Section 7.1: How Sample Proportions Vary Around the Population Proportion

7.1 Simulating the exit poll

a) Answers for the sample proportion will vary. Although the population proportion is 0.53, it is unlikely that exactly 53 out of 100 polled voters will vote yes. Sample proportions close to 0.53 will be more likely than those further from 0.53.

b) Simulations will vary. The graph of the sample proportions should be close to bell-shaped and centered around 0.53.

c) The predicted standard deviation is $\sqrt{0.53(1-0.53)/100} = 0.0499$.

d) The graph should look similar but shifted so that it is centered around 0.70. The standard deviation changes to $\sqrt{0.70(1-0.70)/100} = 0.0458$.

7.2 Simulate condo solicitations

a) Answers for the sample proportion will vary. Although the population proportion is 0.10, it is unlikely that exactly 10% of the customers in the sample will accept the offer. Sample proportions close to 0.10 will be more likely than those further from 0.10.

b) Simulations will vary. The graph of the 100 sample proportion values should be approximately bell shaped and centered around 0.10. Yes, as the simulation shows, almost all sample proportions will fall between 0.05 and 0.15.

7.3 House owners in a district

a) Mean = p = 0.30, standard deviation $= \sqrt{\dfrac{0.30(1-0.30)}{400}} = 0.02291$.

b) Mean = p = 0.30, standard deviation $= \sqrt{\dfrac{0.30(1-0.30)}{1600}} = 0.01146$.

c) Mean = p = 0.30, standard deviation $= \sqrt{\dfrac{0.30(1-0.30)}{100}} = 0.04583$.

d) As the sample size gets larger (i.e., from 400 to 1600), the standard deviation decreases (in fact, it is only half as large). As the sample size gets smaller (i.e., from 400 to 100), the standard deviation increases (in fact, it becomes twice as large).

7.4 iPhone apps

a) The population distribution is the set of all x values for the population of people who own an iPhone, 25% of which are 1 (individual has the given app) and 75% of which are 0 (individual does not have the app). $P(X = 1) = 0.25$, $P(X = 0) = 0.75$

b) Don't have the app: $P(X = 0) = 0.40$, Have the app: $P(X = 1) = 0.60$

c) The mean is $p = 0.25$.

d) The standard deviation is $\sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.25(1-0.25)}{50}} = 0.061$.

e) The standard deviation describes how much the sample proportion varies from one sample (of size 50) to the next. Since the sampling distribution is approximately normal, most of the sample proportions will fall within 3(0.061) = 0.183 of the mean (0.25).

7.5 Other scenario for exit poll

a) The binary variable is whether the voter voted for Whitman (1) or not (0). For each observation, P(1) = 0.409 and P(0) = 0.591.

b) The mean is $p = 0.409$ and the standard deviation is $\sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.409(1-0.409)}{3889}} = 0.0079$.

7.6 Exit poll and *n*

a) The interval of values within the sample proportion will almost certainly fall within three standard deviations of the mean: 0.409 – 3(0.008) = 0.385 to 0.409 + 3(0.008) = 0.433.

b) Since 0.424 falls within the interval calculated in (a), it is one of the plausible values.

7.7 Random variability in baseball

a) The mean is equal to p or 0.30. The standard deviation is $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.30(1-0.30)}{500}} = 0.0205$. The shape is approximately normal with a mean of 0.30 and standard deviation of 0.0205.

b) $\frac{0.32-0.30}{0.020} = 1.0$ and $\frac{0.28-0.30}{0.020} = -1.0$; Since both 0.32 and 0.28 are about one standard deviation from the mean, they would not be considered unusual for this player's year-end batting average.

7.8 Awareness for cancer

a)

Sample Proportion	Probability
0	0.50
1	0.50

b)

Sample Proportion	Probability
0	0.25
½	0.50
1	0.25

c)

Sample Proportion	Probability
0	0.125
1/3	0.375
2/3	0.375
1	0.125

d) The distribution begins to take a bell shape.

7.9 Buying a car

a) The different outcomes of this distribution are 0, 1/5, 2/5, 3/5, 4/5 and 1. Their chances of occurrence are the probability values resulting from a binomial distribution of parameters $n = 5$ trials and $p = 0.3$. P(0 out of 5) = 0.16807, P(1 out of 5) = 0.36015, P(2 out of 5) = 0.3087, P(3 out of 5) = 0.1323, P(4 out of 5) = 0.02835, P(5 out of 5) = 0.00243.

b) Mean = 0.3, standard error = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3(1-0.3)}{5}} = 0.205$.

c) $n = 10$: mean = 0.3, standard error = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3(1-0.3)}{10}} = 0.145$.

$n = 100$: mean = 0.3, standard error = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3(1-0.3)}{100}} = 0.046$.

d) The mean does not change as n increases but the standard deviation decreases.

7.10 Effect of *n* on sample proportion

a) (i) The standard deviation is $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.50(1-0.50)}{100}} = 0.05$.

(ii) The standard deviation is $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.50(1-0.50)}{1000}} = 0.016$.

b) The sample proportion is likely to fall within three standard deviation of the mean.

