CHAPTER 4
The Normal Distribution

4.3.1 **(a)** 0.9332 - 0.0668 = 0.8664 or 86.64%

(b) 1 - 0.9938 = 0.0062 or 0.62%

(c) 2(0.0013) = 0.0026 or 0.26%

4.3.2 **(a)** 1.28

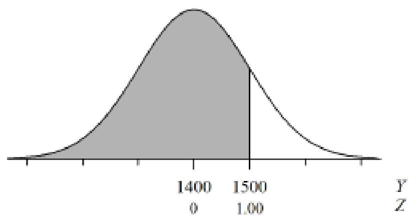
(b) 1.28

• **4.3.3** $\mu = 1400$; $\sigma = 100$.

(a) For y = 1500,

$z = \frac{y-\mu}{\sigma} = \frac{1500-1400}{100} = 1.00.$

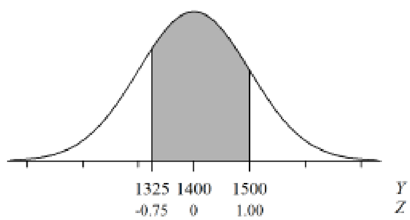
From Table 3, the area is 0.8413 or 84.13%.



(b) For y = 1325,

$z = \frac{y-\mu}{\sigma} = \frac{1325-1400}{100} = -0.75.$

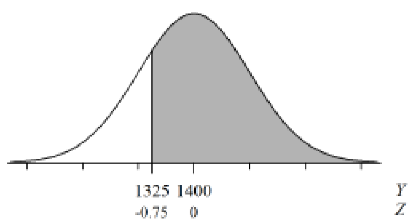
From Table 3, the area below 1325 is 0.2266.
From part (a), the area below 1500 is 0.8413.
Thus, the percentage with $1325 \leq Y \leq 1500$ is
0.8413 - 0.2266 = 0.6147 or 61.47%.



(c) For y = 1325,

$z = \frac{y-\mu}{\sigma} = \frac{1325-1400}{100} = -0.75.$

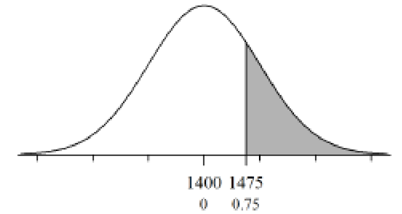
From Table 3, the area below 1325 is 0.2266.
Thus, the percentage with $Y \geq 1325$ is
1 - 0.2266 = 0.7734 or 77.34%.



(d) For y = 1475,

$z = \frac{y-\mu}{\sigma} = \frac{1475-1400}{100} = 0.75.$

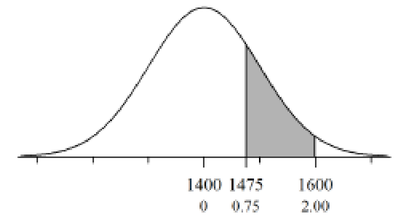
From Table 3, the area below 1475 is 0.7734.
Thus, the percentage with $Y \geq 1475$ is
1 - 0.7734 = 0.2266 or 22.66%.



(e) For y = 1600,

$z = \frac{y-\mu}{\sigma} = \frac{1600-1400}{100} = 2.00.$

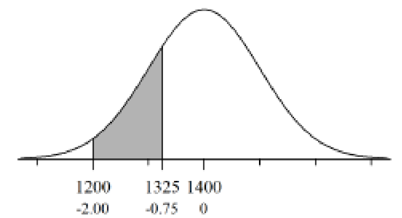
From Table 3, the area below 1600 is 0.9772.
In part (d) we found that the area below 1475 is 0.7734.
Thus, the percentage with $1475 \leq Y \leq 1600$ is
0.9772 - 0.7734 = 0.2038 or 20.38%.



