Chapter 7

Discussion Problem Solutions

D1. The agent can increase his sample size to a value greater than 10. The larger the sample size, the smaller the spread of the distribution of means and the more precise his 95% range for the mean will be.

D2. Make sure that students understand the difference between Display 7.2 and Display 7.3. Display 7.2 shows the distribution of the population of salaries of all NBA basketball players for the given season. Display 7.3 shows the mean salaries of 200 samples of size 10 taken from that population of all players.

a. The largest value in the population has a value of $20-$22 million. These salaries appear to occur with a frequency of about 10. To get a value of $22 million in the sampling distribution for samples of size 10 in Display 7.3, you would have to pick all ten of the players having this salary (if there are ten such players), which is not likely to happen in 200 samples. Similarly, the smallest salary seems to be around $1 million and a sample of 10 players having this same low salary is unlikely.

b. If you continued to add more samples to the sampling distribution of Display 7.3 or constructed the exact sampling distribution using all possible samples, then the means would be equal. This fact is not completely obvious, especially with skewed distributions, but it is not difficult to prove. Random sampling leads to sample means that tend to center at the population mean. As students discovered in Statistics In Action 4.1 involving random rectangles, this is not the case with judgment samples. In Section 7.2, students will learn that random sampling also leads to a predictable amount of variability in the distribution of sample means.

c. No, comparing the displays shows that the standard deviation of the sampling distribution is smaller than that of the population. The reason for this can be seen from the discussion in part a.
Averages tend to be clustered closer to the center of the distribution than are the original population values.

d. The shape of the simulated sampling distribution of the sample mean is more normal than the distribution of the population but is still slightly skewed to the right.

D3. As we have seen, the sampling distribution is approximately normal, with less variation than the population. We will see the standard deviation of the sampling distribution can be computed as you can estimate that the reasonably likely events are those in the interval $4.7/\sqrt{10} \approx 1.5$ so
reasonably likely events would be means in the interval $4.6 \pm 1.96(1.5)$ or 1.66 to 7.54 and rare events are means outside this interval. You can also estimate these values using the plot in Display 7.3. Rare events are those in the upper 2.5% of the distribution and the lower 2.5% of the distribution. Because there are 200 samples, this would be the largest 5

means and the smallest 5 means. Five means out of 200 corresponds to a relative frequency of 0.025. It appears in Display 7.3, that a cumulative frequency of 0.025 occurs between 1 and 2 on the right and around 7.5 from the left.
D4. The similarity is that the method of taking a sample for a survey is biased if the method produces a summary statistic that tends to be, on average, larger or smaller than the population parameter being estimated. A summary statistic is a biased estimator of a population parameter if it is larger or smaller, on average, than the parameter when sampling is random. The difference is in where the bias comes from. In sampling, the bias comes from the method of getting the sample. In estimation, the bias comes from the properties of the estimator (summary statistic) chosen.
D5. The relative frequency for a particular value in a population can be thought of as the probability of that value occurring in a single random selection from that population. (If 30% of adults have no medical insurance, then the chance of randomly selecting an adult without medical insurance is 0.30.) Because the mean of a sample of size 1 is the sampled value itself, the sampling distribution for the mean is the same as the relative frequency distribution of the population. So each has the same shape, mean, and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{1}} = \sigma$.
D6. Picture a curve drawn through the points, then for a sample of size 30, the graph appears to pass through an *SE* value of 0.2. This agrees with the given data since the *SE* for a sample size of 20 is 0.25 and 0.17 for a sample of size 40. The value for 30 must be between 0.25 and 0.17. The equation of the curve is $SE = \frac{1.1}{\sqrt{n}}$ since the population standard deviation is 1.1.
D7. This is a “plain English” version of the rule given in the second bullet in the box on page 330 of the student book. As the sample size increases, the spread of the sampling distribution of the sample mean decreases. So the sample means cluster together more tightly around the population mean. Fortunately, this is also common sense: If you want your estimate to be close to the true mean, take as large a sample as possible.
D8. This problem should not be solved using a *z*-score as in the Average Number of Children example because the sampling distribution is not approximately normal. Display 7.28 shows that the sampling distribution for samples of size 10 still has a skew to the right. Using the plot in Display 7.28 for $n = 10$ as an estimate, the probability that a random sample of nine U.S. families will have an average of 1.5 children or fewer is about 0.97. This was found by summing up the bars in the histogram after 1.5 and subtracting that from 1.
D9. For all practical purposes, it would be about the same. However, the standard error is actually a bit larger if $N = 10,000$ than if $N = 1,000$. Using the formulas you will see in

E29, if $N = 10{,}000,$ then $SE = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{1.1}{10}\sqrt{\frac{9900}{9999}} \approx 0.1095.$ If $N = 1000$ then $SE = \frac{1.1}{10}\sqrt{\frac{900}{999}} \approx 0.1044.$
D10. a. It does not change; for all n, the mean is μ. **b.** It decreases by a factor of $1/\sqrt{n}$. **c.** The mean will be μ because each sample mean will be equal to μ. The standard error will be 0; there is no variation in the sample means because the only sample contains the entire population.
D11. a. The distribution for $n = 40$ is the most normal, and the distribution for $n = 10$ is the least normal. Notice that for $n = 10,$ $np = 10 \cdot 0.6 = 6 < 10,$ and $n(1-p) = 10 \cdot 0.4 = 4 < 10.$ The guideline does not hold and the distribution for $n = 10$ is quite skewed. For $n = 20,$ $np = 12 > 10,$ but $n(1-p) = 8 < 10.$ The guideline does not hold. This distribution is less skewed, but the skew is still noticeable. When $n = 40,$ both $np = 24$ and $n(1-p) = 16$ are more than 10. The guideline does hold and the distribution looks very symmetric and fairly smooth. It appears that the guideline works quite well. **b.** We would expect 60% of any size sample to have a cell phone. **c.** The distribution for $n = 10$ has the largest spread, and the distribution for $n = 40$ has the smallest spread. **d.** Getting a sample with 75% or more cell phone users would be more likely with a sample size of 10. The area of the bars in the histogram to the right of 0.75 gets smaller as the sample size increases because with larger sample sizes, the sample proportions tend to cluster closer to 0.6.
D12. **a.** For $n = 10,$ $\mu_{\hat{p}} = 0.6,$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.6 \cdot 0.4}{10}} \approx 0.155$ For $n = 20,$ $\mu_{\hat{p}} = 0.6,$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.6 \cdot 0.4}{20}} \approx 0.110$ For $n = 40,$ $\mu_{\hat{p}} = 0.6,$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.6 \cdot 0.4}{40}} \approx 0.077$ **b.** Estimates will vary, but the answers from part a should match quite well. In particular, the sampling distributions of the sample proportion for all sample sizes have a center that

is close to 0.6. For the sampling distribution of the sample proportion for samples of size 10, the spread is large, with reasonably likely sample proportions ranging from about 0.3 to 0.9. For $n = 20$, the spread is smaller, with reasonably likely sample proportions ranging from about 0.4 to 0.8. For $n = 40$, the spread is small, with reasonably likely sample proportions ranging from about 0.45 to 0.75.

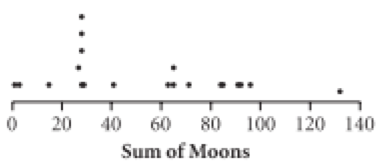
c. The mean using these formulas is 0.6000006, the difference due to rounding error. The standard deviation is 0.155, which matches the result from the previous formula.

Practice Problem Solutions

P1. a. The smallest number of moons would be 1 by describing the moons of Mercury, Venus, and Earth. The largest number of moons would be 146 moons, by describing the moons of Uranus, Saturn, and Jupiter.

b. I. Randomly select three planets.
II. Calculate the sum of the numbers of moons.
III. Repeat steps I and II many times.
IV. Display the distribution of the sums.

c. Answers will vary. One sample plot is shown below:



This can be done fairly easily on a TI-83 Plus or TI-84 Plus with a short program.

```
FnOff
{0,0,1,2,63,56,27,13}→L1
ClrList L3
For(I,1,20)
rand (9)→L2
SortA(L2,L1)
L1(1)+L2(2)+L3(3)→L3(I)
End
Plot1(Histogram,L3)
ZoomStat
```

The For(I,1,20) begins a loop. The first time through, I = 1. Every command is executed until the line End is reached. At that point, the loop is repeated, this time with I = 2. This program takes the algorithm for selecting without replacement and repeats it twenty times, recording the result in the appropriate position in L3.

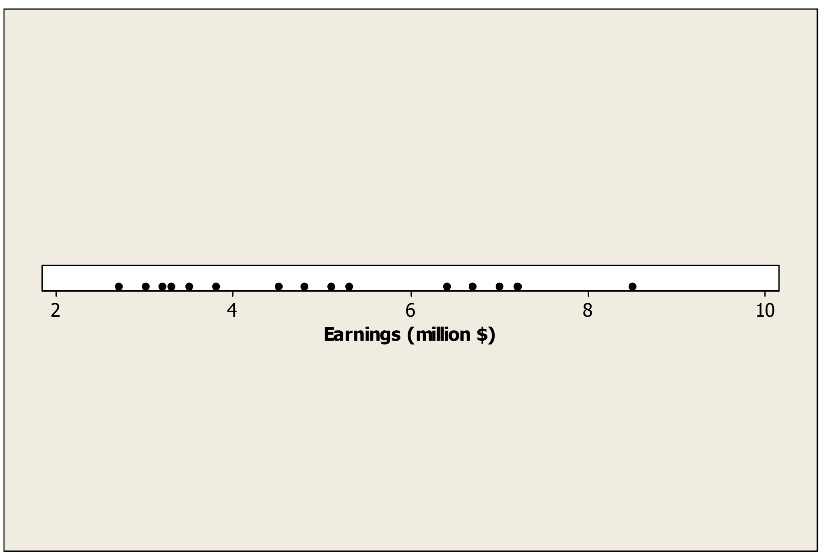
P2. Histogram C corresponds to I, given that all the values are skewed to the left. Likewise, Histogram A corresponds to II for similar reason. Histogram B corresponds to III since the values all tend to be near the center of the given distribution in the display.

P3. a. The most you could be paid is 10% of Elvis Presley's and Charles Schulz's earnings, which is $8,500,000. The least you could be paid is $2,700,000.

b. The possible pairs, along with your possible earnings, are

52 + 33	$8,500,000
52 + 20	$7,200,000
52 + 18	$7,000,000
52 + 15	$6,700,000
52 + 12	$6,400,000
33 + 20	$5,300,000
33 + 18	$5,100,000
33 + 15	$4,800,000
33 + 12	$4,500,000
20 + 18	$3,800,000
20 + 15	$3,500,000
20 + 12	$3,200,000
18 + 15	$3,300,000
18 + 12	$3,000,000
15 + 12	$2,700,000

The sampling distribution is shown below.



c. Only three of the 15 combinations were $7 million or more, so the probability is 3/15, or 0.2,

P4. a. The mean is $\frac{0+1+2+\cdots+5}{6}=2.5$ and the standard deviation is

$$\sqrt{\frac{(0-2.5)^2+(1-2.5)^2+(2-2.5)^2+\cdots+(5-2.5)^2}{6}}\approx 1.71.$$

b. Label the cars as 0, 1, 2, 3, 4, and 5, which corresponds to their respective number of defects. The 15 possible 2-samples, with their mean number of defects, are:

2-sample	Mean number of defects
0,1	0.5
0,2	1
0,3	1.5
0,4	2.0
0,5	2.5
1,2	1.5
1,3	2.0
1,4	2.5
1,5	3.0
2,3	2.5
2,4	3.0
2,5	3.5
3,4	3.5
3,5	4.0
4,5	4.5

c. The mean of the means listed in column 2 of the table in (b) is obtained by summing all 15 values and dividing by 15. This yields 2.5, which coincides with the population mean.

d. The standard deviation is given by

$$\sqrt{\frac{\sum_x (x-2.5)^2}{15-1}}\approx 1.12<1.71,$$

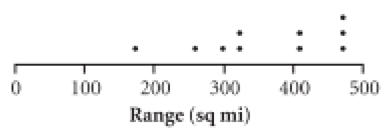
where the sum is taken over the values in the right column in the table in (b).

