml or more?

(c) What is the probability that the mean volume of an SRS of 6 cans differs from the target value by 1 ml or more?

5.16 Number of friends on Facebook. Facebook provides a variety of statistics on their Web site that detail the growth and popularity of the site.[4] One such statistic is that the average user has 130 friends. This distribution only takes integer values, so it is certainly not Normal. We'll also assume it is skewed to the right with a standard deviation $\sigma = 85$. Consider an SRS of 30 Facebook users.

(a) What are the mean and standard deviation of the total number of friends in this sample?

(b) What are the mean and standard deviation of the mean number of friends per user?

(c) Use the central limit theorem to find the probability that the average number of friends in 30 Facebook users is greater than 140.

5.17 Cholesterol levels of teenagers. A study of the health of teenagers plans to measure the blood cholesterol level of an SRS of 13- to 16-year olds. The researchers will report the mean $\bar{x}$ from their sample as an estimate of the mean cholesterol level μ in this population.

(a) Explain to someone who knows no statistics what it means to say that $\bar{x}$ is an "unbiased" estimator of μ.

(b) The sample result $\bar{x}$ is an unbiased estimator of the population truth μ no matter what size SRS the study chooses. Explain to someone who knows no statistics why a large sample gives more trustworthy results than a small sample.

5.18 ACT scores of high school seniors. The scores of high school seniors on the ACT college entrance examination in a recent year had mean $\mu = 19.2$ and standard deviation $\sigma = 5.1$. The distribution of scores is only roughly Normal.

(a) What is the approximate probability that a single student randomly chosen from all those taking the test scores 23 or higher?

(b) Now take an SRS of 25 students who took the test. What are the mean and standard deviation of the sample mean score $\bar{x}$ of these 25 students?

(c) What is the approximate probability that the mean score $\bar{x}$ of these students is 23 or higher?

(d) Which of your two Normal probability calculations in parts (a) and (c) is more accurate? Why?

5.19 Gypsy moths threaten oak and aspen trees. The gypsy moth is a serious threat to oak and aspen trees. A state agriculture department places traps throughout the state to detect the moths. When traps are checked periodically, the mean number of moths trapped is only 0.5, but some traps have several moths. The distribution of moth counts is discrete and strongly skewed, with standard deviation 0.7.

(a) What are the mean and standard deviation of the average number of moths $\bar{x}$ in 50 traps?

(b) Use the central limit theorem to find the probability that the average number of moths in 50 traps is greater than 0.6.

5.20 Grades in an English course. North Carolina State University posts the grade distributions for its courses online.[5] Students in one section of English 210 in the Fall 2008 semester received 33% A's, 24% B's, 18% C's, 16% D's, and 9% F's.

(a) Using the common scale A = 4, B = 3, C = 2, D = 1, F = 0, take X to be the grade of a randomly chosen English 210 student. Use the definitions of the mean (page 261) and standard deviation (page 269) for discrete random variables to find the mean μ and the standard deviation σ of grades in this course.

(b) English 210 is a large course. We can take the grades of an SRS of 50 students to be independent of each other. If $\bar{x}$ is the average of these 50 grades, what are the mean and standard deviation of $\bar{x}$?

(c) What is the probability $P(X \geq 3)$ that a randomly chosen English 210 student gets a B or better? What is the approximate probability $P(\bar{x} \geq 3)$ that the grade point average for 50 randomly chosen English 210 students is a B or better?

5.21 Diabetes during pregnancy. Sheila's doctor is concerned that she may suffer from gestational diabetes (high blood glucose levels during pregnancy). There is variation in both the actual glucose level and blood test that measures the level. A patient is classified as having gestational diabetes if the glucose level is above 140 milligrams per deciliter (mg/dl) one hour after a sugary drink is ingested. Sheila's measured glucose level one hour after ingesting the sugary drink varies according to the Normal distribution with $\mu = 125$ mg/dl and $\sigma = 10$ mg/dl.

(a) If a single glucose measurement is made, what is the probability that Sheila is diagnosed as having gestational diabetes?