(f) For y = 1200,

$z = \frac{y-\mu}{\sigma} = \frac{1200-1400}{100} = -2.00.$

From Table 3, the area below 1200 is 0.0228.
In part (c) we found that the area below 1325 is 0.2266.
Thus, the percentage with $1200 \leq Y \leq 1325$ is
0.2266 - 0.0028 = 0.2038 or 20.38%.

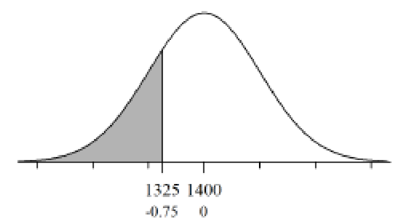


• **4.3.4** $\mu = 1400$; $\sigma = 100$.

(a) For y = 1325,

$z = \frac{y-\mu}{\sigma} = \frac{1325-1400}{100} = -0.75.$

From Table 3, the area is 0.2266 or 22.66%.



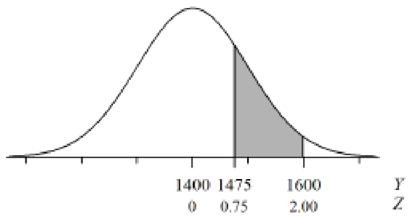
(b) This is the same as part (e) of Exercise 4.3.3.
For y = 1600,

$$z = \frac{y-\mu}{\sigma} = \frac{1600-1400}{100} = 2.00.$$

From Table 3, the area below 1600 is 0.9772.
For y = 1475,

$$z = \frac{y-\mu}{\sigma} = \frac{1475-1400}{100} = 0.75.$$

From Table 3, the area below 1475 is 0.7734.
Thus, the percentage with $1475 \le Y \le 1600$ is 0.9772 - 0.7734 = 0.2038 or 20.38%.



4.3.5 **(a)** 1 - 0.1271 = 0.8729 or 87.29%

(b) 1 - 0.6141 = 0.3859 or 38.59%

(c) 0.0314 or 3.14%

(d) 0.6141 - 0.0314 = 0.5827 or 58.27%

(e) 0.9564 - 0.6141 = 0.3423 or 34.23%

(f) 0.1271 - 0.0314 = 0.0957 or 9.57%

4.3.6 **(a)** 1 - 0.6141 = 0.3859

(b) 0.6141 - 0.0314 = 0.5827

4.3.7 **(a)** 0.67

(b) 1.28

(c) 1.64

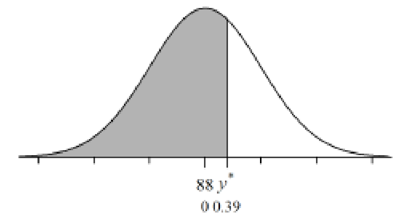
(d) 2.33

• **4.3.8** $\mu = 88$; $\sigma = 7$.

(a) The 65th percentile is the value that is larger than 65% of the observations. Thus, the area under the curve below this value is 0.65.
In Table 3, the area closest to 0.65 is 0.6517, which corresponds to z = 0.39.
Thus, the 65th percentile y* must satisfy the equation

$$z = \frac{y^*-\mu}{\sigma} \text{ or } 0.39 = \frac{y^*-88}{7}.$$

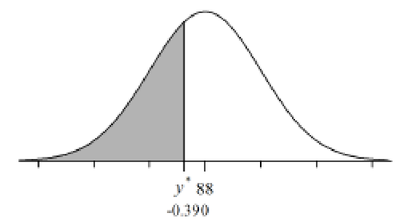
The solution of this equation is
y* = (7)(0.39) + 88 = 90.7.
Thus, the 65th percentile is 90.7 lb.



(b) The 35th percentile is the value that is larger than 35% of the observations. Thus, the area under the curve below this value is 0.35.
In Table 3, the area closest to 0.35 is 0.3483, which corresponds to z = -0.39.
(Note that this is the negative of the value found in part (a).)
Thus, the 35th percentile y* must satisfy the equation

$$z = \frac{y^*-\mu}{\sigma} \text{ or } -0.39 = \frac{y^*-88}{7}.$$

The solution of this equation is
y* = (7)(-0.39) + 88 = 85.3.
Thus, the 35th percentile is 85.3 lb.



