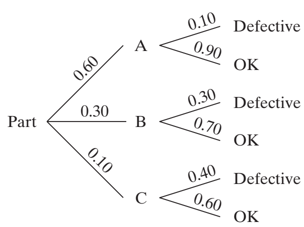
Answers to Chapter 5 AP® Statistics Practice Test

T5.1 e
T5.2 d
T5.3 c
T5.4 b
T5.5 b
T5.6 c
T5.7 e
T5.8 e
T5.9 b
T5.10 c
T5.11 **(a)** Here is a completed table, with T indicating that the teacher wins and Y indicating that you win. $P(\text{teacher wins}) = \dfrac{27}{48} = 0.5625$.

	1	2	3	4	5	6	7	8
1	—	T	T	T	T	T	T	T
2	Y	—	T	T	T	T	T	T
3	Y	Y	—	T	T	T	T	T
4	Y	Y	Y	—	T	T	T	T
5	Y	Y	Y	Y	—	T	T	T
6	Y	Y	Y	Y	Y	—	T	T

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \dfrac{27}{48} + \dfrac{8}{48} - \dfrac{5}{48} = \dfrac{30}{48}$.
(c) Not independent. $P(A) = \dfrac{27}{48} = 0.5625$ does not equal $P(A \mid B) = \dfrac{5}{8} = 0.625$.

T5.12 **(a)**



(b) $P(\text{defective}) = (0.60)(0.10) + (0.30)(0.30) + (0.10)(0.40) = 0.19$. **(c)** Machine B. $P(A \mid \text{defective}) = \dfrac{0.06}{0.19} = 0.3158$. $P(B \mid \text{defective}) = \dfrac{0.09}{0.19} = 0.4737$. $P(C \mid \text{defective}) = \dfrac{0.04}{0.19} = 0.2105$.

T5.13 **(a)** Here is a two-way table that summarizes this information:

	Smokes	Does not smoke	Total
Cancer	0.08	0.04	**0.12**
No cancer	0.17	0.71	**0.88**
Total	**0.25**	**0.75**	**1.00**

$P(\text{gets cancer} \mid \text{smoker}) = \dfrac{0.08}{0.25} = 0.32$.

(b) $P(\text{smokes} \cup \text{gets cancer}) = 0.25 + 0.12 - 0.08 = 0.29$.

(c) P(cancer) = 0.12, so $P(\text{at least one gets cancer}) = 1 - P(\text{neither gets cancer}) = 1 - 0.88^2 = 0.2256$

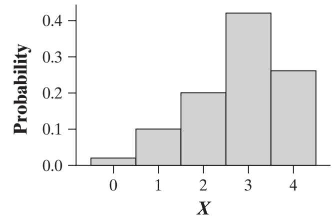
T5.14 **(a)** Let 00–16 represent out-of-state and 17–99 represent in-state. Read two-digit numbers until you have found two numbers between 00 and 16. Record how many 2-digit numbers you had to read. **(b)** The first sample is 41 **05 09** (it took three cars). The second sample is 20 31 **06** 44 90 50 59 59 88 43 18 80 53 **11** (it took 14 cars). The third sample is 58 44 69 94 86 85 79 67 **05** 81 18 45 **14** (it took 13 cars).

Chapter 6

Section 6.1

Answers to Check Your Understanding

page 350: **1.** $P(X \geq 3)$ is the probability that the student got either an A or a B. $P(X \geq 3) = 0.68$.
2. $P(X < 2) = 0.02 + 0.10 = 0.12$
3. The histogram below is skewed to the left. Higher grades are more likely, but there are a few lower grades.



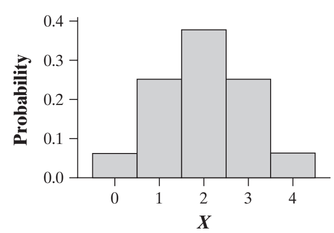
page 355: **1.** $\mu_X = 1.1$. If many, many Fridays are randomly selected, the average number of cars sold will be about 1.1. **2.** $\sigma_X = \sqrt{0.89} = 0.943$. The number of cars sold on a randomly selected Friday will typically vary from the mean (1.1) by about 0.943 cars.

Answers to Odd-Numbered Section 6.1 Exercises

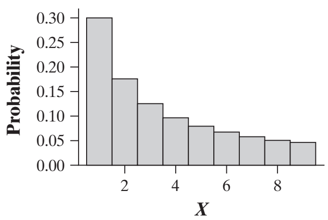
6.1 **(a)**

Value	0	1	2	3	4
Probability	1/16	4/16	6/16	4/16	1/16

(b) The histogram below shows that this distribution is symmetric with a center at 2.



(c) $P(X \leq 3) = 15/16 = 0.9375$. There is a 0.9375 probability that you will get three or fewer heads in 4 tosses of a fair coin.
6.3 **(a)** $P(X \geq 1) = 0.9$. **(b)** The event $X \leq 2$ is "at most two nonword errors." $P(X \leq 2) = 0.6$. $P(X < 2) = 0.3$.
6.5 **(a)** All of the probabilities are between 0 and 1 and they sum to 1. **(b)** The histogram below is unimodal and skewed to the right.



(c) The event $X \geq 6$ is the event that "the first digit in a randomly chosen record is a 6 or higher." $P(X \geq 6) = 0.222$. **(d)** $P(X \leq 5) = 0.778$.

6.7 **(a)** The outcomes that make up the event A are 7, 8, and 9. $P(A) = 0.155$. **(b)** The outcomes that make up the event B are 1, 3, 5, 7, and 9. $P(B) = 0.609$. **(c)** The outcomes that make up the event "A or B" are 1, 3, 5, 7, 8, and 9. $P(A \text{ or } B) = 0.660$. This is not the same as $P(A) + P(B)$ because the outcomes 7 and 9 are included in both events.

6.9 **(a)**

X	−\$1	\$2
Probability	0.75	0.25

(b) $E(X) = -\$0.25$. If the player makes many \$1 bets, he will lose about \$0.25 per \$1 bet, on average.

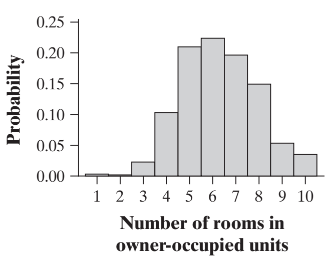
6.11 $\mu_X = 2.1$. If many, many undergraduates performed this task, they would make about 2.1 nonword errors, on average.

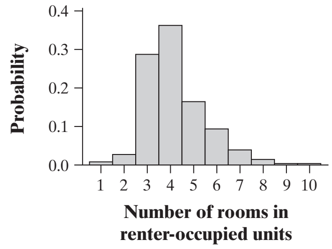
6.13 **(a)** This distribution is symmetric and 5 is located at the center. **(b)** According to Benford's law, $E(X) = 3.441$. To detect a fake expense report, compute the sample mean of the first digits. A value closer to 5 suggests a fake report and a value near 3.441 is consistent with a truthful report. **(c)** $P(Y > 6) = 3/9 = 0.333$. Under Benford's law, $P(X > 6) = 0.155$. To detect a fake expense report, compute the proportion of first digits that begin with 7, 8, or 9. A value closer to 0.333 suggests a fake report and a value closer to 0.155 is consistent with a truthful report.

6.15 $\sigma_X = \sqrt{1.29} = 1.1358$. The number of nonword errors in a randomly selected essay will typically differ from the mean (2.1) by about 1.14 words.

6.17 **(a)** $\sigma_Y = \sqrt{6.667} = 2.58$. **(b)** $\sigma_X = \sqrt{6.0605} = 2.4618$. This would not be the best way to tell the difference between a fake and a real expense report because the standard deviations are similar.

6.19 **(a)** See the following histograms. The distribution of the number of rooms is roughly symmetric for owners and skewed to the right for renters. Renter-occupied units tend to have fewer rooms than owner-occupied units. There is more variability in the number of rooms for owner-occupied units.





(b) Owner: $\mu_X = 6.284$ rooms. Renter: $\mu_Y = 4.187$ rooms. Single people and younger people are more likely to rent and need less space than people with families. **(c)** $\sigma_X = \sqrt{2.68934} = 1.6399$. The number of rooms in a randomly selected owner-occupied unit will typically differ from the mean (6.284) by about 1.6399 rooms. $\sigma_Y = \sqrt{1.71003} = 1.3077$. The number of rooms in a randomly selected renter-occupied unit will typically differ from the mean (4.187) by about 1.3077 rooms.

