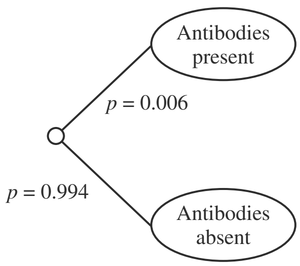
Chapter 10
Statistics and Probability

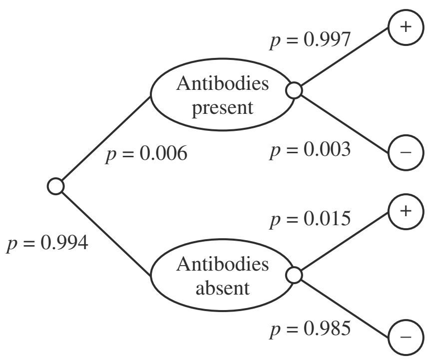
■ Section 10.1 Probability

Exploration 1

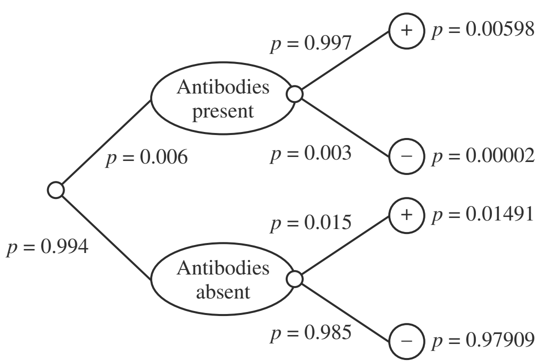
1.



2.



3.



4. $P(+) = 0.00598 + 0.01491 = 0.02089$

5. $P(\text{antibody present } |+) = \dfrac{P(\text{antibody present and } +)}{P(+)}$

$\dfrac{0.00598}{0.02089} \approx 0.286$ (A little more than 1 chance in 4)

Quick Review 10.1

1. 2

2. 6

3. $2^3 = 8$

4. $6^3 = 216$

5. ${}_{52}C_5 = 2{,}598{,}960$

6. ${}_{10}C_2 = 45$

7. $5! = 120$

8. ${}_5P_3 = 60$

9. $\dfrac{{}_5C_3}{{}_{10}C_3} = \dfrac{\frac{5!}{3!2!}}{\frac{10!}{3!7!}} = \dfrac{1}{12}$

10. $\dfrac{{}_5C_2}{{}_{10}C_2} = \dfrac{\frac{5!}{2!3!}}{\frac{10!}{2!8!}} = \dfrac{2}{9}$

Section 10.1 Exercises

For #1–8, consider ordered pairs (a, b) where a is the value of the red die and b is the value of the green die.

1. $E = \{(3, 6), (4, 5), (5, 4), (6, 3)\}$: $P(E) = \dfrac{4}{36} = \dfrac{1}{9}$

2. $E = \{\text{both dice even, both dice odd}\}$:
$P(E) = \dfrac{3 \cdot 3 + 3 \cdot 3}{36} = \dfrac{18}{36} = \dfrac{1}{2}$

3. $E = \{(2, 1), (3, 1), (3, 2), (4, 1), (4, 2), (4, 3), (5, 1), (5, 2), (5, 3), (5, 4), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5)\}$:
$P(E) = \dfrac{15}{36} = \dfrac{5}{12}$

4. $E = \{(1, 1), (1, 2), \ldots, (6, 2), (6, 3)\}$
$P(E) = \dfrac{30}{36} = \dfrac{5}{6}$

5. $P(E) = \dfrac{3 \cdot 3}{36} = \dfrac{1}{4}$

6. $P(E) = \dfrac{3 \cdot 3}{36} = \dfrac{1}{4}$

7. $E = \{(1, 1), (1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5), (3, 2), (3, 4), (4, 1), (4, 3), (5, 2), (5, 6), (6, 1), (6, 5)\}$
$P(E) = \dfrac{15}{36} = \dfrac{5}{12}$

8. $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (5, 6), (6, 1), (6, 5)\}$:
$P(E) = \dfrac{8}{36} = \dfrac{2}{9}$

9. (a) No. $0.25 + 0.20 + 0.35 + 0.30 = 1.1$. The numbers do not add up to 1.

(b) There is a problem with Alrik's reasoning. Since the gerbil must always be in exactly one of the four rooms, the proportions must add up to 1, just like a probability function.

10. Since $4 + 3 + 2 + 1 = 10$, we can divide each number in the ratio by 10 and get the proportions relative to the whole. The table then becomes

Compartment	A	B	C	D
Proportion	0.4	0.3	0.2	0.1

Yes, this is a valid probability function.

11. $P(B \text{ or } T) = P(B) + P(T) = 0.3 + 0.1 = 0.4$

12. $P(R \text{ or } G \text{ or } O) = P(R) + P(G) + P(O)$
$= 0.2 + 0.1 + 0.1 = 0.4$

13. $P(R) = 0.2$

14. $P(\text{not } R) = 1 - P(R) = 1 - 0.2 = 0.8$

15. $P[\text{not } (O \text{ or } Y)] = 1 - P(O \text{ or } Y)$
$= 1 - (0.2 + 0.1) = 0.7$

16. $P[\text{not } (B \text{ or } T)] = 1 - P(B \text{ or } T) = 1 - (0.3 + 0.1)$
$= 0.6$

17. $P(B_1 \text{ and } B_2) = P(B_1) \cdot P(B_2) = (0.3)(0.3) = 0.09$

18. $P(O_1 \text{ and } O_2) = P(O_1) \cdot P(O_2) = (0.1)(0.1) = 0.01$

19. $P[(R_1 \text{ and } G_2) \text{ or } (G_1 \text{ and } R_2)] = P(R_1) \cdot P(G_2)$
$+ P(G_1) \cdot P(R_2) = (0.2)(0.2) + (0.2)(0.2) = 0.08$

20. $P(B_1 \text{ and } Y_2) = P(B_1) \cdot P(Y_2) = (0.3)(0.2) = 0.06$

21. P (neither is yellow) $= P(\text{not } Y_1 \text{ and not } Y_2)$
$= P(\text{not } Y_1) \cdot P(\text{not } Y_2) = (0.8)(0.8) = 0.64$

22. $P(\text{not } R_1 \text{ and not } O_2) = P(\text{not } R_1) \cdot P(\text{not } O_2)$
$= (0.8)(0.9) = 0.72$

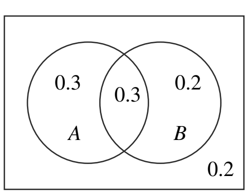
23. There are ${}_{24}C_6 = 134{,}596$ possible hands; of these, only one consists of all spades, so the probability is $\dfrac{1}{134{,}596}$.

24. Of the ${}_{24}C_6 = 134{,}596$ possible hands, one consists of all spades, one consists of all clubs, one consists of all hearts, and one consists of all diamonds, so the probability is $\dfrac{4}{134{,}596} = \dfrac{1}{33{,}649}$.

25. Of the ${}_{24}C_6 = 134{,}596$ possible hands, there are ${}_4C_4 \cdot {}_{20}C_2 = 190$ hands with all the aces, so the probability is $\dfrac{190}{134{,}596} = \dfrac{5}{3542}$.

26. There are ${}_2C_2 \cdot {}_{22}C_4 = 7315$ ways to get both black jacks and 4 "other" cards. Similarly, there are 7315 ways to get both red jacks. These two numbers together count twice the ${}_2C_2 \cdot {}_2C_2 \cdot {}_{20}C_2 = 190$ ways to get all four jacks. Therefore, altogether we have $2 \cdot 7315 - 190 = 14{,}440$ distinct ways to have both bowers, so the probability is $\dfrac{14{,}440}{134{,}596} = \dfrac{190}{1771}$.

27. **(a)**



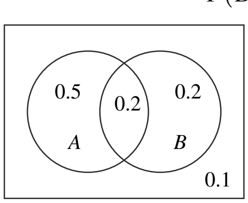
(b) 0.3

(c) 0.2

(d) 0.8

(e) Yes. $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)} = \dfrac{0.3}{0.5} = 0.6 = P(A)$

28. **(a)**



(b) 0.5

(c) 0.1

(d) 0.9

(e) No. $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)} = \dfrac{0.2}{0.4} = 0.5 \neq P(A)$

29. $P(\text{John will practice}) = (0.6)(0.8) + (0.4)(0.4) = 0.64$

30. **(a)** $P(\text{meatloaf is served}) = \dfrac{1}{5} = 0.20$

(b) $P(\text{meatloaf and peas are served}) = (0.20)(0.70)$
$= 0.14$

(c) $P(\text{peas are served}) = (0.20)(0.70) + (0.80)(0.30)$
$= 0.38$

31. If all precalculus students were put on a single list and a name then randomly chosen, the probability $P(\text{from Mr. Abel's class} \mid \text{girl})$ would be $\dfrac{12}{22} = \dfrac{6}{11}$. But when one of the two classes is selected at random, and then a student from this class is selected, $P(\text{from Mr. Abel's class} \mid \text{girl})$

$$= \frac{P(\text{girl from Mr. Abel's class})}{P(\text{girl})} = \frac{\left(\frac{1}{2}\right)\left(\frac{12}{20}\right)}{\left(\frac{1}{2}\right)\left(\frac{12}{20}\right) + \left(\frac{1}{2}\right)\left(\frac{10}{25}\right)} = \frac{3}{5}$$

32. Within each box, any of the coins is equally likely to be chosen and either side is equally likely to be shown. But a head in the 2-coin box is more likely to be displayed than a head in the 3-coin box.

$$P(\text{from 2-coin box} \mid \text{H}) = \frac{P(\text{H from 2-coin box})}{P(\text{H})} = \frac{\left(\frac{1}{2}\right)\left(\frac{3}{4}\right)}{\left(\frac{1}{2}\right)\left(\frac{3}{4}\right) + \left(\frac{1}{2}\right)\left(\frac{3}{6}\right)} = \frac{3}{5}$$

33. $\dfrac{{}_{20}C_2}{{}_{25}C_2} = \dfrac{190}{300} = \dfrac{19}{30}$

34. $P(\text{none are defective}) = {}_4C_0 \cdot (0.037)^0(0.963)^4$
$= (0.963)^4 \approx 0.860$

35. **(a)** $P(\text{cardiovascular disease or cancer}) = 0.45 + 0.22$
$= 0.67$

(b) $P(\text{other cause of death}) = 1 - 0.67 = 0.33$

36. $P(\text{Yahtzee}) = 6 \cdot \left(\dfrac{1}{6}\right)^5 = \dfrac{1}{1296}$

37. The sum of the probabilities is greater than 1 – an impossibility, since the events are mutually exclusive.

38. $P(\text{at least one false positive})$
$= 1 - P(\text{no false positives}) = 1 - (0.993)^{60} \approx 0.344.$

39. **(a)** $P(\text{a woman}) = \dfrac{172}{254} = \dfrac{86}{127}$

(b) $P(\text{went to graduate school}) = \dfrac{124 + 58}{254} = \dfrac{91}{127}$

(c) $P(\text{a woman who went to graduate school})$
$= \frac{124}{254} = \frac{62}{127}$

40. (a) $\frac{{}_{14}C_8}{{}_{20}C_8} = \frac{3003}{125{,}970} = \frac{77}{3230}$

(b) $\frac{{}_{14}C_5 \cdot {}_6C_3}{{}_{20}C_8} = \frac{40{,}040}{125{,}970} = \frac{308}{969}$

(c) $\frac{{}_{14}C_6 \cdot {}_6C_2 + {}_{14}C_7 \cdot {}_6C_1 + {}_{14}C_8 \cdot {}_6C_0}{{}_{20}C_8}$
$= \frac{45{,}045 + 20{,}592 + 3003}{125{,}970} = \frac{68{,}640}{125{,}970} = \frac{176}{323}$

41. $\frac{1}{{}_9C_2} = \frac{1}{36}$

42. This cannot be true. Let A be the event that it is cloudy all day, B be the event that there is at least 1 hour of sunshine, and C be the event that there is some sunshine, but less than 1 hour. Then A, B, and C are mutually exclusive events, so $P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$ $= 0.22 + 0.78 + P(C) = 1 + P(C)$. Then it must be the case that $P(C) = 0$. This is absurd; there must be some probability of having more than 0 but less than 1 hour of sunshine.

43. Answers will vary. One possible answer: The 90% probability comes from the doctor's knowledge of outcomes of similar surgeries in the past.

44. Answers will vary. One possible answer: Predictions about weather come from observing the outcomes of similar weather conditions in the past.