(i) With a mean of 0.50 and a standard deviation of 0.05, this would be between 0.50 – 3(0.05) = 0.35 and 0.50 + 3(0.05) = 0.65.

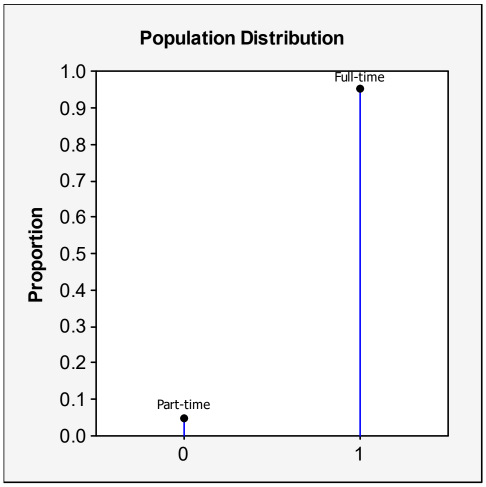
(ii) With a mean of 0.50 and a standard deviation of 0.016, this would be between 0.50 –3(0.016) = 0.45 and 0.50 + 3(0.016) = 0.55.

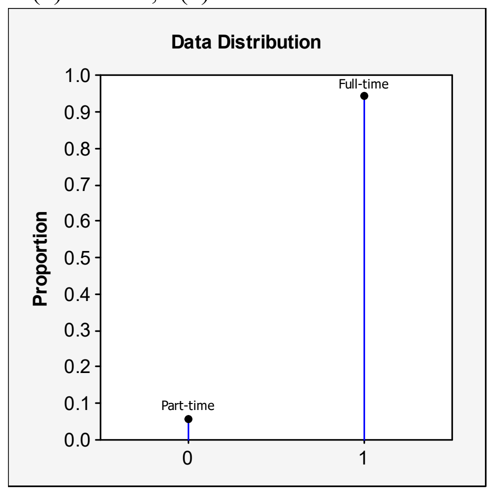
7.10 (continued)

c) When the sample size is larger, the standard deviation is smaller (a reflection of the fact that a larger representative sample is likely to be more accurate than a smaller representative sample). Three standard deviation from a larger sample, therefore, will be smaller than three standard deviation from a smaller sample. The interval will be smaller, an indication of a more precise estimate of the population proportion.

7.11 Syracuse full-time students

a) The population distribution (see graph below left) is based on the *x* values of the 14,201 students, 95.1% of which are 1s and 4.9% of which are 0s, so P(0) = 0.049, P(1) = 0.951.





b) The data distribution (see graph above right) is based on the *x* values in the sample of size 350, 330 of which are 1s and 20 of which are 0s, so the data distribution has P(0) = 20/350 = 0.0571 and P(1) = 330/350 = 0.9429.

c) The mean is $p = 0.951$ and the standard deviation is $\sqrt{p(1-p)/n} = \sqrt{0.951(1-0.951)/350} = 0.0115$.

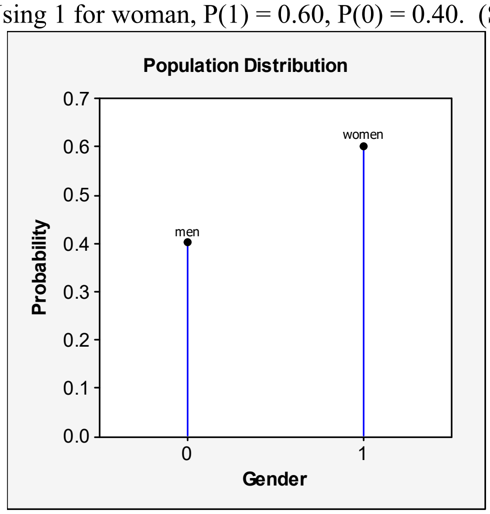
The sampling distribution represents the probability distribution of the sample proportion of full-time students in a random sample of 100 students. In this case, the sampling distribution is bell shaped and centered at 0.951.

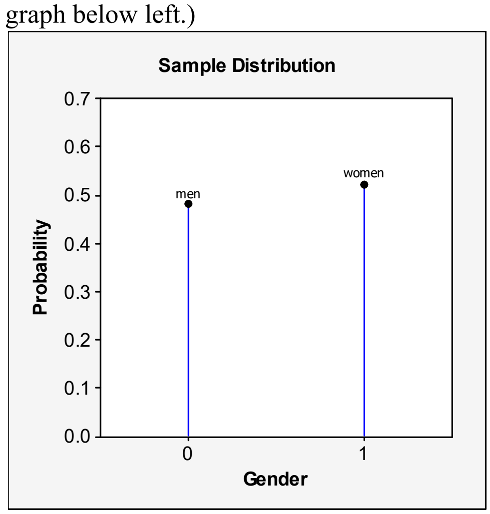
d) The population and sampling distribution graphs will look the same. Data distribution graphs will vary, since they are based on a different sample.

7.12 Gender distributions

a) We can set up a binary random variable by assigning a number, such as 1, to women, and another number, such as 0, to men.

b) Using 1 for woman, P(1) = 0.60, P(0) = 0.40. (See graph below left.)





