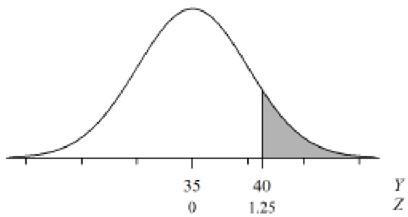
• **4.S.15** The distribution of readings is a normal distribution with mean μ (the true concentration) and standard deviation σ. A reading of 40 or more is considered "unusually high." Suppose that $\mu = 35$ and $\sigma = 4$.

For $y = 40$,

$$z = \frac{40-35}{4} = 1.25.$$

From Table 3, the area below 1.25 is 0.8944, which means that the area above 1.25 is 1 - 0.8944 = 0.1056. Thus,

Pr{specimen is flagged as "unusually high} = 0.1056.



4.S.16 (a) 1 - 0.5948 = 0.4052

(b) 1 - 0.8729 = 0.1271

(c) 0.7549 - 0.4168 = 0.3381

4.S.17 0.2546

• **4.S.18** $Pr\{0 < Y < 15\} = 0.7549 - 0.2546 = 0.5003$. Thus we expect (400)(0.5003), or about 200 observations to fall between 0 and 15.

4.S.19 The IQR is 14.74 – (-0.14) = 14.88. An outlier on the high end of the distribution is any point greater than 14.74 + (1.5)(14.88) = 37.06.

4.S.20 Any data for which the Shapiro_Wilk's test P-value ≥ 0.05 would be evidence that is consistent with the claim of normality. Thus, only (b) provides evidence that is consistent with normality. (a), (c), and (d) all provide evidence that the data has come from an abnormal population. Note that evidence that is consistent with the claim of normality is not the same as evidence for normality.

4.S.21 Histogram II is skewed to the right, so it goes with plot (a). Histogram I is closer to normal, so it goes with plot (b). Histogram III has long tails, so it goes with plot (c).

CHAPTER 5
Sampling Distributions

5.1.1 There are three possible total costs in excess of \$125,000 when sampling n = 3 patients as listed in Table 5.1.2: \$130,000, \$155,000, and \$180,000. Thus,

Pr{total cost > \$125,000} = Pr{total cost = \$130,000} + Pr{total cost = \$155,000} + Pr{total cost = \$180,000}
= 12/64 + 6/64 + 1/64
= 19/64 = 0.2969

• **5.1.2** There are three possible total costs in between \$80,000 and \$125,000 when sampling n = 3 patients as listed in Table 5.1.2: \$95,000, \$105,000, and \$120,000. Thus,

Pr{total cost > \$125,000} = Pr{total cost = \$95,000} + Pr{total cost = \$105,000} + Pr{total cost = \$120,000}
= 12/64 + 8/64 + 3/64
= 23/64 = 0.3594

5.1.3 There are three possible mean costs between \$30,000 and \$45,000 when sampling n = 3 patients as listed in Table 5.1.2: 31.7, 35.0, 40.0, and 43.4 (where units are \$1,000). Thus,

Pr{30 < mean cost < 45} = Pr{mean cost = 31.7} + Pr{mean cost = 35.0} + Pr{mean cost = 40.0} + Pr{mean cost = 43.3}
= 12/64 + 8/64 + 3/64 + 12/64
= 35/64 = 0.547

5.1.4 (a) There are nine possible samples of size n = 2. They are (None, None), (None, Single), (None, Double), (Single, None), (Single, Single), (Single, Double), (Double, None), (Double, Single), and (Double, Double).

(b) First we make a table of total knee replacement costs for all possible samples of size n=2.

Sample	Costs (in units of \$1,000)	Sample Total	Probability
None, None	0,0	0	1/16
None, Single	0,35	35	2/16
None, Double	0,60	60	1/16
Single, None	35,0	35	2/16
Single, Single	35,35	70	4/16
Single, Double	35,60	95	2/16
Double, None	60,0	60	1/16
Double, Single	60,35	95	2/16
Double, Double	60,60	120	1/16

Sampling distribution of total knee replacement costs for samples of size n=2

Sample total	Probability
0	1/16
35	4/16
60	2/16
70	4/16
95	4/16
120	1/16

(c) Pr(total costs exceed 75) = 4/16 + 1/16 = 5/16.

5.1.5 To compute the sampling distribution for the total weight of samples of two dogs, we first must list all possible samples of two dogs and compute the total weight for each possible sample. The following table displays the weights of two dogs for all 16 possible samples of two dogs and their total weights.

Dog 1	Dog 2	Total Weight
42	42	84
42	48	90
42	52	94
42	58	100
48	42	90
48	48	96
48	52	100
48	58	106
52	42	94
52	48	100
52	52	104
52	58	110
58	42	100
58	48	106
58	52	110
58	58	116

Tabulating the unique total weights, the sampling distribution of the total weight is summarized in the following table.

Total Weight	Frequency	Probability
84	1	1/16
90	2	2/16
94	2	2/16
96	1	1/16
100	4	4/16
104	1	1/16

106	2	2/16
110	2	2/16
116	1	1/16
	16	16/16

5.2.1 – 5.2.3 See Section III of this Manual.

• 5.2.4

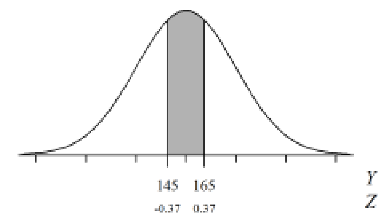
(a) In the population, $\mu = 155$ and $\sigma = 27$.
For y = 165,

$$z = \frac{y-\mu}{\sigma} = \frac{165-155}{27} = 0.37..$$

From Table 3, the area below 0.37 is 06443.
For y = 145,

$$z = \frac{y-\mu}{\sigma} = \frac{145-155}{27} = -0.37.$$

From Table 3, the area below -0.37 is 0.3557.
Thus, the percentage with $145 \le y \le 165$ is 0.6443 – 0.3557 = 0.2886, or 28.86%.



(b) We are concerned with the sampling distribution of $\bar{Y}$ for n = 9. From Theorem 5.2.1, the mean of the sampling distribution of $\bar{Y}$ is

$$\mu_{\bar{Y}} = \mu = 155,$$

the standard deviation is

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{27}{\sqrt{9}} = 9,$$

and the shape of the distribution is normal because the population distribution is normal (part 3a of Theorem 5.2.1).