6.21 **(a)** $P(X > 0.49) = 0.51$. **(b)** $P(X \geq 0.49) = 0.51$.
(c) $P(0.19 \leq X < 0.37 \text{ or } 0.84 < X \leq 1.27) = 0.18 + 0.16 = 0.34$

6.23 The time Y of a randomly chosen student has the $N(7.11, 0.74)$ distribution. We want to find $P(Y < 6)$. $z = \dfrac{6 - 7.11}{0.74} = -1.50$ and $P(Z > -1.50) = 0.0668$. *Using technology:* `normalcdf(lower:−1000,upper:6,µ:7.11,σ:0.74)` $= 0.0668$. There is about a 7% chance that this student will run the mile in under 6 minutes.

6.25 **(a)** The speed Y of a randomly chosen serve has the $N(115, 6)$ distribution. We want to find $P(Y > 120)$. $z = \dfrac{120 - 115}{6} = 0.83$ and $P(Z > 0.83) = 0.2033$. *Using technology:* `normalcdf(lower:120,upper:1000,µ:115,σ:6)` $= 0.2023$. There is a 0.2023 probability of selecting a serve that is greater than 120 mph. **(b)** The line above 120 has no area, so $P(Y \geq 120) = P(Y > 120) = 0.2023$. **(c)** We want to find c such that $P(Y \leq c) = 0.15$. Solving $-1.04 = \dfrac{c - 115}{6}$ gives $c = 108.76$. *Using technology:* `invNorm(area:0.15,µ:115,σ:6)` $=$ 108.78. Fifteen percent of Nadal's serves will be less than or equal to 108.78 mph.

6.27 b

6.29 c

6.31 Yes. The mean difference (post − pre) was 5.38 and the median difference was 3. This means that at least half of the students improved their reading scores.

6.33 predicted post-test = 17.897 + 0.78301(pretest).

Section 6.2

Answers to Check Your Understanding

page 367:
1. $Y = 500X$. $\mu_Y = 500(1.1) = \$550$. $\sigma_Y = 500(0.943) = \$471.50$.
2. $T = Y - 75$. $\mu_T = 550 - 75 = \$475$. $\sigma_T = \$471.50$.

page 376: **1.** $\mu_T = 1.1 + 0.7 = 1.8$. Over many Fridays, this dealership sells or leases about 1.8 cars in the first hour of business, on average.
2. $\sigma_T^2 = (0.943)^2 + (0.64)^2 = 1.2988$, so $\sigma_T = \sqrt{1.2988} = 1.14$.
3. $\mu_B = 500(1.1) + 300(0.7) = \$760$. $\sigma_B^2 = (500)^2(0.943)^2 + (300)^2(0.64)^2 = 259{,}176.25$, so $\sigma_B = \sqrt{259{,}176.25} = \$509.09$.

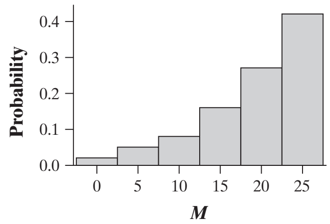
page 378: **1.** $\mu_D = 1.1 - 0.7 = 0.4$. Over many Fridays, this dealership sells about 0.4 cars more than it leases during the first hour of business, on average.

2. $\sigma_D^2 = (0.943)^2 + (0.64)^2 = 1.2998$, so $\sigma_D = \sqrt{1.2998} = 1.14$. 3. $\mu_B = 500(1.1) - 300(0.7) = \$340$. $\sigma_B^2 = (500)^2(0.943)^2 + (300)^2(0.64)^2 = 259{,}176.25$, so $\sigma_B = \sqrt{259{,}176.25} = \$509.09$.

Answers to Odd-Numbered Section 6.2 Exercises

6.35 $\mu_Y = 2.54(1.2) = 3.048$ cm and $\sigma_Y = 2.54(0.25) = 0.635$ cm.

6.37 **(a)** The distribution shown below is skewed to the left. Most of the time, the ferry makes \$20 or \$25.



(b) $\mu_M = \$19.35$. If many ferry trips were selected at random, the ferry would collect about \$19.35 per trip, on average. **(c)** $\sigma_M = \$6.45$. The amounts collected on randomly selected ferry trips will typically vary by about \$6.45 from the mean (\$19.35).

6.39 **(a)** $\mu_G = 5(7.6) + 50 = 88$. **(b)** $\sigma_G = 5(1.32) = 6.6$. **(c)** $\sigma_G^2 = (5\sigma_X)^2 = 25\sigma_X^2$. The variance of G is 25 times the variance of X.

6.41 **(a)** $\mu_Y = -\$0.65$. If many ferry trips were selected at random, the ferry would lose about \$0.65 per trip, on average. **(b)** $\sigma_Y = \$6.45$. The amount of profit on randomly selected ferry trips will typically vary by about \$6.45 from the mean ($-\$0.65$).

6.43 $\mu_Y = 6(3.87) - 20 = \$3.22$. $\sigma_Y = 6(1.29) = \$7.74$.

6.45 **(a)** $\mu_Y = 47.3$°F. $\sigma_Y = 4.05$ °F. **(b)** Y has the N(47.3, 4.05) distribution. We want to find $P(Y < 40)$. $z = \dfrac{40 - 47.3}{4.05} = -1.80$ and $P(Z < -1.80) = 0.0359$. *Using technology:* `normalcdf(lower:–1000,upper:40,µ:47.3,σ:4.05)` = 0.0357. There is a 0.0357 probability that the midnight temperature in the cabin is below 40°F.

6.47 **(a)** Yes. The mean of a sum is always equal to the sum of the means. **(b)** No, because it is not reasonable to assume that X and Y are independent.

6.49 $\mu_{Y_1+Y_2} = (-0.65) + (-0.65) = -\$1.30$. $\sigma_{Y_1+Y_2}^2 = 6.45^2 + 6.45^2 = 83.205$, so $\sigma_{Y_1+Y_2} = \sqrt{83.205} = \$9.12$.

6.51 $\mu_{3X} = 3(2.1) = 6.3$ and $\sigma_{3X} = 3(1.136) = 3.408$. $\mu_{2Y} = 2(1.0) = 2.0$ and $\sigma_{2Y} = 2(1.0) = 2.0$. Thus, $\mu_{3X+2Y} = 6.3 + 2.0 = 8.3$ and $\sigma_{3X+2Y}^2 = 3.408^2 + 2.0^2 = 15.6145$, so $\sigma_{3X+2Y} = \sqrt{15.6145} = 3.95$.

6.53 **(a)** $\mu_{Y-X} = 1.0 - 2.1 = -1.1$. If you were to select many essays, there would be about 1.1 fewer word errors than nonword errors, on average. $\sigma_{Y-X}^2 = (1.0)^2 + (1.136)^2 = 2.2905$, so $\sigma_{Y-X} = \sqrt{2.2905} = 1.51$. The difference in the number errors will typically vary by about 1.51 from the mean (-1.1). **(b)** The outcomes that make up this event are $1 - 0 = 1$, $2 - 0 = 2$, $2 - 1 = 1$, $3 - 0 = 3$, $3 - 1 = 2$, $3 - 2 = 1$. There is a 0.15 probability that a randomly chosen student will have more word errors than nonword errors.

6.55 The difference in score deductions for a randomly selected essay is $3X - 2Y$. $\mu_{3X} = 3(2.1) = 6.3$ and $\sigma_{3X} = 3(1.136) = 3.408$. $\mu_{2Y} = 2(1.0) = 2.0$ and $\sigma_{2Y} = 2(1.0) = 2.0$. Thus, $\mu_{3X-2Y} = 6.3 - 2.0 = 4.3$ and $\sigma_{3X-2Y}^2 = 3.408^2 + 2.0^2 = 15.6145$, so $\sigma_{3X-2Y} = \sqrt{15.6145} = 3.95$.

6.57 $\mu_{X_1+X_2} = 303.35 + 303.35 = \$606.70$ and $\sigma_{X_1+X_2}^2 = 9707.57^2 + 9707.57^2 = 188{,}473{,}830.6$, so $\sigma_{X_1+X_2} = \sqrt{188{,}473{,}830.6} = \$13{,}728.58$. $W = \frac{1}{2}(X_1 + X_2)$, so $\mu_W = \frac{1}{2}(606.70) = \$303.35$ and $\sigma_W = \frac{1}{2}(13{,}728.58) = \$6864.29$.