7.12 (continued)

c) Sample proportions of 0.52 for $x = 1$ (women) and 0.48 for $x = 0$ (men). (See graph above right.)

d) The sampling distribution of the sample proportion of women in the sample is approximately a normal distribution. Its mean is 0.60 and its standard deviation is $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.60(1-0.60)}{50}} = 0.069$.

e) The population and sampling distribution graphs will look the same. Data distribution graphs will vary, since they are based on a different sample.

7.13 Shapes of distributions

a) With random sampling, the data distribution would more closely resemble the population distribution. In both cases, the distributions are based on individual scores, not means of samples.

b) For a population proportion of 0.9 or 0.95, with $n = 30$, the sampling distribution does not look bell shaped but rather skewed to the left.

c) It is inappropriate because we expect the bell shape to occur only when np and $n(1 - p)$ are larger than 15, which is not the case here.

7.14 Beauty contest election

a) The mean is $p = 0.52$ and the standard deviation is $\sqrt{\frac{0.52(1-0.52)}{400}} = 0.02498$.

b) $n = 400$ and $p = 0.52$; $n(p) = 400(0.52) = 208$ and $n(1 - p) = 400(0.48) = 192$, number of successes and number of failures both larger than 15. Thus, we can reasonably assume a normal shape for this sampling distribution.

c) The probability of observing a sample proportion of 0.5 or lower under a normal distribution is P(sample propotion < 50%) = $P(Z < -0.8)$ = 21.17%, where Z is the standard normal distribution and –0.8 is the z-score of 0.5 for a mean = 0.52 and a standard deviation = 0.02498.

d) If $n = 2000$ then the mean of sample proportion is 0.52 and its standard deviation is 0.01117. Thus P(sample propotion < 50%) = P(Z < –1.79) = 3.67%, where Z is the standard normal distribution and –1.79 is the z-score of 0.5 for a mean = 0.52 and a standard deviation = 0.01117.

Section 7.2: How Sample Means Vary Around the Population Mean

7.15 Simulate taking midterms

a) Although the mean of the population distribution is 70, because of random variability (as expressed by the population standard deviation), the sample mean of a sample of size 12 will sometimes be smaller or larger than 70.

b) The simulated sampling distribution is bell shaped and centered at 70. Almost all sample means fall between a score of about 60 to 80, or within three sample standard deviations of $10/\sqrt{12} = 2.89$.

c) It is still bell shaped and centered at 70 but now has smaller variability. Now, almost all sample means fall within about 65 to 75, or within three sample standard deviations of $5/\sqrt{12} = 1.44$.

7.16 Education of the self-employed

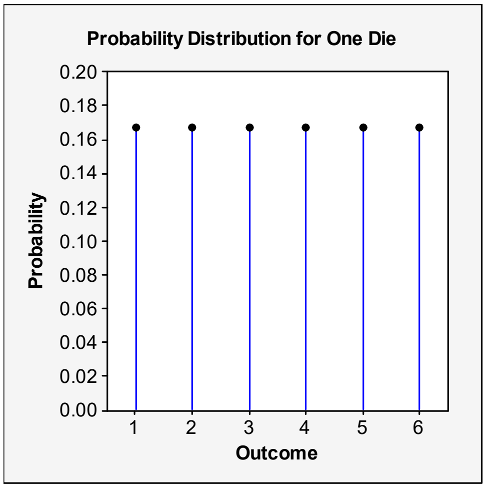
a) The random variable X is years of education of self-employed U.S. citizens.

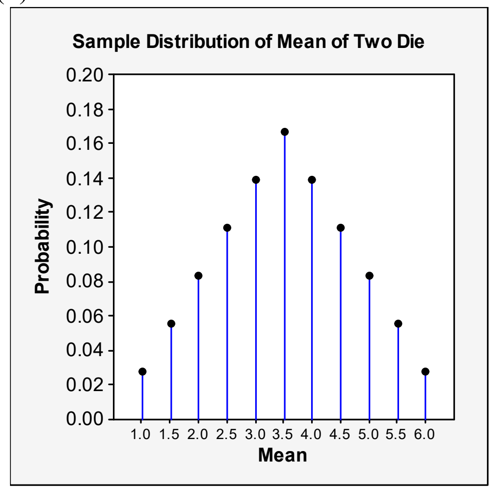
b) The mean is 13.6 and the standard deviation is $\sigma/\sqrt{n} = 3.0/\sqrt{100} = 0.30$. The mean of the sampling distribution is 13.6, the same as the mean of the population. The standard deviation is smaller than the standard deviation of the population. It reflects the variability among all possible samples of a given size taken from this population.

c) The mean is 13.6 and the standard deviation is $\sigma/\sqrt{n} = 3.0/\sqrt{400} = 0.15$. As n increases, the mean of the sampling distribution stays the same, but the standard deviation gets smaller.

7.17 Rolling one die

a) (i) (ii)





b) (i) The mean for n = 2 is 3.50 and the standard deviation is $\sigma/\sqrt{n} = 1.71/\sqrt{2} = 1.21$.

(ii) The mean for n = 30 is 3.50 and the standard deviation is $\sigma/\sqrt{n} = 1.71/\sqrt{30} = 0.31$.