(b) If measurements are made instead on three separate days and the mean result is compared with the criterion 140 mg/dl, what is the probability that Sheila is diagnosed as having gestational diabetes?

5.22 A lottery payoff. A $1 bet in a state lottery's Pick 3 game pays $500 if the three-digit number you choose exactly matches the winning number, which is drawn at random. Here is the distribution of the payoff X:

Payoff X	$0	$500
Probability	0.999	0.001

Each day's drawing is independent of other drawings.

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

(b) Joe buys a Pick 3 ticket twice a week. What does the law of large numbers say about the average payoff Joe receives from his bets?

(c) What does the central limit theorem say about the distribution of Joe's average payoff after 104 bets in a year?

(d) Joe comes out ahead for the year if his average payoff is greater than $1 (the amount he spent each day on a ticket). What is the probability that Joe ends the year ahead?

5.23 Defining a high glucose reading. In Exercise 5.21, Sheila's measured glucose level one hour after ingesting the sugary drink varies according to the Normal distribution with $\mu = 125$ mg/dl and $\sigma = 10$ mg/dl. What is the level L such that there is probability only 0.05 that the mean glucose level of three test results falls above L for Sheila's glucose level distribution?

5.24 Flaws in carpets. The number of flaws per square yard in a type of carpet material varies with mean 1.3 flaws per square yard and standard deviation 1.5 flaws per square yard. This population distribution cannot be Normal, because a count takes only whole-number values. An inspector studies 200 square yards of the material, records the number of flaws found in each square yard, and calculates $\bar{x}$, the mean number of flaws per square yard inspected. Use the central limit theorem to find the approximate probability that the mean number of flaws exceeds 2 per square yard.

5.25 Weights of airline passengers. In response to the increasing weight of airline passengers, the Federal Aviation Administration told airlines to assume that passengers average 190 pounds in the summer, including clothing and carry-on baggage. But passengers vary: the FAA gave a mean but not a standard deviation. A reasonable standard deviation is 35 pounds. Weights are not Normally distributed, especially when the population includes both men and women, but they are not very non-Normal. A commuter plane carries 25 passengers. What is the approximate probability that the total weight of the passengers exceeds 5200 pounds? (*Hint:* To apply the central limit theorem, restate the problem in terms of the mean weight.)

5.26 Risks and insurance. The idea of insurance is that we all face risks that are unlikely but carry high cost. Think of a fire destroying your home. So we form a group to share the risk: we all pay a small amount, and the insurance policy pays a large amount to those few of us whose homes burn down. An insurance company looks at the records for millions of homeowners and sees that the mean loss from fire in a year is $\mu = \$250$ per house and that the standard deviation of the loss is $\sigma = \$1000$. (The distribution of losses is extremely right-skewed: most people have $0 loss, but a few have large losses.) The company plans to sell fire insurance for $250 plus enough to cover its costs and profit.

(a) Explain clearly why it would be unwise to sell only 12 policies. Then explain why selling many thousands of such policies is a safe business.

(b) If the company sells 10,000 policies, what is the approximate probability that the average loss in a year will be greater than $275?

5.27 Treatment of cotton fabrics. "Durable press" cotton fabrics are treated to improve their recovery from wrinkles after washing. Unfortunately, the treatment also reduces the strength of the fabric. The breaking strength of untreated fabric is Normally distributed with mean 57 pounds and standard deviation 2.2 pounds. The same type of fabric after treatment has Normally distributed breaking strength with mean 30 pounds and standard deviation 1.6 pounds.[6] A clothing manufacturer tests 6 specimens of each fabric. All 12 strength measurements are independent.

(a) What is the probability that the mean breaking strength of the 6 untreated specimens exceeds 50 pounds?

(b) What is the probability that the mean breaking strength of the 6 untreated specimens is at least 25 pounds greater than the mean strength of the 6 treated specimens?