P5. a. The range is 527 – 56 = 471 sq. mi.

b.

Sample of Three Parks	Range of the Areas (sq mi)
A, B, C	527 − 56 = 471
A, B, R	378 − 56 = 322
A, B, Z	229 − 56 = 173
A, C, R	527 − 119 = 408
A, C, Z	527 − 119 = 408
A, R, Z	378 − 119 = 259
B, C, R	527 − 56 = 471
B, C, Z	527 − 56 = 471
B, R, Z	378 − 56 = 322
C, R, Z	527 − 229 = 298

c.



d. No. The mean of the sampling distribution of the sample range is 360.3 sq. mi. This is much less than the actual range of 471 sq. mi. You could tell this without calculating because the greatest possible range for a sample occurs when the population maximum and minimum are contained in the sample. All other samples will have a smaller range.

e. There is much variation. There is a difference of 278 between the greatest sample range and the smallest sample range, and a standard error of 96.68 sq. mi.

P6. a. **A.** II **B.** I **C.** III

b. The mean of all sampling distributions is 30, which can easily be seen from the symmetry of the distributions.

c. Distribution A has the smallest standard error since its values deviate from 30 by the least amount.

P7. a. Plot A is the population, starting out with only five different values and a slight skew. Plot C is for a sample size of 4, having more values, a smaller spread, and a more nearly normal shape. Plot B is for a sample size of 10, which has even more values, an even smaller spread, and a shape that is closest to normal of all the distributions. The mean appears to be about 1.7 for each distribution. The standard deviation of plots A, B, and C appear to be about 1, 0.3, and 0.5, respectively.

b. The mean will be the 1.7 for each distribution, which is consistent with the estimates.

c. For a sample size of 4, $\sigma_{\bar{x}} = 1/\sqrt{4} = 0.5$. For $n = 10$, $\sigma_{\bar{x}} = 1/\sqrt{10} \approx 0.316$. These match the estimates very closely.

d. The population distribution is roughly mound-shaped with a slight skew. The two sampling distributions however are approximately normal with the distributions becoming more nearly normal as the sample size increases.

P8. The shape will be approximately normal because the population's distribution is approximately normal; the mean will be the mean of the population, or 0.266, and the standard error will be

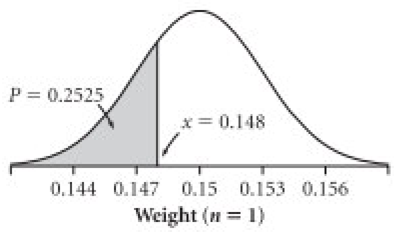
$$\frac{\sigma}{\sqrt{n}} = \frac{0.037}{\sqrt{15}} \approx 0.009553 .$$

P9. Because the weights are normally distributed, we can use the normal approximation with all sample sizes.

a. The z-score is

$$z = \frac{0.148 - 0.15}{0.003} = -0.667$$

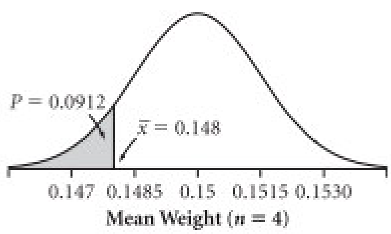
which gives a probability of 0.2525. Alternatively, on a TI-84 the result of the command **normalcdf(–9999999,0.148,0.15,0.003)** is approximately 0.2525.



b. The z-score is

$$z = \frac{0.148 - 0.15}{0.003/\sqrt{4}} = -1.3333$$

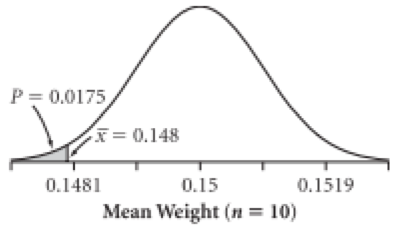
which gives a probability of 0.0912.



c. The z-score is

$$z = \frac{0.148 - 0.15}{0.003/\sqrt{10}} = -2.1082$$

which gives a probability of 0.0175.

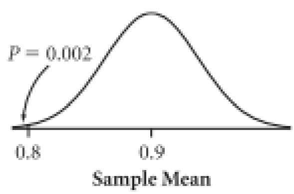


P10. a. The sampling distribution should be approximately normal because this is a very big sample size. With random samples of 1000 families, the potential values of the sample mean are centered at 0.9—the mean number of children per family—and have a standard error of 1.1 divided by the square root of 1000, or about 0.035.

b. The middle 95% of the sample means would lie between 0.9 ± 1.96(0.035) or 0.8314 to 0.9686. So, the distribution of potential values of the sample mean is concentrated very tightly around the expected value of 0.9.

c. The *z*-score for a mean of 0.8 children is

$$z = \frac{0.8 - 0.9}{1.1/\sqrt{1000}} \approx -2.87 .$$



The probability that the mean is 0.8 or less is about 0.002.

d. We must compute $P(0.8 < \bar{x} < 1.0)$. The corresponding *z*-score for 0.8 is -2.87 (from (c)), and the *z*-score for a mean of 1.0 children is

$$z = \frac{1.0 - 0.9}{1.1/\sqrt{1000}} \approx 2.87 .$$

So,

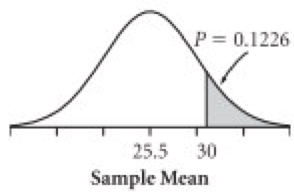
$$P(0.8 < \bar{x} < 1.0) = P(-2.87 < z < 2.87) = 0.9960 ,$$

or 0.9959 using Table A.

P11. a. Since the distribution of the population is roughly symmetric, a sample size of 15 should be large enough to assume the sampling distribution of the total will be approximately normal. The mean of the sampling distribution, $\mu_{\bar{x}}$, would be 1.7 • 15 = 25.5, and the standard error of the sum, σ_{sum}, is $\sigma\sqrt{n}$ = 1 • $\sqrt{15} \approx$ 3.873. The z-score is

$z = \frac{30-25.5}{3.873} \approx 1.16$ and the probability is P= 0.1226.

Alternatively, **normalcdf**(30, 1E99, 1.7•15, $\sqrt{(15)}$) = 0.1226. Or, using the means, **normalcdf**(30/15, 1E99, 1.7, 1/$\sqrt{15}$) = 0.1226.



b. The two z-scores are

$$z = \frac{sum - n\mu}{\sigma\sqrt{n}} = \frac{25-20(1.7)}{1.0\sqrt{20}} \approx -2.012$$

$$z = \frac{sum - n\mu}{\sigma\sqrt{n}} = \frac{30-20(1.7)}{1.0\sqrt{20}} \approx -0.894,$$

so the probability is approximately 0.1635 (or 0.1645 using Table A).

P12. a. $\sigma_{\bar{x}} = \frac{4.7}{\sqrt{5}} \approx \$2.102$ million . The z-score for a mean salary of \$1.56 million is

$$z = \frac{1.5-4.6}{2.102} \approx -1.475 .$$

So, $P(\bar{x} < 1.5) = P(z < -1.475) = 0.0708$.

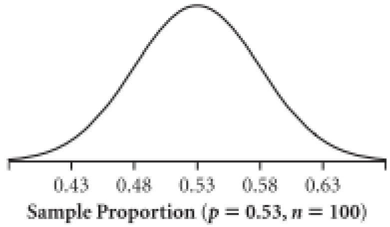
b. This probability corresponds to the leftmost 2 bars, the sum of which is approximately 0.005 + 0.02 = 0.025.

c. No, the sampling distribution isn't approximately normal for a sample size of 5, so using the normal distribution to approximate it won't be very accurate.

P13. The analysis would not change because the sample size in both cases is a very small fraction of the population size.

P14. a. This sampling distribution can be considered approximately normal because both np = 53 and n(1 – p) = 47 are 10 or greater. The sampling distribution has a mean of 0.53 and a standard error of

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.53(1-0.53)}{100}} \approx 0.05$$



b. No. As you can see from the sampling distribution in part a, the probability of getting

9 (or 9%) or fewer women just by chance is almost 0. That there are so few women in the U.S. Senate cannot reasonably be attributed to chance alone, and so we should look for other explanations (for example, that fewer women go into politics, that women are not able to raise as much money for their campaigns, or that voters are reluctant to elect a woman senator).

P15. a. First, note that $p = 0.15$ and $\sigma_p = \sqrt{\frac{(0.15)(0.85)}{400}} \approx 0.018$. Also, the sampling distribution of $\hat{p}$ is approximately normal because $np = 0.15(400) > 10$ and $n(1-p) = 0.85(400) > 10$. Hence, the reasonably likely proportions are

$$0.15 \pm 1.960(0.018),$$

or 0.115 to 0.185. This means that there is a 95% chance it is within 3.5% of 15%.

b. First, note that $p = 0.15$ and $\sigma_p = \sqrt{\frac{(0.15)(0.85)}{4000}} \approx 0.0056$. Hence, the reasonably likely proportions are

$$0.15 \pm 1.960(0.0056),$$

or 0.139 to 0.161. This means that there is a 95% chance it is within 1.1% of 15%.

P16. a. Here, $n = 435$ and $p = 0.38$. This sampling distribution can be considered approximately normal because both $np = 165.3$ and $n(1 - p) = 269.7$ are 10 or greater. The sampling distribution has a mean of 0.38 and a standard error of

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.38(1-0.38)}{435}} \approx 0.023$$

A sample proportion of 0.59 is about 95 standard deviations above the mean of the sampling distribution, and therefore it would be very unusual to see such a result. The probability is close to 0.

b. Yes, $\hat{p} = 257/435 \approx 0.59$. As you saw in part a, it is almost impossible for this many representatives to be Democrats if they were chosen at random from the general population. Possible reasons include the fact that people do not always vote for their party's candidate. In addition, the population that year identified themselves as 28% Republican, 34% Democrat, and 38% Independent, yet there was only one Independent in the House. So one possible explanation is that Independents tended to vote Democrat in this particular election.

c. For example, people don't always vote for the candidate of their party.

P17. First check the guideline: Both $np = 100 \bullet 0.61 = 61$ and $n(1 - p) = 100 \bullet 0.39 = 39$ are greater than 10, so the shape of the distribution will be approximately normal. $\mu_{sum} = np = 61$. $\sigma_{sum} = \sqrt{np(1-p)} = \sqrt{100 \cdot 0.61 \cdot 0.39} \approx 4.88$. We must find z-scores and probabilities for 50 freshmen.

$$z = \frac{50-61}{4.88} \approx -2.25,\ P(sum \geq 50) \approx 0.9879$$

Yes, because there is about a 98.8% probability that a sample of 100 freshmen will contain at least 50 freshmen who are attending their first choice college.

P18. a. First check the guideline: Both np = 1000• 0.23 = 230 and $n(1-p)$ = 1000 • 0.77 = 770 are greater than 10, so the shape of the distribution will be approximately normal. $\mu_{sum} = np = 230$. $\sigma_{sum} = \sqrt{np(1-p)} = \sqrt{1000 \cdot 0.23 \cdot 0.77} \approx 13.31$. So, the z-score corresponding to 18% (or 0.18(1000)=180 people) is

$$z = \frac{180-230}{13.31} \approx -3.757,\ P(sum \leq 180) = P(z \leq -3.757) \approx 0.0001.$$

So, there is about a 0.01% chance that the sum will not exceed 180 people.

b. Similar to (a), the z-score corresponding to 24% (or 240 people) is

$$z = \frac{240-230}{13.31} \approx 0.751,\ P(sum \geq 240) = P(z \geq 0.751) \approx 0.2261.$$

c. The reasonably likely proportions are given by the interval

$$0.23 \pm 1.96(0.013) = 0.23 \pm 0.026.$$

So, there is a 95% chance that it is no farther than 2.6% from 23%.