As n increases, the sampling distribution becomes more normal in shape and less variable.

7.18 Performance of airlines

a)

X	P(X)
0	0.14
1	0.86

b) The mean would be the same as the population mean of 0.86. The standard error would be $0.1/\sqrt{n} = 0.1/\sqrt{100} = 0.01$.

c) z = (0.88 – 0.86)/0.01 = 2, i.e. 0.88 is two standard deviations above the mean. The central limit theorem tells us that 0.9544 of observations fall within two standard deviations of the mean. That indicates that 0.0446 of observations fall beyond two standard deviations from the mean. Half of this (0.0223) would be below two standard deviations below the mean and half (0.0223) would be above two standard deviation above the mean. Thus, the probability that with this number of flights observed, you get a mean of at least 0.88 flights on time is 0.0223.

7.19 Simulate rolling dice

a) Judging by the histogram, the simulated sampling distribution is not bell shaped but, rather, has a triangular shape.

b) Values will vary. One run of the simulation yielded a mean of 3.49 and a standard deviation of 1.21 for the simulated sampling distribution of the 10,000 sample means. These are very close to the theoretical mean and standard deviation of the sampling distribution, which are 3.5 and $\sigma/\sqrt{n}$ = $1.71/\sqrt{2} = 1.21$.

c) With n = 30, the histogram representing the sampling distribution is now bell shaped and shows a much smaller standard deviation compared to the case with n = 2.

7.20 Canada lottery

a) The mean would be 0.10 and the standard deviation would be $\sigma/\sqrt{n} = 100/\sqrt{1{,}000{,}000} = 0.10$.

b) The z-score at \$1 would be $\frac{1.00-0.10}{0.10} = 9.0$. When we look this up in the Table A, this z-score is not on the table. We find that an area of 0.0 of the curve falls above this z-score. Thus, it's exceedingly unlikely that Joe's average winnings would exceed \$1.

7.21 Shared family phone plan

a) Since σ is almost as large as μ and 0 is a lower bound that is only $(0-2.8)/2.1=-1.33,$ or 1.33 standard deviations below the mean, the distribution is right skewed.
b) With a random sample, the data distribution picks up the characteristics from the population distribution, so we anticipate it as right skewed.
c) With $n = 45$, the sampling distribution of sample mean will be bell shaped by the central limit theorem.

7.22 Dropped from plan

a) The population distribution will be right-skewed with a mean of 2.8 minutes and a standard deviation of 2.1 minutes.
b) The data distribution also has a mean of 3.40 and a standard deviation of 2.9.
c) The sampling distribution has a mean of 3.40 and a standard deviation of $\sigma/\sqrt{n}=2.9/\sqrt{45}=0.31.$
d) For 3.4 minutes, $z=\dfrac{3.4-2.8}{0.31}=1.936.$ Since the distribution is bell shaped, this is not unusually high since it is within 2 standard deviations of the mean.
e) For 3.5 minutes, $z=\dfrac{3.5-2.8}{0.31}=2.258,$ which corresponds to a cumulative probability of 0.9980. The probability of sample mean being larger than 3.5 is 1 – 0.9980 = 0.0119, or 1.2%.

7.23 Restaurant profit?

a) The mean would be 8.20 and the standard deviation would be $\sigma/\sqrt{n}=3/\sqrt{100}=0.30.$
b) $z=\dfrac{8.95-8.20}{0.3}=2.5;$ Table A tells us that 0.994 is the cumulative proportion below this z-score. Thus, the probability that the restaurant makes a profit that day is 0.994.

7.24 Survey accuracy

a) Standard error $=\sigma/\sqrt{n}=36,961/\sqrt{1000}=1168.81.$
z = 2000/1168.81 = 1.71; the proportion below this z-score is 0.9564 (2000 is used in the numerator of the z-score equation because it would be the difference between a given sample mean and the population mean no matter what these means actually were), and the proportion below –1.71 is 0.0436. Then, 0.9564 – 0.0436 = 0.9128 is the proportion that falls within two thousand of the mean income.
b) If the standard deviation were 25,000, the standard error would be 790.57. Thus, the sampling distribution of the mean is less variable and the probability of being within two thousand of the mean income is higher.

7.25 Blood pressure

a) The mean would be 130 and the standard deviation would be $\sigma/\sqrt{n}=6/\sqrt{3}=3.46.$
b) If the probability distribution of the blood pressure reading is normal, the sampling distribution of the sample mean would also be normal. If the population distribution is approximately normal, then the sampling distribution is approximately normal for all sample sizes.
c) $z=\dfrac{140-130}{3.46}=2.89;$ Table A tells us that 0.998 of the sampling distribution falls below this z-score. The probability that the sample mean exceeds 140 is 1.00 – 0.998 = 0.002.

7.26 Average price of an ebook

a) The random variable, X = price of a bestselling ebook, is quantitative.
b) The center of the population distribution is the mean of the population, \$8.05. The spread is the standard deviation of the population, \$1. The distribution is skewed slightly to the right due to some relatively high priced ebooks.
c) The center of the data distribution is \$7.8. The spread is the standard deviation of \$0.95. The shape is similar to the population distribution.
d) The center of the sampling distribution of the sample mean is \$8.05. The spread, or standard error, is $\sigma/\sqrt{n}=1/\sqrt{20}=0.224.$ The shape is approximately normal due to the Central Limit Theorem.