5.28 Advertisements and brand image. Many companies place advertisements to improve the image of their brand rather than to promote specific products. In a randomized comparative experiment, business students read ads that cited either the *Wall Street Journal* or the *National Enquirer* for important facts about a fictitious company. The students then rated the trustworthiness of the source on a 7-point scale. Suppose that in the population of all students scores for the *Journal* have mean 4.8 and standard deviation 1.5, while scores for the *Enquirer* have mean 2.4 and standard deviation 1.6.[7]

(a) There are 28 students in each group. Although individual scores are discrete, the mean score for a group of 28 will be close to Normal. Why?

(b) What are the means and standard deviations of the sample mean scores $\overline{y}$ for the *Journal* group and $\overline{x}$ for the *Enquirer* group?

(c) We can take all 56 scores to be independent because students are not told each other's scores. What is the distribution of the difference $\overline{y} - \overline{x}$ between the mean scores in the two groups?

(d) Find $P(\overline{y} - \overline{x} \geq 1)$.

5.29 Treatment and control groups. The two previous exercises illustrate a common setting for statistical inference. This exercise gives the general form of the sampling distribution needed in this setting. We have a sample of n observations from a treatment group and an independent sample of m observations from a control group. Suppose that the response to the treatment has the $N(\mu_X, \sigma_X)$ distribution and that the response of control subjects has the $N(\mu_Y, \sigma_Y)$ distribution. Inference about the difference $\mu_Y - \mu_X$ between the population means is based on the difference $\overline{y} - \overline{x}$ between the sample means in the two groups.

(a) Under the assumptions given, what is the distribution of $\overline{y}$? Of $\overline{x}$?

(b) What is the distribution of $\overline{y} - \overline{x}$?

5.30 Investments in two funds. Linda invests her money in a portfolio that consists of 70% Fidelity 500 Index Fund and 30% Fidelity Diversified International Fund. Suppose that in the long run the annual real return X on the 500 Index Fund has mean 9% and standard deviation 19%, the annual real return Y on the Diversified International Fund has mean 11% and standard deviation 17%, and the correlation between X and Y is 0.6.

(a) The return on Linda's portfolio is $R = 0.7X + 0.3Y$. What are the mean and standard deviation of R?

(b) The distribution of returns is typically roughly symmetric but with more extreme high and low observations than a Normal distribution. The average return over a number of years, however, is close to Normal. If Linda holds her portfolio for 20 years, what is the approximate probability that her average return is less than 5%?

(c) The calculation you just made is not overly helpful, because Linda isn't really concerned about the mean return $\overline{R}$. To see why, suppose that her portfolio returns 12% this year and 6% next year. The mean return for the two years is 9%. If Linda starts with \$1000, how much does she have at the end of the first year? At the end of the second year? How does this amount compare with what she would have if both years had the mean return, 9%? Over 20 years, there may be a large difference between the ordinary mean $\overline{R}$ and the *geometric mean*, which reflects the fact that returns in successive years multiply rather than add.

5.31 Concrete blocks and mortar. You are building a wall from precast concrete blocks. Standard 8-inch blocks are $7\frac{5}{8}$ inches high to allow for a $\frac{3}{8}$-inch layer of mortar under each row of blocks. In practice, the height of a block-plus-mortar row varies according to a Normal distribution with mean 8 inches and standard deviation 0.1 inch. Heights of successive rows are independent. Your wall has four rows of blocks. What is the distribution of the height of the wall? What is the probability that the height differs from the design height of 32 inches by more than half an inch?

Chapter 5 Solutions

5.1. The population is iPhone users (or iPhone users who use the AppsFire service). The statistic is an average of 65 apps per device. Likely values will vary, in part based on how many apps are on student phones (which they might consider "typical").

5.2. With $\mu = 240$, $\sigma = 18$, and $n = 36$, we have mean $\mu_{\bar{x}} = \mu = 240$ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 3$.

5.3. When $n = 144$, the mean is $\mu_{\bar{x}} = \mu = 240$ (unchanged), and the standard deviation is $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.5$. Increasing n does not change $\mu_{\bar{x}}$ but decreases $\sigma_{\bar{x}}$, the variability of the sampling distribution. (In this case, because n was increased by a factor of 4, $\sigma_{\bar{x}}$ was halved.)