45. $P(\text{no preference}) = 1 - P(\text{favor right}) - P(\text{favor left}) = 1 - 0.60 - 0.10 = 0.70$

(a) $P(\text{none show preference})$
$= P(N_1 \text{ and } N_2 \text{ and } N_3) =$
$= (0.30)(0.30)(0.30) = 0.027$

(b) $P(\text{one hand or the other})$
$= P(3R \text{ or } 3L \text{ or } 2R \text{ and } 1L \text{ or } 2R \text{ and } 1L) =$
$= (0.60)^3 + (0.10)^3 + {}_3C_2 \cdot (0.60)^2(0.10)^1 +$
${}_3C_2 \cdot (0.10)^2(0.60)^1$
$= 0.343$

(c) $P(\text{same hand})$
$= P(R_1 \text{ and } R_2 \text{ and } R_3 \text{ or } L_1 \text{ and } L_2 \text{ and } L_3) =$
$= (0.60)^3 + (0.10)^3 = 0.217$

46. (a) $P(\text{no reds}) = P(\text{not } R_1 \text{ and not } R_2 \text{ and not } R_3) =$
$= \left(\frac{8}{20}\right)\left(\frac{7}{19}\right)\left(\frac{6}{18}\right) = \frac{14}{285}$

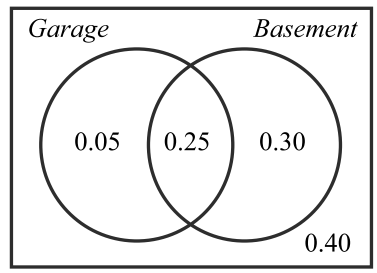
(b) $P(3 \text{ of same color})$
$= P(3R \text{ or } 3W \text{ or } 3B)$
$= \left(\frac{12}{20}\right)\left(\frac{11}{19}\right)\left(\frac{10}{18}\right) + \left(\frac{5}{20}\right)\left(\frac{4}{19}\right)\left(\frac{3}{18}\right) +$
$\left(\frac{3}{20}\right)\left(\frac{2}{19}\right)\left(\frac{1}{18}\right) = \frac{77}{380}$

(c) There are $3 \cdot 2 \cdot 1 = 6$ ways to select the three colors.
$= P(\text{a set of red, white, and blue})$
$= 6 \cdot P(1R \text{ and } 1W \text{ and } 1B)$
$= 6\left(\frac{12}{20}\right)\left(\frac{5}{19}\right)\left(\frac{3}{18}\right) = \frac{3}{19}$

47. $P(\text{first type A is fourth in line}) = P(\text{not } A \text{ and not } A \text{ and not } A \text{ and } A) = (0.60)(0.60)(0.60)(0.40) = 0.0864$

48. $P(\text{buy } 5) = P(\text{not } B_1 \text{ and not } B_2 \text{ and not } B_3 \text{ and not } B_4 \text{ and } B_5) = \left(\frac{17}{20}\right)\left(\frac{16}{19}\right)\left(\frac{15}{18}\right)\left(\frac{14}{17}\right)\left(\frac{3}{16}\right) = \frac{7}{76}$

49. Use the Venn diagram below to answer the exercise.



(a) $P(\text{neither}) = 1 - 0.05 - 0.25 - 0.30 = 0.40$

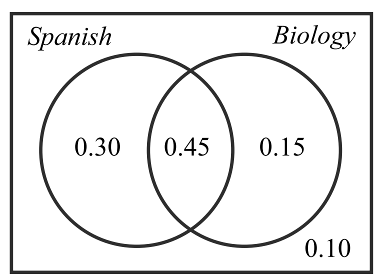
(b) $P(\text{garage or basement}) = 0.05 + 0.25 + 0.30 = 0.60$

(c) $P(\text{garage} \mid \text{basement}) = \frac{0.25}{0.55} \approx 0.455$

(d) $P(\text{basement} \mid \text{garage}) = \frac{0.25}{0.30} \approx 0.833$

(e) No, $P(\text{garage} \mid \text{basement}) \neq P(\text{garage})$

50. Use the Venn diagram below to answer the exercise.



(a) $P(\text{neither course}) = 1 - 0.30 - 0.45 - 0.15 = 0.1$

(b) $P(\text{Spanish or Biology}) = 0.30 + 0.45 + 0.15 = 0.9$

(c) $P(\text{Biology} \mid \text{Spanish}) = \frac{0.45}{0.75} = 0.6$

(d) $P(\text{Spanish} \mid \text{Biology}) = \frac{0.45}{0.60} = 0.75$

(e) Yes, $P(\text{Biology} \mid \text{Spanish}) = P(\text{Biology})$

51. False. A sample space consists of outcomes, which are not necessarily equally likely.

52. False. All probabilities are between 0 and 1, inclusive.

53. Of the 36 different, equally likely ways the dice can land, 4 ways have a total of 5. So the probability is $4/36 = 1/9$. The answer is D.

54. A probability must always be between 0 and 1, inclusive. The answer is E.

55. $P(B \text{ and } A) = P(B)\,P(A|B)$, and for independent events, $P(B \text{ and } A) = P(B)P(A)$. It follows that the answer is A.

56. A specific sequence of one "heads" and two "tails" has probability $(1/2)^3 = 1/8$. There are three such sequences. The answer is C.

57. (a)

Type of Bagel	Probability
Plain	0.37
Onion	0.12
Rye	0.11
Cinnamon Raisin	0.25
Sourdough	0.15

(b) $(0.37)(0.37)(0.37) \approx 0.051$

(c) No. They are more apt to share bagel preferences if they arrive at the store together.

58. (a) $P(\text{at least one king}) = 1 - P(\text{no kings})$

$$= 1 - \frac{{}_{48}C_5}{{}_{52}C_5} = \frac{18{,}472}{54{,}145} \approx 0.34 = 34.0\%$$

(b) The number of ways to choose, e.g., 3 fives and 2 jacks, is ${}_4C_3 \cdot {}_4C_2$. There are ${}_{13}P_2 = 13 \cdot 12$ different combinations of cards that can make up the full house, so

$$P(\text{full house}) = \frac{13 \cdot 12 \cdot {}_4C_3 \cdot {}_4C_2}{{}_{52}C_5} = \frac{6}{4165} \approx 0.0014$$

$= 0.14\%$.

59. (a) $P(\text{all Republicans}) = P(R_1 \text{ and } R_2 \text{ and } R_3 \text{ and } R_4) =$

$$= \left(\frac{10}{17}\right)\left(\frac{9}{16}\right)\left(\frac{8}{15}\right)\left(\frac{7}{14}\right) = \frac{3}{34} \approx 0.088 = 8.8\%$$

The chosen group all being Republicans is plausible, but not likely.

(b) The calculations were based on the assumption that all names were equally likely to be chosen, which may not be the case if the selection was indeed rigged in some way.

60. (a) $P(\text{5 red lights}) = P(R_1 \text{ and } R_2 \text{ and } R_3 \text{ and } R_4 \text{ and } R_4) =$

$$= \left(\frac{40}{60}\right)\left(\frac{40}{60}\right)\left(\frac{40}{60}\right)\left(\frac{40}{60}\right)\left(\frac{40}{60}\right) = \frac{32}{243} \approx 0.132 = 13.2\%$$

Hitting 5 red lights in a row is plausible, but not likely, so it could have been bad luck.

(b) The calculations were based on the assumption of the independence of hitting red lights.

61. (a) \$1.50

(b) $3 \cdot \frac{2}{6} + (-1) \cdot \frac{4}{6} = \frac{6}{6} - \frac{4}{6} = \frac{1}{3}$

62. (a) $\frac{10}{13{,}983{,}816} \approx 0.000000715$

(b)

Value	Probability
-10	$\frac{13{,}983{,}806}{13{,}983{,}816}$
$+4{,}999{,}990$	$\frac{10}{13{,}983{,}816}$

(c) $4{,}999{,}990 \cdot \frac{10}{13{,}983{,}816} + (-10) \cdot \left(1 - \frac{10}{13{,}983{,}816}\right) \approx -6.42$

(d) In the long run, Gladys is losing \$6.42 every time she buys the 10 tickets. Given the low probability of a positive payoff, she stands to lose a lot of money if she does this often.

■ Section 10.2 Statistics (Graphical)

Exploration 1

1. We observe that the numbers seem to be centered a bit below 13. We would need to take into account the different state populations (not given in the table) in order to compute the national average exactly; but, just for the record, it was about 12.8 percent.
2. We observe in the stemplot that five states have percentages above 15.
3. We observe in the stemplot that the bottom five states are all below 10%. Returning to the table, we pick these out as Alaska, Colorado, Georgia, Texas, and Utah.
4. The low outlier is Alaska, where older people would be less willing or able to cope with the harsh winter conditions. The high outlier is Florida, where the mild weather and abundant retirement communities attract older residents.

Quick Review 10.2

1. $\approx 15.48\%$
2. $\approx 20.94\%$
3. $\approx 14.44\%$
4. $\approx 27.22\%$
5. ≈ 1723
6. 9200
7. \$235 thousand
8. 238 million
9. 1 million
10. 1 billion

Section 10.2 Exercises

1. (a) $\frac{292}{1008} \approx 29.0\%$ **(b)** $\frac{132}{1008} \approx 13.1\%$

(c) $\frac{132}{492} \approx 26.8\%$ **(d)** $\frac{132}{292} \approx 45.2\%$

2. (a) $\frac{818}{1316} \approx 62.1\%$ **(b)** $\frac{202}{1316} \approx 15.3\%$

(c) $\frac{123}{818} \approx 15.0\%$ **(d)** $\frac{123}{325} \approx 37.8\%$

3. (a) Pie charts; stemplots are not appropriate for categorical variables.

(b) No. Men are more likely than women to be interested in the game (57% to 39%); women are more likely than men to be interested in the commercials (30% to 16.5%) or not to watch at all (31% to 27%).

4. (a) Bar graph; histograms are not appropriate for categorical variables.

(b) No; while 62% of first-class passengers survived, only 41% of second-class and 25% of third-class passengers did.

5. (a) $\frac{210}{240} \cdot 48 = 42$ **(b)** $\frac{30}{240} \cdot 48 = 6$

(c) $\frac{210}{240} \cdot 192 = 168$ **(d)** $\frac{30}{240} \cdot 192 = 24$

6. (a) $\frac{210}{600} \cdot 220 = 77$ **(b)** $\frac{390}{600} \cdot 220 = 143$

(c) $\frac{210}{600} \cdot 240 = 84$ **(d)** $\frac{390}{600} \cdot 240 = 156$

(d) $\frac{210}{600} \cdot 140 = 49$ **(e)** $\frac{390}{600} \cdot 140 = 91$

7.

0	5 8 9
1	3 4 6
2	3 6 8
3	3 9
4	
5	
6	1

61 is an outlier.

8.

0	3 9 9
1	
2	2 9
3	2 2 3 9 9
4	2 9
5	2 8
6	5
7	0

The seasons with only 3, 9, and 9 home runs may be outliers.

9.

Maris		Aaron
9 8 5	0	
6 4 3	1	0 2 3
8 6 3	2	0 4 6 7 9
9 3	3	0 2 4 4 8 9 9
	4	0 0 4 4 4 4 5 7
	5	
1	6	

Except for Maris's one record-breaking year, his home run output falls well short of Aaron's.

10.

Bonds		McGwire
	0	3 9 9
9 6	1	
8 6 5 5 4	2	2 9
7 7 4 4 3 3	3	2 2 3 9 9
9 6 6 5 5 2 0	4	2 9
	5	2 8
	6	5
3	7	0

Although Bonds hit more home runs than McGwire, he played more seasons, hitting over 50 only once compared to 4 times for McGwire.

11. Males

6	0
6	5 9
7	1 2 2 3 3 3
7	5 5 6

This stemplot shows the life expectancies of males in the nations of South America are clustered near 70, with one lower value clustered near 60.

12. Females

6	7 9
7	
7	6 8 8 8 9 9 9
8	0 1 2

This stemplot shows the life expectancies of females in the nations of South America are clustered in the high 70s and at 80, with two lower values in the high 60s and low 70s.

13.

Males		Females
3 0	6	
9 5	6	7 9
3 3 3 2 2 1	7	
6 5 5	7	6 8 8 8 9 9 9
	8	0 1 2

This stemplot shows that the life expectancies of the women in the nations of South America are about 5–6 years higher than that of the men in the nations of South America.

14. Female-Male Difference

3	2
4	1
5	6 9
6	0 1 1 6 8
7	0 0 3

This stemplot of differences shows that the women in the nations of South America have higher life expectancies than the men by about 5–6 years.

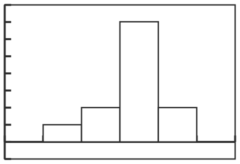
15.

Life expectancy (years)	Frequency
60.0–64.9	1
65.0–69.9	2
70.0–74.9	7
75.0–79.9	2

16.

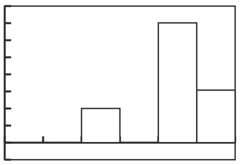
Life expectancy (years)	Frequency
65.0–69.9	2
70.0–74.9	0
75.0–79.9	7
80.0–84.9	3

17.



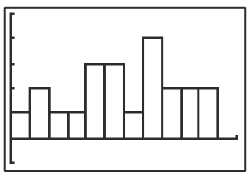
[55, 85] by [–1, 8]

18.



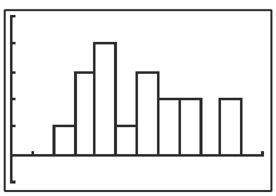
[55, 85] by [–1, 8]

19.



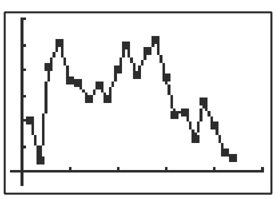
[0, 60] by [–1, 5]

20.



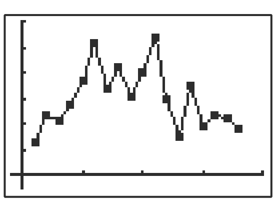
[0, 60] by [–1, 5]

21.



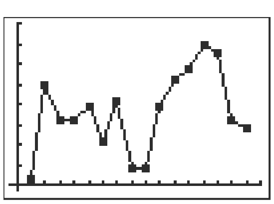
[−1, 25] by [−5, 60]

22.