7.27 Average time to fill job positions

a) The center of the population distribution is the mean of the population, 37 days. The spread is the standard deviation of the population, 12 days. The shape is probably skewed to the right due to a few relatively long time needed to fill a job position.

b) The center of the data distribution is 39. The spread is the standard deviation of 13. The shape is similar to the shape of the population distribution.

c) The center of the sampling distribution of the sample mean is 37. The spread, or standard error, is $\sigma/\sqrt{n} = 12/\sqrt{100} = 1.2.$ With a sample size of 100, this would have approximately a normal distribution due to the Central Limit Theorem.

d) It would not be unusual to observe an individual who earns more than 55 days since this is only (55 – 37)/12 = 1.5 standard deviations above the mean. It would be highly unusual to observe a sample mean over 50 for a random sample of 100 job positions because this is more than 9 standard errors above the mean.

7.28 Central limit theorem for uniform population

a) The shape is triangular (you may see this better by decreasing the bin size), although it is already pretty close to bell shaped.

b) The shape is already close to bell shaped. Compared to n = 2, the variability is smaller.

c) The shape is bell shaped, and the variability is much smaller compared to n = 2. Most sample means fall between 0.35 and 0.65. Results are similar to the ones indicated in Figure 7.11.

d) As the sample size increases, the sampling distribution increasingly resembles a normal distribution.

7.29 CLT for skewed population

a) The shape is decisively right skewed.

b) The shape is still right skewed but begins to resemble a bell shape. Compared to n = 2, the variability is a bit smaller because the right tail's extend is smaller.

c) The shape is bell shaped, and the variability is much smaller compared to n = 2. Most sample means fall between about 5 and 10. Results are similar to the ones indicated in Figure 7.11.

d) As the sample size increases, the sampling distribution increasingly resembles a normal distribution.

7.30 Sampling distribution for normal population

The sampling distribution is normal even for n = 2. If the population distribution is approximately normal, the sampling distribution is approximately normal for all sample sizes.

Chapter Problems: Practicing the Basics

7.31 Exam performance

a) The mean is p = 0.70 and the standard deviation is $\sqrt{p(1-p)/n} = \sqrt{0.70(1-0.70)/50} = 0.0648.$

b) Since n = 50, by the Central Limit Theorem we would expect the shape of the sampling distribution to be approximately normal with mean = 0.70 and standard deviation = 0.0648.

c) The z-score for 0.60 is $\dfrac{0.60-0.70}{0.0648} = -1.54,$ which corresponds to a cumulative probability of 0.06. It would not be too surprising to only get 60% of the answers correct.

7.32 Cardiovascular diseases

a) The mean is 1/3 = 0.3333, and the standard error is $\sqrt{p(1-p)/n} = \sqrt{0.3333(1-0.3333)/100} = 0.04714.$

b) Yes. The z-score for 0.50 is (0.50 – 0.3333)/0.04714 = 3.54. It would be somewhat unusual to obtain a sample proportion that falls 3.54 standard deviations above the population proportion.

c) The population distribution is the set of 0s and 1s describing whether a death occurred in America in 2011 was from cardiovascular diseases (1) or not (0). The population distribution consists of roughly 33.33% 1s and 66.67% 0s. The sample distribution is the set of 50 0s and 50 1s describing whether deaths in the hospital occurred from cardiovascular diseases or not. The sampling distribution of the sample proportion is the probability distribution of the sample proportion. It has a mean of 0.3333 and standard error $= \sqrt{0.3333(1-0.3333)/100} = 0.04714.$

7.33 Alzheimer's

Since $np = 200(1/9) = 22.2$ and $n(1 - p) = 200(8/9) = 177.8$, which are both greater than 15, the shape of the sampling distribution is approximately normal with a mean of $p = 1/9$. For $n = 800$, the number of successes and failures will be even larger.

a) For a sample size of 200, the standard deviation is $\sqrt{p(1-p)/n} = \sqrt{(1/9)(1-1/9)/200} = 0.0222.$

b) For a sample size of 800, the standard deviation is $\sqrt{p(1-p)/n} = \sqrt{(1/9)(1-1/9)/800} = 0.0111.$

7.34 Basketball shooting

a) The mean is $p = 0.45$ and the standard deviation is $\sqrt{p(1-p)/n} = \sqrt{0.45(1-0.45)/12} = 0.144.$

b) Since $(3/12 - 0.45)/0.144 = -1.39,$ this game's result is 1.39 standard deviations below the mean.

c) If the population proportion is 0.45, it would not be that unusual to obtain a sample proportion of 3/12 = 0.25 since this value is only 1.39 standard deviations below the mean.