5.4. When $n = 144$, $\sigma_{\bar{x}} = \sigma/\sqrt{144} = 18/12 = 1.5$. The sampling distribution of $\bar{x}$ is approximately $N(240, 1.5)$, so about 95% of the time, $\bar{x}$ is between 237 and 243.

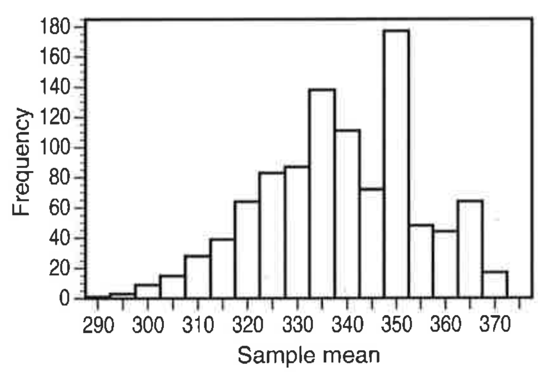
5.5. When $n = 1296$, $\sigma_{\bar{x}} = \sigma/\sqrt{1296} = 18/36 = 0.5$. The sampling distribution of $\bar{x}$ is approximately $N(240, 0.5)$, so about 95% of the time, $\bar{x}$ is between 239 and 241.

5.6. With $\sigma/\sqrt{50} \doteq 3.54$, we have $P(\bar{x} < 28) = P\left(\frac{\bar{x}-25}{3.54} < \frac{28-25}{3.54}\right) \doteq P(Z < 0.85) \doteq 0.8023$.

5.7. (a) Either change "variance" to "standard deviation" (twice), or change the formula at the end to $10^2/30$. **(b)** Standard deviation decreases with increasing sample size. **(c)** $\mu_{\bar{x}}$ always equals μ, regardless of the sample size.

5.8. (a) The distribution of $\bar{x}$ is approximately Normal. (The distribution of observed values—that is, the population distribution—is unaffected by the sample size.) **(b)** $\bar{x}$ is within $\mu \pm 2\sigma/\sqrt{n}$ about 95% of the time. **(c)** The (distribution of the) sample mean $\bar{x}$ is approximately Normal. (μ is not random; it is just a number, albeit typically an unknown one.)

5.9. (a) $\mu = 3388/10 = 338.8$. **(b)** The scores will vary depending on the starting row. The smallest and largest possible means are 290 and 370. **(c)** Answers will vary. Shown on the right is a histogram of the (exact) sampling distribution. With a sample size of only 3, the distribution is noticeably non-Normal. **(d)** The center of the exact sampling distribution is μ, but with only 10 values of $\bar{x}$, this might not be true for student histograms.



Note: *This histograms were found by considering all 1000 possible samples.*

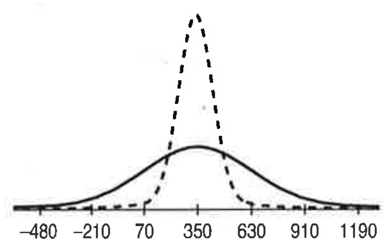
5.10. **(a)** $\sigma_{\bar{x}} = \sigma/\sqrt{200} \doteq 0.08132$. **(b)** With $n = 200$, $\bar{x}$ will be within ± 0.16 (about 10 minutes) of $\mu = 7.02$ hours. **(c)** $P(\bar{x} \le 6.9) = P\left(Z \le \frac{6.9-7.02}{0.08132}\right) \doteq P(Z \le -1.48) \doteq 0.0694$.

5.11. **(a)** With $n = 200$, the 95% probability range was about ± 10 minutes, so need a larger sample size. (Specifically, to halve the range, we need to roughly quadruple the sample size.) **(b)** We need $2\sigma_{\bar{x}} = \frac{5}{60}$, so $\sigma_{\bar{x}} \doteq 0.04167$. **(c)** With $\sigma = 1.15$, we have $\sqrt{n} = \frac{1.15}{0.04167} = 27.6$, so $n = 761.76$—use 762 students.