6.59 **(a)** Normal with mean = 11 + 20 = 31 seconds and standard deviation = $\sqrt{2^2 + 4^2} = 4.4721$ seconds. **(b)** We want to find the probability that the total time is less than 30 seconds. $z = \dfrac{30 - 31}{4.4721} = -0.22$ and $P(Z < -0.22) = 0.4129$. *Using technology:* `normalcdf(lower:–1000,upper:30,µ:31,σ:4.4721)` = 0.4115. There is a 0.4115 probability of completing the process in less than 30 seconds for a randomly selected part.

6.61 Let T = the total team swim time. $\mu_T = 55.2 + 58.0 + 56.3 + 54.7 = 224.2$ seconds and $\sigma_T^2 = (2.8)^2 + (3.0)^2 + (2.6)^2 + (2.7)^2 = 30.89$, so $\sigma_T = \sqrt{30.89} = 5.56$ seconds. Thus, T has the $N(224.2, 5.56)$ distribution. We want to find $P(T < 220)$. $z = \dfrac{220 - 224.2}{5.56} = -0.76$ and $P(Z < -0.76) = 0.2236$. *Using technology:* `normalcdf(lower:–1000,upper:220,µ:224.2,σ:5.56)` = 0.2250. There is a 0.2250 probability that the total team time is less than 220 seconds in a randomly selected race.

6.63 Let $D = X_1 - X_2$ = the difference in NOX levels. $\mu_D = 1.4 - 1.4 = 0$ and $\sigma_{X_1-X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 = 0.3^2 + 0.3^2 = 0.18$, so $\sigma_{X_1-X_2} = \sqrt{0.18} = 0.4243$. Thus, D has the $N(0, 0.4243)$ distribution. We want to find $P(D > 0.8 \text{ or } D < -0.8) = P(D > 0.8) + P(D < -0.8)$. $z = \dfrac{0.8 - 0}{0.4243} = 1.89$ and $z = \dfrac{-0.8 - 0}{0.4243} = -1.89$ and $P(Z < -1.89 \text{ or } Z > 1.89) = 0.0588$. *Using technology:* `1 – normalcdf(lower:–0.8,upper:0.8,µ:0,σ:0.4243)` = 0.0594. There is a 0.0594 probability that the difference is at least as large as the attendant observed.

6.65 c

6.67 **(a)** Fidelity Technology Fund, because its correlation is larger. **(b)** No, the correlation doesn't tell us anything about the values of the variables, only about the strength of the linear relationship between them.

Section 6.3

Answers to Check Your Understanding

page 389: **1.** Binomial. Binary? "Success" = get an ace. "Failure" = don't get an ace. Independent? Because you are replacing the card in the deck and shuffling each time, the result of one trial does not tell you anything about the outcome of any other trial. Number? $n = 10$. Success? The probability of success is $p = 4/52$ for each trial. **2.** Not binomial. Binary? "Success" = over 6 feet. "Failure" = not over 6 feet. Independent? Because we are selecting without replacement from a small number of students, the observations are not independent. Number? $n = 3$. Success? The probability of success will not change from trial to trial. **3.** Not binomial. Binary? "Success" = roll a 5. "Failure" = don't roll a 5. Independent? Because you are rolling a die, the outcome of any one trial does not tell you anything about the outcome of any other trial. Number? $n = 100$. Success? No. The probability of success changes when the corner of the die is chipped off.

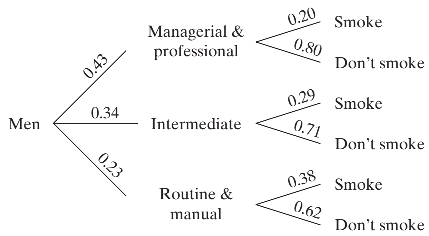
page 397: **1.** Binary? "Success" = question answered correctly. "Failure" = question not answered correctly. Independent? The computer randomly assigned correct answers to the questions, so knowing the result of one trial (question) should not tell you anything about the result on any other trial. Number? $n = 10$. Success? The probability of success is $p = 0.20$ for each trial.

2. $P(X = 3) = \binom{10}{3}(0.2)^3(0.8)^7 = 0.2013$. There is a 20% chance that Patti will answer exactly 3 questions correctly.
3. $P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.9936 = 0.0064$. There is only a 0.0064 probability that a student would get 6 or more correct, so we would be quite surprised if Patti was able to pass.
page 400: 1. $\mu_X = 10(0.20) = 2$. If many students took the quiz, we would expect students to get about 2 answers correct, on average. 2. $\sigma_X = \sqrt{10(0.20)(0.80)} = 1.265$. If many students took the quiz, we would expect individual students' scores to typically vary from the mean of 2 correct answers by about 1.265 correct answers.
3. $P(X > 2 + 2(1.265)) = P(X > 4.53) = 1 - P(X \leq 4) = 1 - 0.9672 = 0.0328$.
page 408: 1. Die rolls are independent, the probability of getting doubles is the same on each roll (1/6), and we are repeating the chance process until we get a success (doubles).
2. $P(T = 3) = \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) = 0.1157$. There is a 0.1157 probability that you will get the first set of doubles on the third roll of the dice. 3. $P(T \leq 3) = \frac{1}{6} + \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) = 0.4213$.

Answers to Odd-Numbered Section 6.3 Exercises

6.69 Binomial. Binary? "Success" = seed germinates and "Failure" = seed does not germinate. Independent? Yes, because the seeds were randomly selected, knowing the outcome of one seed shouldn't tell us anything about the outcomes of other seeds. Number? $n = 20$ seeds. Success? $p = 0.85$.
6.71 Not binomial. Binary? "Success" = person is left-handed and "Failure" = person is right-handed. Independent? Because students are selected randomly, their handedness is independent. Number? There is not a fixed number of trials for this chance process because you continue until you find a left-handed student. Success? $p = 0.10$.
6.73 (a) Binomial. Binary? "Success" = reaching a live person and "Failure" = any other outcome. Independent? Knowing whether or not one call was completed tells us nothing about the outcome on any other call. Number? $n = 15$. Success? $p = 0.2$. (b) This is not a binomial setting because there are not a fixed number of attempts. The Binary, Independent, and Success conditions are satisfied, however, as in part (a).
6.75 $P(X = 4) = \binom{7}{4}(0.44)^4(0.56)^3 = 0.2304$. There is a 0.2304 probability that exactly 4 of the 7 elk survive to adulthood.
6.77 $P(X > 4) = \binom{7}{5}(0.44)^5(0.56)^2 + \cdots = 0.1402$. Because this probability isn't very small, it is not surprising for more than 4 elk to survive to adulthood.
6.79 (a) $P(X = 17) = \binom{20}{17}(0.85)^{17}(0.15)^3 = 0.2428$.
(b) $P(X \leq 12) = \binom{20}{0}(0.85)^0(0.15)^{20} + \cdots + \binom{20}{12}(0.85)^{12}(0.15)^8 = 0.0059$. Because this is such a low probability, Judy should be suspicious.
6.81 (a) $\mu_X = 15(0.20) = 3$. If we watched the machine make many sets of 15 calls, we would expect about 3 calls to reach a live person, on average. (b) $\sigma_X = \sqrt{15(0.20)(0.80)} = 1.55$. If we watched the machine make many sets of 15 calls, we would expect the number of calls that reach a live person to typically vary by about 1.55 from the mean (3).
6.83 (a) $\mu_Y = 15(0.80) = 12$. Notice that $\mu_X = 3$ and $12 + 3 = 15$ (the total number of calls). (b) $\sigma_Y = \sqrt{15(0.80)(0.20)} = 1.55$. This is the same value as σ_X, because $Y = 15 - X$ and adding a constant to a random variable doesn't change the spread.
6.85 (a) Binary? "Success" = win a prize and "Failure" = don't win a prize. Independent? Knowing whether one bottle wins or not should not tell us anything about the caps on other bottles. Number? $n = 7$. Success? $p = 1/16$. (b) $\mu_X = 1.167$. If we were to buy many sets of 7 bottles, we would get 1.167 winners per set, on average. $\sigma_X = 0.986$. If we were to buy many sets of 7 bottles, the number of winning bottles would typically differ from the mean (1.167) by 0.986. (c) $P(X \geq 3) = 1 - P(X \leq 2) = 0.0958$. Because 0.0958 isn't a very small probability, the clerk shouldn't be surprised. It is plausible to get 3 or more winners in a sample of 7 bottles by chance alone.
6.87 No. Because we are sampling without replacement and the sample size (10) is more than 10% of the population size (76), we should not treat the observations as independent.
6.89 If the sample is a small fraction of the population (less than 10%), the make-up of the population doesn't change enough to make the lack of independent trials an issue.
6.91 (a) Binary? "Success" = visit an auction site at least once a month and "Failure" = don't visit an auction site at least once a month. Independent? We are sampling without replacement, but the sample size (500) is far less than 10% of all males aged 18 to 34. Number? $n = 500$. Success? $p = 0.50$. (b) $np = 250$ and $n(1 - p) = 250$ are both at least 10. (c) $\mu_X = 250$ and $\sigma_X = 11.18$. Thus, X has approximately the $N(250, 11.18)$ distribution. We want to find $P(X \geq 235)$. $z = \frac{235 - 250}{11.18} = -1.34$ and $P(Z \geq -1.34) = 0.9099$ *Using technology:* `normalcdf(lower:235, upper:1000,µ:250,σ:11.18)` = 0.9102. There is a 0.9102 probability that at least 235 of the men in the sample visit an online auction site.
6.93 Let X be the number of 1s and 2s. Then X has a binomial distribution with $n = 90$ and $p = 0.477$ (in the absence of fraud). $P(X \leq 29) = 0.0021$. Because the probability of getting 29 or fewer invoices that begin with the digits 1 or 2 is quite small, we have reason to be suspicious that the invoice amounts are not genuine.
6.95 (a) Not geometric. We can't classify the possible outcomes on each trial (card) as "success" or "failure" and we are not selecting cards until we get a *single* success. (b) Games of 4-Spot Keno are independent, the probability of winning is the same in each game ($p = 0.259$), and Lola is repeating a chance process until she gets a success. X = number of games needed to win once is a geometric random variable with $p = 0.259$.
6.97 (a) Let X = the number of bottles Alan purchases to find one winner. $P(X = 5) = (5/6)^4(1/6) = 0.0804$.
(b) $P(X \leq 8) = (1/6) + \cdots + (5/6)^7(1/6) = 0.7674$.
6.99 (a) $\mu_X = \frac{1}{0.097} = 10.31$.
(b) $P(X \geq 40) = 1 - P(X \leq 39) = 0.0187$. Because the probability of not getting an 8 or 9 before the 40th invoice is small, we may begin to worry that the invoice amounts are fraudulent.
6.101 b