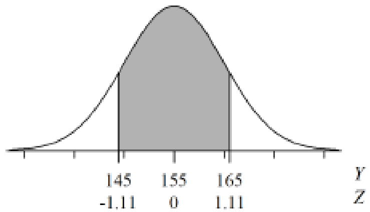
We need to find the shaded area in the figure.
For $\bar{y} = 165$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{165 - 155}{9} = 1.11.$$

From Table 3, the area below 1.11 is 0.8665.
For $\bar{y} = 145$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{145 - 155}{9} = -1.11.$$

From Table 3, the area below -1.11 is 0.1335.
Thus, the percentage with $145 \le \bar{y} \le 165$ is 0.8665 - 0.135 = 0.7330, or 73.30%.



(c) The probability of an event can be interpreted as the long-run relative frequency of occurrence of the event (Section 3.2). Thus, the question in part (c) is just a rephrasing of the question in part (b). It follows from part (b) that

$\Pr\{145 \le \bar{Y} \le 165\} = 0.7330$.

5.2.5

(a) We are concerned with the sampling distribution of $\bar{Y}$ for n = 16. From Theorem 5.2.1, the mean of the sampling distribution of $\bar{Y}$ is

$\mu_{\bar{Y}} = \mu = 155$,

the standard deviation is

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{27}{\sqrt{16}} = 6.75,$$

and the shape of the distribution is normal because the population distribution is normal (part 3a of Theorem 5.2.1).

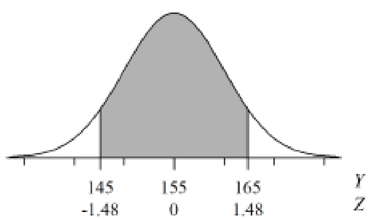
We need to find the shaded area in the figure.
For $\bar{y} = 165$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{165 - 155}{6.75} = 1.48.$$

From Table 3, the area below 1.48 is 0.9306.
For $\bar{y} = 145$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{145 - 155}{6.75} = -1.48.$$

From Table 3, the area below -1.48 is 0.0694.
Thus, the percentage with $145 \le \bar{y} \le 165$ is 0.9306 - 0.0694 = 0.8612, or 86.12%.



(b)

We need to find the shaded area in the figure.

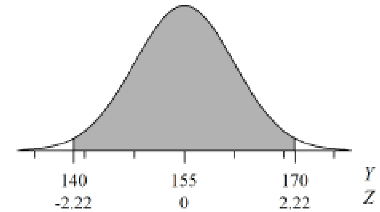
For $\bar{y} = 170$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{170 - 155}{6.75} = 2.22.$$

From Table 3, the area below 2.22 is 0.9868.
For $\bar{y} = 140$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{140 - 155}{6.75} = -2.22.$$

From Table 3, the area below -1.48 is 0.0132.
Thus, the percentage with 140 This $\le \bar{y} \le 170$ is 0.9868 - 0.0132 = 0.9736, or 97.36%.



• **5.2.6** (a) $\mu = 3000$; $\sigma = 400$.

The event E occurs if $\bar{Y}$ is between 2900 and 3100. We are concerned with the sampling distribution of $\bar{Y}$ for n = 15. From Theorem 5.2.1, the mean of the sampling distribution of $\bar{Y}$ is

$\mu_{\bar{Y}} = \mu = 3000$,

the standard deviation is

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{400}{\sqrt{15}} = 103.3,$$

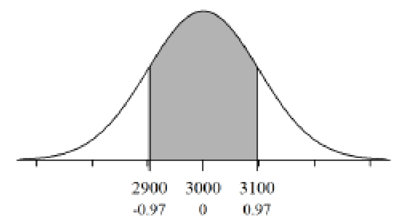
and the shape of the distribution is normal because the population distribution is normal (part 3a of Theorem 5.2.1).
For $\bar{y} = 3100$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{3100 - 3000}{103.3} = 0.97.$$

From Table 3, the area below 0.97 is 0.8340.
For $\bar{y} = 2900$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{2900 - 3000}{103.3} = -0.97.$$

From Table 3, the area below -0.97 is 0.1660.
Thus, $\Pr\{2900 \le \bar{Y} \le 3100\}$
= Pr{E} = .8340 - .1660 = .6680.



(b) n = 60; $\sigma_{\bar{Y}} = 400 / \sqrt{60} = 51.64$

$z = \frac{\pm 100}{51.64} = \pm 1.94$; Table 3 gives 0.9738 and 0.0262, so Pr{E} = 0.9738 - 0.0262 = 0.9476.

(c) As n increases, Pr{E} increases.

5.2.7 $\sigma_{\bar{Y}} = 400 / \sqrt{15} = 103.3$

(a) $z = \frac{2900 - 2800}{103.3} = 0.97$. From Table 3, the area below 0.97 is 0.8340.

$z = \frac{2700-2800}{103.3} = -0.97$. From Table 3, the area below -0.97 is 0.1660.

Thus, Pr{E} = 0.8340 - 0.1660 = 0.6680

(b) $z = \frac{2700-2600}{103.3} = 0.97$. From Table 3, the area below 0.97 is 0.8340.

$z = \frac{2500-2600}{103.3} = -0.97$. From Table 3, the area below -0.97 is 0.1660.

Thus, Pr{E} = 0.8340 - 0.1660 = 0.6680

(c) For fixed n and s, Pr{E} does not depend on μ.

5.2.8 $\mu = 145$; $\sigma = 22$.

(a) $z = \frac{155-145}{22} = 0.45$; Table 3 gives 0.6736.

$z = \frac{135-145}{22} = -0.45$; Table 3 gives 0.3264.

Thus, 0.6736 - 0.3264 = .3472 or 34.72% of the plants.

(b) n = 16; $\sigma_{\bar{Y}} = 22/\sqrt{16} = 5.5$.

$z = \frac{155-145}{5.5} = 1.82$; Table 3 gives 0.9656.

$z = \frac{135-145}{5.5} = -1.82$; Table 3 gives 0.0344.

Thus, 0.9656 - 0.0344 = 0.9312 or 93.12% of the groups.