Exercise Solutions

E1. a. $\bar{x} = \frac{12.500+11.625+18.275+13.225}{4} = 13.90625$

b. $\bar{x} = \frac{6.625+10.375+13.225+8.800}{4} = 9.75625$

c. $P(\bar{x} \geq 13.90625) = \frac{1}{70}$ (because just 1 bar of height 1 occurs to the right of 13.90625).

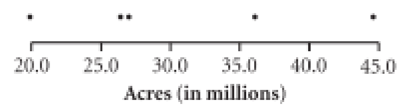
d. No, the high mean enzyme level probably was not due to chance, but because the shorter days cause higher enzyme levels.

E2. a. I is B and. II is A.

b. No, a mean of 86 or greater for a class of size 30 did not occur even once in the 100 samples represented in the sampling distribution of Histogram A. Thus, a random sample of 30 students is very unlikely to have a mean score as high as 86. If the class does have a mean score of 86, it cannot reasonably be attributed to chance variation. Perhaps the teacher is very good or the harder-working students had to enroll in this course during the second hour.

E3. a. The distribution is highly skewed toward the larger values. There are two outliers, Alaska and Texas.

b. Number the states from 01 to 50 and then use a table of random digits or the TI-84 entry **randInt(1,50,5)** to select five of them, discarding any duplicates.
Here is a dot plot for one random sample of size five. Although results will vary, it is unlikely that a single sample of size five will include Texas or Alaska, so these values, ranging from 19.9 to 44.6, do appear to capture the essence of where most of the areas lie. The mean of these five values is 30.80.



c. The samples will occasionally pick up Texas or Alaska, so the sampling distribution for these small samples is still highly skewed toward the larger values but with a much smaller spread than the population. The mean is still in the low 40's.

d. To get a sample mean between 95 and 105, one of the states must be either Alaska or Texas. The five states could be, for example, Alaska, Minnesota, Florida, Iowa, and Kentucky. If Texas is chosen, the other four states must all be fairly large, say California, Arizona, Idaho, and Montana.

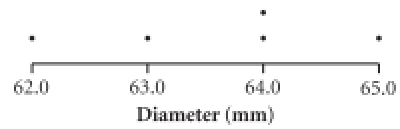
E4. a. For example, the midrange of the sample of rectangle areas from steps 2–3 of the activity—4, 10, 4, 4, and 16—is $\frac{(4+16)}{2}$, or 10.

b. The midrange also has a mound-shaped, nearly symmetric sampling distribution. (Do not be misled by the one short bar; we could change the scale on the *x*-axis to eliminate this.) The center of this distribution is well above the population mean but a little below the population midrange—which is $\frac{(1+18)}{2}$, or 9.5. (A larger simulation should produce a center much closer to the population midrange, however.) The spread of this distribution is comparable to that of the mean and appears to be a little less than that of the median. (The spread of the midrange is, in theory, a little larger than the spread of the mean for samples from nearly normal populations.)

c. The midrange of the sample is a good estimator of the midrange of the population of rectangle areas, 9.5. However, this is not a particularly good measure of center for this skewed population, as most of the areas are below 9.5.

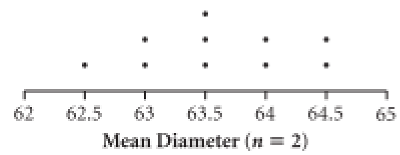
Note: In general, the midrange is greatly influenced by outliers—which are not present in this data set—and may produce a very poor estimate in some cases. The mean is influenced by outliers but to a lesser degree. The median is not influenced much by outliers in a sample.

E5. a.



b. 62 and 63; 62 and 64; 62 and 64; 62 and 65; 63 and 64; 63 and 64; 63 and 65; 64 and 64; 64 and 65; 64 and 65.

c. In the same order as the samples listed in part b, their means are 62.5, 63, 63, 63.5, 63.5, 63.5, 64, 64, 64.5, and 64.5. The exact sampling distribution of the sample mean for samples of size 2 is shown in this dot plot.



The sampling distribution has a mean of 63.6 and a standard error of about 0.6245. The population has a mean of 63.6 and a standard deviation of about 1.02. The two means are exactly the same and, as usual, the standard error is smaller than the standard deviation of the population.

Note: *You may have noticed that the standard error is not equal to*

$$\sigma / \sqrt{n} = 1.02 / \sqrt{2} = 0.72.$$

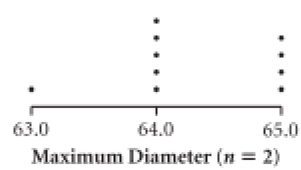
That is because this formula is true only if there is an infinite population or if sampling with replacement is used, which we did not do in this exercise. Because we are taking a sample of size 2 from a population of only five tennis balls, sampling with or without replacement makes quite a difference in the standard error.

If the tennis balls had been sampled with replacement, this formula would give the correct SE. But because there is no replacement a different formula is needed:

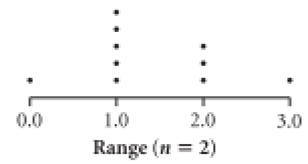
$$SE = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{1.02}{\sqrt{2}} \cdot \sqrt{\frac{3}{4}} = 0.6245$$

This also would be the correct short-cut formula to use in E6, where the sampling distribution for the mean is computed by sampling without replacement in part c.]

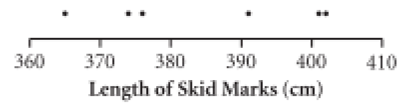
d. The sample maximums for the 10 possible samples are 63, 64, 64, 65, 64, 64, 65, 64, 65, 65. A dot plot is shown here.



e. The sample ranges are 1, 2, 2, 3, 1, 1, 2, 0, 1, and 1. A dot plot is shown here.



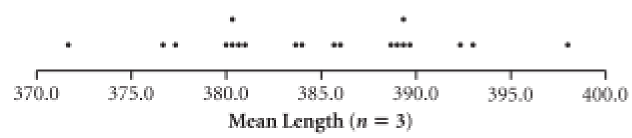
E6. a. The sample mean is $\bar{x} = 384.83$.



b. The 20 possible samples are listed here.

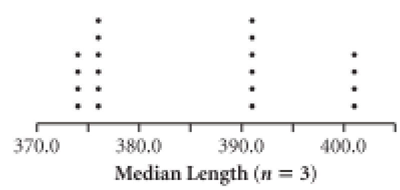
365, 374, 376;	374, 376, 391;
365, 374, 391;	374, 376, 401;
365, 374, 401;	374, 376, 402;
365, 374, 402;	374, 391, 401;
365, 376, 391;	374, 391, 402;
365, 376, 401;	374, 401, 402;
365, 376, 402;	376, 391, 401;
365, 391, 401;	376, 391, 402;
365, 391, 402;	376, 401, 402;
365, 401, 402;	391, 401, 402

c. The 20 sample means are 371.667, 376.667, 380.000, 380.333, 377.333, 380.667, 381.000, 385.667, 386.000, 389.333, 380.333, 383.667, 384.000, 388.667, 389.000, 392.333, 389.333, 389.667, 393.000, and 398.000. A dot plot is shown here.



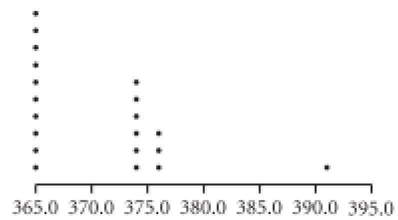
This exact sampling distribution has mean 384.83, which is the same as the population mean. It has a standard error of about 6.30—smaller than that of the standard deviation of the population.

d. The 20 sample medians are 374, 374, 374, 374, 376, 376, 376, 391, 391, 401, 376, 376, 376, 391, 391, 401, 391, 391, 401, and 401. A dot plot is shown here.



The median of this sampling distribution is 383.5. The median of the six measurements is also 383.5. (This does not always occur in distributions that are not symmetric.)

e. The dot plot of the sampling distribution of the sample minimum is shown here. As is typical, it is strongly skewed to the right. The mean of this sampling distribution is 370.65, whereas the minimum in the population of measurements is 365 cm. The sample minimum is not a good estimator of the population minimum; it tends to be too big.



E7. a. The sample mean is an unbiased estimator of the population mean because the mean of the sampling distribution of the sample mean is the same as the population mean.

b. The sample maximum is not an unbiased estimator of the population maximum because the mean of the sampling distribution of the sample maximum is less than the population maximum. The sample maximum tends to be too low.

c. The sample range is not an unbiased estimator of the population range because the mean of the sampling distribution of the sample range is less than the population range. The sample range tends to be too low.

E8. a. The sample mean is an unbiased estimator of the population mean because the mean of the sampling distribution of the sample mean is the same as the population mean.

b. The sample median is an unbiased estimator of the population median because the mean of the sampling distribution of the sample median is the same as the population median. Note: *This is not generally the case in distributions that are not symmetric.*

c. The sample minimum is not an unbiased estimator of the population minimum because the mean of the sampling distribution of the sample minimum is greater than the population minimum. The sample minimum tends to be too high.

E9. a. The population range is 100 – 0 = 100.

b. The sample range is always less than or equal to 100, so the mean of these values is less than 100. In this case, the mean of the values was 78.725.

c. Joel's estimate is more symmetric, but it is still centered too low. The mean value of doubling the IQR for his 10,000 samples was 78.725. This is still a biased estimator. Additionally, the spread is too large to be able to have much confidence in the result of any one sample.

E10. a. Because the inspector is selecting only two cartons and the cartons are either all good or all bad, there are only three possible choices for the number of pounds of spoiled fish. The sample of 48 packages will contain 0, 24, or 48 pounds of spoiled fish. We cannot attach a relative frequency to these three values with the information given.

b. The inspector thought he had a sample of size 48, but he only had a sample of size 2. One way he can improve his plan is to have a larger sample size, which means randomly selecting more cartons. Because all fish within a carton are in the same state, sampling any one fish from a carton will tell the whole story for that carton.

E11. a. – e.

Sample	Mean	Variance, Dividing by $n = 2$	Variance, Dividing by $n - 1 = 1$
2, 2	2	0	0
2, 4	3	1	2
2, 6	4	4	8
4, 2	3	1	2
4, 4	4	0	0
4, 6	5	1	2
6, 2	4	4	8
6, 4	5	1	2
6, 6	6	0	0
Average	4	$1\frac{1}{3}$	$2\frac{2}{3}$

f. The variance of the population {2, 4, 6} is $\frac{8}{3}$ which is equal to the average of the sample variances when you divide by $n - 1$.

g. Dividing by $n - 1$, because the average of the sample variances is equal to the population variance. Although this phenomenon has been shown for only one example, it is true in all cases.

h. Too small. When you divide by $n - 1$, the sample variance is an unbiased estimator of the population variance. This means that although the sample variance is not always equal to the population variance, the average of the variances of all possible samples is exactly equal to the population variance. If you divide by n, the sample variance tends to be too small, on average. Note that the sample mean, when you divide by n, is an unbiased estimator of the population mean. If you divided by $n - 1$, it would be too big.

E12. a. Results from the student's samples will vary, but in each case the effect of dividing by 4 rather than by 5 makes the standard deviation larger.

b. When we divide by $n - 1$, the center of the sampling distribution is much closer to the population standard deviation than when we divide by n. Because we always like the results from our sample to be as close as possible to the population parameter, we choose to divide by $n - 1$.
Note that, even dividing by $n - 1$, the sample standard deviation is a biased estimator of the population standard deviation—it tends to be a bit too small.