4.3.9 **(a)** 1 - 0.6293 = 0.3707 or 37.07%

(b) 0.2514 or 25.14%

(c) 0.8749 or 87.49%

(d) 1 - 0.0207 = 0.9793 or 97.93%

(e) 0.8749 - 0.5596 = 0.3153 or 31.53%

(f) 0.1977 - 0.0207 = 0.1770 or 17.7%

(g) 0.5596 - 0.1977 = 0.3619 or 36.19%

4.3.10 **(a)** 1 - 0.5596 = 0.4404

(b) 0.8749 - 0.5596 = 0.3153

4.3.11 **(a)** In Table 3, the area closest to 0.8 is 0.7995, corresponding to z = 0.84. Thus, the 80th percentile y* must satisfy the equation

$$0.84 = \frac{y^* - 155}{27}$$

which yields y* = (27)(0.84) + 155= 177.7 mg/dl.

(b) The 20th percentile y* satisfies the equation

$$-0.84 = \frac{y^* - 155}{27}$$

which yields y* = (27)(-0.84) + 155 = 132.3 mg/dl.

• **4.3.12** The distribution of readings is a normal distribution with mean μ (the true value) and standard deviation σ = .008μ (0.8% of the true value).

(a) μ = 5,000,000 which means that σ = (.008)(5,000,000) = 40,000.

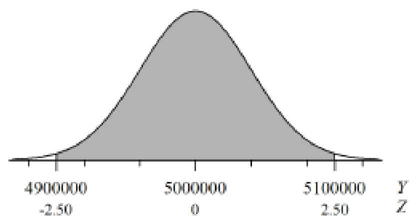
For y = 4,900,000,

$$z = \frac{4{,}900{,}000 - 5{,}000{,}000}{40{,}000} = -2.5.$$

For y = 5,100,000,

$$z = \frac{4{,}900{,}000 - 5{,}000{,}000}{40{,}000} = 2.5.$$

Thus, Pr{4,900,000 < Y < 5,100,000}
= Pr{-2.5 < Z < 2.5}
= 0.9938 - 0.0062 = 0.9876 or 98.76%.



(b) $Pr\{0.98\mu < Y < 1.02\mu\} = Pr\{\frac{0.98\mu - \mu}{0.008\mu} < \frac{Y - \mu}{\sigma} < \frac{1.02\mu - \mu}{0.008\mu}\}$
= Pr{-2.5 < Z < 2.5} = 0.9938 - 0.0062 = 0.9876 or 98.76%

(c) A specimen reading Y differs from the correct value by 2% or more if it does not satisfy 0.98μ < Y < 1.02μ. Using the answer from part (b), this probability is 1 - 0.9876 = 0.0124 or 1.24%.

4.3.13 **(a)** 1 - 0.9394 = 0.0606

(b) 0.3669

(c) 0.7257 - 0.1003 = 0.6254

4.3.14 From 2.31 cm to 4.05 cm

4.3.15 $-0.67 = \frac{y^* - 3.18}{0.53}$, so y* = 3.18 - 0.67*.53 = 2.82 cm

4.3.16 The 10th percentile is 2.4 – 1.28x0.45 = 1.824. Thus 1.75 is below the 10th percentile.

4.3.17 **(a)** 1 - 0.2033 = 0.7967 or 79.67%

(b) In Table 3, the area closest to 0.6 is 0.5987, corresponding to z = 0.25. Thus, the 60th percentile y* satisfies the equation