7.35 Defective chips

a) The shape of the sampling distribution is approximately normal with a mean of $p = 0.04$ and a standard deviation of $\sqrt{p(1-p)/n} = \sqrt{0.04(1-0.04)/500} = 0.0088.$

b) Since $np = 500(0.04) = 20$ and $n(1 - p) = 500(0.96) = 480$, which are both greater than 15, the shape of the sampling distribution is approximately normal.

c) The z-score for 5% is $\dfrac{0.05-0.04}{0.0088} = 1.136,$ which corresponds to a cumulative probability of 0.8721. The probability of a shipment being returned is 1 – 0.8721 = 0.1279, or 12.8%.

7.36 Returning shipment

a) Since $np = 380(0.04) = 15.2$ and $n(1 - p) = 380(0.96) = 364.8$, which are both greater than 15, the shape of the sampling distribution is approximately normal with a mean of $p = 0.04$ and a standard deviation of $\sqrt{p(1-p)/n} = \sqrt{0.04(1-0.04)/380} = 0.010.$ The z-score for 5% is $\dfrac{0.05-0.04}{0.010} = 1,$ which corresponds to a cumulative probability of 0.8413. The probability of a shipment being returned is 1 – 0.8413 = 0.1587, or 15.8%.

b) Since $np = 380(0.06) = 22.8$ and $n(1 - p) = 380(0.94) = 357.2$, which are both greater than 15, the shape of the sampling distribution is approximately normal with a mean of $p = 0.06$ and a standard deviation of $\sqrt{p(1-p)/n} = \sqrt{0.06(1-0.06)/380} = 0.012.$ The z-score for 5% is $\dfrac{0.05-0.06}{0.012} = -0.833,$ which corresponds to a cumulative probability of 0.2023. The probability of a shipment being returned is 1 – 0.2023 = 0.7977, or 79.8%.

7.37 Aunt Erma's restaurant

a) The population distribution has a mean of \$900 and a standard deviation of \$300.

b) The data distribution has a mean of \$980 and a standard deviation of \$276. The standard deviation of the data distribution describes the spread of the daily sales values for this past week.

c) The mean of the sampling distribution of the sample mean is \$900; the standard deviation is $\sigma/\sqrt{n} = 300/\sqrt{7} = 113.4$ dollars. The standard deviation describes the spread of the sample means based on samples of seven daily sales.

7.38 Home runs

a) No. X is discrete, positive, and likely skewed to the right given the relative sizes of the mean and standard deviation.

b) Since $n = 162$ is greater than 30, the shape of the sampling distribution of the mean number of runs the team will hit in its 162 games is approximately normal with a mean of 1.0 and a standard deviation of $\sigma/\sqrt{n} = 1/\sqrt{162} = 0.0786.$

c) The z-score for 1.50 is $\dfrac{1.5-1}{0.0786} = 6.36.$ The probability of exceeding 1.5 runs per game is practically 0.

7.39 Student debt

a) The mean would be \$18,367. The standard error would be $\sigma/\sqrt{n} = 4,709/\sqrt{100} = \$470.9$. The sampling distribution would have a bell shape because the sample size is greater than 30.

b) $z = \dfrac{20,000 - 18,367}{470.9} = 3.47$

c) The z-scores are –2.12 and 2.12 enclosing a probability of 0.966.

7.40 Bank machine withdrawals

The sample standard deviation is $\sigma/\sqrt{n} = 50/\sqrt{100} = 5$; \$10 is two standard deviations, so 0.9545 of the observations will fall within two standard deviations of the mean.

7.41 PDI

a) $z = \dfrac{90 - 100}{15} = -0.67$

b) $z = \dfrac{90 - 100}{15/\sqrt{225}} = -10$

c) For an individual PDI value, 90 is only 0.67 standard deviations below the mean and is therefore not surprising. However, it would be unusual for the mean of a sample of size 225 to be 10 standard deviations below the mean of its sampling distribution.

7.42 Number of pets

a) X does not likely have a normal distribution because the standard deviation is as big as the mean, an indication of skew. In fact, the lowest possible value of 0 is only 1.88/1.67 = 1.13 standard deviations below the mean, an indication of right skew.

b) The sampling distribution would have a mean of 1.88, and a standard error of $\sigma/\sqrt{n} = 1.67/\sqrt{100} = 0.167$. Because the samples are large (>30), the sampling distribution approximates a normal curve even though the population is not normally distributed.

♦♦7.43 Using control charts to assess quality

a) For a normal distribution, nearly every sample mean will fall within three standard deviations of the mean of the sampling distribution. Thus, there is a probability of 0.003 that this process will indicate a problem where none exists.

b) The Empirical Rule says that 95% of the sampling distribution falls within two standard deviations of the mean. The probability of falsely indicating a problem would then be 5% if we used two standard deviations.

c) (i) The chance that any one sample mean would fall above the mean is 0.50. Thus, the probability that the next 9 means in a row will all fall above the mean is 0.0002.

$$P(9) = \frac{9!}{9!(9-9)!}(0.50)^9(1-0.50)^{9-9} = 0.002$$

(ii) In this scenario, it does not matter where the first observation falls, only that each succeeding observation falls on the same side as the first. No matter where the first falls, the chance that the next will be on the same side of the mean is 0.50. The probability is twice the probability found in (i), 2(0.002) = 0.004.