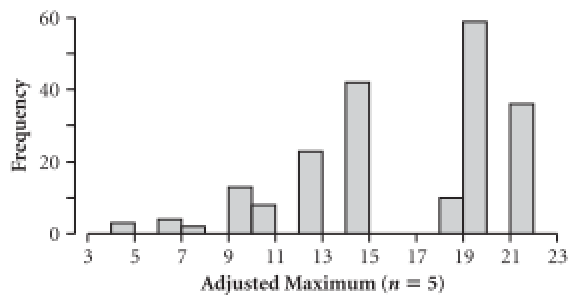
These sampling distributions of the sample standard deviation are skewed left. But that isn't typically the case. Because the standard deviation is always positive, there is a wall at 0. For a mound-shaped population, this limit results in a sampling distribution for the standard deviation that is skewed to the right. As the sample size increases, the skew decreases and the sampling distribution becomes more like a normal distribution.

E13. a. There would be five approximately equal gaps over a distance of N, so the average gap between each number is (1/5)N.

b. The maximum would be 4/5 of the way to N, so it would be at (4/5)N. Alternatively, in terms of the average gap, the sample maximum would occur after 4 average gaps or at 4 (1/5)N = (4/5)N.

c. From (b), *sample maximum* ≈ (4/5)*N*, so *N* ≈ (5/4) · *sample maximum*

d. The sampling distribution of the adjusted maximum from the same batch of 200 random samples of size 5 is shown in the next display. It appears from this simulation, that the adjusted maximum does a better job of estimating the population maximum. However, the mean of this simulated sampling distribution is still a little less than the population maximum of 18. This difference is because the distribution of the areas of the rectangles is not uniform. This sampling distribution has mean 16.8, and standard deviation 4.44.



E14. a. Yes, because the bulk of the values of a skewed distribution would be concentrated on a small set, and so the median of the samples chosen to form the sampling distribution of the median will be close together. However, for the sampling distribution of the mean, the values in the tail of the distribution affect the means of the samples more dramatically, thereby creating more variability in its sampling distribution.

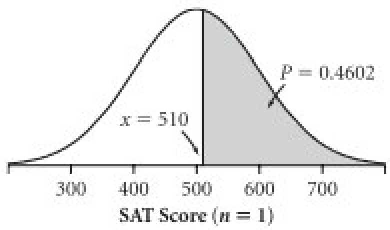
b. Yes.

c. The simulations confirm the observations in (a) and (b), but do not constitute proof because they are a single simulation, and there is no guarantee it is representative of all possible simulations.

E15. a. The z-score is

$$z = \frac{510-500}{100} = 0.10$$

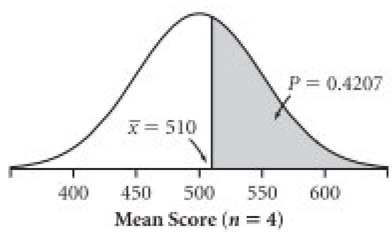
which gives a probability of 0.4602. Alternatively, on a TI-84 the result of **normalcdf (510,9999999,500,100)** is approximately 0.4602.



b. The z-score is

$$z = \frac{510-500}{100/\sqrt{4}} = 0.20$$

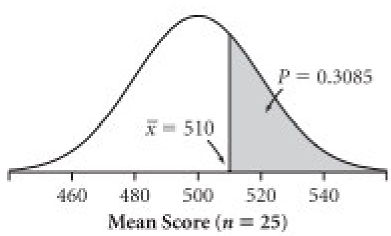
which gives a probability of 0.4207.



c. The z-score is

$$z = \frac{510-500}{100/\sqrt{25}} = 0.50$$

which gives a probability of 0.3085.

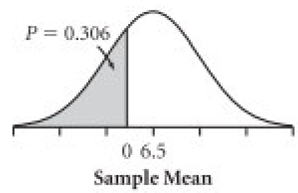


d. The z-score for 490 is

$$z = \frac{490-500}{100/\sqrt{25}} = -0.50 .$$

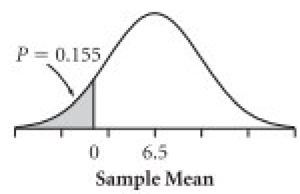
Using this, together with the z-score for 510 given in (c), we find that $P\left(490 < \bar{x} < 510\right) = P(-0.5 < z < 0.5) = 0.383$.

E16. a. $z = \dfrac{0-6.5}{12.8} = -0.507$



Using your calculator, **normalcdf**(–1E99, 0, 6.5, 12.8) = 0.306.

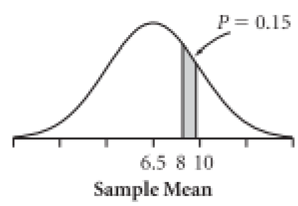
b. Now $\sigma_{\bar{x}} = \frac{12.8}{\sqrt{4}} = 6.4$, so $z = (0 - 6.5) / 6.4 = -1.016$. The probability is about 0.155.



Using your calculator, **normalcdf**(–1E99, 0, 6.5, 6.4) = 0.155.

c. $\sigma_{\bar{x}} = 12.8 / \sqrt{8} \approx 4.53$. We need to find the z-score and probability for both 10% and 8%, then subtract the probabilities.

For 10%, $z = (10 - 6.5) / 4.53 = 0.773$, $P= 0.780$.
For 8%, $z = (8 - 6.5) / 4.53 = 0.331$, $P= 0.630$.
The probability between these two values is 0.15.



Using your calculator, **normalcdf**(8, 10, 6.5, 4.53) $\approx$ 0.15.

d. The middle 95% of price increases are between $6.5 \pm 1.96 \cdot 4.53$, or between –2.38 and 15.38.

E17. a. If the weekends are randomly selected, we would expect to see $2 \cdot 67.4 = 134.8$, or a total of about 135 children. The expected number is 134.8 children.

b. Since the numbers of children seen on a summer weekend were approximately normally distributed, the sampling distribution of the mean, even with $n = 2$, will be approximately normally distributed. The z-score is

$$z = \frac{73 - 134.8}{10.4/\sqrt{2}} = -8.40$$

and the probability is almost 0.

c. We can conclude that the low number of children seen in the emergency room is not due to chance.

E18. a. 16,525,437 • \$120.02 = \$1,983,382,949.

b. With a sample of 100 people, the sampling distribution of the mean will be approximately normal with μ_{sum} = 100 • \$120.02 = \$12,002 and σ_{sum} = \$100 • $\sqrt{100}$ = \$1000.

The *z*-score is $z = \dfrac{12,500 - 12,002}{1000} = 0.498$ and the probability is about 0.3083.

c. The *z*-score is $z = \dfrac{0 - 12,002}{1000} = -12.002$, so that $P(sum \leq 0)$ is very close to zero. There is virtually no chance that the casino would lose money on a randomly selected group of 100 customers.

E19. a. I. Histogram B; $n = 25$
II. Histogram A; $n = 4$
III. Histogram C; $n = 2$

b. The theoretical standard error is $2.402/\sqrt{n}$ which turns out to be 1.698, 1.201, and 0.480 for the respective sample sizes of 2, 4, and 25. All of these are fairly close to the observed standard errors.

c. For samples of size 2 and 4, the simulated sampling distributions of the mean reflect the skewness of the population distribution. For samples of size 25, the skewness is essentially eliminated and the simulated sampling distribution looks like a normal distribution.

d. The rule that about 95% of the observations lie within two standard errors of the population mean works well for $n = 25$, and slightly less well for the skewed distributions occurring for $n = 4$ and $n = 2$.

E20. a. I. Histogram B; $n = 4$
II. Histogram C; $n = 25$
III. Histogram A; $n = 2$

b. The population standard deviation for this distribution is 3.5. This standard deviation divided by the square root of 2, 4, and 25, respectively, yields 2.47, 1.75, and 0.70. These values are quite close to the observed standard deviations of the simulated sampling distributions.

c. Despite the new peak centered at the mean, the simulated sampling distribution for $n = 2$ still reflects much of the pattern of the population, showing the mounds at the extremes. For $n = 4$, a little of the population pattern remains, but by $n = 25$ it disappears and all that is seen is an essentially normal distribution.

d. The rule works well for $n = 25$, but not nearly so well for the smaller sample sizes. As usual, the rule works well for the sampling distribution of the sample mean as long as the sample size is reasonably large.

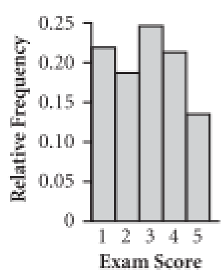
E21. a. No. Exactly two accidents happened in about 15% of the days. Two or more accidents happened in about 25% of the days.

b. In the display, the first plot is the one for eight days, and the second plot is the one for four days.

c. For the four-day averages, two accidents are not very likely. It occurs here about 6 times out of 100. For the second part, yes, if the days can be viewed as a random sample of all days. For the eight-day average, an average of two accidents never occurred and, hence, could be deemed very unlikely.

d. The sampling distributions used in parts a–c are based on random samples of four and eight days. A particular period of four (or eight) days may not look like a random sample at all because of the high dependency from day to day. For example, if all days of a sample are taken from the winter season in a region that has ice and snow, the accident rate might be far higher than what is typical for the rest of the year.

E22. a.

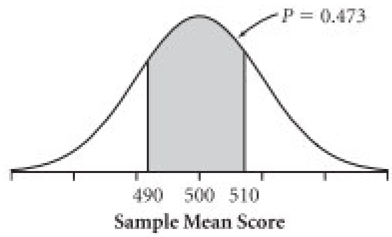


b. A has $n = 25$, B has $n = 1$, and C has $n = 5$. Remind students that repeated sampling with samples of size 1 should produce a distribution that looks very much like the population.

c. If the class can be considered a random sample of the students who took this exam, then an average of 3.6 would be very unusual for a class size of 25. A more reasonable conclusion is that the class size is 5. It is always possible, however, that 25 students in a class would do much better than a random sample of 25 students.

E23. a. You want the probability that the mean score will be between 490 and 510. For a mean score of 490, the z-score is $z = (490 - 500) / (100/\sqrt{40}) \approx -0.632$ and the probability that the mean is less than or equal to 490 is about 0.2635.

For a mean score of 510, the z-score is $z = (510 - 500) / (100/\sqrt{40}) \approx 0.632$ and the probability that the mean is less than or equal to 510 is about 0.7365. So the probability that the mean score is between 490 and 510 is 0.7365 – 0.2635 = 0.473.



Or, **normalcdf**(490, 510, 500, $100/\sqrt{(40)}$) = 0.473

b. We know that if the sampling distribution is approximately normal, about 95% of all sample means are in the interval $\mu_{\bar{x}} \pm 1.96\frac{\sigma}{\sqrt{n}}$. Thus, to be 95% sure that the sample mean is within a value E of the population mean, we must have $E \geq 1.96\frac{\sigma}{\sqrt{n}}$. Solving for the square root gives: $\sqrt{n} \geq 1.96\frac{\sigma}{E}$. Squaring both sides and simplifying gives:

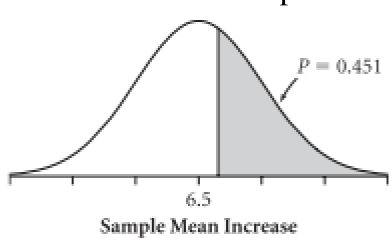
$$n \geq \frac{3.8416\sigma^2}{E^2} = \frac{3.8416 \cdot 100^2}{10^2} \approx 384.16$$

Rounding up, this gives a sample size of 385.