(c) $\Pr\{135 \le \bar{Y} \le 155\} = 0.9312$ (from part (b)).

(d) n = 36; $\sigma_{\bar{Y}} = 22/\sqrt{36} = 3.67$.

$z = \frac{155-145}{3.67} = 2.72$; Table 3 gives 0.9967.

$z = \frac{135-145}{3.67} = -2.72$; Table 3 gives 0.0033.

Thus, 0.9967 - 0.0033 = 0.9934 or 99.34% of the groups.

5.2.9 (a) $\sigma_{\bar{Y}} = 1.4/\sqrt{25} = 0.28$;

$z = \frac{5-4.2}{0.28} = 2.86$; Table 3 gives 0.9979.

$z = \frac{4-4.2}{0.28} = -0.71$; Table 3 gives 0.2389.

$\Pr\{4 \le \bar{Y} \le 5\} = 0.9979 - 0.2389 = 0.7590$.

(b) The answer is approximately correct because the Central Limit Theorem says that the sampling distribution of $\bar{Y}$ is approximately normal if n is large. The same approach is not valid for n = 2, because the Central Limit Theorem does not apply when the sample size is small.

• **5.2.10 (a)** In the population, 65.68% of the fish are between 51 and 60 mm long. To find the probability that four randomly chosen fish are all between 51 and 60 mm long, we let "success" be "between 51 and 60 mm long" and use the binomial distribution with n = 4 and p = 0.6568, as follows:

Pr{all 4 are between 51 and 60} = $_4C_4 p^4(1-p)^0 = (1)0.6568^4(1) = 0.1861$.

(b) The mean length of four randomly chosen fish is $\bar{Y}$. Thus, we are concerned with the sampling distribution of $\bar{Y}$ for a sample of size n = 4 from a population with μ = 54 and σ = 4.5. From Theorem 5.2.1, the mean of the sampling distribution of $\bar{Y}$ is

$\mu_{\bar{Y}} = \mu = 54$,

the standard deviation is

$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{4.5}{\sqrt{4}} = 2.25$,

and the shape of the distribution is normal because the population distribution is normal (part 3a of Theorem 5.2.1).

For $\bar{y} = 60$,

$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{60-54}{2.25} = 2.67$.

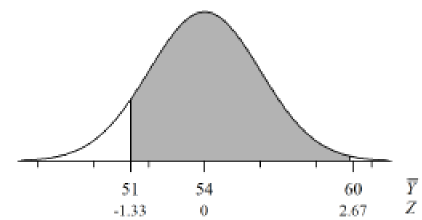
From Table 3, the area below 2.67 is 0.9962.

For $\bar{y} = 51$,

$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{51-54}{2.25} = -1.33$.

From Table 3, the area below -1.33 is 0.0918.

Thus, $\Pr\{51 \le \bar{Y} \le 60\}$
= 0.9962 - 0.0918 = 0.9044.



5.2.11 Let E_1 be the event that all four fish are between 51 and 60 mm long and let E_2 be the event that $\bar{Y}$ is between 51 and 60 mm long. If E_1 occurs, then E_2 must also occur -- the mean of four numbers, each of which is between 51 and 60, must be between 51 and 60 -- but E_2 can occur without E_1

occurring. Thus, in the long run, E_2 will happen more often than E_1, which shows that $Pr\{E_2\} > Pr\{E_1\}$.

5.2.12 $\mu_{\bar{Y}} = 50$ and $\sigma_{\bar{Y}} = \sigma/\sqrt{n} = 9/\sqrt{n}$

An area of 0.68 corresponds to ±1 on the z scale; therefore

$$z = 1.0 = \frac{51.5 - 50}{9/\sqrt{n}}$$

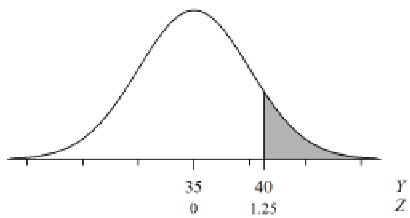
which yields n = 36.

• **5.2.13** The distribution of repeated assays of the patient's specimen is a normal distribution with mean $\mu = 35$ (the true concentration) and standard deviation $\sigma = 4$.

(a) The result of a single assay is like a random observation Y from the population of assays. A value $Y \geq 40$ will be flagged as "unusually high." For y = 40,

$$z = \frac{y - \mu}{\sigma} = \frac{40 - 35}{4} = 1.25.$$

From Table 3, the area below 1.25 is 0.8944, so the area beyond 1.25 is
1 - 0.8944 = 0.1056.



Thus, Pr{specimen will be flagged as "unusually high"}= 0.1056.

(b) The reported value is the mean of three independent assays, which is like the mean $\bar{Y}$ of a sample of size n = 3 from the population of assays. A value $\bar{Y} \geq 40$ will be flagged as "unusually high." We are concerned with the sampling distribution of $\bar{Y}$ for a sample of size n = 3 from a population with mean $\mu = 35$ and standard deviation $\sigma = 4$. From Theorem 5.2.1, the mean of the sampling distribution of $\bar{Y}$ is

$\mu_{\bar{Y}} = \mu = 35$,

the standard deviation is

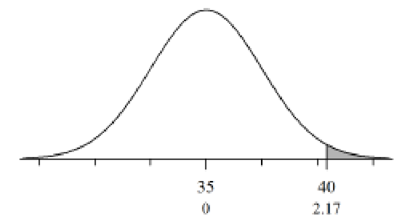
$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{3}} = 2.309,$$

and the shape of the distribution is normal because the population distribution is normal (part 3a of Theorem 5.2.1).

For $\bar{y} = 40$,

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{40 - 35}{2.309} = 2.17.$$

From Table 3, the area below 2.17 is 0.9850, so the area beyond 2.17 is
1 - 0.9850 = 0.0150.



Thus, Pr{mean of three assays will be flagged as "unusually high"}= 1 - 0.9850 = 0.0150.

5.2.14 (a) $\mu_{\bar{Y}} = \mu = 162$.

(b) $\sigma_{\bar{Y}} = 18/\sqrt{9} = 6$

• **5.2.15 (a)** $\mu_{\bar{Y}} = \mu = 41.5$.