6.103 d
6.105 c
6.107 (a)



$P(\text{smoke}) = 0.43(0.20) + 0.34(0.29) + 0.23(0.38) = 0.272 = 27.2\%$

(b) $P(\text{routine and manual} \mid \text{smoke}) = \dfrac{(0.23)(0.38)}{0.272} = 0.321 = 32.1\%$

Answers to Chapter 6 Review Exercises

R6.1 (a) $P(X = 5) = 1 - 0.1 - 0.2 - 0.3 - 0.3 = 0.1$. (b) Discrete, because it takes a fixed set of values with gaps in between. (c) $P(X \le 2) = 0.3$. $P(X < 2) = 0.1$. These are not the same because the outcome $X = 2$ is included in the first calculation but not the second. (d) $\mu_X = 1(0.1) + \cdots + 5(0.1) = 3.1$. $\sigma_X^2 = (1 - 3.1)^2(0.1) + \cdots + (5 - 3.1)^2(0.1) = 1.29$, so $\sigma_X = \sqrt{1.29} = 1.136$.

R6.2 (a) Temperature is a continuous random variable because it takes all values in an interval of numbers—there are no gaps between possible temperatures. (b) $P(X < 540) = P(X \le 540)$ because X is a continuous random variable. In this case, $P(X = 540) = 0$ because the line segment above $X = 540$ has no area. (c) Mean $= 550 - 550 = 0°$C. The standard deviation stays the same, 5.7°C, because subtracting a constant does not change the variability. (d) In degrees Fahrenheit, the mean is $\mu_Y = \dfrac{9}{5}(550) + 32 = 1022°$F and the standard deviation is $\sigma_Y = \left(\dfrac{9}{5}\right)(5.7) = 10.26°$F.

R6.3 (a) If you were to play many games of 4-Spot Keno, you would get a payout of about \$0.70 per game, on average. If you were to play many games of 4-Spot Keno, the payout amounts would typically vary by about \$6.58 from the mean (\$0.70). (b) Let Y be the amount of Jerry's payout. $\mu_Y = 5(0.70) = \$3.50$ and $\sigma_Y = 5(6.58) = \$32.90$. (c) Let W be the amount of Marla's payout. $\mu_W = 0.70 + 0.70 + 0.70 + 0.70 + 0.70 = \$3.50$ and $\sigma_W^2 = 6.58^2 + 6.58^2 + 6.58^2 + 6.58^2 + 6.58^2 = 216.482$, so $\sigma_W = \sqrt{216.482} = \$14.71$. (d) Even though their expected values are the same, the casino would probably prefer Marla since there is less variability in her strategy and her winnings are more predictable.

R6.4 (a) C follows a N(10, 1.2) distribution and we want to find $P(C > 11)$. $z = \dfrac{11 - 10}{1.2} = 0.83$ and $P(Z > 0.83) = 0.2033$. *Using technology:* `normalcdf(lower:11,upper:1000,µ:10,σ:1.2)` = 0.2023. There is a 0.2023 probability that a randomly selected cap has a strength greater than 11 inch-pounds. (b) The machine that makes the caps and the machine that applies the torque are not the same. (c) $C - T$ is Normal with mean $10 - 7 = 3$ inch-pounds and standard deviation $\sqrt{0.9^2 + 1.2^2} = 1.5$ inch-pounds. (d) We want to find $P(C - T < 0)$. $z = \dfrac{0 - 3}{1.5} = -2$ and $P(Z < -2) = 0.0228$. *Using technology:* `normalcdf(lower:−1000,upper:0,µ:3,σ:1.5)` = 0.0228. There is a 0.0228 probability that a randomly selected cap will break when being fastened by the machine.

R6.5 (a) Binary? "Success" = orange and "Failure" = not orange. Independent? The sample of size $n = 8$ is less than 10% of the large bag, so we can assume the outcomes of trials are independent. Number? $n = 8$. Success? $p = 0.20$. (b) $\mu_X = 8(0.2) = 1.6$. If we were to select many samples of size 8, we would expect to get about 1.6 orange M&M'S, on average. (c) $\sigma_X = \sqrt{8(0.2)(0.8)} = 1.13$. If we were to select many samples of size 8, the number of orange M&M'S would typically vary by about 1.13 from the mean (1.6).

R6.6 (a) $P(X = 0) = \binom{8}{0}(0.2)^0(0.8)^8 = 0.1678$. Because the probability is not that small, it would not be surprising to get no orange M&M'S in a sample of size 8. (b) $P(X \ge 5) = \binom{8}{5}(0.2)^5(0.80)^3 + \cdots = 0.0104$ Because the probability is small, it would be surprising to find 5 or more orange M&M'S in a sample of size 8.

R6.7 Let Y be the number of spins to get a "wasabi bomb." Y is a geometric random variable with $p = \dfrac{3}{12} = 0.25$. $P(Y \le 3) = (0.75)^2(0.25) + (0.75)(0.25) + 0.25 = 0.5781$.

R6.8 (a) Let X be the number of heads in 10,000 tosses. $\mu_X = 10{,}000(0.5) = 5{,}000$ and $\sigma_X = \sqrt{10{,}000(0.5)(0.5)} = 50$. (b) $np = 10{,}000(0.5) = 5{,}000$ and $n(1 - p) = 10{,}000(0.5) = 5000$ are both at least 10. (c) We want to find $P(X \le 4933 \text{ or } X \ge 5067)$. $z = \dfrac{4933 - 5000}{50} = -1.34$ and $z = \dfrac{5067 - 5000}{50} = 1.34$ and $P(Z \le -1.34) + P(Z \ge 1.34) = 0.1802$. *Using technology:* `1 − normalcdf(lower:4933,upper:5067,µ:5000,σ:50)` = 0.1802. Because this probability isn't small, we don't have convincing evidence that Kerrich's coin was unbalanced—a difference this far from 5000 could be due to chance alone.