5.12. **(a)** The standard deviation is $\sigma/\sqrt{10} = 280/\sqrt{10} \doteq 88.5438$ seconds. **(b)** In order to have $\sigma/\sqrt{n} = 15$ seconds, we need $\sqrt{n} = \frac{280}{15}$, so $n \doteq 348.4$—use $n = 349$.

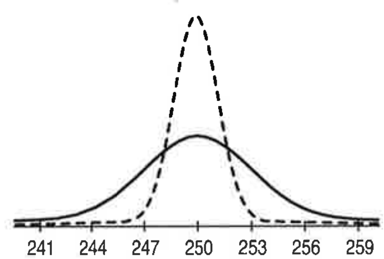
5.13. Mean $\mu = 250$ ml and standard deviation $\sigma/\sqrt{6} = 3/\sqrt{6} \doteq 1.2247$ ml.

5.14. **(a)** For this exercise, bear in mind that the actual distribution for a single song length is definitely *not* Normal; in particular, a Normal distribution with mean 350 seconds and standard deviation 280 seconds extends well below 0 seconds. The Normal curve for $\bar{x}$ should be taller by a factor of $\sqrt{10}$ and skinnier by a factor of $1/\sqrt{10}$ (although that technical detail will likely be lost on most students). **(b)** Using a $N(350, 280)$ distribution, $1 - P(331 < X < 369) \doteq 1 - P(-0.07 < Z < 0.07) \doteq 0.9442$. **(c)** Using a $N(350, 88.5438)$ distribution, $1 - P(331 < X < 369) \doteq 1 - P(-0.21 < Z < 0.21) \doteq 0.8336$.



5.15. In Exercise 5.13, we found that $\sigma_{\bar{x}} \doteq 1.2247$ ml, so $\bar{x}$ has a $N(250 \text{ ml}, 1.2247 \text{ ml})$ distribution. **(a)** On the right. The Normal curve for $\bar{x}$ should be taller by a factor of $\sqrt{6}$ and skinnier by a factor of $1/\sqrt{6}$. **(b)** The probability that a single can's volume differs from the target by at least 1 ml—one-third of a standard deviation—is $1 - P(-0.33 < Z < 0.33) = 0.7414$. **(c)** The probability that $\bar{x}$ is at least 1 ml from the target is

$$1 - P(249 < \bar{x} < 251) = 1 - P(-0.82 < Z < 0.82) = 0.4122.$$



5.16. For the population distribution (the number of friends of a randomly chosen individual), $\mu = 130$ and $\sigma = 85$ friends. **(a)** For the total number of friends for a sample of $n = 30$ users, the mean is $n\mu = 3900$ and the standard deviation is $\sigma\sqrt{n} \doteq 465.56$ friends. **(b)** For the mean number of friends, the mean is $\mu = 130$ and the standard deviation is $\sigma/\sqrt{n} \doteq 15.519$ friends. **(c)** $P(\bar{x} > 140) = P\left(Z > \frac{140-130}{15.519}\right) \doteq P(Z > 0.64) = 0.2611$ (software: 0.2597).

5.17. **(a)** $\bar{x}$ is not systematically higher than or lower than μ; that is, it has no particular tendency to underestimate or overestimate μ. (In other words, it is "just right" on the average.) **(b)** With large samples, $\bar{x}$ is more likely to be close to μ because with a larger sample comes more information (and therefore less uncertainty).

5.18. **(a)** $P(X \geq 23) \doteq P\left(Z \geq \frac{23-19.2}{5.1}\right) = P(Z \geq 0.75) = 0.2266$ (with software: 0.2281). Because ACT scores are reported as whole numbers, we might instead compute $P(X \geq 22.5) \doteq P(Z \geq 0.65) = 0.2578$ (software: 0.2588). **(b)** $\mu_{\bar{x}} = 19.2$ and $\sigma_{\bar{x}} = \sigma/\sqrt{25} = 1.02$. **(c)** $P(\bar{x} \geq 23) \doteq P\left(Z \geq \frac{23-19.2}{1.02}\right) = P(Z \geq 3.73) = 0.0001$. (In this case, it is not appropriate to find $P(\bar{x} \geq 22.5)$, unless $\bar{x}$ is rounded to the nearest whole number.) **(d)** Because individual scores are only roughly Normal, the answer to (a) is approximate. The answer to (c) is also approximate but should be more accurate because $\bar{x}$ should have a distribution that is closer to Normal.