(b) $\sigma_{\bar{Y}} = 4.7/\sqrt{4} = 2.35$

5.2.16 (a) Because the sample size of 2 is small, we would expect the histogram of the sample means to be skewed to the right, as is the histogram of the data. However, the histogram of the sample means will be somewhat symmetric (more so than the histogram of the data).

(b) Because the sample size of 25 is fairly large, we would expect the histogram to have a bell shape.

5.2.17 The sample mean is just an individual observation when n=1. Thus, the histogram of the sample means will be the same as the histogram of the data (and therefore be skewed to the right).

5.2.18 No. The histogram shows the distribution of observations in the sample. Such a distribution would look more like the population distribution for n = 400 than for n = 100, and the population distribution is apparently rather skewed. The Central Limit Theorem applies to the sampling distribution of $\bar{Y}$, which is not what is shown in the histogram.

5.2.19 $\mu_{\bar{Y}} = 38$ and $\sigma_{\bar{Y}} = 9/\sqrt{25} = 1.8$.

(a) $z = \frac{36 - 38}{1.8} = -1.11$. Table 3 gives 0.1335, so $Pr\{\bar{Y} > 36\} = 1 - 0.1335 = 0.8665$.

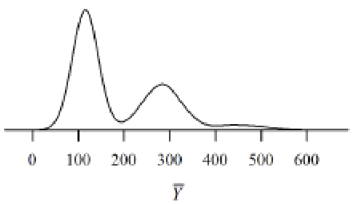
(b) $z = \frac{41 - 38}{1.8} = 1.67$. Table 3 gives 0.9525, so $Pr\{\bar{Y} > 41\} = 1 - 0.9525 = 0.0475$.

• **5.3.1** For each thrust, the probability is 0.9 that the thrust is good and the probability is 0.1 that the thrust is fumbled. Letting "success" = "good thrust," and assuming that the thrusts are independent, we apply the binomial formula with n = 4 and p = 0.9.

(a) The area under the first peak is approximately equal to the probability that all four thrusts are good. To find this probability, we set j = 4; thus, the area is approximately
$_4C_4p^4(1 - p)^0 = (1)(0.9^4)(1) = 0.66.$

(b) The area under the second peak is approximately equal to the probability that three thrusts are good and one is fumbled. To find this probability, we set j = 3; thus, the area is approximately
$_4C_3p^3(1 - p)^1 = (4)(0.9^3)(0.1) = 0.29.$

5.3.2 (a)



The first peak is at 115, the second peak is at $\frac{1}{2}(115+450) = 282.5$, and the third peak is at 450.

(b) First peak: $0.9^2 = 0.81$
Second peak: $(2)(0.9)(0.1) = 0.18$
Third peak: $0.1^2 = 0.01$

5.3.3 When n=1 the sample mean is just an individual observation. Thus, the sampling distribution of the sample mean is the same as the distribution of the individual time scores, as shown in Figure 5.3.3. There are two peaks, one at 115 ms and one at 450 ms.

5.4.1 By Theorem 5.4.1(b), the distribution of $\hat{P}$ will be approximately normal with mean 0.80 and standard deviation $\sqrt{\frac{(0.8)(0.2)}{50}} = 0.0566.$

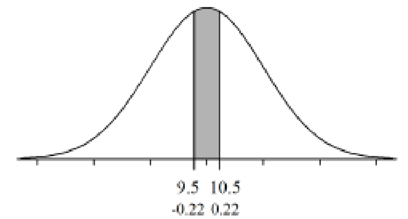
• **5.4.2** Letting "success" = "heads," the probability of ten heads and ten tails is determined by the binomial distribution with n = 20 and p = 0.5.

(a) We apply the binomial formula with j = 10:
Pr{10 heads, 10 tails} = $_{20}C_{10}p^{10}(1 - p)^{10} = (184{,}756)(0.5^{10})(0.5^{10}) = 0.1762.$

(b) According to part (a) of Theorem 5.4.1, the binomial distribution can be approximated by a normal distribution with
mean = np = (20)(0.5) = 10
and
standard deviation = $\sqrt{np(1-p)} = \sqrt{(20)(0.5)(0.5)} = 2.236.$
Applying continuity correction, we wish to find the area under the normal curve between

10 - 0.5 = 9.5 and 10 + 0.5 = 10.5.

The desired area is shaded in the figure.



The boundary 10.5 corresponds to
$z = \frac{10.5-10}{2.236} = 0.22.$
From Table 3, the area below 0.22 is 0.5871.

The boundary 9.5 corresponds to
$z = \frac{9.5-10}{2.236} = -0.22.$
From Table 3, the area below -0.22 is 0.4129.

Thus, the normal approximation to the binomial probability is
Pr{10 heads, 10 tails} ≈ 0.5871 - 0.4129 = 0.1742.

5.4.3 Letting "success" = "type O blood," the probability that 6 of the persons will have type O blood is determined by the binomial distribution with n = 12 and p = 0.44.

(a) We apply the binomial formula with j = 6:
Pr{6 type O blood} = $_{12}C_6p^6(1 - p)^6 = (924)(0.44^6)(0.56^6) = 0.2068.$

(b) According to part (a) of Theorem 5.4.1, the binomial distribution can be approximated by a normal distribution with
mean = np = (12)(0.44) = 5.28
and
standard deviation = $\sqrt{np(1-p)} = \sqrt{(12)(0.44)(0.56)} = 1.72.$
Applying continuity correction, we wish to find the area under the normal curve between 6 - 0.5 = 5.5 and 6 + 0.5 = 6.5.
Thus, Pr{6 type O blood} ≈ Pr{$\frac{5.5-5.28}{1.72} < Z < \frac{6.5-5.28}{1.72}$} = Pr{0.13 < Z < 0.71} = 0.7611 - 0.5517 = 0.2094.