Answers to Chapter 6 AP® Statistics Practice Test

T6.1 b
T6.2 d
T6.3 d
T6.4 e
T6.5 d
T6.6 b
T6.7 c
T6.8 b
T6.9 b
T6.10 c

T6.11 (a) $P(Y \le 2) = 0.96$. (b) $\mu_Y = 0(0.78) + \cdots = 0.38$. If we were to randomly select many cartons of eggs, we would expect about 0.38 to be broken, on average. (c) $\sigma_Y^2 = (0 - 0.38)^2(0.78) + \ldots = 0.6756$. So $\sigma_Y = \sqrt{0.6756} = 0.8219$. If we were to randomly select many cartons of eggs, the number of broken eggs would typically vary by about 0.6756 from the mean (0.38). (d) Let X stand for the number of cartons inspected to find one carton with at least 2 broken eggs. X is a geometric random variable with $p = 0.11$. $P(X \le 3) = (0.11) + (0.89)(0.11) + (0.89)^2(0.11) = 0.2950$.

T6.12 **(a)** Binary? "Success" = dog first and "Failure" = not dog first. Independent? We are sampling without replacement, but 12 is less than 10% of all dog owners. Number? $n = 12$. Success? $p = 0.66$. **(b)** $P(X \le 4) = \binom{12}{0}(0.66)^0(0.34)^{12} + \cdots + \binom{12}{4}(0.66)^4(0.34)^8 = 0.0213$. Because this probability is small, it is unlikely to have only 4 or fewer owners greet their dogs first by chance alone. This gives convincing evidence that the claim by the *Ladies Home Journal* is incorrect.

T6.13 **(a)** $\mu_D = 50 - 25 = 25$ minutes, $\sigma_D^2 = 100 + 25 = 125$, and $\sigma_D = \sqrt{125} = 11.18$ minutes. **(b)** D follows a $N(25, 11.18)$ distribution and we want to find $P(D < 0)$. $z = \frac{0 - 25}{11.18} = -2.24$ and $P(Z < -2.24) = 0.0125$. *Using technology:* `normalcdf (lower:−1000,upper:0,μ:25,σ:11.18)` = 0.0127. There is a 0.0127 probability that Ed spent longer on his assignment than Adelaide did on hers.

T6.14 **(a)** Let X stand for the number of Hispanics in the sample. $\mu_X = 1200(0.13) = 156$ and $\sigma_X = \sqrt{1200(0.13)(0.87)} = 11.6499$. **(b)** 15% of 1200 is 180, so we want to find $P(X \ge 180) = \binom{1200}{180}(0.13)^{180}(0.87)^{1020} + \cdots = 0.0235$. Because this probability is small, it is unlikely to select 180 or more Hispanics in the sample just by chance. This gives us reason to be suspicious about the sampling process.

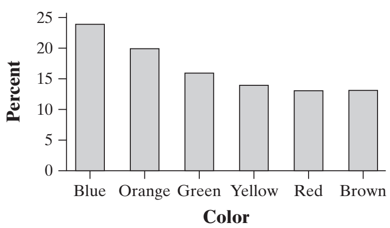
Chapter 7

Section 7.1

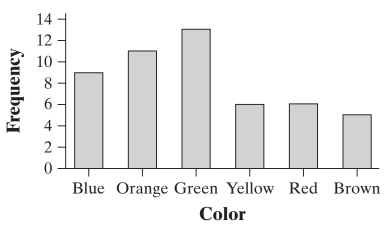
Answers to Check Your Understanding

page 425: **1.** Parameter: $\mu = 20$ ounces. Statistic: $\bar{x} = 19.6$ ounces. **2.** Parameter: $p = 0.10$, or 10% of passengers. Statistic: $\hat{p} = 0.08$, or 8% of the sample of passengers.

page 428: **1.** Individuals: M&M'S Milk Chocolate Candies; variable: color; and parameter of interest: proportion of orange M&M'S. The graph below shows the population distribution.



2. The graph below shows a possible distribution of sample data. For this sample there are 11 orange M&M'S, so $\hat{p} = \frac{11}{50} = 0.22$.



3. The middle graph is the approximate sampling distribution of $\hat{p}$ because the center of the distribution should be at approximately 0.20. The first graph shows the distribution of the colors for one sample and the third graph is centered at 0.40 rather than 0.20.

page 434: **1.** No. The mean of the approximate sampling distribution of the sample median (73.5) is not equal to the median of the population (75). **2.** Smaller. Larger samples provide more precise estimates because larger samples include more information about the population distribution. **3.** Skewed to the left and unimodal.

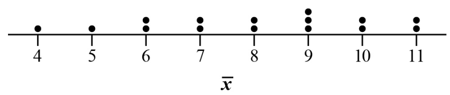
Answers to Odd-Numbered Section 7.1 Exercises

7.1 **(a)** *Population*: all people who signed a card saying that they intend to quit smoking. *Parameter*: the proportion of the population who actually quit smoking. *Sample*: a random sample of 1000 people who signed the cards. *Statistic*: the proportion of the sample who actually quit smoking; $\hat{p} = 0.21$. **(b)** *Population*: all the turkey meat. *Parameter*: minimum temperature in all of the turkey meat. *Sample*: four randomly chosen locations in the turkey. *Statistic*: minimum temperature in the sample of four locations; sample minimum = 170°F.

7.3 $\mu = 2.5003$ is a parameter and $\bar{x} = 2.5009$ is a statistic.

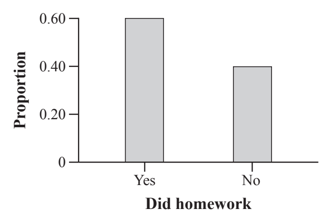
7.5 $\hat{p} = 0.48$ is a statistic and $p = 0.52$ is a parameter.

7.7 **(a)** 2 and 6 ($\bar{x} = 4$), 2 and 8 (5), 2 and 10 (6), 2 and 10 (6), 2 and 12 (7), 6 and 8 (7), 6 and 10 (8), 6 and 10 (8), 6 and 12 (9), 8 and 10 (9), 8 and 10 (9), 8 and 12 (10), 10 and 10 (10), 10 and 12 (11), 10 and 12 (11). **(b)** The sampling distribution of $\bar{x}$ is skewed to the left and unimodal. The mean of the sampling distribution is 8, which is equal to the mean of the population. The values of $\bar{x}$ vary from 4 to 11.

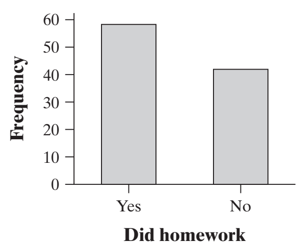


7.9 **(a)** In one simulated SRS of 100 students, there were 73 students who did all their assigned homework. **(b)** The distribution is reasonably symmetric and bell-shaped. It is centered at about 0.60. Values vary from about 0.47 to 0.74. There don't appear to be any outliers. **(c)** Yes, because there were no values of $\hat{p}$ less than or equal to 0.45 in the simulation. **(d)** Because it would be very surprising to get a sample proportion of 0.45 or less in an SRS of size 100 when $p = 0.60$, we should be skeptical of the newspaper's claim.

7.11 **(a)** A graph of the population distribution is shown below.



(b) Answers will vary. An example bar graph is given.