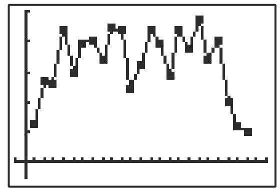
[−1, 20] by [−5, 60]

23.



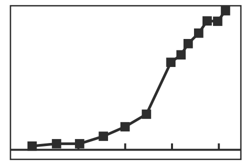
[–1.5, 17] by [–2, 80]

24.



[−1, 25] by [−5, 50]

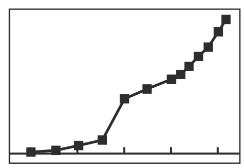
25.



[1965, 2015] by [–100, 1500]

The winner's prize money for the PGA Championship appears to be growing exponentially, though the rate of growth may be slowing down in recent years.

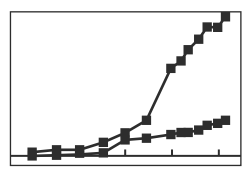
26.



[1965, 2015] by [–25, 400]

The winner's prize money for the LPGA Championship appears to be growing faster than linearly, and at a much more rapid rate since 1990.

27.

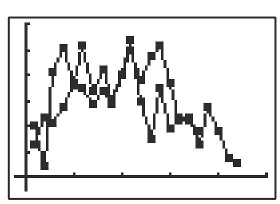


[1965, 2015] by [–100, 1500]

The women's winner's prize money grew at a rate similar to that of the men's until 1995; then the men's PGA purse began increasing rapidly, leaving the LPGA purse far behind.

28. When viewed together, women may have had a larger percentage increase in prize money over the time period, but still make a far lower amount when compared to that of the men.

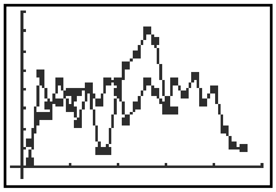
29.



[−1, 25] by [−5, 60]

The two home run hitters enjoyed similar success, with Mays enjoying a bit of an edge in the earlier and later years of his career, and Mantle enjoying an edge in the middle years.

30.



[–1, 25] by [–5, 80]

The two home run hitters were comparable for the first seven years of their careers.

31. **(a)** Bimodal

(b) The group on the left are most likely healthier cereals for adults, while the group on the right are sugary cereals for children.

32. **(a)** Unimodal and skewed to the right

(b) Games between good teams are usually close, but there are occasional blowouts.

33. **(a)** The data are quantitative

(b)

Stem	Leaf
28	2
29	3 7
30	
31	6 7
32	7 8
33	5 5 5 8
34	2 8 8
35	3 3 4
36	3 7
37	
38	5

(c) Unimodal and symmetric

34. **(a)** The data are quantitative

(b)

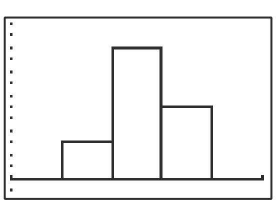
Stem	Leaf
6	2 3 9 9
7	7 9 9 9
8	0 2 3 7 7 8 9 9
9	0 0 0 1 1 2 3 3 3 4 5 6 7
10	2 3 3 6 6 6 7 9
11	0 2 4 6
12	0 5 9

(c) Unimodal and symmetric

35 **(a)**

Interval	Frequency
25.0–29.9	3
30.0–34.9	11
35.0–39.9	6

(b)



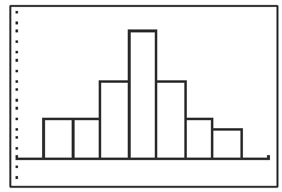
[20, 45] by [−1, 13]

(c) The distribution is unimodal and symmetric, with most salaries in the interval \$30,000−34,999.

36. **(a)**

Interval	Frequency
6.0–6.9	4
7.0–7.9	4
8.0–8.9	8
9.0–9.9	13
10.0–10.9	8
11.0–11.9	4
12.0–12.9	3

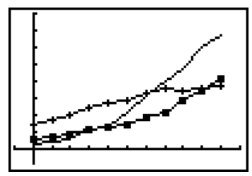
(b)



[5, 14] by [−2, 15]

(c) The distribution is unimodal and symmetric, with most wind speeds averaging between 8 and 11 mph.

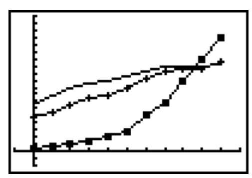
37.



[1890, 2010] by [–4, 40]

___ = CA + = NY ■ = TX

38.



[1890, 2010] by [–4, 40]

___ = PA + = IL ■ = FL

39. False. If the graduation rates are the same, then the likelihood of graduating is independent of a student's gender; there is no association.

40. False. They are outliers only if they are significantly higher or lower than the other numbers in the data set.

41. A time plot uses a continuous line. The answer is C.

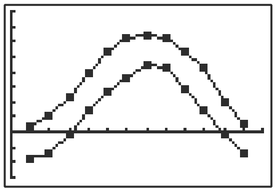
42. Back-to-back stemplots are designed for comparing data sets. The answer is B.

43. The histogram suggests data values clustered near an upper limit — such as the maximum possible score on an easy test. The answer is A.

44. 45° is 1/8, or 12.5%, of 360°. The answer is B.

45. Answers will vary. Possible outliers could be the pulse rates of long-distance runners and swimmers, which are often unusually low. Students who have had to run to class from across campus might have pulse rates that are unusually high.

46. Answers will vary. Female heights typically have a distribution that is uniformly lower than male heights, but the difference might not be apparent from a stemplot, especially if the sample is small.

47. 

$[0, 13]$ by $[-15, 40]$

48. Using the form $f(t) = k + a\sin[b(t - h)]$, we have $b = \pi/6$ for both graphs since the period is 12. a is half the difference between the extremes, and k is the average of the extremes. Finally, h is the offset to the first time this average occurs.

The maxima of both graphs occur in July ($t = 7$), while the minima occur in January ($t = 1$), so we should choose $b = 3.5$ or 4 (if the latter is halfway between the maximum and the minimum, but the former gives a better match for some values of t). Thus, for the high temperature, use $f(t) \approx 17 + 15\sin\left[\frac{\pi}{6}(t - 3.5)\right]$ or $17 + 15\sin\left[\frac{\pi}{6}(t - 4)\right]$, and for the low temperature, use $g(t) \approx 6.5 + 15.5\sin\left[\frac{\pi}{6}(t - 3.5)\right]$ or $6.5 + 15.5\sin\left[\frac{\pi}{6}(t - 4)\right]$.

■ Section 10.3 Statistics (Numerical)

Exploration 1

1. The ranges are approximately the same.
2. Figure (b) has the greater IQR since the data values are more spread out.
3. Figure (b) has the greater variability since the data values are more spread out.

Exploration 2

1. Figure 10.18(b) has a longer "tail" to the right (skewed right), so the values in the tail pull the mean to the right of the (resistant) median. Figure 10.18(c) is skewed left, so the values in the tail pull the mean to the left of the median. Figure 10.18 (a) is symmetric about a vertical line, so the median and the mean are close together.

Quick Review 10.3

1. $\sum_{i=1}^{7} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7$

2. $\sum_{i=1}^{5} (x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + (x_4 - \bar{x}) + (x_5 - \bar{x}) = x_1 + x_2 + x_3 + x_4 + x_5 - 5\bar{x}$.

Note that, since $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i$, this simplifies to 0.

3. $\frac{1}{7}\sum_{i=1}^{7} x_i = \frac{1}{7}(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7)$

4. $\frac{1}{5}\sum_{i=1}^{5} (x_i - \bar{x}) = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5 - 5\bar{x}) = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) - \bar{x}$. Note that, since $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i$, this simplifies to 0.

5. The expression at the end of the first line is a simple expansion of the sum (and is a reasonable answer to the given question). By expanding further, we can also arrive at the final expression below, which is somewhat simpler.

$$\begin{aligned}
\frac{1}{5}\sum_{i=1}^{5}(x_i - \bar{x})^2 &= \frac{1}{5}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2] \\
&= \frac{1}{5}[(x_1^2 - 2x_1\bar{x} + \bar{x}^2) + (x_2^2 - 2x_2\bar{x} + \bar{x}^2) + \dots + (x_5^2 - 2x_5\bar{x} + \bar{x}^2)] \\
&= \frac{1}{5}[x_1^2 + x_2^2 + \dots + x_5^2 - 2\bar{x}(x_1 + x_2 + \dots + x_5)] + \bar{x}^2 \\
&= \frac{1}{5}(x_1^2 + x_2^2 + \dots + x_5^2) - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{5}(x_1^2 + x_2^2 + \dots + x_5^2) - \bar{x}^2
\end{aligned}$$

6. The square root does not allow for further simplication. The final answer is the square root of the expression from Exercise 5: either $\sqrt{\frac{1}{5}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2]}$ or $\sqrt{\frac{1}{5}(x_1^2 + x_2^2 + \dots + x_5^2) - \bar{x}^2}$.

7. $\sum_{i=1}^{8} x_i f_i$

8. $\sum_{i=1}^{10} (x_i - \bar{x})^2$

9. $\frac{1}{50}\sum_{i=1}^{50} (x_i - \bar{x})^2$

10. $\sqrt{\frac{1}{7}\sum_{i=1}^{7} (x_i - \bar{x})^2}$

Section 10.3 Exercises

1. **(a)** Statistic. The number characterizes a set of known data values.

(b) Parameter. The number describes an entire population, on the basis of some statistical inference.

2. **(a)** Parameter.The number describes an entire population, on the basis of some statistical inference.

(b) Statistic. The number is calculated from information about all rats in a small, experimental population.

3. **(a)** Mean. A pitcher's earned run average (ERA) is total earned runs divided by number of nine-inning blocks pitched.

(b) Median. The middle height is considered average.

4. **(a)** Median.The middle time is considered average.

(b) Mean. Grade point averages are total grade points divided by number of units.

5. There are 8 data values, which is an even number, so the median is the middle data value when they are arranged in order. In order, the data values are {0, 0, 1, 2, 14, 27, 62, 67}. The median is $\frac{14 + 2}{2} = 8$.

6. There are 7 data values, which is an odd number, so the median is the middle data value when they are arranged in order. In order, the data values are {8112, 9938, 13,209, 17,819, 24,474, 30,065, 44,579} in thousands of km^2. The median is 17,819,000 km^2.

7. There are 19 data values, which is an odd number, so the median is the middle data value when they are arranged in order. In order, the data values are {28.2, 29.3, 29.7, 31.6, 31.7, 32.7, 32.8, 33.5, 33.5, 33.5, 33.8, 34.2, 34.8, 34.8, 35.3, 35.4, 36.7, 37.3, 38.5}.The median is \$33,500.

8. There are 44 data values. The median is between the 22nd and 23rd data values, so median $= \frac{9.2 + 9.3}{2} = 9.25$ mph

9. For {$\underline{28.2}$, 29.3, 29.7, 31.6, $\underline{31.7}$, 32.7, 32.8, $\underline{33.5}$, 33.5, 33.5, 33.8, 34.2, 34.8, 34.8, $\underline{35.3}$, 35.4, 36.7, 37.3, $\underline{38.5}$}, range = \$38,500 − \$28,200 = \$10,300 and IQR = \$35,300 − \$31,700 = \$3,600.

10. Range = 12.9 − 6.2 = 6.7 mph, $IQR = 10.6 - 8.5 = 2.1$ mph.

11. $IQR = 165 - 143 = 22$ lb, $Q_1 - 1.5 \times IQR = 143 - 1.5 \times 22 = 110$ lb and $Q_3 + 1.5 \times IQR = 165 + 1.5 \times 22 = 198$ lb. Since 210 lb > 198 lb, it is an outlier.

12. $IQR = 72 - 49 = 23$ thousand miles, $Q_1 - 1.5 \times IQR = 49 - 1.5 \times 23 = 14.5$ thousand miles and $Q_3 + 1.5 \times IQR = 72 + 1.5 \times 23 = 106.5$ thousand miles. Since 7000 miles < 14,500 miles, it is an outlier.

13. For {$\underline{16}$, 16, 16, 16, $\underline{18}$, 20, 20, 20, 20, $\underline{20}$, 21, 21, 22, 24, $\underline{24}$, 30, 30, 40, $\underline{42}$}, the five-number summary is {16, 18, 20, 24, 42}. $IQR = 24 - 18 = 6$, $Q_1 - 1.5 \times IQR = 18 - 1.5 \times 6 = 9$ and $Q_3 + 1.5 \times IQR = 24 + 1.5 \times 6 = 33$. Since 40 and 42 are greater than 33, they are outliers.

14. There are 32 data values, so Q_1 is between the 8th and 9th data values so $Q_1 = \frac{6 + 6}{2} = 6$, the median is between the 16th and 17th data values, so the median $= \frac{7 + 8}{2} = 7.5$, and Q_3 is between the 24th and 25th data values, so $Q_3 = \frac{10 + 11}{2} = 10.5$. The five-number summary is {2, 6, 7.5, 10.5, 13}. $IQR = 10.5 - 6 = 4.5$, $Q_1 - 1.5 \times IQR = 6 - 1.5 \times 4.5 = -0.75$ and $Q_3 + 1.5 \times IQR = 10.5 + 1.5 \times 4.5 = 17.25$. There are no outliers.

15. While a stemplot is not needed to answer this question, the sorted stemplot below is more compact than a sorted list of the 30 numbers. The underlined numbers are the ones used for the five-number summary, which is {20, 29, 42, 49, 66}. Median $= \frac{41 + 43}{2} = 42$. $IQR = 49 - 29 = 20$, $Q_1 - 1.5 \times IQR = 29 - 1.5 \times 20 = -1$ and $Q_3 + 1.5 \times IQR = 49 + 1.5 \times 20 = 79$. There are no outliers.