5.4.4 Letting "success" = "type O blood," the probability that at most 6 of the persons will have type O blood is determined by the binomial distribution with n = 12 and p = 0.44. According to part (a) of Theorem 5.4.1, the binomial distribution can be approximated by a normal distribution with
mean = np = (12)(0.44) = 5.28
and
standard deviation = $\sqrt{np(1-p)} = \sqrt{(12)(0.44)(0.56)} = 1.72.$

(a) Without continuity correction:
Pr{at most 6 type O blood} ≈ Pr{ $Z < \frac{6.0-5.28}{1.72}$} = Pr{Z < 0.42} = 0.6628

(b) With continuity correction:

Pr{at most 6 type O blood} ≈ Pr{ $Z < \frac{6.5-5.28}{1.72}$ } = Pr{Z < 0.71} = 0.7611

• **5.4.5** **(a)** Because p = 0.12, the event that $\hat{p}$ will be within ±0.03 of p is the event

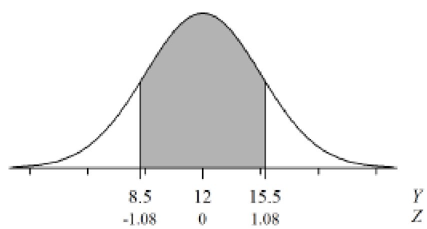
$0.09 \le \hat{p} \le 0.15$,

which, if n = 100, is equivalent to the event

9 ≤ number of success ≤ 15.

Letting "success" = "oral contraceptive user," the probability of this event is determined by the binomial distribution with
mean = np = (100)(0.12) = 12
and
standard deviation = $\sqrt{np(1-p)} = \sqrt{100(0.12)(0.88)}$ = 3.250.
Applying continuity correction, we wish to find the area under the normal curve between
9 - 0.5 = 8.5 and 15 + 0.5 = 15.5.

The desired area is shaded in the figure.



The boundary 15.5 corresponds to
$z = \frac{15.5-12}{3.250} = 1.08$.
From Table 3, the area below 1.08 is 0.8599.

The boundary 8.5 corresponds to
$z = \frac{8.5-12}{3.250} = -1.08$.
From Table 3, the area below -1.08 is 0.1401.

Thus, the normal approximation to the binomial probability is
Pr{ $\hat{p}$ will be within ±0.03 of p} ≈ 0.8599 - 0.1401 = 0.7198.

(Note: An alternative method of solution is to use part (b) of Theorem 5.4.1 rather than part (a). Such a method is illustrated in the solution to Exercise 5.4.8.)

(b) With n = 200, $\hat{p}$ is within ±0.03 of p if and only if the number of successes is between (200)(0.09) = 18 and (200)(0.15) = 30. The mean is (200)(0.12) = 24 and the standard deviation is $\sqrt{200(0.12)(0.88)}$ = 4.60.
Applying continuity correction, we wish to find the area under the normal curve between
18 - 0.5 = 17.5 and 30 + 0.5 = 30.5.

$z = \frac{30.5-24}{4.60} = 1.41$; Table 3 gives 0.9207.

$z = \frac{17.5-24}{4.60} = -1.41$; Table 3 gives 0.0793.
0.9207 - 0.0783 = 0.8414.

5.4.6 **(b)** p = 0.5
For n = 45, 60% boys means 27 boys. For the normal approximation to the binomial, the mean is
np = (45)(0.5) = 22.5
and the SD is
$\sqrt{np(1-p)} = \sqrt{45(0.5)(0.5)} = 3.354$.

$z = \frac{27-22.5}{3.354} = 1.34$; Table 3 gives 0.9099.

1 - 0.9099 = 0.0901.

For n = 15, 60% boys means 9 boys. For the normal approximation to the binomial, the mean is
np = (15)(0.5) = 7.5
and the SD is
$\sqrt{np(1-p)} = \sqrt{15(0.5)(0.5)} = 1.936$.

$z = \frac{9-7.5}{1.936} = 0.77$; Table 3 gives 0.7794.

1 - 0.7794 = 0.2206.

In the larger hospital, 9% of days have 60% or more boys. In the smaller hospital, 22% of days have 60% or more boys. The smaller hospital recorded more such days.

• **5.4.7** p ±0.05 is 0.25 to 0.35.
The normal approximation to the sampling distribution of $\hat{p}$ has mean p = 0.3 and standard deviation
$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.3)(0.7)}{400}} = 0.02291$.

$z = \frac{0.35-0.3}{0.02291} = 2.18$; Table 3 gives 0.9854.

$z = \frac{0.25-0.3}{0.02291} = -2.18$; Table 3 gives 0.0146.

0.9854 - 0.0146 = 0.9708.

5.4.8 Because p = 0.3, the event E, that $\hat{p}$ will be within ±0.05 of p, is equivalent to

$0.25 \le \hat{p} \le 0.35$.

The sample size is n = 40. According to part (b) of Theorem 5.4.1, the sampling distribution of $\hat{p}$ can be approximated by a normal distribution with
mean p = 0.3
and

standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.3)(0.7)}{40}} = 0.07246$.

$z = \frac{0.35 - 0.3}{0.07246} = 0.69$; Table 3 gives 0.7549.

$z = \frac{0.25 - 0.3}{0.07246} = -0.69$; Table 3 gives 0.2451.

Thus, the normal approximation to the probability is
Pr{E} ≈ 0.7549 - 0.2451 = 0.5098.

5.4.9 Because p = 0.3, the event E, that $\hat{p}$ will be within ±0.05 of p, is equivalent to

$0.25 \le \hat{p} \le 0.35$.

The sample size is n = 40. According to part (b) of Theorem 5.4.1, the sampling distribution of $\hat{p}$ can be approximated by a normal distribution with
mean p = 0.3
and

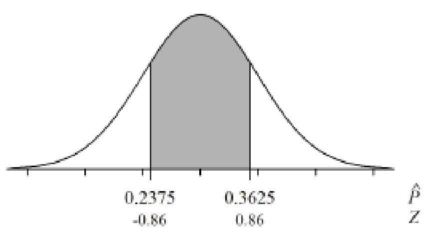
standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.3)(0.7)}{40}} = 0.07246$.

To apply continuity correction, we first calculate the half-width of a histogram bar (on the $\hat{p}$ scale) as

$\left(\frac{1}{2}\right)\left(\frac{1}{40}\right) = 0.1025$.

Thus, we wish to find the area under the normal curve between
0.25 - 0.0125 = 0.2375 and 0.35 + 0.0125 = 0.3625.

The desired area is shaded in the figure.



The boundary 0.3625 corresponds to

$z = \frac{0.3625 - 0.3}{0.07246} = 0.86$.

From Table 3, the area below 0.86 is 0.8051.