E24. a. The z-score is

$$z = \frac{7 - 6.5}{12.8/\sqrt{10}} = 0.1235$$

and the probability that the mean increase is greater than 7 is about 0.451. There is about a 45.1% chance that the mean increase in her stock prices will exceed 7%.



Or, **normalcdf**(7, 1E99, 6.5, $12/\sqrt{10}$) ≈ 0.451.

b. Jenny wants

$$P\left(\overline{X} > 5\right) = P\left(\frac{\overline{X} - 6.5}{12.8/\sqrt{n}} > \frac{5 - 6.5}{12.8/\sqrt{n}}\right) = 0.95 .$$

Because the sample mean is approximately normally distributed and the *z*-score cutting off an area of 0.95 to the right is -1.645, it must be that

$$\frac{5-6.5}{12.8/\sqrt{n}} = -1.645 .$$

Solving this equation for *n* gives *n* = 197.05. Hence, Jenny should choose about 197 randomly selected stocks.

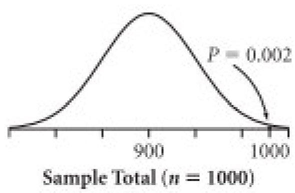
E25. a. The sampling distribution of the sample total should be approximately normal because this is a very big sample size. It has mean and standard error

$$\mu_{sum} = n\mu = 1000(0.9) = 900$$

$$\sigma_{sum} = \sigma\sqrt{n} = 1.1\sqrt{1000} \approx 34.79$$

The *z*-score for 1000 families is then

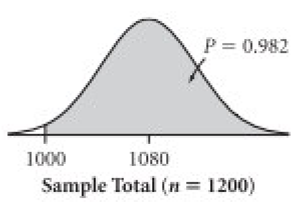
$$z = \frac{\text{sum} - n\mu}{\sigma\sqrt{n}} = \frac{1000-900}{34.785} \approx 2.875 .$$



The probability of getting at least 1000 children is about 0.002. Thus, there is almost no chance the network will get 1000 children.

b. Yes, the probability goes from practically 0 to almost certain. The *z*-score for 1200 families is

$$z = \frac{\text{sum} - n\mu}{\sigma\sqrt{n}} = \frac{1000-1080}{38.105} \approx -2.099 .$$



The probability of getting at least 1000 children in a random sample of 1200 families is 0.982.

E26. a. $0.9 \pm 1.96(1.1/\sqrt{25}) = 0.9 \pm 0.4312$ or (0.469, 1.331)

b. $0.9 \pm 1.96(1.1/\sqrt{100}) = 0.9 \pm 0.2156$ or (0.684, 1.116)

c. $0.9 \pm 1.96(1.1/\sqrt{1000}) = 0.9 \pm 0.0682$ or (0.832, 0.968)

d. $0.9 \pm 1.96(1.1/\sqrt{4000}) = 0.9 \pm 0.0341$ or (0.866, 0.934)

E27. We will do this problem using the sampling distribution of the sample sum, which has a mean of $4.3n$ and a standard error $1.4 \cdot \sqrt{n}$.

a. The point that cuts off the lower 0.02 (so 98% is above it) on the normal distribution is at $z = -2.054$. If the total number of people is above 100, then we must choose the sample size, *n,* so that the point 100 lies 2.054 standard errors below $4.3n$, or

$$100 = 4.3n - 2.054 \cdot 1.4\sqrt{n} \text{ or } 4.3n - 2.8756\sqrt{n} - 100 = 0$$

This equation is quadratic in form. It can be solved for $\sqrt{n}$ by using the quadratic formula and then squaring the positive solution to get *n.* The negative solution for $\sqrt{n}$ can be ignored. Alternatively, the solution can be estimated by plotting the equation on a graphing calculator. Because we only need the nearest integer solution, either method works well. The solution is about $n = 27$, so the director should select about 27 names.

b. However, drawing 27 names, he is likely to get more than 100 people. The expected number of people he will get is 4.3 • 27 = 116.1 people. So 16.1 extra people will cost him 16.1 • $250 = $4025. This was a pretty costly oversight!

E28. Note that $\mu = 2.236$, $\sigma = 1.115$. We will do this problem using the sampling distribution of the sample sum, which has a mean of $2.236n$ and a standard error $1.115 \cdot \sqrt{n}$.

The point that cuts off the lower 0.05 (so 95% is above it) on the normal distribution is at $z = -1.645$. If the total number of color television sets is above 1000, then we must choose the sample size, *n,* so that the point 1000 lies 1.645 standard errors below $2.235n$. That is,

$$1000 = 2.236n - 1.645 \cdot 1.115\sqrt{n}$$

or equivalently

$$2.236n - 1.834\sqrt{n} - 1000 = 0.$$

This equation is quadratic in form. It can be solved for $\sqrt{n}$ by using the quadratic formula and then squaring the positive solution to get *n.* The negative solution for $\sqrt{n}$ can be ignored. Alternatively, the solution can be estimated by plotting the equation on a graphing calculator. Because we only need the nearest integer solution, either method works well. The solution is about $n = 465$ households.

E29. **a.** Observe that

i. 0.9535 **ii.** 0.7107 **iii.** 0

b. For all sample sizes, the *SE* is smaller with the adjustment for non-replacement, not much smaller at first, but decreases to 0 when $n = N$. As the sample size increases, you are sure to have a bigger proportion of the population in the sample and so $\bar{x}$ tends to be closer to μ.

E30. a.

Remaining Test Scores	**Mean**	$(\text{Mean}-82)^2$
68, 75, 82	75	49
68, 75, 90	77.7	18.49
68, 75, 95	79.3	5.29
68, 82, 90	80	4
68, 82, 95	81.7	0.09
68, 90, 95	84.3	5.29
75, 82, 90	82.3	0.09
75, 82, 95	84	4
75, 90, 95	86.7	22.09
82, 90, 95	89	49

The mean of the sampling distribution is the sum of the values in the middle column divided by 10, namely 82. This coincides with the population mean of

$$\frac{68+75+82+90+95}{5}=82\,.$$

The standard error for the sampling distribution is

$$\sqrt{\frac{\sum(\text{rightmost column})}{10-1}}=4.208\,.$$

b. We need σ for population:

$$\sigma=\sqrt{\frac{(68-82)^2+(75-82)^2+(82-82)^2+(90-82)^2+(95-82)^2}{5-1}}\approx 10.93\,.$$

Now, using $\sigma_{\bar{x}}=\dfrac{\sigma}{\sqrt{n}}$, we obtain $\sigma_{\bar{x}}=\frac{10.93}{\sqrt{3}}=6.310$. In contrast, using

$\sigma_{\bar{x}}=\dfrac{\sigma}{\sqrt{n}}\cdot\sqrt{\dfrac{N-n}{N-1}}$, we obtain $\sigma_{\bar{x}}=\frac{10.93}{\sqrt{3}}\cdot\sqrt{\frac{5-3}{5-1}}=4.462$.

c. The second formula is close.

E31. a.

$$\mu_{v+m}=\mu_v+\mu_m=501+515=1016$$

$$\sigma_{v+m}=\sqrt{\sigma_v^2+\sigma_m^2}=\sqrt{112^2+116^2}\approx 161.245$$

b. The sampling distribution of the sum will be approximately normal because both distributions are normal. The *z*-score is

$$z=\frac{800-1016}{161.245}=-1.340$$

The probability is about 0.0901.

c. 1016 ± 1.96(161.245), or between approximately 700 and 1332.

d. We need $v - m \geq 100$. The sampling distribution of the difference will be approximately normal because both distributions are normal. The mean and standard error of the sampling distribution of the difference are

$$\mu_{v-m} = \mu_v - \mu_m = 501 - 515 = -14$$

$$\sigma_{v-m} = \sqrt{\sigma_v^2 + \sigma_m^2} = \sqrt{112^2 + 116^2} \approx 161.245$$

The z-score for a difference of 100 is

$$z = \frac{100 - (-14)}{161.245} \approx 0.707$$

The probability is 0.2969 (or 0.2981 using Table A).

Note: This does not imply that 29.69% of students have an SAT verbal score of at least 100 points higher than their SAT math score. See part e.

e. The mean of the distribution of the sum will still be 1016. The standard error is unpredictable because the two scores are certainly not independent. Students who score high on the verbal portion also tend to score high on the math portion. The shape is also unpredictable.

E32. a.

$$\mu_{v+m} = \mu_v + \mu_m = 69.3 + 64.1 = 133.4$$

$$\sigma_{v+m} = \sqrt{\sigma_v^2 + \sigma_m^2} = \sqrt{2.92^2 + 2.75^2} \approx 4.011$$

b. The sampling distribution of the sum will be approximately normal because both distributions are normal. The z-score is

$$z = \frac{125 - 133.4}{4.011} = -2.094$$

The probability is about 0.0181.

c. 133.4 ± 1.96(4.011), or between approximately 125.54 and 141.26.

d. We need $v - m \geq 2$. The sampling distribution of the difference will be approximately normal because both distributions are normal. The mean and standard error of the sampling distribution of the difference are

$$\mu_{v+m} = \mu_v + \mu_m = 69.3 - 64.1 = 5.2$$

$$\sigma_{v+m} = \sqrt{\sigma_v^2 + \sigma_m^2} = \sqrt{2.92^2 + 2.75^2} \approx 4.011$$

The z-score for a difference of 200 is

$$z = \frac{2 - (5.2)}{4.011} \approx -0.798$$

The probability is 0.242.

e. The mean of the distribution of the sum will still be 133.4. The standard error is unpredictable because the two scores are certainly not independent. The shape is also unpredictable.

E33. a. $\mu_{\text{sum}} = 3(3.5) = 10.5$; $\sigma^2_{sum} = 3(2.917) = 8.751$

b. $\mu_{\text{sum}} = 7(3.5) = 24.5$; $\sigma^2_{sum} = 7(2.917) = 20.419$

c. The sampling distribution of the sum is approximately normal because we were told in E31 that the distribution of verbal scores is approximately normal. It has mean and variance

$$\mu_{sum} = 20 \cdot 501 = 10,020$$

$$\sigma^2_{sum} = n\sigma^2 = 20 \cdot 112^2 = 250,880$$

The z-score for a total of 10,000 is

$$z = \frac{10,000 - 10,020}{\sqrt{250,880}} \approx -0.040$$

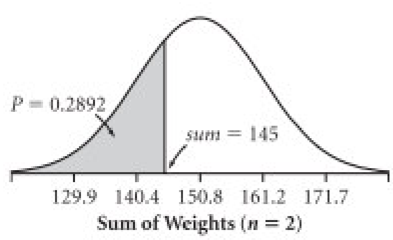
The probability is about 0.4840.

E34. a. Because the distribution of the weights is approximately normal, the sampling distribution of the sum of two (or any number of) weights will be approximately normal as well. The mean and standard error are

$$\mu_{sum} = 75.4 + 75.4 = 150.8$$

$$\sigma_{sum} = \sqrt{7.38^2 + 7.38^2} \approx 10.437$$

b. Using the results from part a, the z-score is –0.556 and the probability is 0.2892.



c. From E33, the mean and standard error of the sampling distribution are

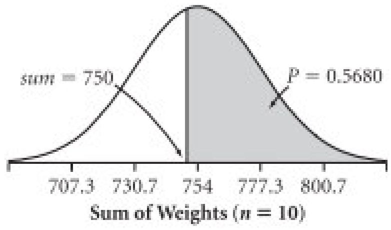
$$\mu_{sum} = n\mu = 10(75.4) = 754$$

$$\sigma_{sum} = \sqrt{n\sigma^2} = \sqrt{10(7.38)^2} \approx 23.338$$

The z-score for a total of 750 is

$$z = \frac{750 - 754}{23.338} \approx -0.1714$$

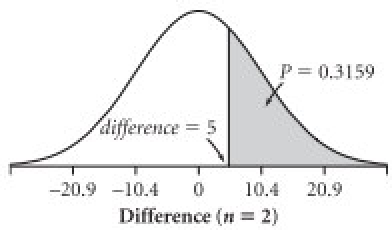
The probability that the total weight is more than 750 kg is 0.5680.



d. The first ram is more than 5 kg heavier than the second ram if $ram_1 - ram_2 > 5$. The sampling distribution of the difference $ram_1 - ram_2$ has mean $\mu_{1-2} = \mu_1 - \mu_2 = 75.4 - 75.4 = 0$.