7.13 **(a)** Skewed to the right with a center at $9(°F)^2$. The values vary from about 2 to $27.5(°F)^2$. **(b)** A sample variance of $25(°F)^2$ provides convincing evidence that the manufacturer's claim is false and that the thermostat actually has more variability than claimed because a value this large was rare in the simulation.
7.15 If we chose many SRSs and calculated the sample mean $\bar{x}$ for each sample, we will not consistently underestimate μ or consistently overestimate μ.
7.17 A larger random sample will provide more information and, therefore, more precise results.
7.19 **(a)** Statistics ii and iii, because the means of their sampling distributions appear to be equal to the population parameter. **(b)** Statistic ii, because it is unbiased and has very little variability.
7.21 c
7.23 a
7.25 **(a)** We are looking for the percentage of values that are 2.5 standard deviations or farther below the mean in a Normal distribution. In other words, we are looking for $P(Z \le -2.5)$. Using Table A, $P(Z \le -2.5) = 0.0062$. *Using technology*: `normalcdf(lower:-1000,upper:-2.50,µ:0,σ:1)` = 0.0062. Less than 1% of healthy young adults have osteoporosis. **(b)** Let X be the BMD for women aged 70 − 79 on the standard scale. Then X follows a N(−2, 1) distribution and we want to find $P(X \le -2.5)$. $z = \frac{-2.5-(-2)}{1} = -0.5$ and $P(Z \le -0.5) = 0.3085$. *Using technology*: `normalcdf(lower:-1000,upper:-2.5,µ:-2,σ:1)` = 0.3085. About 31% of women aged 70 − 79 have osteoporosis.

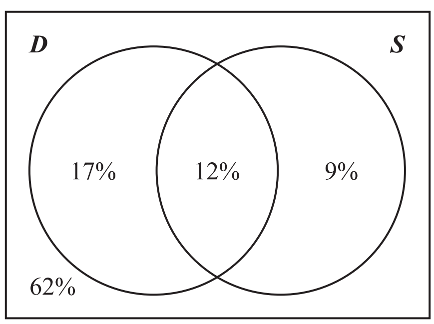
Section 7.2

Answers to Check Your Understanding

page 445: **1.** $\mu_{\hat{p}} = p = 0.75$. **2.** The standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.75(0.25)}{1000}} = 0.0137$. There are more than 10(1000) = 10,000 young adult Internet users, so the 10% condition has been met. **3.** Yes. Both $np = 1000(0.75) = 750$ and $n(1-p) = 1000(0.25) = 250$ are at least 10. **4.** The sampling distribution would still be approximately Normal with mean 0.75. However, the standard deviation would be smaller by a factor of 3: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.75(0.25)}{9000}} = 0.0046$.

Answers to Odd-Numbered Section 7.2 Exercises

7.27 **(a)** We would not be surprised to find 8 (32%) orange candies because values this small happened fairly often in the simulation. However, there were few samples in which there were 5 (20%) or fewer orange candies. So getting 5 orange candies would be surprising. **(b)** A sample of 50, because we expect to be closer to $p = 0.45$ in larger samples.
7.29 **(a)** $\mu_{\hat{p}} = p = 0.45$. **(b)** $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.45(0.55)}{25}} = 0.0995$. The 10% condition is met because there are more than 10(25) = 250 candies in the large machine. **(c)** Yes, because $np = 25(0.45) = 11.25$ and $n(1-p) = 25(0.55) = 13.75$ are both at least 10. **(d)** The sampling distribution would still be approximately Normal with a mean of $\mu_{\hat{p}} = 0.45$. However, the standard deviation decreases to $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.45(0.55)}{100}} = 0.0497$.
7.31 **(a)** No, because more than 10% of the population (10/76 = 13%) was selected. **(b)** No, because the sample size was only $n = 10$. Neither np nor $n(1-p)$ will be at least 10.
7.33 The Large Counts condition is not met because $np = 15(0.3) = 4.5 < 10$.
7.35 **(a)** $\mu_{\hat{p}} = p = 0.70$. **(b)** $\sigma_{\hat{p}} = \sqrt{\frac{0.7(0.3)}{1012}} = 0.0144$. The 10% condition is met because the sample of size 1012 is less than 10% of the population of all U.S. adults. **(c)** Yes, because $np = 1012(0.70) = 708.4$ and $n(1-p) = 1012(0.30) = 303.6$ are both at least 10. **(d)** We want to find $P(\hat{p} \le 0.67)$. $z = \frac{0.67-0.70}{0.0144} = -2.08$ and $P(Z \le -2.08) = 0.0188$. *Using technology*: `normalcdf(lower:-1000,upper:0.67,µ:0.70,σ:0.0144)` = 0.0186. There is a 0.0186 probability of obtaining a sample in which 67% or fewer say they drink the milk. Because this is a small probability, there is convincing evidence against the claim.
7.37 4048, because using $4n$ for the sample size halves the standard deviation ($\sqrt{4n} = 2\sqrt{n}$).
7.39 $\mu_{\hat{p}} = 0.70$. Because 267 is less than 10% of the population of college women, $\sigma_{\hat{p}} = \sqrt{\frac{0.7(0.3)}{267}} = 0.0280$. Because $np = 267(0.7) = 186.9$ and $n(1-p) = 267(0.3) = 80.1$ are both at least 10, the sampling distribution of $\hat{p}$ can be approximated by a Normal distribution. We want to find $P(\hat{p} \ge 0.75)$. $z = \frac{0.75-0.7}{0.0280} = 1.79$ and $P(Z \ge 1.79) = 0.0367$. *Using technology*: `normalcdf(lower:0.75,upper:1000,µ:0.7,σ:0.0280)` = 0.0371. There is a 0.0371 probability that 75% or more of the women in the sample have been on a diet within the last 12 months.
7.41 **(a)** $\mu_{\hat{p}} = 0.90$. Because 100 is less than 10% of the population of orders, $\sigma_{\hat{p}} = \sqrt{\frac{0.90(0.10)}{100}} = 0.03$. Because $np = 100(0.90) = 90$ and $n(1-p) = 100(0.10) = 10$ are both at least 10, the sampling distribution of $\hat{p}$ can be approximated by a Normal distribution. We want to find $P(\hat{p} \le 0.86)$. $z = \frac{0.86-0.90}{0.03} = -1.33$ and $P(Z \le -1.33) = 0.0918$. *Using technology*: `normalcdf(lower:-1000,upper:0.86,µ:0.90,σ:0.03)` = 0.0912. There is a 0.0912 probability that 86% or fewer of orders in an SRS of 100 were shipped within 3 working days. **(b)** Because the probability isn't very small, it is plausible that the 90% claim is correct and that the lower than expected percentage is due to chance alone.
7.43 a
7.45 b
7.47 The Venn diagram is shown below.



62% neither download nor share music files.

Section 7.3

Answers to Check Your Understanding

page 456: **1.** X = length of pregnancy follows a N(266, 16) distribution and we want to find $P(X > 270)$. $z = \frac{270 - 266}{16} = 0.25$ and $P(Z > 0.25) = 0.4013$. *Using technology*: `normalcdf(lower:270,upper:1000,µ:266,σ:16)` = 0.4013. There is a 0.4013 probability of selecting a woman whose pregnancy lasts for more than 270 days. **2.** $\mu_{\bar{x}} = \mu = 266$ days **3.** The sample of size 6 is less than 10% of all pregnant women, so $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{6}} = 6.532$ days. **4.** $\bar{x}$ follows a N(266, 6.532) distribution and we want to find $P(\bar{x} > 270)$. $z = \frac{270 - 266}{6.532} = 0.61$ and $P(Z > 0.61) = 0.2709$. *Using technology*: `normalcdf(lower:270,upper:1000,µ:266,σ:6.532)` = 0.2701. There is a 0.2701 probability of selecting a sample of 6 women whose mean pregnancy length exceeds 270 days.

Answers to Odd-Numbered Section 7.3 Exercises

7.49 $\mu_{\bar{x}} = \mu = 255$ seconds. Because the sample size (10) is less than 10% of the population of songs on David's iPod, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{10}} = 18.974$ seconds.

7.51 $30 = \frac{60}{\sqrt{n}} \rightarrow \sqrt{n} = \frac{60}{30} = 2 \rightarrow n = 4$.

7.53 **(a)** Normal with $\mu_{\bar{x}} = \mu = 188$ mg/dl. Because the sample size (100) is less than 10% of all men aged 20 to 34, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{41}{\sqrt{100}} = 4.1$ mg/dl. **(b)** We want to find $P(185 \le \bar{x} \le 191)$. $z = \frac{185 - 188}{4.1} = -0.73$ and $z = \frac{191 - 188}{4.1} = 0.73$
$P(-0.73 \le Z \le 0.73) = 0.5346$. *Using technology*: `normalcdf(lower:185,upper:191,µ:188,σ:4.1)` = 0.5357. There is a 0.5357 probability that $\bar{x}$ estimates μ within ±3 mg/dl.
(c) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{41}{\sqrt{1000}} = 1.30$ mg/dl. So $\bar{x}$ follows a N(188, 1.30) distribution and we want to find $P(185 \le \bar{x} \le 191)$. $z = \frac{185 - 188}{1.30} = -2.31$ and $z = \frac{191 - 188}{1.30} = 2.31$. $P(-2.31 \le Z \le 2.31) = 0.9792$. *Using technology*: `normalcdf(lower:185,upper:191,µ:188,σ:1.30)` = 0.9790. There is a 0.9790 probability that $\bar{x}$ estimates μ within ±3 mg/dl. The larger sample is better because it is more likely to produce a sample mean within 3 mg/dl of the population mean.