The boundary 0.2375 corresponds to

$z = \frac{0.2375 - 0.3}{0.07246} = -0.86$.

From Table 3, the area below -0.86 is 0.1949.

Thus, the normal approximation to the probability is
Pr{E} ≈ 0.8051 - 0.1949 = 0.6102.

(Note: An alternative method of solution is to use part (a) of Theorem 5.4.1 rather than part (b). Such a method is illustrated in the solution to Exercise 5.4.4.)

5.4.10 Let E be the event that $\hat{p}$ is closer to $\frac{1}{2}$ than to $\frac{9}{16}$.

(a) n = 1. E occurs if the number of purple plants is 0. Pr{E} = 7/16 = 0.4375.

(b) n = 64. E occurs if the number of purple plants is less than or equal to 33. The normal approximation to the binomial has mean np = (64)(9/16) = 36 and standard deviation
$\sqrt{np(1-p)} = \sqrt{64(9/16)(7/16)} = 3.969$.

$z = \frac{33 - 36}{3.969} = -0.76$; Table 3 gives 0.2236 = Pr{E}.

(c) n = 320. E occurs if the number of purple plants is less than or equal to 169. The normal approximation to the binomial has mean np = (320)(9/16) = 180 and standard deviation
$\sqrt{np(1-p)} = \sqrt{320(9/16)(7/16)} = 8.874$.

$z = \frac{169 - 180}{8.874} = -1.24$; Table 3 gives 0.1075 = Pr{E}.

5.4.11 We first note that with n=10, the probability that between 30% and 40% (inclusive) of those sampled will have CMV corresponds to between 3 to 4 persons (inclusive) with CMV in the sample.

(a) Using the binomial formula with n = 10 and p = 0.5 we have

$$\Pr\{3 \text{ or } 4 \text{ persons with CMV}\} = {}_{10}C_3 p^3(1-p)^7 + {}_{10}C_4 p^4(1-p)^6$$
$$= (120)(0.5^3)(0.5^7) + (210)(0.5^3)(0.5^7)$$
$$= 0.1172 + 0.2051$$
$$= 0.3223.$$

(b) Using the normal approximation with continuity correction, n = 10, and p = 0.5, the binomial distribution can be approximated by a normal distribution with
mean = np = (10)(0.50) = 5.0
and
standard deviation = $\sqrt{np(1-p)} = \sqrt{(10)(0.50)(0.50)} = 1.58$.

Applying continuity correction, we wish to find the area under the normal curve between
3 - 0.5 = 2.5 and 4 + 0.5 = 4.5.

Thus, Pr{3 or 4 persons with CMV} ≈ $\Pr\{\frac{2.5 - 5.0}{1.58} < Z < \frac{4.5 - 5.0}{1.58}\} = \Pr\{-1.58 < Z < -0.32\}$ =
0.3745 - 0.0571 = 0.3174.

Note that the answers to (a) and (b) are similar.

5.4.12 For the normal approximation to the binomial, the mean is np = (100)(0.8) = 80 and the SD is $\sqrt{np(1-p)} = \sqrt{100(0.8)(0.2)} = 4$.

$z = \frac{85-80}{4} = 1.25$; Table 3 gives 0.8944.

1 - 0.8944 = 0.1056.

5.4.13 For the normal approximation to the binomial, the mean is np = (100)(0.8) = 80 and the SD is $\sqrt{np(1-p)} = \sqrt{100(0.8)(0.2)} = 4$.

$z = \frac{84.5-80}{4} = 1.13$; Table 3 gives 0.8708.

1 - 0.8708 = 0.1292.

5.4.14 For the normal approximation to the binomial, the mean is np = (50)(0.8) = 40 and the SD is $\sqrt{np(1-p)} = \sqrt{50(0.8)(0.2)} = 2.83$.

(a) $z = \frac{35-40}{2.83} = -1.77$; Table 3 gives 0.0384.

(b) $z = \frac{35.5-40}{2.83} = -1.59$; Table 3 gives 0.0559.

• **5.S.1** μ = 88; σ = 7.

We are concerned with the sampling distribution of $\bar{Y}$ for n = 5. From Theorem 5.2.1, the mean of the sampling distribution of $\bar{Y}$ is

$\mu_{\bar{Y}} = \mu = 88$,

the standard deviation is

$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{7}{\sqrt{5}} = 3.13$,

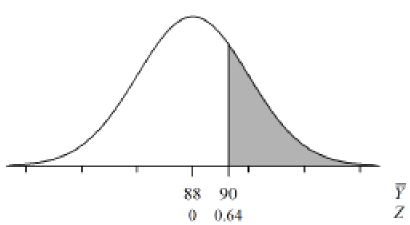
and the shape of the distribution is normal because the population distribution is normal (part 3a of Theorem 5.2.1).

For = 90,

$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{90-88}{3.13} = 0.64$.

From Table 3, the area below 0.64 is 0.7389.

Thus, Pr{$\bar{Y}$ > 90} = 1 - 0.7389 = 0.2611.



5.S.2 The sampling distribution in the context of this problem describes the distribution of sample mean blood pressures for all possible samples of 14 students taken from the population of college students.

5.S.3 Applying Theorem 5.2.1 (2), with σ = 10, we compute $\sigma_{\bar{Y}} = \sigma/\sqrt{n} = 10/\sqrt{14} = 2.67$.

5.S.4 (a) $z = \frac{72-69.7}{2.8} = 0.82$; Table 3 gives 0.7939.

1 - 0.7939 = 0.2061.

(b) (i) Using the binomial distribution,

Pr{both are > 72} = $0.2061^2 = 0.0425$.

(ii) n = 2; $\sigma_{\bar{Y}} = \sigma/\sqrt{n} = 2.8/\sqrt{2} = 1.980$.

$z = \frac{72-69.7}{1.980} = 1.16$; Table 3 gives .8770.

1 - 0.8770 = 0.1230

5.S.5 μ = 800; σ = 90.

(a) $z = \frac{850-800}{90} = 0.56$; Table 3 gives 0.7123.

$z = \frac{750-800}{90} = -0.56$; Table 3 gives 0.2877.

0.7123 - 0.2877 = 0.4246 or 42.46% of the plants.