The standard error is the same as that in part a, 10.437.
The z-score for a difference of 5 is

$$z = \frac{5-0}{10.437} \approx 0.4791$$



The probability that the difference is greater than 5 is about 0.3159.

E35. The sampling distributions for the sample proportion and the number of successes can be considered approximately normal because both $np = 920$ and $n(1 - p) = 80$ are 10 or greater.

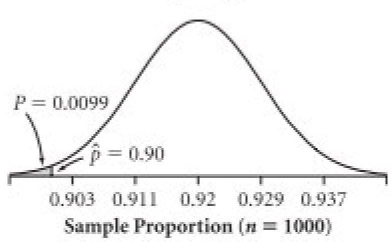
a. This distribution is approximately normal with mean and standard error

$$\mu_{\hat{p}} = 0.92$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.92(1-0.92)}{1000}} \approx 0.00858$$

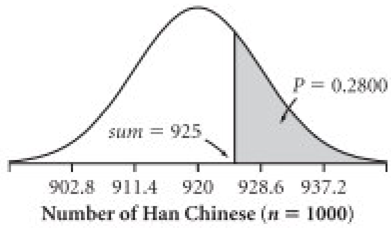
The sketch, with scale on the x-axis, is shown in part (c).

b. The z-score is: $z = \dfrac{0.9-0.92}{\sqrt{\dfrac{0.92(1-0.92)}{1000}}} \approx -2.33$



The probability is about 0.0099.

c. The z-score is: $z = \dfrac{925-920}{\sqrt{1000(0.92)(1-0.92)}} \approx 0.5828$



The probability is about 0.2800.

d. Rare events would be the totals that are outside the interval 920 ± 1.96(8.58); that is, larger than 936.82 or smaller than 903.18. Rare events would be the proportions that are outside the interval 0.92 ± 1.96(0.00858); that is, larger than 0.937 or smaller than 0.903.

E36. a. Because $np = 500 \cdot 0.223 = 111.5$ and $n(1-p) = 500 \cdot 0.777 = 388.5$ are both ten or more, the shape of the distribution will be approximately normal.

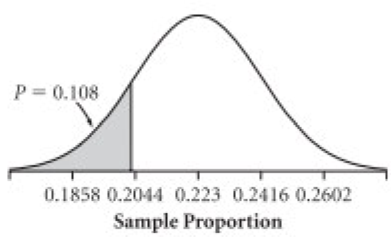
$$\mu_{\hat{p}} = p = 0.223 \text{ and } \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.223 \cdot 0.777}{500}} \approx 0.0186$$

The sketch, with scale on the x-axis, is shown in part c.

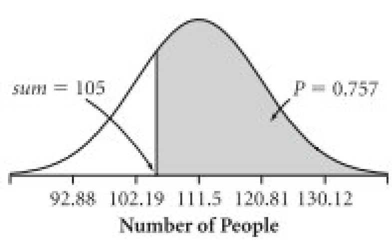
b. Using the sampling distribution for the sample proportion, the z-score for a sample proportions of 0.2 is

$$z = (0.2 - 0.223) / 0.0186 = -1.24.$$

The probability of getting a sample proportion of 0.20 or fewer with one of these surnames is 0.108.



c. Using the sampling distribution for the sample sum, the z-score for 105 successes is $z = (105 - 111.5) / 9.31 = -0.698$. The probability of getting 105 people or more with one of these surnames is 0.757.



d. Rare events would be the totals that are outside the interval 111.5 ± 1.96(9.31); that is, larger than 129.75 or smaller than 93.25. Rare events would be the proportions that are outside the interval 0.223 ± 1.96(0.0186); that is, larger than 0.259 or smaller than 0.187.

E37. a. The shape would be slightly skewed to the left because $np = 100 \cdot 0.92 = 92$ and $n(1-p) = 100 \cdot 0.08 = 8$ are not both at least 10. The mean of the sampling distribution would still be 0.92 and the spread would be greater because of the smaller sample size.

b. With $n = 100$, the distribution is more spread out and skewed left than the distribution for $n = 1000$. This means a larger percentage of the sample proportions would be further from p and, in this case, in the direction of the skew or in the left tail. Because the sample proportion of 0.9 is in the left portion of the distribution or the part that is with the skew, the probability of getting 90% or fewer in the sample will be greater with a sample size of 100 than with a sample size of 1000.

Note: we cannot accurately calculate this probability using our formula because the distribution is not approximately normal.

E38. a. $np = 100 \cdot 0.223 = 22.3$ and $n(1-p) = 100 \cdot 0.777 = 77.7$ are both at least 10 so the shape will still be approximately normal. The mean of the sampling distribution of the sample proportion will still be 0.223 but the standard deviation will be larger by a factor of $\sqrt{5}$.

b. The probability of getting 20% or fewer in the sample will be greater with a sample size of 100 because the distribution is more spread out. This means a larger percentage of the sample proportions would be further from p.

E39. a. 0.5

b. The sampling distribution can be considered approximately normal because both $np = 25$ and $n(1-p) = 25$ are 10 or greater. The sampling distribution has mean and standard error

$$\mu_{\hat{p}} = 0.5 \text{ and } \sigma_{\hat{p}} = \sqrt{\frac{0.5(1-0.5)}{50}} \approx 0.0707$$

The z-score for a sample proportion of 0.2 is

$$z = \frac{0.2 - 0.5}{0.0707} \approx -4.2433$$

The probability of getting 10 or fewer under the median age is 0.000011.

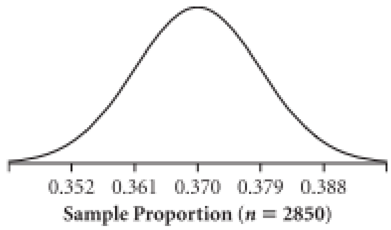
c. These results are not at all what we would expect from a random sample of 50 people. This group is special in that they are all old enough to have a job. Children are included in computing the 33.1 median age, but they do not hold jobs. Thus, we would expect the median age of job holders to be greater than the median age of all people in the United States, as they are at Westvaco.

E40. The sampling distribution for the sample total can also be considered approximately normal because both $np = 1656.86$ and $n(1-p) = 2821.4$ are 10 or greater.

a. This distribution is approximately normal with mean and standard error

$$\mu_{\hat{p}} = 0.37$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.37 \cdot 0.63}{4478}} \approx 0.007$$



b. The z-score for a sample proportion of 0.54 is

$$z = \frac{0.54 - 0.37}{\sqrt{\frac{0.37 \cdot 0.63}{4478}}} \approx 30$$

The probability is very close to 0.

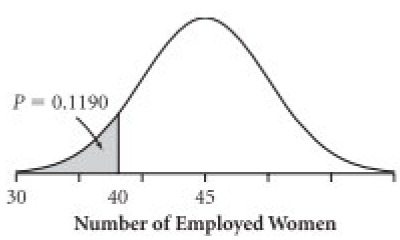
c. The z-score for a sample proportion of 0.58 is 30 from part d. The probability of getting this large of a percentage or larger by chance alone is very close to 0. You should conclude that this group is special in some way. (In fact, many of the students come from urban high schools, and a large percentage does not speak English at home.)

E41. You are assuming that the 75 married women were selected randomly from a population in which 60% are employed. Both np = 75 • 0.6 = 45 and $n(1 - p)$ = 75 • 0.4 = 30 are at least 10, so the distribution will be approximately normal. We will use the distribution of the sum which has

$$\mu_{sum} = np = 45 \text{ and } \sigma_{sum} = \sqrt{np(1-p)} = \sqrt{75 \cdot 0.6 \cdot 0.4} \approx 4.24.$$

For 30 employed women in the sample, $z = (30 - 45) / 4.24 = -3.538$.
$P(sum \le 30) = 0.0002$.

For 40 employed women in the sample, $z = (40 - 45) / 4.24 = -1.179$.
$P(sum \le 40) = 0.1192$.
$P(30 \le \text{sum} \le 40) \approx 0.1192 - 0.0002 = 0.1190$.

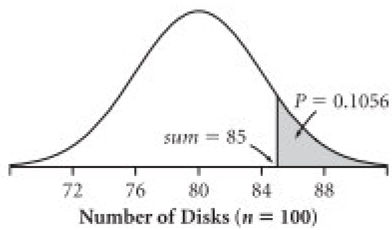


E42. Getting 15 or fewer with bad sectors is the same thing as getting 85 or more without. The sampling distribution for the number of disks that have no bad sectors can be considered approximately normal because both $np = 80$ and $n(1 - p) = 20$ are 10 or greater. The sampling distribution has mean and standard error

$$\mu_{sum} = 100(0.8) = 80 \text{ and } \sigma_{sum} = \sqrt{100(0.8)(1-0.8)} = 4$$

The z-score for 85 with no bad sectors is

$$z = \frac{85-80}{4} = 1.25$$



The probability of getting 85 or more with no bad sectors is about 0.1056. The assumption that underlies this answer is that the 100 sampled disks are a random sample from a population in which 80% of the disks have no bad sectors.

E43. **a.** The distribution for $n = 100$ is the most normal while the distribution for $n = 10$ is the least normal. Notice that for $n = 10$, $np = 10 \bullet 0.1 = 1 < 10$, and $n(1-p) = 10 \bullet 0.9 = 9 < 10$. The guideline does not hold and the distribution for $n = 10$ is quite skewed. For $n = 20$, $np = 2 < 10$, but $n(1 - p) = 18 > 10$. The guideline does not hold and this distribution is less skewed, but the skew is still very noticeable. When $n = 40$, $np = 4 < 10$ but $n(1 - p) = 36 > 10$. The guideline does not hold and this distribution is less skewed, but the skew is still noticeable. For $n = 100$, both $np = 10$ and $n(1 - p) = 90$ are at least 10. The guideline does hold and the distribution looks very symmetric and fairly smooth. The skew is visible for all but $n = 100$, where both np and $n(1 - p)$ are at least 10. It appears that the guideline works quite well.

b. In each of these, the expected proportion of drivers is 10%.

c. The distribution for $n = 10$ has the largest spread, and the distribution for $n = 100$ has the smallest spread.

d. This would be more likely in a sample of size 10 drivers. The area of the bars in the histogram to the right of 0.20 gets smaller as the sample size increases and with larger sample sizes, the proportions tend to cluster closer to 0.1.

E44. **a.** The mean for each distribution is $p = 0.10$.

For $n = 10$, $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.1 \cdot 0.9}{10}} \approx 0.095$

For $n = 20$, $\sigma_{\hat{p}} = \sqrt{\dfrac{0.1 \cdot 0.9}{20}} \approx 0.067$

For $n = 40$, $\sigma_{\hat{p}} = \sqrt{\frac{0.1 \cdot 0.9}{40}} \approx 0.047$

For $n = 100$, $\sigma_{\hat{p}} = \sqrt{\frac{0.1 \cdot 0.9}{100}} \approx 0.03$

b. Estimates will vary, but the answers from part (a) should match quite well. In particular, the sampling distribution of the sample proportion for all sample sizes have a center that is close to 0.1. For the sampling distribution of the sample proportion for samples of size 10, the spread is large, with reasonably likely sample proportions ranging from about 0 to 0.3. For $n = 20$, the spread is smaller, with reasonably likely sample proportions ranging from about 0 to 0.23. For $n = 40$, the spread is small, with reasonably likely sample proportions ranging from about 0.01 to 0.2. For $n = 100$, the spread is smallest, with reasonably likely sample proportions ranging from about 0.04 to 0.16.

c. The mean using the table is 0.0999999, off from 0.1 because of rounding error. The standard error is about 0.095, which matches with the result in part a.