2	$\underline{0}$ 1 4 5 7 8 9 $\underline{9}$
3	1 3 4 4 8
4	1 $\underline{1\ 3}$ 4 5 5 5 7 9 $\underline{9}$
5	4 6 6 7 8
6	0 $\underline{6}$

16. While a stemplot is not needed to answer this question, the sorted stemplot below is more compact than a sorted list of the 24 numbers. The underlined numbers are the ones used for the five-number summary, which is {8.0, 10.55, 11.05, 11.6, 13.4}.
$Q_1 = \frac{10.5 + 10.6}{2} = 10.55$, median $= \frac{11.0 + 11.1}{2} = 11.05$, $Q_3 = \frac{11.6 + 11.6}{2} = 11.6$. $IQR = 11.6 - 10.55 = 1.05$,
$Q_1 - 1.5 \times IQR = 10.55 - 1.5 \times 1.05 \approx 8.98$ and $Q_3 + 1.5 \times IQR = 11.6 + 1.5 \times 1.05 \approx 13.18$. 8.0 and 13.4 are outliers (and possibly 8.9 and 8.9).

8	$\underline{0}$ 9 9
9	5
10	0 $\underline{5}$ $\underline{6}$ 7 7 9 9
11	$\underline{0}$ $\underline{1}$ 3 4 5 5 $\underline{6}$ $\underline{6}$ 7
12	4 6 9
13	$\underline{4}$

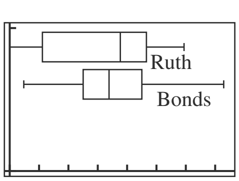
17. In general, NHL teams do win more games at home than on the road; the median for home wins is about 5 games higher than the median for away games and greater than the third quartile for away wins. Although the variability is comparable, over 25% of the teams won more home games than even the best team won on the road.

18. The number of games teams win on the road is less variable and the median is about 7 games lower than wins at home. Over 25% of the teams had more home wins than even the best team had on the road.

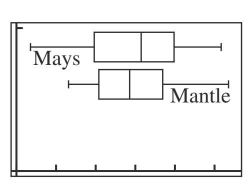
19. The ordered data for Babe Ruth is {$\underline{0}$, 1, 3, 3, 6, $\underline{11}$, 22, 25, 29, 34, $\underline{35}$, $\underline{41}$, 41, 46, 46, 46, $\underline{47}$, 49, 54, 54, 59, $\underline{60}$}, so the median is $\frac{35 + 41}{2} = 38$.
Five-number summary: {0, 11, 38, 47, 60}
Range: $60 - 0 = 60$
IQR: $47 - 11 = 36$; there are no outliers.
The ordered data for Barry Bonds is:
{$\underline{5}$, 16, 19, 24, 25, $\underline{25}$, 26, 28, 33, 33, $\underline{34}$, $\underline{34}$, 37, 37, 40, 42, 45, $\underline{45}$, 46, 46, 49, $\underline{73}$}, so the median is
$\frac{34 + 34}{2} = 34$.
Five-number summary: {5, 25, 34, 45, 73}
Range: $73 - 5 = 68$
IQR: $45 - 25 = 20$; 73 could possibly be an outlier.

20. The ordered data for Willie Mays is:
{$\underline{4}$, 6, 8, 13, 18, $\underline{20}$, 22, 23, 28, 29, $\underline{29}$, $\underline{34}$, 35, 36, 37, 38, $\underline{40}$, 41, 47, 49, 51, $\underline{52}$}, so the median is
$\frac{29 + 34}{2} = 31.5$, $Q_3 = 40$, and $Q_1 = 20$.
Five-number summary: {4, 20, 31.5, 40, 52}
Range: $52 - 4 = 48$
IQR: $40 - 20 = 20$
No outliers
The ordered data for Mickey Mantle is:
{$\underline{13}$, 15, 18, 19, $\underline{21}$, 22, 23, 23, $\underline{27}$, $\underline{30}$, 31, 34, 35, $\underline{37}$, 40, 42, 52, $\underline{54}$} so the median is
$\frac{27 + 30}{2} = 28.5$, $Q_3 = 37$, and $Q_1 = 21$.
Five-number summary: {13, 21, 28.5, 37, 54}
Range: $54 - 13 = 41$
IQR: $37 - 21 = 16$
No outliers

21.



22.



23. $\overline{x} = \frac{1}{9}(0 + 0 + 1 + 2 + 67 + 62 + 27 + 14) = \frac{172}{8} \approx 21.625$ satellites, which is much larger than the median because 2 planets have over 60 moons.

24. $\overline{x} = \frac{1}{7}(30{,}065{,}000 + 13{,}209{,}000 + 44{,}579{,}000 + \cdots + 17{,}819{,}000) = \frac{1}{7}(148{,}196{,}000) = 21{,}171{,}000\ \text{km}^2$, which is a bit larger than the median

25. $\frac{28.2 + 29.3 + 29.7 + \cdots + 37.3 + 38.5}{19} \approx \$33{,}542$, which is approximately the same as the median.

26. $\frac{6.2 + 6.3 + \cdots + 12.5 + 12.9}{44} \approx 9.4$ mph, which is a bit larger than the median.

27. $$\overline{x} = \frac{(5)(67{,}394) + (4)(45{,}523) + (3)(46{,}526) + (2)(27{,}216) + (1)(80{,}335)}{67{,}394 + 45{,}523 + 46{,}526 + 27{,}216 + 80{,}335} \approx 2.97$$

28. $$\overline{x} = \frac{(5)(47{,}553) + (4)(15{,}231) + (3)(14{,}957) + (2)(5076) + (1)(11{,}586)}{47{,}553 + 15{,}231 + 14{,}957 + 5076 + 11{,}586} \approx 3.87$$

29. The mean is $\frac{(16)(4) + (18)(1) + \cdots + (40)(1) + (42)(1)}{4 + 1 + 5 + 2 + 1 + 2 + 2 + 1 + 1} \approx 22.95$. The median is the better summary since the distribution is skewed to the right, which pulls the mean toward the high values.

30. The mean is $\frac{(2)(2) + (4)(3) + \cdots + (12)(2) + (13)(2)}{2 + 3 + \cdots + 2 + 2} \approx 7.97$. Either the mean or median could be used, since the distribution is roughly symmetric.

31. **(a)** Non-weighted: $\overline{x} = \frac{1}{12}(2 + 5 + 12 + 20 + \cdots + 10 + 3) = \frac{221}{12} \approx 18.42°\text{C}$

(b) Weighted: $\overline{x} = \frac{(2)(31) + (5)(28) + (12)(31) + (20)(30) + \cdots + (10)(30) + (3)(31)}{31 + 28 + 31 + 30 + \cdots + 30 + 31} = \frac{6748}{365} \approx 18.49°\text{C}$

(c) The weighted average is the better indicator.

32. **(a)** Non-weighted: $\overline{x} = \frac{1}{12}(-9 - 7 - 1 + 7 + \cdots - 1 - 7) = \frac{77}{12} \approx 6.42°\text{C}$

(b) Weighted: $\overline{x} = \frac{(-9)(31) + (-7)(28) + (-1)(31) + (7)(30) + \cdots + (-1)(30) + (-7)(31)}{31 + 28 + 31 + 30 + \cdots + 30 + 31} = \frac{2370}{365} \approx 6.49°\text{C}$

(c) The weighted average is the better indicator.

For #33–38, the best way to do the computation is with the statistics features of a calcuator.

33. $s \approx 9.71$, $s^2 = 94.3$

34. $s \approx 25.29$, $s^2 = 639.6$

35. $s \approx \$66.8$ billion; $s^2 \approx 4462$

36. $s \approx \$216.3$ billion; $s^2 \approx 46{,}793$ (46,785.7 using technology)

37. $s \approx \$2673$, $s^2 \approx 7{,}150{,}000$ (7,144,929 using technology)

38. $s \approx 1.55$ mph, $s^2 \approx 2.40$

39. **(a)** The standard deviation of the first is smaller; the values are closer to the mean.

(b) For the first data set, $s \approx 2.58$, which is smaller than $s \approx 2.94$ for the second data set.

40. **(a)** The standard deviation of both data sets are the same; the values are the same distance apart.

(b) The standard deviation for both data sets is $s \approx 7.91$.

41. It is possible for the standard deviation of a set to be zero, but all the numbers in the set would have to be the same.

42. The standard deviation of a set can never be negative, since it is the (positive) square root of the variance.

43. **(a)** 68%

(b) 2.5%

(c) A parameter, since it applies to the entire population.

44. **(a)** 16%

(b) 33

(c) No. The mean would have to be *weighted* according to the number of people in each state who took the ACT.

45. **(a)** 16%

(b) 13.5%

(c) Over 101 g

(d) Individuals more than 3 standard deviations below the mean are very rare.

46. **(a)** 0.025

(b) 0.815

(c) Falling, or rising by less than 3%

(d) Occurrences more than 3 standard deviations above the mean are very rare.

47. False. The median is a resistant measure. The *mean* is strongly affected by outliers.

48. True. The box extends from the first quartile, Q_1, to the third quartile, Q_3, and $Q_3 - Q_1$ is the interquartile range.

49. The plot of an ideal normal distribution is a symmetric "bell curve." The answer is A.

50. $$\bar{x} = \frac{10(3) + 9(3) + 8(5) + 7(6) + 6(4) + 5(3) + 4(1) + 3(0) + 2(0) + 1(0)}{25} = 7.28$$

The answer is B.

51. The total number of points from all 30 exams combined is $30 \times 81.3 = 2439$. Adding 9 more points and recalculating produces a new mean of $(2439 + 9)/30 = 81.6$. The median will be unaffected by an adjustment in the top score. The answer is B.

52. In a normal distribution, 95% of the data values lie within 2 standard deviations of the mean. The answer is C.

53. There are many possible answers; examples are given.

(a) $\{2, 2, 2, 3, 6, 8, 20\}$ — median $= 3$,

$\bar{x} = \dfrac{43}{7} \approx 6.14.$

(b) $\{1, 2, 3, 4, 6, 48, 48\}$ — median $= 4$, $\bar{x} = 16$.

(c) $\{-20, 1, 1, 1, 2, 3, 4, 5, 6\}$ — $\bar{x} = \dfrac{1}{3}$, median $= 2$.

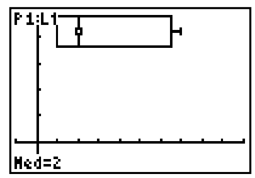
54. There are many possible answers; examples are given.

(a) $\{2, 4, 6, 8\}$ — $\sigma \approx 2.24$ and IQR $= 7 - 3 = 4$.

(b) $\{1, 5, 5, 6, 6, 9\}$ — IQR $= 6 - 5 = 1$, and $\sigma \approx 2.36$.

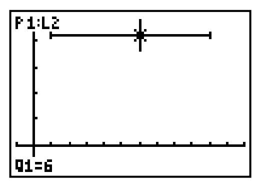
55. There are many possible answers; example data sets are given.

(a) $\{1, 1, 2, 6, 7\}$ — median $= 2$ and $\bar{x} = 3.4$.



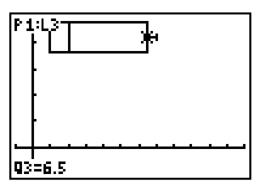
[–1, 10] by [–1, 5]

(b) $\{1, 6, 6, 6, 6, 10\}$ — $2 \times$ IQR $= 2(6 - 6) = 0$ and range $= 10 - 1 = 9$.



[–1, 12] by [–1, 5]

(c) $\{1, 1, 2, 6, 7\}$ — range $= 7 - 1 = 6$ and $2 \times$ IQR $= 2(6 - 1) = 10$.



[–1, 12] by [–1, 5]

56. One possible answer: $\{1, 2, 3, 4, 5, 6, 6, 6, 30\}$.

57. For women living in South American nations, the mean life expectancy is

$$\bar{x} = \frac{(79.1)(42.2) + (69.2)(10.3) + \ldots + (80.0)(3.3) + (79.3)(28.0)}{42.2 + 10.3 + \ldots + 3.3 + 28.0} = \frac{31{,}190.86}{397.9} \approx 78.4 \text{ years.}$$

58. For men living in South American nations, the mean life expectancy is

$$\bar{x} = \frac{(72.5)(42.2) + (65.1)(10.3) + \ldots + (73.0)(3.3) + (72.0)(28.0)}{42.2 + 10.3 + \ldots + 3.3 + 28.0} = \frac{28{,}685.43}{397.9} \approx 72.1 \text{ years.}$$

59. Since $\sigma = 0.05$ mm, we have $2\sigma = 0.1$ mm, so 95% of the ball bearings will be acceptable. Therefore, 5% will be rejected.

60. Use $\mu = 12.08$ and $\sigma = 0.04$.

Then $\mu - 2\sigma = 12.00$ and $\mu + 2\sigma = 12.16$, so 95% of the cans contain 12 to 12.16 oz of cola, 2.5% contain less than 12 oz, and 2.5% contain more than 12 oz. Therefore, 2.5% of the cans contain less than the advertised amount.

■ Section 10.4 Random Variables and Probability Models

Exploration 1

Plan A: $E(\text{cost}) = \$350$

Plan B: $E(\text{cost}) = 0(0.2) + 250(0.4) + 500(0.3) + 750(0.1) = \$325$

Plan C: $E(\text{cost}) = 300 + 0(0.2) + 0(0.4) + 50(0.3) + 100(0.1) = \$325$

Either plans B or C should cost the least.