(b) n = 4; $\sigma_{\bar{Y}} = \sigma/\sqrt{n} = 90/\sqrt{4} = 45$.

$z = \frac{850-800}{45} = 1.11$; Table 3 gives 0.8665.

$z = \frac{750-800}{45} = -1.11$; Table 3 gives 0.1335.

0.8665 - 0.1335 = 0.7330 or 73.30% of the groups will have means in this range.

5.S.6 Two possible factors are: (a) environmental variation from one pot (or location) to another; (b) competition between plants in a pot (for instance, overlapping leaves).

5.S.7 For the normal approximation to the sampling distribution of $\hat{p}$, the mean is p = 0.42 and the SD is $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.42)(0.58)}{25}} = 0.0987$.

Continuity correction: $\left(\frac{1}{2}\right)\left(\frac{1}{25}\right) = 0.02$.

$z = \frac{0.46-0.42}{0.0987} = 0.405$; Table 3 gives 0.6590.

1 - 0.6590 = 0.3410.

5.S.8 $\mu = 1{,}200$; $\sigma = 35$.
For $\Pr\{1175 \le Y \le 1225\}$

$z = \dfrac{1225 - 1200}{35} = 0.71$; Table 3 gives 0.7611.

$z = \dfrac{1175 - 1200}{35} = -0.71$; Table 3 gives 0.2389.

$0.7611 - 0.2389 = 0.5222$.

For $\Pr\{1175 \le \bar{Y} \le 1225\}$, $\sigma_{\bar{Y}} = \sigma/\sqrt{n} = 35/\sqrt{6} = 14.29$.

$z = \dfrac{1225 - 1200}{14.29} = 1.75$; Table 3 gives 0.9599.

$z = \dfrac{1175 - 1200}{14.29} = -1.75$; Table 3 gives 0.0401.

$0.9599 - 0.0401 = 0.9198$.

Comparison: $0.9189 > 0.5222$; this shows that the mean of 6 counts is more likely to be near the correct value (1200) than is a single count.

• **5.S.9** $\mu = 8.3$; $\sigma = 1.7$.
If the total weight of 10 mice is 90 gm, then their mean weight is

$\dfrac{90}{10} = 9.0$ gm.

Thus, we wish to find the percentage of litters for which $\bar{y} \ge 9.0$ gm. We are concerned with the sampling distribution of $\bar{Y}$ for $n = 10$. From Theorem 5.2.1, the mean of the sampling distribution of $\bar{Y}$ is

$\mu_{\bar{Y}} = \mu = 8.3$,

the standard deviation is

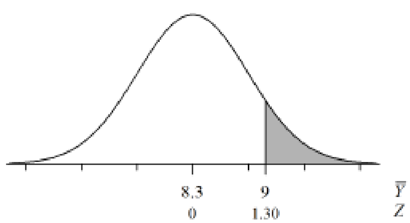
$\sigma_{\bar{Y}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{1.7}{\sqrt{10}} = 0.538$,

and the shape of the distribution is normal because the population distribution is normal (part 3a of Theorem 5.2.1).

We need to find the shaded area in the figure.
For $\bar{y} = 9.0$,

$z = \dfrac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \dfrac{9.0 - 8.3}{0.538} = 1.30$.

From Table 3, the area below 1.30 is 0.9032.
Thus, the percentage with $\bar{y} \ge 9.0$ is
$1 - 0.9032 = 0.0968$, or 9.68%.



5.S.10 Two possible factors are: (a) environmental and genetic differences between litters; (b) competition between mice in a litter.

5.S.11 The sampling distribution in the context of this problem describes the distribution of sample total weights for all possible samples of 25 plants taken from the target population of plants.

5.S.12 $\sigma_{\bar{Y}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{10}{\sqrt{64}} = 1.25$.

(a) $z = \dfrac{52 - 50}{1.25} = 1.6$; Table 3 gives 0.9452.

$z = \dfrac{48 - 50}{1.25} = -1.6$; Table 3 gives 0.0548.

$0.9452 - 0.0548 = 0.8904$.

(b) $z = \dfrac{102 - 100}{1.25} = 1.6$; Table 3 gives 0.9452.

$z = \dfrac{98 - 100}{1.25} = -1.6$; Table 3 gives 0.0548.

$0.9452 - 0.0548 = 0.8904$.

(c) $z = \dfrac{\mu + 2 - \mu}{1.25} = 1.6$; Table 3 gives 0.9452.

$z = \dfrac{\mu - 2 - \mu}{1.25} = -1.6$; Table 3 gives 0.0548.

$0.9452 - 0.0548 = 0.8904$.

5.S.13 (a) Each sampling distribution will be bell-shaped.

(b) The two distributions will both be centered at the population mean number of calories consumed per day by the students.

(c) We can expect that Susan's distribution will have a smaller spread due to her larger sample size. The means of samples of size 30 are less variable than the means of samples of size 20.

UNIT I SUMMARY

• **I.1** **(a)** For Graph II, the mean – 2(SD) is 1.7 – 2.2, which is negative. But precipitation totals can't be negative; moreover we know that the minimum is 0.3. Thus, Graph II can't be right. Graph III is even worse. Thus, the answer must be Graph I (which fits having min ≈ mean – 1*SD).

(b) The median is *less than* the mean because the distribution is skewed to the right. The long right tail pulls up the mean.

I.2 The sample SD, s, is given by $s = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}}$ so the point farthest from the mean of 70 has the largest contribution. This is 62.

I.3 Pr{FF} + Pr{MM} = $\frac{10}{18} * \frac{9}{17} + \frac{8}{18} * \frac{7}{17}$ which is about 0.48.

• **I.4** The sampling distribution of the sample percentage is the distribution of $\hat{P}$, the sample percentage of mice weighing more then 26 gm, as it varies from one sample of 20 mice to another in repeated samples.

• **I.5** **(a)** $\Pr\{16 < Y < 23\} = \Pr\left\{\frac{16-20}{5} < Z < \frac{23-20}{5}\right\} = \Pr\{-0.80 < Z < 0.60\} = 0.514$.

(b) $\bar{Y}$ has a normal distribution with mean 20 and SD $5/\sqrt{5} = 2.24$. $\Pr\{16 < \bar{Y} < 23\} = \Pr\left\{\frac{16-20}{5/\sqrt{5}} < Z < \frac{23-20}{2.24}\right\} = \Pr\{-1.79 < Z < 1.34\} = 0.9099 - 0.0367 = 0.873$.