5.19. **(a)** $\mu_{\bar{x}} = 0.5$ and $\sigma_{\bar{x}} = \sigma/\sqrt{50} = 0.7/\sqrt{50} \doteq 0.09899$. **(b)** Because this distribution is only approximately Normal, it would be quite reasonable to use the 68–95–99.7 rule to give a rough estimate: 0.6 is about one standard deviation above the mean, so the probability should be about 0.16 (half of the 32% that falls outside ±1 standard deviation). Alternatively, $P(\bar{x} > 0.6) \doteq P\left(Z > \frac{0.6-0.5}{0.09899}\right) = P(Z > 1.01) = 0.1562$.

5.20. **(a)** $\mu = (4)(0.33) + (3)(0.24) + (2)(0.18) + (1)(0.16) + (0)(0.09) = 2.56$ and

$$\sigma^2 = (4-2.56)^2(0.33) + (3-2.56)^2(0.24) + (2-2.56)^2(0.18) + (1-2.56)^2(0.16) + (0-2.56)^2(0.09) = 1.7664,$$

so $\sigma = \sqrt{1.7664} \doteq 1.3291$. **(b)** $\mu_{\bar{x}} = \mu = 2.56$ and $\sigma_{\bar{x}} = \sigma/\sqrt{50} \doteq 0.1880$. **(c)** $P(X \geq 3) = 0.33 + 0.24 = 0.57$, and $P(\bar{x} \geq 3) \doteq P\left(Z \geq \frac{3-2.56}{0.1880}\right) = P(Z \geq 2.34) = 0.0096$.

5.21. Let X be Sheila's measured glucose level. **(a)** $P(X > 140) = P(Z > 1.5) = 0.0668$. **(b)** If $\bar{x}$ is the mean of three measurements (assumed to be independent), then $\bar{x}$ has a $N(125, 10/\sqrt{3})$ or $N(125 \text{ mg/dl}, 5.7735 \text{ mg/dl})$ distribution, and $P(\bar{x} > 140) = P(Z > 2.60) = 0.0047$.

5.22. **(a)** $\mu_X = (\$500)(0.001) = \$0.50$ and $\sigma_X = \sqrt{249.75} \doteq \$15.8035$. **(b)** In the long run, Joe makes about 50 cents for each \$1 ticket. **(c)** If $\bar{x}$ is Joe's average payoff over a year, then $\mu_{\bar{x}} = \mu = \$0.50$ and $\sigma_{\bar{x}} = \sigma_X/\sqrt{104} \doteq \$1.5497$. The central limit theorem says that $\bar{x}$ is approximately Normally distributed (with this mean and standard deviation). **(d)** Using this Normal approximation, $P(\bar{x} > \$1) \doteq P(Z > 0.32) = 0.3745$ (software: 0.3735).

Note: *Joe comes out ahead if he wins at least once during the year. This probability is easily computed as* $1 - (0.999)^{104} \doteq 0.0988$. *The distribution of* $\bar{x}$ *is different enough from a Normal distribution so that answers given by the approximation are not as accurate in this case as they are in many others.*

5.23. The mean of three measurements has a $N(125 \text{ mg/dl}, 5.7735 \text{ mg/dl})$ distribution, and $P(Z > 1.645) = 0.05$ if Z is $N(0, 1)$, so $L = 125 + 1.645 \cdot 5.7735 \doteq 134.5$ mg/dl.

5.24. $\bar{x}$ is approximately Normal with mean 1.3 and standard deviation $1.5/\sqrt{200} \doteq 0.1061$ flaws/yd^2, so $P(\bar{x} > 2) \doteq P(Z > 6.6) = 0$ (essentially).