Quick Review 10.4

1. There are two possible outcomes: {H, T}.
2. There are eight possible outcomes: {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}.
3. There are four possible outcomes: {0, 1, 2, 3}.
4. There are $\binom{44}{5} = \dfrac{44!}{5! \cdot (44-5)!} = 1{,}086{,}008$ possible outcomes.
5. There are $\binom{52}{3} = \dfrac{52!}{3! \cdot (52-3)!} = 22{,}100$ possible outcomes.
6. There are $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ possible outcomes.
7. $\binom{4}{2} = \dfrac{4!}{2! \cdot (4-2)!} = \dfrac{4!}{2! \cdot 2!} = \dfrac{24}{4} = 6$
8. $\binom{10}{0} = \dfrac{10!}{0! \cdot (10-0)!} = \dfrac{10!}{1 \cdot 10!} = 1$
9. $\binom{8}{3} = \dfrac{8!}{3! \cdot (8-3)!} = \dfrac{8 \cdot 7 \cdot 6 \cdot 5!}{3! \cdot 5!} = \dfrac{336}{6} = 56$
10. $\binom{100}{98} = \dfrac{100!}{98! \cdot (100-98)!} = \dfrac{100 \cdot 99 \cdot 98!}{98! \cdot 2!} = \dfrac{9900}{2} = 4950$

Section 10.4 Exercises

1. $E(X) = 10(0.5) + 20(0.3) + 30(0.2) = 17$
2. $E(X) = 2(0.1) + 4(0.2) + 6(0.3) + 8(0.4) = 6$
3. $E(X) = 5\left(\dfrac{3}{6}\right) + 10\left(\dfrac{2}{6}\right) + 25\left(\dfrac{1}{6}\right) = 10$
4. $P(X=1) = P(X=2) = P(X=3) = P(X=4) = P(X=5) = \dfrac{1-0.60}{5} = 0.08$

   $E(X) = 1(0.08) + 2(0.08) + 3(0.08) + 4(0.08) + 5(0.08) + 6(0.60) = 4.8$
5. $P(X=1) = P(X=2) = P(X=3) = \dfrac{1}{4}\, P(X=4) = P(X=5) = \dfrac{1-3/4}{2} = \dfrac{1}{8}$

   $E(X) = 1\left(\dfrac{1}{4}\right) + 2\left(\dfrac{1}{4}\right) + 3\left(\dfrac{1}{4}\right) + 4\left(\dfrac{1}{8}\right) + 5\left(\dfrac{1}{8}\right) = \dfrac{21}{8} = 2.625$
6. $E(X) = 1\left(\dfrac{7}{12}\right) + 5\left(\dfrac{3}{12}\right) + 10\left(\dfrac{2}{12}\right) = \dfrac{7}{2} = 3.5$
7. For each person, the expected value the carnival earns is $\$2(0.72) + (\$2 - \$5)(0.21) + (\$2 - \$10)(0.07) = \$0.25$. So, each day, the carnival can expect to win $300 \times \$0.25 = \$75$.
8. The expected payout of the game is $\$0\left(\dfrac{12}{52}\right) + \$1\left(\dfrac{4}{52}\right) + \$2\left(\dfrac{4}{52}\right) + \$3\left(\dfrac{4}{52}\right) + \$4\left(\dfrac{4}{52}\right) + \$5\left(\dfrac{4}{52}\right) + \$6\left(\dfrac{4}{52}\right) +$

   $\$7\left(\dfrac{4}{52}\right) + \$8\left(\dfrac{4}{52}\right) + \$9\left(\dfrac{4}{52}\right) + \$10\left(\dfrac{4}{52}\right) \approx \$4.23$. Since this is less than the \$5 to play, it would not be a good idea to play the game.

9. (a)

Y	1	2	3	4	5
$P(Y)$	$\frac{5}{20}$	$\frac{3}{20}$	$\frac{7}{20}$	$\frac{4}{20}$	$\frac{1}{20}$

(b) $E(Y) = 1\left(\frac{5}{20}\right) + 2\left(\frac{3}{20}\right) + 3\left(\frac{7}{20}\right) + 4\left(\frac{4}{20}\right) + 5\left(\frac{1}{20}\right) = \frac{53}{20} = 2.65$

10. The expected payout of the game is $\$0\left(\frac{4}{8}\right) + \$6\left(\frac{3}{8}\right) + \$10\left(\frac{1}{8}\right) = \$3.50$. Since it costs \$5.00 to play, you would expect to lose \$1.50 per game, so it would not be a good idea to play.

11. $P(X = \$0) = 1 - 0.04 - 0.01 = 0.95$. The expected loss is $\$0(0.95) + -\$200(0.04) + -\$300(0.01) = -\$11$. Since this is less than the expected loss of $-\$79$ to purchase the warranty, puchasing the warranty is not a good idea.

12. The expected cost for repairs is $\$0(0.94) + \$300(0.05) + \$500(0.01) = \$20$. Since this is less than the \$49 to purchase the warranty, then you should probably not buy the extended coverage.

13. (a)

Family	G	BG	BBG or BBB
X = children	1	2	3
$P(X)$	$\frac{1}{2}$	$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$	$1 - \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

(b) $E(X) = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{4}\right) = 1.75$

14. (a)

Shots good?	N	YN	YY
X	0	1	2
$P(X)$	0.20	$(0.80)(0.20) = 0.16$	$(0.80)(0.80) = 0.64$

(b) $E(X) = 0(0.20) + 1(0.16) + 2(0.64) = 1.44$

15. For the game to be fair, its expected value $(-\$5)\left(\frac{1}{2}\right) + (-\$5 + \$5)\left(\frac{1}{4}\right) + (-\$5 + \$10)\left(\frac{12}{52}\right) + (-\$5 + x)\left(\frac{1}{52}\right) = \$0$ so $\frac{1}{52}x = \frac{75}{52}$ so $x = \$75$.

16. The expected winnings on one ticket are $(\$0.50)\left(\frac{999}{1000}\right) + (-\$250)\left(\frac{1}{1000}\right) \approx \$0.25$. Since the state sells 1,000,000 tickets, the expecting daily profit for the state is $\$0.25(1{,}000{,}000) = \$250{,}000$.

17. (a) $P(X = 1) = \binom{4}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3 \approx 0.386$

(b) $P(X = 1) = \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 \approx 0.116$

(c) $P(X \geq 2) = \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 + \binom{4}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^1 + \binom{4}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^0 \approx 0.132$

18. (a) $P(X = 5) = \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 \approx 0.246$

(b) $P(X = 8) = \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 \approx 0.044$

(c) $P(X \geq 8) = \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^1 + \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0 \approx 0.055$

19. (a) $P(X = 5) = \binom{6}{5}(0.8)^5(0.2)^1 \approx 0.393$

(b) $P(X \leq 2) = \binom{6}{0}(0.8)^0(0.2)^6 + \binom{6}{1}(0.8)^1(0.2)^5 + \binom{6}{2}(0.8)^2(0.2)^4 \approx 0.017$

20. (a) $P(X = 3) = \binom{15}{3}(0.12)^3(0.88)^{12} \approx 0.170$

(b) $P(X \leq 2) = \binom{15}{0}(0.12)^0(0.88)^{15} + \binom{15}{1}(0.12)^1(0.88)^{14} + \binom{15}{2}(0.12)^2(0.88)^{13} \approx 0.735$

21. **(a)** $P(X = 5) = \binom{10}{5}(0.40)^5(0.60)^5 \approx 0.200$

(b) $P(X \geq 8) = \binom{10}{8}(0.40)^8(0.60)^2 + \binom{10}{9}(0.40)^9(0.60)^1 + \binom{10}{10}(0.40)^{10}(0.60)^0 \approx 0.012$

22. **(a)** $P(X = 1) = \binom{12}{1}(0.11)^1(0.89)^{11} \approx 0.366$

(b) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - \binom{12}{0}(0.11)^0(0.89)^{12} - \binom{12}{1}(0.11)^1(0.89)^{11} \approx 0.387$

23. **(a)** The value of X can take on more than two values.

(b) Let $X =$ the numbers of times you roll a "5" in 10 rolls.

24. **(a)** The value of X can take on more than two values.

(b) Let $X =$ the numbers aces drawn if five cards are drawn from the deck, with replacement.

25. **(a)** $\mu = np = 75(0.300) = 22.5$

(b) $\sigma = \sqrt{npq} = \sqrt{75(0.300)(0.700)} \approx 3.97$

26. **(a)** $\mu = np = 50(0.85) = 42.5$

(b) $\sigma = \sqrt{npq} = \sqrt{50(0.85)(0.15)} \approx 2.525$

27. For the math test, $z = \dfrac{x - \mu}{\sigma} = \dfrac{98 - 82}{7} \approx 2.29$. For the English test, $z = \dfrac{x - \mu}{\sigma} = \dfrac{98 - 86}{4} = 3.0$. You did better on the English test.

28. For the 100-m dash, $z = \dfrac{x - \mu}{\sigma} = \dfrac{10.2 - 10.4}{0.3} \approx -0.67$. For the long jump, $z = \dfrac{x - \mu}{\sigma} = \dfrac{22 - 20}{2.5} = 0.80$. The athlete did better in the long jump.

29. **(a)** $P(z > 1.5) \approx 0.067$

(b) $P(z < 0.75) \approx 0.773$

(c) $P(-1.8 < z < -0.5) \approx 0.273$

30. **(a)** $P(z > -1.2) \approx 0.885$

(b) $P(z < -0.6) \approx 0.274$

(c) $P(-2 < z < 1.25) \approx 0.872$

31. **(a)** InvNorm(0.97) ≈ 1.88

(b) InvNorm(0.60) ≈ 0.25

(c) InvNorm(0.10) ≈ -1.28, so ± 1.28

32. **(a)** InvNorm(0.15) ≈ -1.04

(b) InvNorm(0.92) ≈ 1.41

(c) InvNorm(0.05) ≈ -1.645, so ± 1.645

33. **(a)** $5'10'' = 5 \cdot 12 + 10 \approx 70''$. $z = \dfrac{70 - 63.8}{2.8} = 2.214$ and $P(z > 2.214) \approx 0.013$

(b) $5'6'' = 5 \cdot 12 + 6 \approx 66''$. $z = \dfrac{66 - 63.8}{2.8} = 0.786$ and $P(z < 0.786) \approx 0.784$

34. **(a)** $z = \dfrac{650 - 500}{100} = 1.5$ and $P(z \geq 1.5) \approx 0.067$

(b) $z = \dfrac{420 - 500}{100} = -0.8$ and $P(z \leq -0.8) \approx 0.212$

35. Using the grapher's inverse Normal function, $z \approx -1.645$, so $x = 63.8 - 1.645(2.8) = 59.2$ inches.

36. Using the grapher's inverse Normal function, $z \approx 2.054$, so $x = 500 + 2.054(100) = 705$.

37. **(a)** $z = \dfrac{10 - 13}{1.2} = -2.5$. $P(z \geq -2.5) \approx 0.994 = 99.4\%$

(b) $z = \dfrac{15 - 13}{1.2} \approx 1.667$ and $z = \dfrac{14 - 13}{1.2} \approx 0.833$. $P(0.833 < z < 1.667) \approx 0.155$

(c) Using the grapher's inverse Normal function, $z \approx -1.75$, so $x = 13 - 1.75(1.2) = 10.9$ hours.

38. **(a)** $z = \dfrac{40{,}000 - 37{,}200}{2650} \approx 1.06$. $P(z < 1.06) \approx 0.855 = 85.5\%$

(b) $z = \dfrac{35{,}000 - 37{,}200}{2650} \approx -0.83$ and $z = \dfrac{32{,}000 - 37{,}200}{2650} \approx -1.96$. $P(-1.96 < z < -0.83) \approx 0.178 = 17.8\%$

(c) Using the grapher's inverse Normal function, $z \approx -1.28$, so $x = 37{,}200 - 1.28(2650) \approx 33{,}800$ miles.

39. **(a)** Yes; $np = nq = 250(0.5) = 125 > 10$

(b) $\mu = np = 250(0.5) = 125$
$\sigma = \sqrt{npq} = \sqrt{250(0.5)(0.5)} \approx 7.91$

(c) No; $z = \dfrac{140 - 125}{7.91} = 1.90 < 2$

40. **(a)** Yes; $np = 320(0.04) = 12.8 > 10$ and $nq = 320(0.96) = 307.2 > 10$

(b) $\mu = np = 320(0.04) = 12.8$
$\sigma = \sqrt{npq} = \sqrt{320(0.04)(0.96)} \approx 3.505$

(c) Yes; $z = \dfrac{26 - 12.8}{3.505} = 3.77 > 3$

41. We would expect someone to pick the right card $\frac{1}{4}$ of the time, so $\mu = np = 100(0.25) = 25$ and $\sigma = \sqrt{npq} = \sqrt{100(0.25)(0.75)} \approx 4.33$.

An unusually high number of correct choices would need $z > 3$. If we let $x =$ the number of correct choices, then

$$\begin{aligned} z &> 3 \\ \frac{x-\mu}{\sigma} &> 3 \\ \frac{x-25}{4.33} &> 3 \\ x &> 25 + 3(4.33) \\ x &\geq 38 \end{aligned}$$

We would need to see at least 38 correct choices to convince us ESP really exists.