7.44 Too little or too much cola?

a) With a sample size of 4, the standard deviation would be $\sigma/\sqrt{n} = 4/\sqrt{4} = 2$. The target line would be at the mean of 500 within upper and lower control limits at 506 and 494, respectively (each three standard deviations from the mean).

b) The standard deviation is now $\sigma/\sqrt{n} = 6/\sqrt{4} = 3$. The z-score at the lower control limit would be $\frac{494-491}{3} = 1$. The proportion falling below this z-score would be 0.84. The z-score at the upper control limit would be $\frac{506-491}{3} = 5$. The proportion falling above this z-score is essentially 0.Thus, the probability that the next value plotted on the control chart indicates a problem with the process is 0.84. We assume a normally distributed population in making these calculations because we do not have a sample size of more than 30.

Chapter Problems: Concepts and Investigations

7.45 CLT for custom population

Answers will vary. However, as n increases, the sampling distribution will become more bell shaped and have less variability around the mean, no matter what distribution the student chose.

7.46 What is a sampling distribution?

Each time a poll is conducted, there will be a sample proportion that is calculated (How many of the 1000 polled Canadians think the prime minister is doing a good job?). The distribution of sample proportions is called a sampling distribution. The sample proportions fluctuate from one poll to the next around the population proportion value, with the degree of spread around the population proportion being smaller when the sample size is larger.

7.47 What good is a standard deviation?

The standard deviation of the sampling distribution describes how closely a sample statistic will be to the parameter it is designed to estimate, in this case how close the sample proportion will be to the unknown population proportion.

7.48 Purpose of sampling distribution

This standard deviation tells us how close a typical sample proportion is to the population proportion.

7.49 Sampling distribution for small and large *n*

a)

Sample	Number Who Prefer Pizza A	Proportion Who Prefer Pizza A
AAA	3	1
AAD	2	2/3
ADA	2	2/3
DAA	2	2/3
ADD	1	1/3
DAD	1	1/3
DDA	1	1/3
DDD	0	0

b) If the proportion of the population who favors Aunt Erma's pizza is 0.50, each of the eight outcomes listed in (a) are equally likely. Thus, the probability of obtaining a sample proportion of 0 is 1/8, a sample proportion of 1/3 has a (1 + 1 + 1)/8 = 3/8 chance, a sample proportion of 2/3 has a (1 + 1 + 1)/8 = 3/8 chance and a sample proportion of 1 has a 1/8 chance. This describes the sampling distribution of the sample proportion for $n = 3$ with $p = 0.50$.

c) For $p = 0.5$ and $n = 50$, the mean of the sampling distribution is 0.5 and the standard deviation is $\sqrt{p(1-p)/n} = \sqrt{0.5(1-0.5)/50} = 0.0707$.

7.50 Sampling distribution via the binomial

a) Let X = the number of people, out of 3, that prefer pizza A.

$$P(X=1)=\binom{3}{1}(0.5)^1(0.5)^{3-1}=0.375=3/8$$

b) $P(X=0)=\binom{3}{0}(0.5)^0(0.5)^{3-0}=0.125=1/8,\ P(X=2)=\binom{3}{2}(0.5)^2(0.5)^{3-2}=0.375=3/8,$

$$P(X=3)=\binom{3}{3}(0.5)^3(0.5)^{3-3}=0.125=1/8$$

c) Let X = the number of people, out of 4, that prefer pizza A.

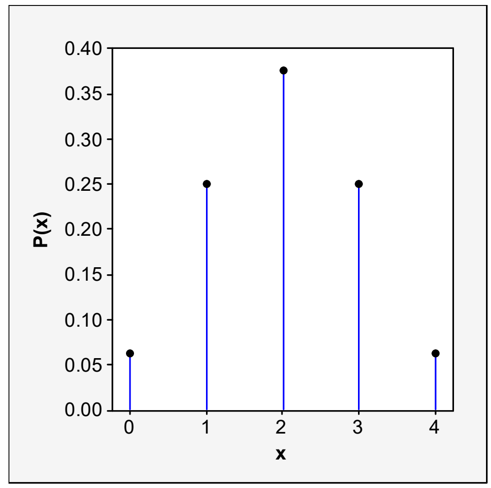
$$P(X=0)=\binom{4}{0}(0.5)^0(0.5)^{4-0}=0.0625=1/16$$

$$P(X=1)=\binom{4}{1}(0.5)^1(0.5)^{4-1}=0.25=1/4$$

$$P(X=2)=\binom{4}{2}(0.5)^2(0.5)^{4-2}=0.375=3/8$$

$$P(X=3)=\binom{4}{3}(0.5)^3(0.5)^{4-3}=0.25=1/4$$

$$P(X=4)=\binom{4}{4}(0.5)^4(0.5)^{4-4}=0.0625=1/16$$



7.51 Pizza preference with p = 0.6

a) Let X = the number of people, out of 3, that prefer pizza A.

$$P(X=0)=\binom{3}{0}(0.6)^0(0.4)^{3-0}=0.064,\quad P(X=1)=\binom{3}{1}(0.6)^1(0.4)^{3-1}=0.288,$$

$$P(X=2)=\binom{3}{2}(0.6)^2(0.4)^{3-2}=0.432,\quad P(X=3)=\binom{3}{3}(0.6)^3(0.4)^{3-3}=0.216$$

b) The mean number of people preferring pizza A is np = 100(0.6) = 60.

c) The proportion of people preferring pizza A is 60/100 = 0.60.