7.55 **(a)** Let X = amount of cola in a randomly selected bottle. X follows the N(298, 3) distribution and we want to find $P(X < 295)$. $z = \frac{295 - 298}{3} = -1$ and $P(Z < -1) = 0.1587$. *Using technology*: `normalcdf(lower:−1000,upper:295,µ:298,σ:3)` = 0.1587. There is a 0.1587 probability that a randomly selected bottle contains less than 295 ml. **(b)** $\mu_{\bar{x}} = \mu = 298$ ml. Because 6 is less than 10% of all bottles produced, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{6}} = 1.2247$ ml. We want to find $P(\bar{x} < 295)$ using the N(298, 1.2247) distribution. $z = \frac{295 - 298}{1.2247} = -2.45$ and $P(Z < -2.45) = 0.0071$. *Using technology*: `normalcdf(lower:−1000,upper:295,µ:298,σ:1.2247)` = 0.0072. There is a 0.0072 probability that the mean contents of six randomly selected bottles are less than 295 ml.

7.57 No. The histogram of the sample values will look like the population distribution. The CLT says that the histogram of the sampling distribution of the *sample mean* will look more and more Normal as the sample size increases.

7.59 **(a)** Because the distribution of the play times of the population of songs is heavily skewed to the right and $n = 10 < 30$. **(b)** Because $n = 36 \ge 30$, the CLT applies. $\mu_{\bar{x}} = \mu = 225$ seconds. Because 36 is less than 10% of all songs on David's iPod, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{36}} = 10$ seconds. We want to find $P(\bar{x} > 240)$ using the N(225, 10) distribution. $z = \frac{240 - 225}{10} = 1.50$ and $P(Z > 1.50) = 0.0668$. *Using technology*: `normalcdf(lower:240,upper:1000,µ:225,σ:10)` = 0.0668. There is a 0.0668 probability that the mean play time is more than 240 seconds.

7.61 **(a)** We do not know the shape of the distribution of passenger weights. **(b)** We want to find $P(\bar{x} > 6000/30) = P(\bar{x} > 200)$. Because the sample size is large ($n = 30 \ge 30$), the distribution of $\bar{x}$ is approximately Normal with $\mu_{\bar{x}} = \mu = 190$ pounds. Because $n = 30$ is less than 10% of all possible passengers, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{30}} = 6.3901$ pounds. $z = \frac{200 - 190}{6.3901} = 1.56$ and $P(Z > 1.56) = 0.0594$. *Using technology*: `normalcdf(lower:200,upper:1000,µ:190,σ:6.3901)` = 0.0588. There is a 0.0588 probability that the mean weight exceeds 200 pounds.

7.63 Because the sample size is large ($n = 10{,}000 \ge 30$), the sampling distribution of $\bar{x}$ is approximately Normal. $\mu_{\bar{x}} = \mu = \$250$. Assuming 10,000 is less than 10% of all homeowners with fire insurance, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1000}{\sqrt{10{,}000}} = \$10$. We want to find $P(\bar{x} \le 275)$ using the N(250, 10) distribution. $z = \frac{275 - 250}{10} = 2.50$ and $P(Z \le 2.50) = 0.9938$. *Using technology*: `normalcdf(lower:−1000,upper:275,µ:250,σ:10)` = 0.9938. There is a 0.9938 probability that the mean annual loss from a sample of 10,000 policies is no greater than $275.

7.65 b

7.67 b

7.69 Didn't finish high school: $\frac{1062}{12{,}470} = 0.0852$; high school but no college: $\frac{1977}{37{,}834} = 0.0523$, less than a bachelor's degree: $\frac{1462}{34{,}439} = 0.0425$, college graduate: $\frac{1097}{40{,}390} = 0.0272$. The unemployment rate decreases with additional education.

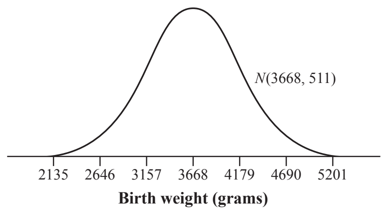
7.71 $P(\text{in labor force} \mid \text{college graduate}) = \frac{40{,}390}{51{,}582} = 0.7830$.

Answers to Chapter 7 Review Exercises

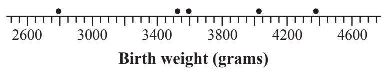
R7.1 The population is the set of all eggs shipped in one day. The sample consists of the 200 eggs examined. The parameter is the proportion $p = 0.03$ of eggs shipped that day that had salmonella.

The statistic is the proportion $\hat{p} = \frac{9}{200} = 0.045$ of eggs in the sample that had salmonella.

R7.2 **(a)** A sketch of the population distribution is given below.



(b) Answers will vary. An example dotplot is given. **(c)** The dot at 2750 represents one SRS of size 5 from this population where the sample range was 2750 grams.



R7.3 **(a)** No, because sample range is always less than the actual range (3417). If it were unbiased, the distribution would be centered at 3417. **(b)** Take larger samples.

R7.4 **(a)** $\mu_{\hat{p}} = p = 0.15$. **(b)** Because the sample size of $n = 1540$ is less than 10% of the population of all adults, $\sigma_{\hat{p}} = \sqrt{\frac{0.15(0.85)}{1540}} = 0.0091$. **(c)** Yes, because $np = 1540(0.15) = 231$ and $n(1-p) = 1540(0.85) = 1309$ are both at least 10. **(d)** We want to find $P(0.13 \le \hat{p} \le 0.17)$. $z = \frac{0.13 - 0.15}{0.0091} = -2.20$ and $z = \frac{0.17 - 0.15}{0.0091} = 2.20$. The desired probability is $P(-2.20 \le Z \le 2.20) = 0.9722$. *Using technology*: `normalcdf(lower:0.13,upper:0.17,µ:0.15,σ:0.0091)` $= 0.9720$. There is a 0.9720 probability of obtaining a sample in which between 13% and 17% are joggers.

R7.5 **(a)** $\mu_{\hat{p}} = p = 0.30$. Because 100 is less than 10% of the population of travelers, $\sigma_{\hat{p}} = \sqrt{\frac{0.30(0.70)}{100}} = 0.0458$. Because $np = 100(0.30) = 30$ and $n(1-p) = 100(0.70) = 70$ are both at least 10, the sampling distribution of $\hat{p}$ can be approximated by a Normal distribution. We want to find $P(\hat{p} \le 0.20)$. $z = \frac{0.20 - 0.30}{0.0458} = -2.18$ and $P(Z \le -2.18) = 0.0146$. *Using technology*: `normalcdf(lower:–1000,upper:0.20,µ:0.30,σ:0.0458)` $= 0.0145$. There is a 0.0145 probability that 20% or fewer of the travelers get a red light. **(b)** Because this is a small probability, there is convincing evidence against the agents' claim—it isn't plausible to get a sample proportion of travelers with a red light this small by chance alone.

R7.6 **(a)** X = WAIS score for a randomly selected individual follows a N(100, 15) distribution and we want to find $P(X \ge 105)$. $z = \frac{105 - 100}{15} = 0.33$ and $P(Z \ge 0.33) = 0.3707$. *Using technology*: `normalcdf(lower:105,upper:1000,µ:100,σ:15)` $= 0.3694$. There is a 0.3694 probability of selecting an individual with a WAIS score of at least 105. **(b)** $\mu_{\bar{x}} = \mu = 100$. Because the sample of size 60 is less than 10% of all adults, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{60}} = 1.9365$. **(c)** $\bar{x}$ follows a N(100, 1.9365) distribution and we want to find $P(\bar{x} \ge 105)$. $z = \frac{105 - 100}{1.9365} = 2.58$ and $P(Z \ge 2.58) = 0.0049$. *Using technology*: `normalcdf(lower:105,upper:1000,µ:100,σ:1.9365)` $= 0.0049$. There is a 0.0049 probability of selecting a sample of 60 adults whose mean WAIS score is at least 105. **(d)** The answer to part (a) could be quite different depending on the shape of the population distribution. The answer to part (b) would be the same because the mean and standard deviation do not depend on the shape of the population distribution. Because of the large sample size ($60 \ge 30$), the answer for part (c) would still be fairly reliable due to the central limit theorem.