42. Since $\frac{1}{12}$ of the frogs showed the defect, $\mu = np = 150\left(\frac{1}{12}\right) = 12.5$ and $\sigma = \sqrt{npq} = \sqrt{150\left(\frac{1}{12}\right)\left(\frac{11}{12}\right)} \approx 3.385$.

An unusually high number of correct choices would need $z > 3$. If we let $x =$ defects found, then

$$\begin{aligned} z &> 3 \\ \frac{x-\mu}{\sigma} &> 3 \\ \frac{x-12.5}{3.385} &> 3 \\ x &> 12.5 + 3(3.385) \\ x &\geq 23 \end{aligned}$$

We would need to see at least 23 defects to convince us the condition is becoming more common.

43. False; The expected value is calculated from the probability model, not sample data.

44. False; such percentages apply only to Normal distributions.

45. $E(X) = 50(0.1) + 20(0.3) + 5(0.6) = 14$. The correct answer is D.

46. All other options violate at least one requirement of the binomial probability model. The correct answer is B.

47. $P(X = 1) = \binom{3}{1}(0.5)^1(0.5)^2 = 0.375 = \frac{3}{8}$. The correct answer is C.

48. $P(-0.5 < z < 0.5) \approx 0.38 = 38\%$. The correct answer is C.

49. **(a)** We expect $(8)(0.23) = 1.84$ (about 2) to be married.

(b) Yes, this would be an unusual sample.

(c) $P(5 \text{ or more are married}) = {}_8C_5 \cdot (0.23)^5(0.77)^3 + {}_8C_6 \cdot (0.23)^6(0.77)^2 + {}_8C_7 \cdot (0.23)^7(0.77)^1 + {}_8C_8 \cdot (0.23)^8(0.77)^0 \approx 0.01913 = 1.913\%$.

50. **(a)** ${}_{32}C_{17} \cdot (0.75)^{17}(0.25)^{15} \approx 0.00396 = 0.396\%$

(b) $\sum_{k=0}^{17} {}_{32}C_k \cdot (0.75)^k(0.25)^{17-k} \approx 0.00596 = 0.596\%$

(c) The university's graduation rate seems to be exaggerated; at least, this particular class did not fare as well as the university claims.

51. Choice 1: $E(X) = 10(0.20) + 20(0.20) + 50(0.20) + 100(0.20) + 5000(0.20) = \$1036$

Choice 2: $E(X) = 0\left(\frac{2}{3}\right) + 240 \cdot 20 \cdot \left(\frac{1}{3}\right) = \$1600$

Choice 3: $E(X) = 1000(1) = \$1000.$

Opinions will vary. Choice 2 has the highest expected value, so it is the option most contestants would choose, but many would opt for the guaranteed \$1000. Choice 3 has the lowest expected value, so that is the choice the sponsors would prefer contestants take.

52. **(a)** Choose green.

$P(\text{green wins}) = P(\text{red shows 2 and green shows 3}) = \left(\frac{5}{6}\right)\left(\frac{4}{6}\right) = \frac{20}{36} = \frac{5}{9}$ and $P(\text{red wins}) = 1 - P(\text{green wins}) = \frac{4}{9}.$

(b) Choose red.

For each roll of the red die, $E(\text{roll}) = 2\left(\frac{5}{6}\right) + 6\left(\frac{1}{6}\right) = 2.7$, while for each roll of the green die

$E(\text{roll}) = 1\left(\frac{2}{6}\right) + 3\left(\frac{4}{6}\right) \approx 2.3.$

53. **(a)** $E(X) = 0(0.80) + 10(0.15) + 50(0.04) + 100(0.01) = \$4.50$

(b) $\sigma = \sqrt{(0 - 4.50)^2(0.80) + (10 - 4.50)^2(0.15) + (50 - 4.50)^2(0.04) + (100 - 4.50)^2(0.01)}$
$= \$13.96$

(c) The standard deviation is large because the values of \$50 and \$100 are far from the mean of \$4.50.

(d) Opinions will vary. The expected winnings of the game are larger than \$5.00, so in the long run, playing the game would result in a net gain.

54. **(a)**

X = heads	$P(X)$
0	$\binom{4}{0}(0.5)^0(0.5)^4 = 0.0625$
1	$\binom{4}{1}(0.5)^1(0.5)^3 = 0.2500$
2	$\binom{4}{2}(0.5)^2(0.5)^2 = 0.3750$
3	$\binom{4}{3}(0.5)^3(0.5)^1 = 0.2500$
4	$\binom{4}{4}(0.5)^4(0.5)^0 = 0.0625$

(b) $E(X) = 0(0.0625) + 1(0.2500) + 2(0.375) + 3(0.2500) + 4(0.0625) = 2$

(c) $\sigma = \sqrt{(0 - 2)^2 \cdot (0.0625) + (1 - 2)^2 \cdot (0.2500) + (2 - 2)^2 \cdot (0.3750) + (3 - 2)^2 \cdot (0.25) + (4 - 2)^2 \cdot (0.0625)}$
$= 1$

(d) $np = (4)(0.5) = 2.0$ and $\sqrt{npq} = \sqrt{4(0.5)(0.5)} = 1$, which match the results for **(a)** and **(b)**.

55. **(a)** $\frac{325}{625} = 0.52 = 52\%$

(b) $\sqrt{625\left(\frac{325}{625}\right)\left(\frac{300}{625}\right)} \approx 12.49$

(c) Yes; $np = 325 > 0$ and $nq = 300 > 10$.

(d) $2\left(\frac{12.49}{625}\right) \approx 0.04$, so the margin of error is $\pm 4\%$.

(e) The confidence interval would be $52\% \pm 4\%$ or 48% to 56%. Since the interval contains 50%, it is not possible to determine if Candidate A has a majority of the votes.

56. If $np > 10$ an $nq > 0$, then $\sqrt{np} > 3$ abd $\sqrt{nq} > 3$.

For 0, $z = \dfrac{0 - np}{\sqrt{npq}} = -\dfrac{np}{\sqrt{npq}} = -\dfrac{\sqrt{np}}{\sqrt{q}} < -\dfrac{3}{\sqrt{q}} < -\dfrac{3}{\sqrt{1}}$, so $z < -3$.

For n, $z = \dfrac{n - np}{\sqrt{npq}} = \dfrac{n(1 - p)}{\sqrt{npq}} = \dfrac{nq}{\sqrt{npq}} = \dfrac{\sqrt{nq}}{\sqrt{p}} > \dfrac{3}{\sqrt{p}} > \dfrac{3}{\sqrt{1}}$, so $z > 3$.

Section 10.5 Statistical Literacy

Exploration 1

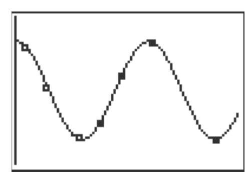
1. Correlation begins with a scatter plot, which requires numerical data from two quantitative variables (like height and weight). "Gender" is a categorical variable.
2. The sample did not represent the population of all voters very well, as many subgroups were underrepresented (for example, office workers). It was also not a good idea for the mayor to use his own staff to gather this kind of data.
3. The doctor's "experiment" proves nothing about the effect of vanilla gum on headache pain unless we can compare these subjects with a similar group that does not use vanilla gum. Many headaches are gone in two hours anyway.
4. The negative correlation simply showed that students with more absences tended to have lower GPAs. This does not imply that the absences caused the grades to go down. Perhaps getting low grades *caused* students to skip school. Or another variable, such as parental involvement, might have influenced both grades and absences. Correlation does not imply causation.
5. The percentages make the difference seem large, but it actually amounts to only 6 of the 50 people. The results may not be statistically significant.

Quick Review 10.5

1. $\dfrac{1}{6}$
2. $\dfrac{5}{36}$
3. $\dfrac{4}{52} = \dfrac{1}{13}$
4. $\dfrac{1}{10}$
5. $\dfrac{1}{10}$
6. $\left(\dfrac{1}{10}\right)\left(\dfrac{1}{10}\right) = \dfrac{1}{100}$
7. $\left(\dfrac{1}{10}\right)^5 = 0.00001$
8. $\left(\dfrac{9}{10}\right)^5 = 0.59049$
9. $1 - \left(\dfrac{9}{10}\right)^5 = 0.40951$
10. $\left(\dfrac{1}{10}\right)^5 + 5\left(\dfrac{1}{10}\right)^4\left(\dfrac{9}{10}\right) = 0.00046$

Section 10.5 Exercises

1. Correlation is being used incorrectly. Intelligence might be associated with some quantitative variable, but beauty is categorical.
2. Correlation is being used incorrectly. The correlation coefficient is usually not the slope of the regression line.
3. Correlation is being used incorrectly. The high correlation coefficient does nothing to support Sean's crazy theory, because the great blue whale (with a long name and a huge weight) is an unusual point that lies far away from the other three.
4. Correlation is being used correctly. Notice that Jenna's first observation does not commit her to a linear model, but her second sentence does. Her observation about the model is then appropriate.
5. Correlation is being used incorrectly. Marcus is OK with his first observation, but not with his second. While his linear model is a bad fit, he should not conclude that "there is no significant mathematical relationship." In fact, check out this sinusoidal fit:



6. Correlation is being used incorrectly. Correlation does not imply causation! (If you are wondering, though, the rainwater theory has been confirmed for several sea snake species through controlled laboratory experiments.)
7. This is a random sample (technically pseudo-random, but it should suffice).
8. This is not a random sample, since the students will come in five chunks of ten, each chunk probably being alphabetical. This could result in an unusual number of students in the sample who are related to one another.
9. This is not a random sample of all Reno citizens. All 50 selected are likely to be from early in the alphabet.
10. This is not a random sample, although there is quite a bit of chance involved. Notice, for example, that the five chosen students will probably not include any pairs of best friends, since they would probably have lined up together.
11. This is not a random sample, nor does it apparently try to be.
12. This is not a random sample, because it relies on human choice. It might appear to be random, but what would happen if one of the winners turned out (even by chance) to be related to one of the ushers?

13. Voluntary response bias. The students most likely to respond were those who felt strongly about suggestions for improvement, so the rate of negative responses was probably higher than the parameter. He could have gotten a less biased response with an in-class census of all his students (ideally in multiple-choice form so that their handwriting would not betray their identities).

14. Response bias. By giving free samples and money to the respondents, the company was influencing them to respond positively. The support for the new cereal was likely to be higher in this sample than in the population. The company would have gotten more reliable data with a comparative study of the new flavor against an established product.

15. Undercoverage bias. The survey systematically excluded the students who were not actually eating in the dining hall, so the sample statistic was bound to be higher than the population parameter. A better method would have been to choose a random sample from the student body first, then seek them out for the survey (perhaps in their homerooms).

16. Response bias and undercoverage bias. The question itself was clearly designed to elicit a positive response, and the sampling frame (PTA parents) was designed to survey only people who actually had children. The support in the sample was probably higher than in the population. Getting a truly random sample for polling purposes is notoriously difficult, but even a random sample from the phone book would improve on this one.

17. Response bias. The question was designed to elicit a negative response, and it never even mentioned stop signs. The 97% was much higher than it would have been with a simple question like, "Should citizens be allowed to ignore stop signs?"

18. Undercoverage bias. Mrs. Bohackett (who never attempted to sample randomly) is making a global conclusion on the basis of her own back yard, which might reflect white-throated sparrow migration in unusual ways. The evidence would be more persuasive if similar observations were made at many other feeders, but causation could not be established without experimentation.

19. This is an observational study since no treatment was imposed.

20. This is an observational study since no treatment was imposed.

21. This is an observational study since no treatment was imposed.

22. This is an experiment since a treatment was imposed.

23. This is an experiment since a treatment was imposed.

24. This is an observational study since no treatment was imposed.

25. Using random numbers, select 12 of the 24 plots to get the new fertilizer. Use the original fertilizer on the other 12 plots. Compare the yields at harvest time.

26. The plots can be arranged in pairs of similar productivity, then one member of each pair can be randomly assigned to receive the new fertilizer.

27. This requires three treatments. Split the 24 plots randomly into three groups of 8: new fertilizer 1, new fertilizer 2, and original fertilizer.

28. This requires four treatments. Split the 24 plots randomly into four groups of 6: crop 1 with new fertilizer, crop 1 with original fertilizer, crop 2 with new fertilizer, and crop 2 with original fertilizer.

29. Fatigue may be a factor after they have driven 20 golf balls. They could gather the data on different days, or they could randomly choose half the golfers to drive the new ball first.

30. The tasters should be blinded to the containers. Ideally, the drinks should be poured into identical cups for presentation (being careful not to lose the fizz).

31. The music assignment should be randomized, not left to the choice of the mother. Otherwise, the mother's music preference (with possible lifestyle implications) becomes a potentially significant confounding variable.

32. Because of the nature of the experiment, it should be blocked for gender to be sure that each group contains an equal mix of male and female listeners.

33. One possible solution: Use the command "randInt(1, 500, 50)" to choose 50 random numbers from 1 to 500. If there are any repeat numbers in the list, use "randInt(1, 500)" to pick additional numbers until you have a sample of 50.

34. One possible solution: Use the command "randInt(1, 400, 100)" to choose 100 random numbers from 1 to 400. If there are any repeat numbers in the list, use "randInt(1, 400)" to pick additional numbers until you have a sample of 100.

35. One possible solution: Enter the numbers 1 to 32 in list L1 using the command "seq(X, X, 1, 32) → L1" and enter 32 random numbers in list L2 using the command "rand(32) → L2 ." Then sort the random numbers into ascending order, bringing L1 along for the ride, using the command "SortA(L2,L1)." The numbers in list L1 are now in random order.