I.6 This table shows the 10 possible total weights (with redundancies not listed):

	40	*50*	*65*	*70*
40	80	90	105	110
50		100	115	120
65			130	135
70				140

I.7 (i) 63.1 ± 5 covers 75% of the data and 63.1 ± SD covers 68%, so the SD is less than 5.

I.8 **(a)** This should be an experiment. An experiment controls for lurking variables by assigning treatments to the subjects.

(b) The placebo effect is an improvement in forced expiratory volume that is caused not by an active drug but by the psychological benefit of the subject thinking that he or she is being treated.

I.9 66.1 – 64.3 = 1.8 and the z score for the 75th percentile is 0.67. Thus, $Z = 0.67 = 1.8/\sigma$, so $\sigma = 1.8/0.67 = 2.69$.

• **I.10** $\Pr\{Y \geq 54\} = \Pr\left\{Z \geq \frac{54 - 100*9/16}{\sqrt{100*(9/16)*(7/16)}}\right\} = \Pr\left\{Z \geq \frac{54-56.25}{4.96}\right\} = \Pr(Z \geq -0.45) = 0.674$.

I.11 **(a)** (i) This should be an experiment so that we can control for confounding variables. For example, it might be that people who choose to take aspirin regularly are less healthy than others or are older than others. Maybe they exercise more than others. By randomizing subjects into two groups (aspirin/placebo), we can balance these characteristics across groups and see directly the effect of aspirin.

(ii) The experiment should be run double-blind. We don't want the aspirin group to know they are taking aspirin, lest they get a psychological boost that the comparison group doesn't get. We don't want the physicians who evaluate patients to know if a subject was in the aspirin versus control group, lest their expectations lead them to see evidence of strokes that didn't really happen or to ignore evidence of a stroke if they don't expect a subject to be prone to strokes.

(b) (i) This should be an observational study because it is not practical to impose regular attendance at religious services nor to prohibit such attendance. [However, one *can* argue for an experiment: If people sign consent forms and agree to abide by the results of randomization, then an experiment could be conducted (but probably the duration of the study would need to be limited).]

(ii) The study should be single-blind: People know whether or not they are attending religious services, so they cannot be blinded, but those who measure blood pressure can and should be blinded so they don't read blood pressure in a biased way due to their expectations.

I.12 **(a)** No, a binomial model will not work here. The number of trials, n, is not fixed in advance.

(b) Yes, a binomial model will work here. Each trial has a binary outcome (exercise every day or not). There is a fixed number of independent trials (45), each with the same probability of success (whatever the proportion is of people who exercise every day). Note that there are three categories of answers (every day, occasional, never) but we can collapse this into two categories (every day or not) and get a binomial random variable.

I.13 **(a)** The variable is numeric. The amount of thiocyanate will be recorded numerically in the units mg/L.

(b) The variable is continuous, though it could be recorded discretely. Inherently the amount of thiocyanate in the blood is a continuous variable—the values are not restricted to be elements of a countable set of numbers.

I.14 **(a)** We can tell that the study is an experiment because the diets are being *assigned* to the rats. The researchers are actively manipulating which experimental units (rats) get which treatment (diet). If the researches simply put out two bowls of food (raw and heated) and let the rats choose their own diet, then the study would be an observational study.

(b) As an *experiment* the researchers can establish cause–effect relationships. If there is a difference in the blood thiocyanate levels between the two diet groups, the difference can be attributed to the diets.

• **I.15** **(a)** $z_{.15} = 1.04$

$y = 53.3 + 1.04 \times 14.6 = 68.48$ mg/L

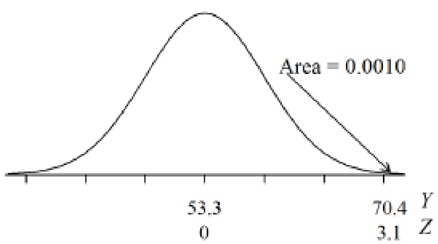
(b) 85th percentile

(c) Note that this question pertains to the sampling distribution of the sample mean for samples of size 7. Thus, we wish to compute Pr $(\bar{Y} > 70.4)$ where $\bar{Y} \sim N(\mu_{\bar{Y}} = 53.3, \sigma_{\bar{Y}} = \frac{14.6}{\sqrt{7}})$.

$$z = \frac{70.4 - 53.3}{14.6/\sqrt{7}} \approx 3.10$$

$$\Pr(Z > 3.10) = 1 - \Pr(Z < 3.10) = 1 - 0.9990 = 0.0010$$

This probability is very small; hence, it would be very unusual to observe a sample of seven rats with a sample mean thiocyanate level above 53.3 mg/L.



I.16 **(a)** The *P*-value for the Shapiro-Wilk's normality test greater than 0.10; thus, there is no evidence that the data aren't coming from a normal distribution (i.e., no evidence that the data are nonnormal). Furthermore, the points in the probability plot are roughly linear and fall within the dotted guidelines for normal data.

(b) If the population is normally distributed, the sampling distribution of the sample mean will also be normally distributed regardless of the sample size. In part (a) we noted that the data are consistent with what we would expect from a normal population, so it seems reasonable to regard the sampling distribution of the mean for even this small sample size of 7 to be normal.

(c) False. Although the sample data might be consistent with what we would expect from a normal population, these data do not provide evidence that the population is normal. It's quite possible that the population is not normal, but our sample is so small we cannot detect the nonnormality; the population may also be only slightly different from a normal population.

I.17 **(a)** iii

(b) i

(c) i

(d) 2

1.18 The probability of a short bill is $\Pr\{Y < 24.0\} = \Pr\left\{Z < \frac{24.0 - 25.4}{0.8}\right\} = \Pr\{Z < -1.75\} = 0.0401$.

Thus, the probability of not having a short bill is 1 – 0.0401 = 0.9959. If 10 blue jays are sampled, then the chance that none of them will have a short bill is $0.9599^{10} = 0.664$. Thus, the probability of at least one short bill is 1 – 0.664 = 0.336. This means that on 33.6% of the days there should be at least one blue jay with a short bill.