R7.7 **(a)** Because $n = 50 \ge 30$. **(b)** $\mu_{\bar{x}} = \mu = 0.5$. Because 50 is less than 10% of all traps, the standard deviation is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.7}{\sqrt{50}} = 0.0990$. Thus, $\bar{x}$ follows a N(0.5, 0.0990) distribution and we want to find $P(\bar{x} \ge 0.6)$. $z = \frac{0.6 - 0.5}{0.0990} = 1.01$ and $P(Z \ge 1.01) = 0.1562$. *Using technology*: `normalcdf(lower:0.6,upper:1000,µ:0.5,σ:0.0990)` $= 0.1562$. There is a 0.1562 probability that the mean number of moths is greater than or equal to 0.6. **(c)** No. Because this probability is not small, it is plausible that the sample mean number of moths is this high by chance alone.

Answers to Chapter 7 AP® Statistics Practice Test

T7.1 c
T7.2 c
T7.3 c
T7.4 a
T7.5 b
T7.6 b
T7.7 b
T7.8 e
T7.9 c
T7.10 e

T7.11 A. Both A and B appear to be unbiased, and A has less variability than B.

T7.12 **(a)** We do not know the shape of the population distribution of monthly fees. **(b)** $\mu_{\bar{x}} = \mu = \$38$. Because the sample of size 500 is less than 10% of all households with Internet access, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{500}} = 0.4472$. **(c)** Because the sample size is large ($n = 500 \ge 30$), the distribution of $\bar{x}$ will be approximately Normal. **(d)** We want to find $P(\bar{x} > 39)$. $z = \frac{39 - 38}{0.4472} = 2.24$ and $P(Z > 2.24) = 0.0125$. *Using technology*: `normalcdf(lower:39,upper:1000,µ:38,σ:0.4472)` $= 0.0127$. There is a 0.0127 probability that the mean monthly fee exceeds \$39.

T7.13 $\mu_{\hat{p}} = p = 0.22$. Because 300 is less than 10% of children under the age of 6, $\sigma_{\hat{p}} = \sqrt{\frac{0.22(0.78)}{300}} = 0.0239$. Because $np = 300(0.22) = 66$ and $n(1-p) = 300(0.78) = 234$ are both at least 10, the sampling distribution of $\hat{p}$ can be approximated by a Normal distribution. We want to find $P(\hat{p} > 0.20)$. $z = \frac{0.20 - 0.22}{0.0239} = -0.84$ and $P(Z > -0.84) = 0.7995$. *Using technology*: `normalcdf(lower:0.20,upper:1000,µ:0.22,σ:0.0239)` $= 0.7987$. There is a 0.7987 probability that more than 20% of the sample are from poverty-level households.

Answers to Cumulative AP® Practice Test 2

AP2.1 a
AP2.2 d
AP2.3 e
AP2.4 b
AP2.5 c
AP2.6 e
AP2.7 c
AP2.8 a
AP2.9 d
AP2.10 c
AP2.11 b
AP2.12 c
AP2.13 d
AP2.14 c
AP2.15 d
AP2.16 c
AP2.17 e
AP2.18 a
AP2.19 c
AP2.20 b
AP2.21 a
AP2.22 **(a)** Observational study, because no treatments were imposed on the subjects. **(b)** Two variables are confounded when their effects on the cholesterol level cannot be distinguished from one another. For example, people who take omega-3 fish oil might also exercise more. Researchers would not know whether it was the omega-3 fish oil or the exercise that was the real explanation for lower cholesterol. **(c)** No. Even though the difference was statistically significant, this wasn't an experiment and taking fish oil is possibly confounded with exercise.
AP2.23 **(a)** P(type O or Hawaiian-Chinese) = 65,516/145,057 = 0.452. **(b)** P(type AB | Hawaiian) = 99/4670 = 0.021.
(c) P(Hawaiian) = 4670/145,057 = 0.032; P(Hawaiian | type B) = 178/17,604 = 0.010. Because these probabilities are not equal, the two events are not independent. **(d)** P(type A and white) = 50,008/145,057 = 0.345. P(at least one type A and white) = 1 − P(neither are type A and white) = $1 - (1 - 0.345)^2 = 0.571$.
AP2.24 **(a)** The distribution of seed mass for the cicada plants is roughly symmetric, while the distribution for the control plants is skewed to the left. The median seed mass is the same for both groups. The cicada plants had a bigger range in seed mass, but the control plants had a bigger *IQR*. Neither group had any outliers. **(b)** The cicada plants. The distribution of seed mass for the cicada plants is roughly symmetric, which suggests that the mean should be about the same as the median. However, the distribution of seed mass for the control plants is skewed to the left, which will pull the mean of this distribution below its median toward the lower values. Because the medians of both distributions are equal, the mean for the cicada plants is greater than the mean for the control plants. **(c)** The purpose of the random assignment is to create two groups of plants that are roughly equivalent at the beginning of the experiment. **(d)** Benefit: controlling a source of variability. Different types of flowers will have different seed masses, making the response more variable if other types of plants were used. Drawback: we can't make inferences about the effect of cicadas on other types of plants, because other plants might respond differently to cicadas.
AP2.25 **(a)** We want to find $P(\bar{x} < 25{,}000/50) = P(\bar{x} < 500)$. Because the sample size is large ($n = 50 \geq 30$), the distribution of $\bar{x}$ is approximately Normal with $\mu_{\bar{x}} = \mu = 525$ pages. Because $n = 50$ is less than 10% of all novels in the library, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{200}{\sqrt{50}} = 28.28$ pages. $z = \frac{500 - 525}{28.28} = -0.88$ and $P(Z < -0.88) = 0.1894$. *Using technology*: `normalcdf(lower:−1000,upper:500,μ:525,σ:28.28)` = 0.1883. There is a 0.1883 probability that the total number of pages in 50 novels is fewer than 25,000. **(b)** Let X be the number of novels that have fewer than 400 pages. X is a binomial random variable with $n = 50$ and $p = 0.30$. We want to find $P(X \geq 20)$. *Using technology*: $P(X \geq 20) = 1 - P(X \leq 19) =$ `1 − binomcdf(trials:50, p:0.30, x value:19)` = 0.0848. There is a 0.0848 probability of selecting at least 20 novels that have fewer than 400 pages. *Note*: Using the Normal approximation, $P(X \geq 20) = 0.0614$.

Chapter 8

Section 8.1

Answers to Check Your Understanding

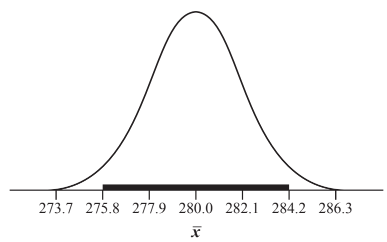
page 485: **1.** We are 95% confident that the interval from 2.84 to 7.55 g captures the population standard deviation of the fat content of Brand X hot dogs. **2.** If this sampling process were repeated many times, approximately 95% of the resulting confidence intervals would capture the population standard deviation of the fat content of Brand X hot dogs. **3.** False. Once the interval is calculated, it either contains σ or it does not contain σ.

Answers to Odd-Numbered Section 8.1 Exercises

8.1 Sample mean, $\bar{x} = 30.35$.

8.3 Sample proportion, $\hat{p} = \frac{36}{50} = 0.72$.

8.5 **(a)** Approximately Normal with mean $\mu_{\bar{x}} = 280$ and standard deviation $\sigma_{\bar{x}} = \frac{60}{\sqrt{840}} = 2.1$. **(b)** See graph below. **(c)** About 95% of the $\bar{x}$ values will be within 2 standard deviations of the mean. Therefore, $m = 2(2.1) = 4.2$. **(d)** About 95%.



8.7 The sketch is given below. The interval with the value of $\bar{x}$ in the shaded region will contain the population mean (280), while the interval with the value of $\bar{x}$ outside the shaded region will not contain the population mean (280).