36. One possible solution: Enter the numbers 1 to 28 in list L1 using the command "seq(X, X, 1, 28) → L1" and enter 28 random numbers in list L2 using the command "rand(28) → L2 ." Then sort the random numbers into ascending order, bringing L1 along for the ride, using the command "SortA(L2,L1)." Read off the randomly sorted numbers in L1 in order by pairs.

37. Number the plants 1–16. Use "randInt(1, 16)" to generate 8 distinct random numbers. Grow those 8 plants with the plant food and the other 8 without it.

38. Number the bowls 1–10. Use "randInt(1, 10)" to generate 5 distinct random numbers. Fire those 5 bowls in her friend's kiln and the others in her own.

39. One possible solution: Use the command "randInt(1, 8)" to generate random numbers between 1 and 8.

40. One possible solution: Use the command "randInt(1, 6)" to generate random numbers between 1 and 6. Push ENTER twice to get a roll of two dice. (Note that you do not want to generate random totals between 2 and 12. You learned in Section 9.3 that those totals are not equally likely.)

41. One possible solution: Use the command "randInt(1, 5, 20)" to generate 20 random numbers from 1 to 5. Let 1 and 2 designate donors with O-positive blood. Do this nine times and keep track of how many strings have fewer than four numbers that are 1 or 2.

42. One possible solution: Number the cards from 1 to 52. Use the command "randInt(1, 52,5)" to choose 5 random numbers from 1 to 52. If there are any repeat numbers in the list, use "randInt(1, 52)" to pick additional numbers until you have a sample "hand" of 5.

43. Results will vary. Use "randInt(1, 6)" to generate a series of random rolls of the die. Keep a running total, but don't add rolls that would make the sum greater than 21. Stop when the total equals 21. Report the number of rolls.

44. Results will vary. Use "randInt(1, 4)" to generate a series of random spins. Keep a running total, but subtract any spin that would make the sum greater than 10. Stop when the total equals 10. Report the number of spins.

45. Yes, there is enough evidence to warrant suspicion. In only 10 of the 500 simulated trials did 8 or more 6's show up. There's only a 2% chance that rolling a die fairly would produce a result like this.

46. **(a)** A difference at least this high happened often just by chance.

(b) Based on the dotplot, it appears differences of 1.5 and above happen by chance less than 5% of the time.

47. False. Observational studies can show strong associations, but experiments would be required to establish causation.

48. False. The underlying relationship might not be linear, in which case the magnitude of the r value is not a relevant measure.

49. Freshman science grades is the only quantitative variable among the choices. The answer is B.

50. The answer is A.

51. The answer is C.

52. People late in the alphabet are less likely to be chosen. The answer is D.

53. Answers will vary. Note that you should not expect all the counts to be exactly the same (that would suggest nonrandomness in itself), but "randomness" would predict an approximately equal distribution, especially for a large class.

54. Answers will vary, but the count is likely to be low. For example, a class of 20 would be looking at 500 two-digit numbers, about 50 of which should be double digit.

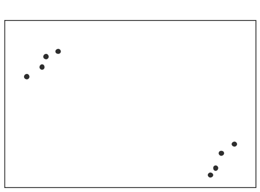
55. **(a)** The correlation coefficient will increase and the slope will remain about the same.

(b) The correlation coefficient will increase and the slope will increase.

(c) The correlation coefficient will decrease and the slope will decrease.

56. The points in **(b)** and **(c)** are influential points.

57. One possible picture:



58. One possible picture:

59. **(a)** The size of the hospital is not affecting the death rates of the patients. The lurking variable is the patient's condition. Bigger hospitals tend to get the more critical cases, and critical cases have a higher death rate.

(b) The number of seats is not affecting the speed of the jet. The lurking variable is the size of the aircraft. Larger jets generally have more seats and go faster.

(c) The size of the shoe does not affect reading ability. The lurking variable is the age of the student. In general, older students have larger feet and read at a higher level.

(d) The extra firemen are not causing more damage. The lurking variable is the size of the fire. Larger fires cause more damage and require more firefighters.

(e) The salary is not generally affected by the player's weight. The lurking variable is the player's position on the team. Linemen weigh more and tend to earn less money than the (usually lighter) players in the so-called "skill" positions (e.g., quarterbacks, running backs, receivers, and defensive backs).

60. **(a)** Prospective

(b) Retrospective

(c) Retrospective

(d) Prospective

(e) Retrospective

■ Chapter 10 Review

1. This is not a valid probability function. The sum of the probabilities $0.45 + 0.25 + 0.15 + 0.05 = 0.9 \neq 1$.

2.
$$P(B) = 2P(A);\ P(C) = \frac{1}{3}P(A)$$
$$P(A) + P(B) + P(C) = 1$$
$$P(A) + 2P(A) + \frac{1}{3}P(A) = 1$$
$$\frac{10}{3}P(A) = 1$$
$$P(A) = \frac{3}{10} = 0.3$$
$$P(B) = 2P(A) = 2(0.3) = 0.6$$
$$P(C) = \frac{1}{3}P(A) = \frac{1}{3}(0.3) = 0.1$$

3. There are ${}_{40}C_5 = 658{,}008$ different possible outcomes, so $P(\text{Winning}) = \dfrac{12}{658{,}008} \approx 0.00001824$

4. There are ${}_{52}C_5 = 2{,}598{,}960$ possible poker hands and ${}_{13}C_5 = 1287$ ways to draw all hearts, so $P(\text{all hearts}) = \dfrac{1287}{2{,}598{,}960} \approx 0.0005.$

5. $P(\text{caramel on third pick}) = \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{6}{8} = \frac{1}{10}$

6. $P(\text{red, white, blue}) = \frac{5}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{24}$

7. **(a)** $P(\text{All}) = (0.85)^5 \approx 0.444$

(b) $P(\text{4th}) = (0.85)(0.85)(0.85)(0.15) \approx 0.092$

(c) $P(\text{At least one is not wearing seatbelt}) = 1 - P(\text{All are wearing seatbelts}) = 1 - (0.85)^5 \approx 0.556$

8. **(a)** $P(\text{First strike in third frame}) = (0.60)(0.60)(0.40) = 0.144$

(b) $P(\text{At least one strike in 5 frames}) = 1 - P(\text{No strikes}) = 1 - (0.60)^5 \approx 0.922$

(c) $P(\text{No strikes in 10 frames}) = (0.60)^{10} \approx 0.006$

9. **(a)** $P(\text{Refrigerator or TV}) = P(\text{Refrigerator}) + P(\text{TV}) - P(\text{Refrigerator and TV})$
$= 0.66 + 0.41 - 0.32 = 0.75$

(b) $P(\text{TV} \mid \text{Refrigerator}) = \dfrac{P(\text{Refrigerator and TV})}{P(\text{Refrigerator})}$
$= \dfrac{0.32}{0.66} \approx 0.48$

10. There are not independent, $P(\text{TV} \mid \text{Refrigerator}) \neq P(\text{TV})$.

11. $P(\text{campylobacter or salmonella})$
$= P(\text{campylobacter}) + P(\text{salmonella}) - P(\text{campylobacter and salmonella})$
$= 0.81 + 0.15 - 0.13 = 0.83$

$P(\text{bacteria free})$
$= 1 - P(\text{campylobacter or salmonella}) = 1 - 0.83 = 0.17$

12. Possibly; $P(\text{salmonella} \mid \text{campylobacter}) = \frac{0.13}{0.81} = 0.16$, which is close to $P(\text{salmonella}) = 0.15$.

13. **(a)** $P(\text{brand } A) = 0.5$

(b) $P(\text{cashews from brand } A) = (0.5)(0.3) = 0.15$

(c) $P(\text{cashew}) = (0.5)(0.3) + (0.5)(0.4) = 0.35$

(d) $P(\text{brand } A \mid \text{cashew}) = \frac{0.15}{0.35} \approx 0.43$

14. **(a)** $P(\text{track wet and Mudder Earth wins}) = (0.80)(0.70) = 0.56$

(b) $P(\text{track dry and Mudder Earth wins})$
$= (0.20)(0.40) = 0.08$

(c) $0.56 + 0.08 = 0.64$

(d) $P(\text{track wet} \mid \text{Mudder Earth wins}) = \frac{0.56}{0.64} = 0.875$

15. **(a)** $P(\text{high blood pressure and high cholesterol}) = \frac{22}{88} = 0.25$

(b) $P(\text{high cholesterol} \mid \text{high blood pressure}) = \frac{22}{34} \approx 0.647$

(c) No; $P(\text{high cholesterol} \mid \text{high blood pressure}) \approx 0.647$, but $(\text{high cholesterol}) = \frac{28}{88} \approx 0.318$

16.

	Boys	Girls
At or above	$\frac{84}{180}(150) = 70$	$\frac{96}{180}(150) = 80$
Below	$\frac{84}{180}(30) = 14$	$\frac{96}{180}(30) = 16$

17. (a)

1	1
2	3 3 5 7 8
3	1 3 4 4 6 9
4	2 2 6 6
5	0 6

(b) Unimodal and symmetric.

18. (a) For the data {1112, 2327, 2382, 2521, 2709, 2806, 3127, 3338, 3485, 3489, 3631, 3959, 4228, 4264, 4689, 4690, 5000, 5648}, so the median $= \dfrac{3485 + 3489}{2} = 3487$

Range $= 5648 - 1112 = 4536$

$IQR = 4264 - 2709 = 1555$

Five-number summary is {1112, 2709, 3487, 4264, 5648}

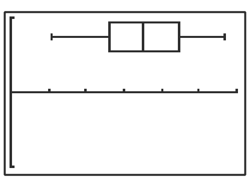
(b) $\bar{x} = 3522.5$ yards; $s \approx 1126.8$ yards

(c) The mean and standard deviation; the distribution is symmetric and there are no outliers.

19.

Yardage	Frequency
1000–1999	1
2000–2999	5
3000–3999	6
4000–4999	4
5000–5999	2

20.



[0, 6000] by [–1, 1]

$Q_1 - 1.5 \times IQR = 2709 - 1.5 \times 1555 = 376.5$ and
$Q_3 + 1.5 \times IQR = 4264 + 1.5 \times 1555 = 6596.5$;
there are no outliers.

21. (a)

12	0 0 4 4
13	1 1 2 6 7 9
14	0 3 4 8
15	6
16	3
17	7 9
18	0
19	0 1 7
20	2
21	
22	
23	0

(b) Unimodal and skewed to the right.

22. (a) For the data {120, 120, 124, 124, 131, 131, 132, 136, 137, 139, 140, 143, 144, 148, 156, 163, 177, 179, 180, 190, 191, 197, 202, 230}, so the median $= \dfrac{143 + 144}{2} = \dfrac{287}{2} = 143.5$

Range $= 230 - 120 = 110$

$Q_1 = \dfrac{131 + 132}{2} = 131.5$; $Q_3 = \dfrac{179 + 180}{2} = 179.5$;

$IQR = 179.5 - 131.5 = 48$

Five-number summary is {120, 131.5, 143.5, 179.5, 230}.

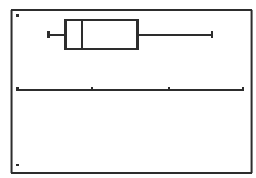
(b) $\bar{x} = 155.6$ sec; $s \approx 30.5$ sec

(c) The median and IQR; the distribution is skewed

23.

Length (in seconds)	Frequency
120–129	4
130–139	6
140–149	4
150–159	1
160–169	1
170–179	2
180–189	1
190–199	3
200–209	1
210–219	0
220–229	0
230–239	1

24.



[100, 250] by [–1, 1]

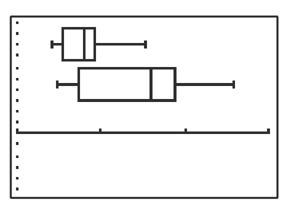
$Q_1 - 1.5 \times IQR = 131.5 - 1.5 \times 48 = 59.5$ and
$Q_3 + 1.5 \times IQR = 179.5 + 1.5 \times 48 = 251.5$; there are no outliers.

25.

4 0 0	12	4
9 2 1	13	1 6 7
8 4 3 0	14	
	15	6
3	16	
7	17	9
	18	0
	19	0 1 7
	20	2
	21	
	22	
	23	0

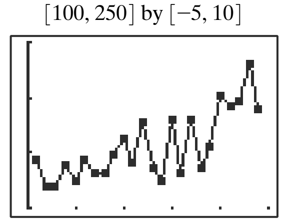
The songs released in the earlier years tended to be shorter.

26. Earlier years are in the upper box plot. The range and interquartile range are both greater in the lower graph, which shows the times for later years.



[100, 250] by [−5, 10]

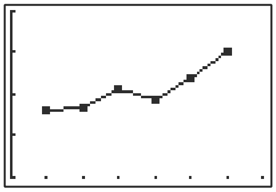
27.



[−1, 25] by [100, 250]

Again, the data demonstrates that songs appearing later tended to be longer in length.

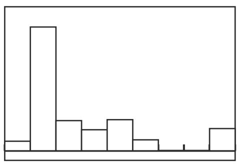
28.



[0, 7] by [50, 250]

The average times are {130.5, 132.75, 157, 142.5, 168.75, 202}. The trend is clearly increasing overall, with less fluctuation than the time plot for Exercise 27.

29. (a)



[0, 900] by [–1, 15]

(b) Unimodal and skewed to the right.