5.25. If W is total weight, and $\bar{x} = W/25$, then:

$$P(W > 5200) = P(\bar{x} > 208) \doteq P\left(Z > \tfrac{208-190}{5/\sqrt{25}}\right) = P(Z > 2.57) = 0.0051$$

5.26. (a) Although the probability of having to pay for a total loss for one or more of the 12 policies is very small, if this were to happen, it would be financially disastrous. On the other hand, for thousands of policies, the law of large numbers says that the average claim on many policies will be close to the mean, so the insurance company can be assured that the premiums they collect will (almost certainly) cover the claims. **(b)** The central limit theorem says that, in spite of the skewness of the population distribution, the average loss among 10,000 policies will be approximately Normally distributed with mean \$250 and standard deviation $\sigma/\sqrt{10{,}000} = \$1000/100 = \$10$. Since \$275 is 2.5 standard deviations above the mean, the probability of seeing an average loss over \$275 is about 0.0062.

5.27. (a) The mean of six untreated specimens has a standard deviation of $2.2/\sqrt{6} \doteq 0.8981$ lbs, so $P(\bar{x}_u > 50) = P\left(Z > \frac{50-57}{0.8981}\right) = P(Z > -7.79)$, which is basically 1.
(b) $\bar{x}_u - \bar{x}_t$ has mean $57 - 30 = 27$ lbs and standard deviation $\sqrt{2.2^2/6 + 1.6^2/6} \doteq 1.1106$ lbs, so $P(\bar{x}_u - \bar{x}_t > 25) = P\left(Z > \frac{25-27}{1.1106}\right) = P(Z > -1.80) \doteq 0.9641$.

5.28. (a) The central limit theorem says that the sample means will be roughly Normal. Note that the distribution of individual scores cannot have extreme outliers because all scores are between 1 and 7. **(b)** For *Journal* scores, $\bar{y}$ has mean 4.8 and standard deviation $1.5/\sqrt{28} \doteq 0.2835$. For *Enquirer* scores, $\bar{x}$ has mean 2.4 and standard deviation $1.6/\sqrt{28} \doteq 0.3024$. **(c)** $\bar{y} - \bar{x}$ has (approximately) a Normal distribution with mean 2.4 and standard deviation $\sqrt{1.5^2/28 + 1.6^2/28} \doteq 0.4145$.
(d) $P(\bar{y} - \bar{x} \geq 1) = P\left(Z \geq \frac{1-2.4}{0.4145}\right) = P(Z \geq -3.38) \doteq 0.9996$.

5.29. (a) $\bar{y}$ has a $N(\mu_Y, \sigma_Y/\sqrt{m})$ distribution and $\bar{x}$ has a $N(\mu_X, \sigma_X/\sqrt{n})$ distribution.
(b) $\bar{y} - \bar{x}$ has a Normal distribution with mean $\mu_Y - \mu_X$ and standard deviation $\sqrt{\sigma_Y^2/m + \sigma_X^2/n}$.

5.30. We have been given $\mu_X = 9\%$, $\sigma_X = 19\%$, $\mu_Y = 11\%$, $\sigma_Y = 17\%$, and $\rho = 0.6$.
(a) Linda's return $R = 0.7X + 0.3Y$ has mean $\mu_R = 0.7\mu_X + 0.3\mu_Y = 9.6\%$ and standard deviation $\sigma_R = \sqrt{(0.7\sigma_X)^2 + (0.3\sigma_Y)^2 + 2\rho(0.7\sigma_X)(0.3\sigma_Y)} \doteq 16.8611\%$. **(b)** $\bar{R}$, the average return over 20 years, has approximately a Normal distribution with mean 9.6% and standard deviation $\sigma_R/\sqrt{20} \doteq 3.7703\%$, so $P(\bar{R} < 5\%) \doteq P(Z < -1.22) \doteq 0.1112$. **(c)** After a 12% gain in the first year, Linda would have \$1120; with a 6% gain in the second year, her portfolio would be worth \$1187.20. By contrast, two years with a 9% return would make her portfolio worth \$1188.10.