7.52 Simulating pizza preference with p = 0.5

a) The mean of the sampling distribution is 0(1/8) + (1/3)(3/8) + (2/3)(3/8) + 1(1/8) = 1/2 = 0.5.

b) Values will vary. Yes, one run of the simulation yielded a mean of 0.496 and a standard deviation of 0.289 for the 10,000 simulated sample proportions, almost identical to the theoretical values.

7.53 Simulating pizza preference with p = 0.6

Values will vary. Yes, one run of the simulation yielded a mean of 0.60 and a standard deviation of 0.282 for the 10,000 simulated sample proportions. These are almost identical to the theoretical mean and standard deviation of the sampling distribution, which are 0.60 and 0.283.

7.54 Winning at roulette

a) In order to win at least \$100, the average winnings per spin must be at least \$100/40 = \$2.50.

b) To win at least \$100, you must win on at least 25 of the spins (note that if you win 25 of the spins, you lose 15 of the spins so that your total winnings are 25(\$10) +15(–\$10) = \$100). The sampling distribution of the sample proportion of winning spins has a mean of 18/38 and standard deviation of $\sqrt{p(1-p)/n} = \sqrt{(18/38)(1-18/38)/40} = 0.0789$. Winning at least 25/40 = 5/8 of the time has a *z*-score of $\frac{5/8-18/38}{0.0789} = 1.92$, which corresponds to a cumulative probability of 0.9726. The probability of winning at least \$100 is 1 – 0.9726 = 0.0274, or about 2.74%.

c) You must win at least 25 of the spins.

d) $\mathrm{P}(X \geq 25) = \mathrm{P}(X = 25) + \mathrm{P}(X = 26) + \cdots + \mathrm{P}(X = 40) = 0.0392$. The estimate using the normal approximation is about 0.0118 smaller.

7.55 True or false

False, as the sample size increases, the denominator of the equation for standard deviation increases. A larger denominator leads to a smaller solution. This reflects the fact that a larger sample size is likely to lead to a more accurate estimate of the population mean.

7.56 Multiple choice: Standard deviation

The best answer is (b).

7.57 Multiple choice: CLT

The best answer is (c).

7.58 Multiple choice: Sampling distribution of sample proportion

The best answer is (c). The mean of the sampling distribution is *p* = 0.5 and the standard deviation is $\sqrt{p(1-p)/n} = \sqrt{(0.5)(1-0.5)/150} = 0.408$ and 95% of sample proportions should be within two standard deviations of the mean.

7.59 Multiple choice: Sampling distribution

The best answer is (a).

♦♦7.60 Sample = population

a) If we sampled everyone in the population, the sample mean would always be the same as the population mean, and we would not have any variability of sample means.

b) The sampling distribution would look exactly like the population distribution with a sample size of one.

♦♦7.61 Standard deviation of a proportion

The standard deviation is $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$.

♦♦7.62 Finite populations

a) The standard deviation is $\sqrt{\frac{30{,}000-300}{30{,}000-1}} \cdot \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{29{,}700}{29{,}999}} \cdot \frac{\sigma}{\sqrt{n}} = 0.995\frac{\sigma}{\sqrt{n}}$.

b) The standard deviation is $\sqrt{\frac{30{,}000-30{,}000}{30{,}000-1}} \cdot \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{0}{29{,}999}} \cdot \frac{\sigma}{\sqrt{n}} = 0 \cdot \frac{\sigma}{\sqrt{n}} = 0$.

Chapter Problems: Student Activities

7.63 Simulate a sampling distribution

a) Answers will vary.
b) Answers will vary.
c) You would expect the mean to be close to the population mean of 196.
d) You would expect a standard deviation of about $\sigma/\sqrt{n} = 57.4/\sqrt{9} = 19.1$, which is the standard deviation of the sampling distribution when $n = 9$.

7.64 Coin tossing distributions

a) The mean is 0(0.5) + 1(0.5) = 0.5.

b)

x	**P(*x*)**
0	0.5
1	0.5

x	**P(*x*)**
0	0.4
1	0.6

c) The results will be different each time this exercise is conducted but should be roughly bell-shaped around 0.5.
d) If we performed the experiment in (c) an indefinitely large number of times, we'd expect to get a mean that's very close to the population proportion and we'd expect the standard error to approach the standard deviation we'd calculate using this formula: $\sqrt{p(1-p)/n}$. The population proportion is 0.5 and the standard error is 0.158.

7.65 Sample versus sampling

a) The histograms will be different each time this exercise is conducted, but we should have a positively skewed distribution.
b) This is a sampling distribution, and with ten coins in each sample, the distribution should be closer to a normal distribution than that in (a). It should also be less spread out than the distribution in (a).
This exercise illustrates the floor effect (no coin can have a score below 0, but the scores can go quite high). It also illustrates the Central Limit Theorem in that the increase in sample size makes the distribution somewhat closer to a normal distribution.