30. (a) For the data {98, 110, 110, 110, 110, 120, 120, 130, 140, 140, 160, 160, 160, 190, 230, 250, 250, 300, 340, 410, 440, 490, 590, 800, 880}, so the median = 160.

Range = 880 − 98 = 782

$Q_1 = \dfrac{120 + 120}{2} = 120$; $Q_3 = \dfrac{340 + 410}{2} = 375$;

$IQR = 375 - 120 = 255$

Five-number summary is {98, 120, 160, 375, 880}.

(b) $\bar{x} = 273.52$ million visitors; $s \approx 217.3$ million visitors

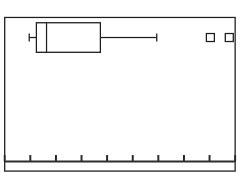
(c) The mean is much larger because there are some very high values.

(d) The median and IQR; the distribution is skewed.

31.

Visitors (in millions)	Frequency
0–99	1
100–199	13
200–299	3
300–399	2
400–499	3
500–599	1
600–699	0
700–799	0
800–899	1

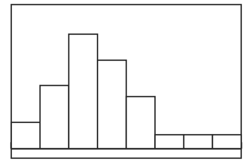
32.



[0, 900] by [–1, 15]

$Q_1 - 1.5 \times IQR = 120 - 1.5 \times 255 = -262.5$ and $Q_3 + 1.5 \times IQR = 375 + 1.5 \times 255 = 757.5$; 800 million and 880 million are outliers.

33. (a)



[75, 275] by [–1, 11]

(b) Unimodal and slightly skewed.

34. (a) Median $= \dfrac{149.1 + 149.2}{2} = 149.15 = \$149{,}150$

Range $= 252.4 - 80.4 = 172 = \$172{,}000$

$IQR = 173.9 - 126.1 = 47.8 = \$47{,}800$

Five-number summary {\$80,400, \$126,100, \$149,150, \$173,900, \$252,400}

(b) $\bar{x} = \$151{,}131$; $s \approx \$38{,}803$

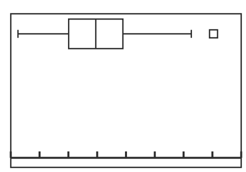
(c) The mean and median are very similar because the distribution is nearly symmetric.

(d) The mean and standard deviation; the distribution is nearly symmetric

35.

Median price (in \$1000s)	Frequency
75–99	2
100–124	5
125–149	9
150–174	7
175–199	4
200–224	1
225–249	1
250–274	1

36.



[75, 275] by [–1, 11]

$Q_1 - 1.5 \times IQR = 126{,}100 - 1.5 \times 47{,}800 = \$54{,}400$ and $Q_3 + 1.5 \times IQR = 173{,}900 + 1.5 \times 47{,}800 = \$245{,}600$; \$252,400 is an outlier.

37. (a) $0.301 = 0.274 + 1(0.027)$, so $100\% - \dfrac{32\%}{2} = 84\%$

(b) Between $0.274 - 2(0.027) = 0.220$ and $0.274 + 2(0.027) = 0.328$

(c) $z = \dfrac{0.356 - 0.274}{0.027} = 3.037$. This batting average is unusually high.

38. (a) Between $1309 - 1(157) = 1152$ lb and $1309 + 1(157) = 1466$ lb.

(b) Above $1309 + 2(157) = 1623.0$ lb.

(c) $z = \dfrac{2527 - 1309}{157} = 7.76$ This weight is extremely unusual.

39. $E(X) = 100(0.4) + 150(0.3) + 200(0.2) + 500(0.1) = 175$

40. $E(X) = -10(0.60) + 1(0.25) + 5(0.10) + x(0.05) = 0$
$\Rightarrow 0.05x - 5.25 = 0 \Rightarrow 0.05x = 5.25 \Rightarrow x = 105.$
The missing prize should be worth \$105.

41. $P(\text{Calipers need cleaning}) = 0.35,$
$P(\text{Calipers need replacing}) = (0.65)(0.90) = 0.585$, and
$P(\text{Calipers need overhaul}) = (0.65)(0.10) = 0.065.$
$E(\text{Cost}) = 50(0.35) + 160(0.585) + 350(0.065) = \$133.85.$

42. $P(\text{Gets neither contract})$
$= 1 - P(\text{Gets 1st or 2nd contract})$
$= 1 - (P(\text{Gets 1st contract}) + P(\text{Gets 2nd contract}) - P(\text{Gets 1st and 2nd contract}))$
$= 1 - (0.60 + 0.35 - 0.10) = 0.15$
$P(\text{Only gets 1st contract}) = 0.60 - 0.10 = 0.5,$
$P(\text{Only gets 2nd contract}) = 0.35 - 0.10 = 0.25.$
$E(\text{Revenue}) = 0(0.15) + (8000)(0.50) + (12000)(0.25) + (8000 + 12000)(0.10) = \$9000$, so $E(\text{Profit}) = 9000 - 500 = \$8500.$

43. $P(\text{2 H and 3 T}) = \binom{5}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5}{16}$

44. $P(\text{1 H and 3 T}) = \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{1}{4}$

45. (a) $P(\text{no defective bats}) = (0.98)^4 \approx 0.922$

(b) $P(\text{one defective bat}) = {}_4C_1 \cdot (0.98)^3(0.02) = 4(0.98)^3(0.02) \approx 0.075$

46. $\mu = np = 240(0.02) = 4.8$ and
$\sigma = \sqrt{npq} = \sqrt{240(0.02)(0.98)} \approx 2.17$

47. No; $np = 240(0.02) = 4.8 < 10.$

48. (a) $P(\text{no defective light bulbs}) = (0.9996)^{10} \approx 0.996$

(b) $P(\text{two defective light bulbs}) = {}_{10}C_2 \cdot (0.9996)^8(0.0004)^2 \approx 7.18 \times 10^{-6}$

49. $\mu = np = 100{,}000(0.0004) = 40$ and
$\sigma = \sqrt{npq} = \sqrt{100{,}000(0.0004)(0.9996)} \approx 6.23$

50. Yes; $np = 100{,}000(0.0004) = 40 > 10$ and
$nq = 100{,}000(0.9996) = 99960 > 10$

51. (a) $\mu = np = 50(0.65) = 32.5$ and
$\sigma = \sqrt{npq} = \sqrt{50(0.65)(0.35)} = 3.37$

(b) Yes; $np = 50(0.65) = 32.5 > 10$ and
$nq = 50(0.35) = 17.5 > 10$

(c) Yes; $z = \frac{41 - 32.5}{3.37} = 2.52 > 2$, which is statistically significant.

52. Since $np = 450(0.08) = 36 > 10$ and
$nq = 450(0.92) = 414 > 10$ a Normal model can be used.
$\sigma = \sqrt{npq} = \sqrt{450(0.08)(0.92)} \approx 5.75$ and
$z = \frac{29 - 36}{5.75} = -1.2 < 2$, which is not statistically significant, so the 8% estimate was not misleading.

53. (a) $z = \frac{1500 - 1309}{157} \approx 1.22$ and $P(z \geq 1.22) \approx 0.11$

(b) $z = \frac{1100 - 1309}{157} \approx -1.331,$
$z = \frac{1250 - 1309}{157} \approx -0.376$, and
$P(-1.331 \leq z \leq -0.376) \approx 0.26$

(c) Using the grapher's inverse Normal function, $z \approx -1.282$, so 10% of steers would weigh less than $1309 - 1.282(157) \approx 1108$ pounds.

54. Using the grapher's inverse Normal function, for Q_1, $z \approx -0.6745$, so $Q_1 = 1309 - 0.6745(157) = 1203.1$ pounds.
Using the grapher's inverse Normal function, for Q_3, $z \approx 0.6745$, so $Q_3 = 1309 + 0.6745(155) = 1413.5$ pounds.
$IQR = Q_3 - Q_1 = 1413.5 - 1203.1 = 210.4$ pounds.
(Using unrounded values, the answer is 211.8 pounds.)

55. (a) $z = \frac{275 - 287}{9} \approx -1.33$ and $P(z \geq -1.33) \approx 0.91$

(b) $z = \frac{290 - 287}{9} \approx 0.333$, $z = \frac{300 - 287}{9} \approx 1.444$, and
$P(0.333 \leq z \leq 1.444) \approx 0.295$

(c) Using the grapher's inverse Normal function, $z \approx 2.326$, so 1% of all drives would be less than $287 + 2.326(9) \approx 308$ yards.

56. Using the grapher's inverse Normal function, for Q_1, $z \approx -0.6745$, so $Q_1 = 287 - 0.6745(9) = 281.0$ yards.
Using the grapher's inverse Normal function, for Q_3, $z \approx 0.6745$, so $Q_3 = 287 + 0.6745(9) = 293.1$ yards.
$IQR = Q_3 - Q_1 = 293.1 - 281.0 \approx 12.1$ yards.

57. Correlation is incorrect, because color is categorical.

58. (a) Answers will vary. One possible answer is: Correlation does not imply causation.

(b) Answers will vary. One possible answer is: Perhaps people who are slower drivers tend to buy white cars.

59. Correlation does not measure straightness.

60. The student should not have calculated the correlation, because the relationship appears to be curved.

61. Yes, using a random number generator will result in a random sample.

62. No, this is not a random sample of all Atlanta citizens. All people in the sample are likely to be those whose names come early in the alphabet.

63. This sampling method will have voluntary response bias.

64. This is an observational study since no treatment was imposed.

65. Randomly divide the students into two groups of 20 (replication). Have one group take the course and the other group study independently (control). Compare improvement in scores.

66. Assign students numbers 1–40. Use randInt(1, 40) to generate 20 distinct random numbers. Assign those students to take the course and the others to study independently.

67. Randomly divide the swatches into two groups of 10 (replication). Wash one group with the old detergent and the other with the new additive. Wash each in the same machine for the same length of time using the same temperature water (control). Compare the cleanliness of the swatches.

68. Number the swatches 1–20. Use randInt(1, 20) to generate 10 distinct random numbers. Wash those swatches with the additive and the others without.

69. Use randInt(0, 9) to generate random digits, letting $0 - 4 =$ red, $5 - 7 =$ white, and $8 - 9 =$ blue. Generate a series of digits until one marble of each color is seen. Record the number of marbles not drawn. Repeat many times and find the mean.

70. Use randInt(1, 6, 5) to generate 5 random digits representing the dice. Repeat many times, then calculate the percent of times a full house appeared.

71. **(a)** No; 21 of the 374 streaks were runs of 5 or more in a row, which is not unusual.

(b) Answers will vary. One possible answer is: 8 or more in a row seems unusual, but even very long runs can occur by chance.

72. Yes; differences at least this large are statistically significant, because they occurred by chance only twice in 500 trials.

Chapter 10 Project

Answers are based on the sample data shown in the table.

1.

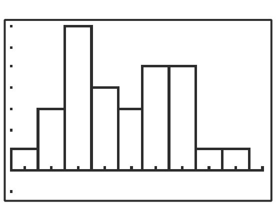
Stem	Leaf
5	
5	9
6	1 1 2 3 3 3 4 4 4 4
6	5 6 6 6 7 8 8 9 9 9
7	0 0 1 1 1 2 2 3
7	5

The average is about 66 or 67 inches.

2. A large number of students are between 63 and 64 inches and also between 69 and 72 inches.

Height	Frequency
59–60	1
61–62	3
63–64	7
65–66	4
67–68	3
69–70	5
71–72	5
73–74	1
75–76	1

3.



[59, 78] by [−1, 7]

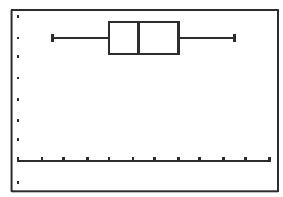
Again, the average appears to be about 66 or 67 inches. Since the data are not broken out by gender, one can only speculate about average heights for males and females separately. Possibly the two peaks within the distribution represent an average height of 63–64 inches for females and 70–71 inches for males.

4. Mean $= 66.9$ in.; median $= 66.5$ in.; mode $= 64$ in. The mean and median both appear to be good measures of the average, but the mode is too low. Still, the mode might well be similar in other classes.

5. The data set is somewhat symmetric and probably does not have outliers.

6. The stem-and-leaf plot puts the data in order.
Minimum value: 59
Maximum value: 75
Median: $\dfrac{66 + 67}{2} = 66.5$
Q_1: 64
Q_3: 70
The five-number summary is $\{59, 64, 66.5, 70, 75\}$.

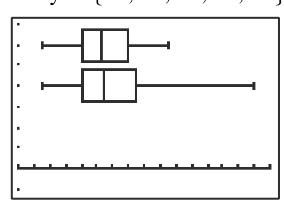
7.



[56, 78] by [−1, 7]

The box plot visually represents the five-number summary. The whisker-to-whisker size of the box plot represents the range of the data, while the width of the box represents the interquartile range.

8. Mean $= 67.5$; median $= 67$; The new five-number summary is $\{59, 64, 67, 71, 86\}$.



[56, 88] by [−1, 7]

The minimum and first quartile are unaffected, but the median, third quartile, and maximum are shifted to varying degrees.

9. The new student's height, 86 inches, lies 15 inches away from Q_3, and that is more than $1.5(Q_3 - Q_1) = 10.5$. The height of the new student should probably be tossed out during prediction calculations.

10. $s = 4.1$.

Interval	Heights	Expected	Observed
$\bar{x} \pm 1s$	62.8–70.9	68%	63.3%
$\bar{x} \pm 2s$	58.7–75.0	95%	96.7%
$\bar{x} \pm 3s$	54.6–79.1	99.7%	100%