Note: *As the text suggests, the appropriate average for this situation is (a variation on) the* geometric mean, *computed as* $\sqrt{(1.12)(1.06)} - 1 \doteq 8.9587\%$. *Generally, if the sequence of annual returns is* $r_1, r_2, \ldots, r_k$ *(expressed as decimals), the mean return is*

$\sqrt[k]{(1+r_1)(1+r_2)\cdots(1+r_k)} - 1$. *It can be shown that the geometric mean is always smaller than the arithmetic mean, unless all the returns are the same.*

5.31. The total height H of the four rows has a Normal distribution with mean $4 \times 8 = 32$ inches and standard deviation $0.1\sqrt{4} = 0.2$ inch. $P(H < 31.5 \text{ or } H > 32.5) = 1 - P(31.5 < H < 32.5) = 1 - P(-2.50 < Z < 2.50) = 1 - 0.9876 = 0.0124$.

5.32. $n = 250$ (the sample size), $\hat{p} = 45\% = 0.45$, and $X = n\hat{p} = 112.5$. (Because X must be a whole number, it was either 112 or 113, and the reported value of $\hat{p}$ was rounded.)

5.33. **(a)** $n = 1500$ (the sample size). **(b)** The "Yes" count seems like the most reasonable choice, but either count is defensible. **(c)** $X = 825$ (or $X = 675$). **(d)** $\hat{p} = \frac{825}{1500} = 0.55$ (or $\hat{p} = \frac{675}{1500} = 0.45$).

5.34. Assuming no multiple births (twins, triplets, quadruplets), we have four independent trials, each with probability of success (type O blood) equal to 0.25, so the number of children with type O blood has the $B(4, 0.25)$ distribution.

5.35. We have 15 independent trials, each with probability of success (heads) equal to 0.5, so X has the $B(15, 0.5)$ distribution.

5.36. Assuming each free-throw attempt is an independent trial, X has the $B(10, 0.8)$ distribution, and $P(X \le 4) = 0.0064$.

5.37. **(a)** For the $B(5, 0.4)$ distribution, $P(X = 0) = 0.0778$ and $P(X \ge 3) = 0.3174$. **(b)** For the $B(5, 0.6)$ distribution, $P(X = 5) = 0.0778$ and $P(X \le 2) = 0.3174$. **(c)** The number of "failures" in the $B(5, 0.4)$ distribution has the $B(5, 0.6)$ distribution. With 5 trials, 0 successes is equivalent to 5 failures, and 3 or more successes is equivalent to 2 or fewer failures.

5.38. **(a)** For the $B(100, 0.5)$ distribution, $\mu_{\hat{p}} = p = 0.5$ and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \frac{1}{20} = 0.05$. **(b)** No; the mean and standard deviation of the sample count are both 100 times bigger. (That is, $\hat{p} = X/100$, so $\mu_{\hat{p}} = \mu_X/100$ and $\sigma_{\hat{p}} = \sigma_X/100$.)

5.39. **(a)** $\hat{p}$ has approximately a Normal distribution with mean 0.5 and standard deviation 0.05, so $P(0.3 < \hat{p} < 0.7) = P(-4 < Z < 4) \doteq 1$. **(b)** $P(0.35 < \hat{p} < 0.65) = P(-3 < Z < 3) \doteq 0.9974$.

Note: *For the second, the 68–95–99.7 rule would give 0.997—an acceptable answer, especially since this is an approximation anyway. For comparison, the exact answers (to four decimal places) are* $P(0.3 < \hat{p} < 0.7) \doteq 0.9999$ *or* $P(0.3 \le \hat{p} \le 0.7) \doteq 1.0000$, *and* $P(0.35 < \hat{p} < 0.65) \doteq 0.9965$ *or* $P(0.35 \le \hat{p} \le 0.65) \doteq 0.9982$. *(Notice that the "correct" answer depends on our understanding of "between.")*

5.40. **(a)** $P(X \ge 3) = \binom{4}{3}0.53^3 0.47 + \binom{4}{4}0.53^4 \doteq 0.3588$. **(b)** If the coin were fair, $P(X \ge 3) \doteq \binom{4}{3}0.5^3 0.5 + \binom{4}{4}0.5^4 = 0.3125$.