E45. a. The means of the sampling distributions definitely depend upon *p,* as the first three center close to $p = 0.2$ and the second three center close to $p = 0.4$. The sampling distributions have centers close to *p* regardless of the sample size; the centers do not depend upon the sample size.

b. The spreads of the sampling distributions decrease as *n* increases for both values of *p.* The spreads do depend upon the value of *p,* however. For each sample size, the spread for $p = 0.4$ has a larger standard error than the one for $p = 0.2$.

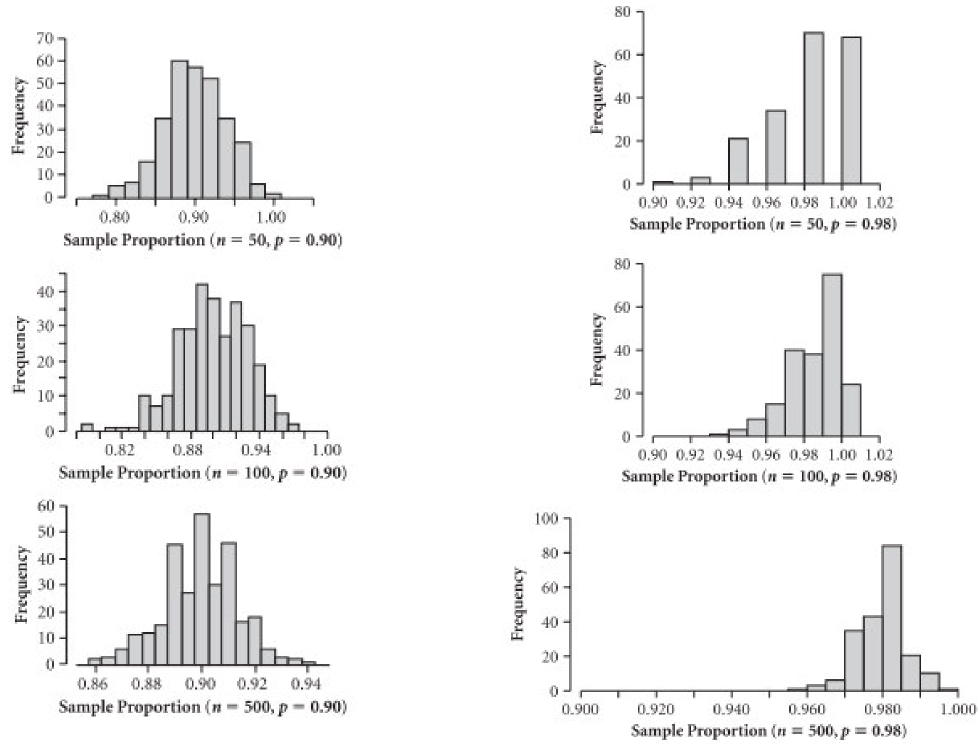
c. For $p = 0.2$, the shape is quite skewed for $n = 5$ and some slight skewness remains at $n = 25$. For $n = 100$, the shape is basically symmetric. For $p = 0.4$, the shape shows a slight skewness at $n = 5$ but is fairly symmetric at $n = 25$ and beyond. The farther *p* is from 0.5, the more skewness in the distribution of the sample proportion.

d. The rule does not work well for samples of size 5 for either value of *p* but works well for samples of size 25 or more for both values of *p.*

E46. The three histograms in the left column below are simulated sampling distributions for $p = 0.90$, and the last three are for $p = 0.98$. The sample sizes for each proportion are 50, 100, and 500.

For $p = 0.90$, the shape is skewed toward the lower values for $n = 50$ but is approximately normal when $n = 100$ and $n = 500$. For $p = 0.98$, the shape is highly skewed for $n = 50$ and remains highly skewed for $n = 100$. It becomes approximately normal at $n = 500$. The rule that np and $n(1 - p)$ must both be at least 10 seems to work quite well.
The important message: Situations where the normal model works well as an approximation depend upon both *n* and *p.* If *p* is close to 0 or 1, the sample size must be

quite large in order to produce approximately normal sampling distributions for sample proportions.



E47. a. $\overline{x} = \dfrac{\sum xf}{n} = \dfrac{0 \cdot 14 + 1 \cdot 26}{40} = 0.65$.

This sample mean is the same as $\hat{p}$ because each success adds 1 and each failure adds 0, so it counts the number of successes, then divides by the sample size. This is exactly the procedure for calculating $\hat{p}$.

b. $\mu = \sum x \cdot P(x) = 0 \cdot 0.4 + 1 \cdot 0.6 = 0.6$. The box says $\mu = p = 0.6$, which is the same.

c. The sample proportion is a type of sample mean, if the successes are given the value 1 and failures the value 0.

E48. a.

$$\sigma = \sqrt{\sum (x-\mu)^2 \cdot P(x)} = \sqrt{(0-0.6)^2 \cdot 0.4 + (1-0.6)^2 \cdot 0.6}$$
$$= \sqrt{0.6^2 \cdot 0.4 + 0.4^2 \cdot 0.6} = \sqrt{0.6 \cdot 0.4(0.6+0.4)} = \sqrt{0.6 \cdot 0.4} = \sqrt{0.24}$$

b. $\sigma = \sqrt{p(1-p)} = \sqrt{0.6 \cdot 0.4} = \sqrt{0.24}$.

Both formulas give the same result.

E49. a. Because the sample size is less than 10% of the population size, it does not really matter that we sampled without replacement, and so we can use the following formulas for the mean and standard error:

$$\mu_{\bar{x}} = \mu = 140.9$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{119.4}{\sqrt{3}} \approx 68.94$$

(The formula for the mean holds with or without replacement. Although the formula for the standard error holds exactly when sampling is done with replacement, it is only approximately true when sampling is done without replacement. The larger N is compared to n, the better the approximation.)

b. No. This is a very small sample from a highly skewed population. The sampling distribution of the sample mean will be skewed right.

E50. a. Because 18 out of the 32 cities have "good" air quality, this is the same as sampling from a population with proportion of successes $p = 18/32 = 0.5625$. Because the sample size is less than 10% of the population size, it does not really matter if we sampled with or without replacement, and so we can use the formulas shown below for the mean and standard error.

$$\mu_{sum} = 3p = 3(18/32) = 1.6875$$

$$\sigma_{sum} = \sqrt{np(1-p)} = \sqrt{3(0.5625)(1-0.5625)} \approx 0.8592$$

b. No, both $np = 3(0.5625) = 1.6875$ and $n(1-p) = 3(1-0.5625) = 1.3125$ are less than 10.

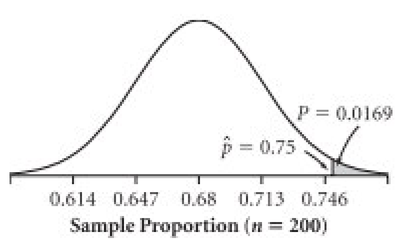
E51. a. The shape can be considered approximately normal because both $np = 200(0.68) = 136$ and $n(1-p) = 200(0.32) = 64$ are 10 or greater. The mean and standard error are

$$\mu_{\hat{p}} = 0.68$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.68 \cdot (1-0.68)}{200}} \approx 0.0330$$

b. The z-score for $\hat{p} = 0.75$ is

$$z = \frac{0.75 - 0.68}{0.0330} \approx 2.121$$



The probability is about 0.0170.

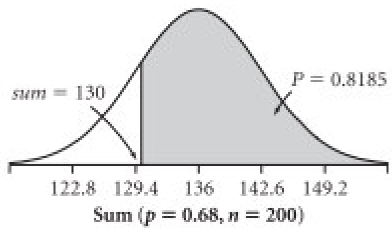
c. Values less than 0.68 – 1.96(0.0330) = 0.615, or larger than 0.68 + 1.96(0.0330) = 0.745.

d. The mean and standard error of the sampling distribution of the sum are

$$\mu_{sum} = 200(0.68) = 136 \text{ and } \sigma_{sum} = \sqrt{200(0.68)(1-0.68)} \approx 6.597$$

The z-score for 130 people is

$$z = \frac{130-136}{6.597} \approx -0.9095$$



The probability is about 0.8185.

Alternatively, this problem can be done using the mean and standard deviation in part a and with $\hat{p} = \frac{130}{200} = 0.65$.

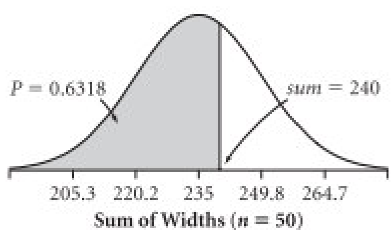
E52. a. The distribution of the sum of the spine widths of 50 randomly selected books should be approximately normal because the population is only slightly skewed right. The mean and standard error of the sampling distribution of the sum are

$$\mu_{sum} = 50(4.7) = 235$$

$$\sigma_{sum} = \sqrt{50}(2.1) = 14.849$$

The books will fit on the shelf if the sum of their widths is less than 240 cm. The z-score for 240 is

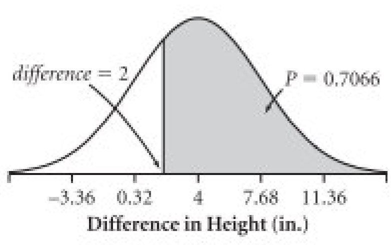
$$z = \frac{240-235}{14.849} \approx 0.3367$$



The probability is about 0.6318.

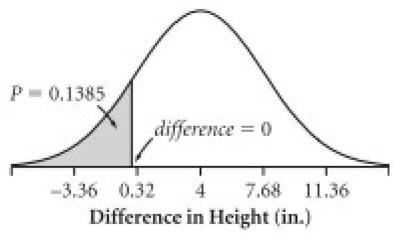
Alternatively, this problem can be done using the sampling distribution of the mean. The spine widths would have to average less than $\frac{240}{50} = 4.8$ cm.

b. No. These are not a random sample of books. If their reputation is deserved, philosophy books may be thicker than the average book, so fewer philosophy books are likely to fit on the shelf.

E53. a. One interesting feature is the regular increase from 40 to 55 followed by a sharp drop after 55 and another sharp drop after 65. **b.** Since the distribution of ages has little skew, a sampling distribution of the sample mean with n = 10 should be approximately normal. The mean and standard deviation of the distribution are $$\mu_{\bar{x}} = 56, \quad \sigma_{\bar{x}} = \frac{6.3}{\sqrt{10}} \approx 1.99$$ For a sample mean of 55, $z = \frac{55-56}{1.99} \approx -0.503$, $P(\bar{x} < 55) = 0.3085$ For a sample mean of 60, $z = \frac{60-56}{1.99} = 2.010$, $P(\bar{x} < 60) = 0.9772$ As such, $$P(55 < \bar{x} < 60) = 0.9772 - 0.3085 = 0.6687.$$
E54. The sampling distribution of the sum will be approximately normal with mean and variance $$\mu_{sum} = n\mu = 50(3.11) = 155.5$$ $$\sigma^2_{sum} = n\sigma^2 = 50(0.43^2) = 9.245$$ Thus, the reasonably likely weights of a roll of 50 pennies are those that fall in the interval $155.5 \pm 1.96\sqrt{9.245}$, or 149.54 to 161.46 grams.
E55. a. The shape is approximately normal because each of the two distributions is approximately normal. The mean and standard error of the sampling distribution of the difference are $$\mu_{m-w} = \mu_m - \mu_w = 68 - 64 = 4$$ $$\sigma_{m-w} = \sqrt{\sigma^2_m + \sigma^2_w} = \sqrt{2.7^2 + 2.5^2} \approx 3.680$$ **b.** The z-score for a difference of 2 is $$z = \frac{2-4}{3.680} \approx -0.5435$$  The probability is 0.7066.

c. She is taller than he is if the difference is negative (less than 0). The z-score for a difference of 0 is

$$z = \frac{0-4}{3.680} = -1.087$$



The probability is 0.1385.

E56. a. The students must generate five arrival times for Bus 1 paired with five arrival times for Bus 2, where the times are from a uniform distribution over the interval [0, 10]. This procedure can be done with a graphing calculator or computer. On the TI-84, enter **10rand** and then repeatedly press ENTER. Place the values into two columns of five values each. Then find the absolute value of the difference between these times. We use the absolute difference because there is no restriction on the order in which the buses arrive—we are just looking at the interarrival times. The statistic of interest is the mean of these five differences.

b. Answers will vary. The majority should get an average difference between 2.5 and 4. Reasonably likely values for the average difference range between approximately 1.25 and 5.5. No one should get a negative number.

c. The largest mean is about 6.1 minutes. Sets of arrival times will vary. One possible set of pairs, the ones we got in the simulation, is the following:

Bus 1 Arrival Time	Bus 2 Arrival Time	Absolute Difference
8.50189	4.43855	4.06334
7.81519	0.57796	7.23723
7.01766	8.97818	1.96052
0.56625	9.02486	8.45861
8.79205	0.06360	8.72845

d. No. Of the 100 repetitions in the simulation, none of the average differences are 7 or larger. Perhaps the sample was not random. Perhaps there is something in the scheduling that prevents the two buses from arriving close to the same time.

E57. Select two integers at random from 0 through 9 and compute their mean. Repeat this. Subtract the second mean from the first. Students should do this three times.

Because the distribution of the results of a single selection from the population {0, 1, 2, 3, 4, 5, 6, 7, 8, 9} has a mean of 4.5 and a standard deviation of 2.87, the sampling distribution of the mean of two selections will have a mean of 4.5 and standard error

$$\frac{\sigma}{\sqrt{n}} = \frac{2.87}{\sqrt{2}} \approx 2.03$$

Then, the mean of the sampling distribution of the difference between the two players' averages is $\mu_{you-opponent} = \mu_{you} - \mu_{opponent} = 4.5 - 4.5 = 0$.

The standard error is

$$\sigma_{you-opponent} = \sqrt{\sigma^2_{you} + \sigma^2_{opponent}} = \sqrt{2.03^2 + 2.03^2} \approx 2.87$$

The shape will still be basically triangular, but rounding out slightly.

E58. a. Histogram A has a sample size of 20.
Histogram B has a sample size of 1.
Histogram C has a sample size of 2, and
Histogram D has a sample size of 50.

b. The means all appear to be in the neighborhood of 4.5. Sample size does not appear to affect the mean.

c. The larger the sample size, the smaller the spread.

E59. The shape becomes more approximately normal, the mean stays fixed at the population mean, and the standard error decreases with (is proportional to) the reciprocal of the square root of *n*.

E60. The mean of the sampling distribution of the sample mean remains at the population mean μ for all sample sizes. However, as the sample size increases, the standard error of the sampling distribution of the sample mean decreases by a factor of 1 divided by the square root of *n*. Specifically, $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma / \sqrt{n}$.

E61. Because the distribution from which the samples were selected is highly skewed, the skewness persists in the simulated sampling distribution of the sample mean for samples of size 5 and 10. However, with samples of size 20, most of the skewness disappears and you see a sampling distribution that is nearly normal in shape. The centers of the sampling distributions lie close to the population mean of 42,520 passengers. Observe that the standard deviations of the sampling distributions shrink faster, with increasing sample size, than the theory used earlier for sampling *with replacement* would predict. For $n = 20$, the observed standard deviation of 1,909 is much smaller than $14758 / \sqrt{20} = 3299.99$ This is because we are taking a sample of size 20 without replacement from a small population, with only 30 values. However, if we use the adjusted formula from E30, for sampling *without replacement*

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{14758}{\sqrt{20}} \sqrt{\frac{30-20}{30-1}} = 1937.8$$

we get very close to the SE of the simulated sampling distribution. Recall that the simpler formula is generally 'close enough' if the sample size is less than 10% of the population size.

E62. a. The two students are independent. So, sampling with replacement, the probability that both students have exactly one term paper to write is

$$\left(\tfrac{20}{100}\right)\left(\tfrac{20}{100}\right) = 0.04.$$

b. Sampling without replacement, the probability that both students have exactly one term paper to write is

$$\left(\tfrac{20}{100}\right)\left(\tfrac{19}{99}\right) = 0.038.$$

c. $\left(\tfrac{53}{100}\right)\left(\tfrac{53}{100}\right) = 0.2809$

d. $\left(\tfrac{53}{100}\right)\left(\tfrac{52}{99}\right) = 0.278$

E63. The cap fits on the bottle if $d_c > d_b$, where d_c is the inside diameter of the cap and d_b is the outside diameter of the bottle. So the problem asks for the probability that $d_c - d_b > 0$. Because both sampling distributions are approximately normal, the sampling distribution of $d_c - d_b$ is approximately normal. This sampling distribution has mean and variance

$$\mu_{difference} = \mu_1 - \mu_2 = 36 - 35 = 1 \text{ mm}$$
$$\sigma^2_{difference} = \sigma_1^2 + \sigma_2^2 = 1^2 + 1.2^2 = 2.44 \text{ mm}^2$$

Taking the square root, the standard deviation is about 1.562 mm. The area to the right of zero is shaded in the sampling distribution of $d_c - d_b$.

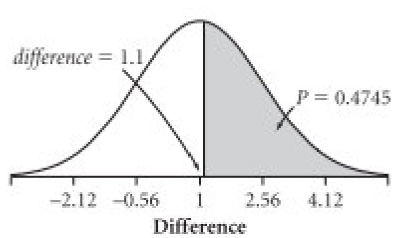
The z-score for a difference of zero is

$$z = \frac{\text{sample difference} - \mu_{difference}}{\sigma_{difference}} = \frac{0-1}{1.562} \approx -0.640$$

The area to the right of this z-score is 0.739. The chance the cap fits on the bottle is only 0.739 – not very good for a manufacturing process.

E64. The cap is too loose if $d_c - d_b > 1.1$. From E59, the sampling distribution of the difference is approximately normal and has a mean of 1 and a standard error of 1.562. A difference of 1.1 has a z-score of

$$z = \frac{1.1-1}{1.562} \approx 0.064$$



The probability that the cap is too loose is about 0.4745.

E65. a. $\mu_{morning} = 13.2$ and $\sigma^2_{morning} = 7.76$; $\mu_{afternoon} = 8$ and $\sigma^2_{afternoon} = 4$. **b.** $\mu_{total} = 21.2$ and $\sigma^2_{total} = 10.96$. **c.** The sum of the means 13.2 + 8 equals 21.2, as we would expect. But the sum of the variances 7.76 + 4 is not equal to 10.96, and we do not expect it to be because the morning and afternoon times were not selected independently when we computed 10.96.
E66. The sampling distribution based on 200 samples is somewhat skewed towards lower values. The 20-25 salaries at the upper end (that is, from \$10-\$18) constitute more than 10% of the salaries and are likely due to a couple of large salaries. As such, the agent can get a reasonably good idea about the 75th percentile by looking at the bulk of salaries between \$5 and \$10 million.

Concept Review Solutions

C1. C. There are three possible samples with a median of 1, six samples with a median of 1.5, and one sample with a median of 2. So, the median of the sampling distribution is the median of those ten values, which is 1.5.
C2. B. The sample maximum is always less than or equal to the population maximum, never greater. So, it is a biased estimator.
C3. D. See E11.
C4. D. The mean of the sampling distribution is always equal to the mean of the population. The standard deviation of the sampling distribution, often called the *standard error*, is the population standard deviation divided by the square root of *n*. So, it gets smaller as the sample size increases. The Central Limit Theorem guarantees that the shape will get closer to normal.
C5. B. The mean of the sampling distribution is 500 with a standard deviation of $\frac{110}{\sqrt{100}} = 11$. The probability that the mean is larger than 510 is then the probability that z is larger than $\frac{510-500}{11}$, which is approximately 0.1817.
C6. D. The expected number correct is 10, and the standard deviation is $\sqrt{40(0.25)(0.75)} \approx 2.74$. So, $z = \frac{15-10}{2.74} \approx 1.83$.
C7. B. Here, $\frac{0.24-0.20}{\sqrt{\frac{(0.2)(0.8)}{1200}}} \approx 3.46$, and because the distribution of the sample proportion is approximately normal, this leads to a probability of about 0.0003.
C8. D. If the sample size is large enough or if the population is close enough to normal, then there will be an approximately normal shape for the sampling distribution of the mean (or the total: they are the same shape, because the only difference is a

multiplication by n). However, here n is only 2, and the population of sleep times might not be very close to normal. In fact, it is probably left-skewed because the mean is so close to the maximum of 360 minutes.
Note: Some may argue that "randomly selected" does not imply independent selection (although it usually does) and so, C would be a correct answer as well.

C9. a. The best estimate of the number going straight on an average day—under the assumption that one-third go each direction—is

$$246{,}000(1/3) = 82{,}000$$

The sampling distribution of the sample total should be approximately normal with a mean of 82,000 and a standard error

$$\sqrt{246{,}000(1/3)(2/3)} \approx 233.809$$

b. The range of reasonably likely outcomes includes values within approximately two standard errors of the mean:

$$82{,}000 \pm 1.96(233.809)$$

In other words, the probability is 0.95 that between 81,542 and 82,548 vehicles will go straight through on an average day if it is true that a randomly selected vehicle has a $\frac{1}{3}$ chance of going straight through the interchange.

c. It is not a reasonably likely outcome to have an average of 138,300 vehicles go straight through the interchange under the assumption that vehicles are equally likely to go in each of the three directions. This assumption cannot be even approximately true.

C10. a. This problem can be done using either the sampling distribution of the mean weight or the sampling distribution of the total weight. We will use the mean weight. The sampling distribution for the mean weight of n people has a mean of 150 and standard error $\sigma_{\bar{x}} = 20/\sqrt{n}$.

If you choose n = 14: The mean weight must be less than $\frac{2000}{14} \approx 142.86$. This is 1.34 standard errors below the expected mean weight of 150. Thus, it will be exceeded about 91% of the time. The elevator is almost sure to be overloaded.

If you choose n = 13: The mean weight must be less than $\frac{2000}{13} \approx 153.85$. This is 0.694 standard errors above the expected mean weight of 150. Thus, the elevator will be overloaded about 24% of the time.

If you choose n = 12: The mean weight must be less than $\frac{2000}{12} \approx 166.67$. This is 2.89 standard errors above the expected mean weight of 150, and the elevator should be overloaded about 0.19% of the time.

It appears that n = 12 or less would be a good choice for the maximum occupancy if the consequences could be severe with an overloaded elevator.

This reasoning assumes that the sample of people who get on the elevator are selected randomly from the population. If you expect for some reason that the sample of people who get on the elevator may not be random, then these results will not apply. For example, if there is a weight loss clinic for men on the upper floor and large groups of men tend to leave together after an exercise class, the “sample” of people would not be random and the elevator would be overloaded every time they left.

b. These are almost the same numbers that we got in part a. It’s somewhat surprising that their conclusion is this close to ours as their estimates probably were made using different data and different assumptions. For example, Mitsubishi uses an average weight of 65 kg (about 143 pounds) per person for its elevators.