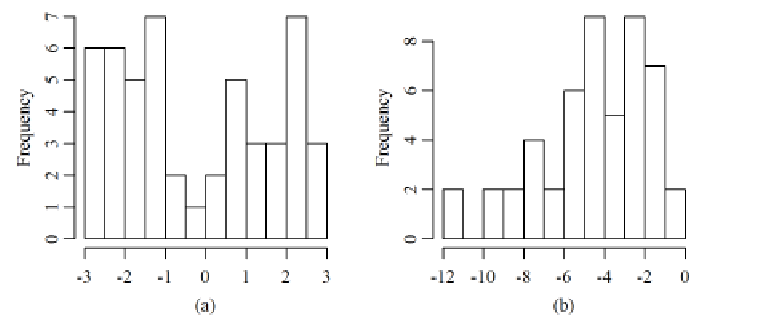
$0.25 = \frac{y^* - 245}{40}$ which yields y* = (0.25)(40) + 245 = 255 min.

(c) 240 minutes = 4 hours. More runners than expected finished in just under 4 hours; fewer than expected finished in just over 4 hours. This may be due to some runners pushing themselves to break the 4 hour mark.

4.4.1 Intervals that include 68%, 95%, and 99.7% of the bill lengths in the distribution are 25.4± 0.8 mm – which is (24.6,26.2), 25.4± 1.6mm – which is (23.8,27.0), and 25.4± 2.4mm – which is (23.0,27.8), respectively.

4.4.2 Normal probability plot (a) goes with histogram I, since that histogram is skewed to the right and plot (a) shows larger Y values than would be expected for a symmetric, bell-shaped distribution; plot (b) goes with histogram III, because that histogram is skewed to the left; and plot (c) goes with histogram II, which is reasonably symmetric and bell-shaped, which corresponds to a linear normal probability plot.

• **4.4.3** The histograms below are histograms of the data used to generate the normal probability plots. Your sketched histogram may look different than these, but should contain similar features: (a) heavy tails, and (b) left skew.



4.4.4 The rainfall data cannot be normally distributed given these summaries. If the data were normal, we would expect about 68% of the rainfall values to fall between -0.09 and 0.13 inches (0.02 ±

0.11). This result implies negative rainfall values. Since rainfall cannot be negative, the distribution cannot be normal.

4.4.5 **(a)** Yes, it is reasonable to believe that the February 1 daily high temperatures in Juneau, Alaska follow a normal distribution. That is to say, the summary statistics do not provide evidence to the contrary. If the temperatures were normally distributed, then about 95% of the time the February 1 daily high temperature would be between -2.7 and 4.9 °C [1.1 ± 2(1.9)]. Negative temperatures (i.e. below freezing) are perfectly reasonable for this time of the year in Alaska.

(b) No, a lack of evidence for abnormality is not the same as evidence for normality.

4.4.6 **(a)** At the low end of the distribution the normal probability plot is fairly straight, indicating that the data agree with what one would expect from a normal distribution. Thus, the times for the fastest riders are roughly equal to the times one would expect if the data came from a truly normal distribution.

(b) At the high end of the distribution the normal probability plot bends upward, indicating that the times are greater than what one would expect from a normal distribution. Thus, the times for the slowest riders are worse than the times one would expect.

4.4.7 **(a)** The P-value for this plot is 0.00015. This small P-value indicates there is compelling evidence for abnormality. Correspondingly, normal probability plot (a) shows strong curvature indicating abnormality in the population.

(b) The P-value for this plot is 0.235. This P-value is rather large, which indicates a lack of evidence for abnormality in the population. Normal probability plot (b) shows a fairly linear pattern, particularly in the middle of the data set.

• **4.4.8** **(a)** No, it does not seem reasonable to believe that these data came from a normal population, since the P-value = 0.039 is small (indeed, it is less than 0.05). Thus, we have moderate evidence that the population from which this data comes is abnormal.

(b) Yes, it seems reasonable to believe that these data came from a normal population, since the P-value = 0.770 $\geq$ 0.10. That is, there is not even weak evidence for abnormality in the population.

4.S.1 **(a)** 1 - 0.9236 = 0.8268

(b) 0.2389

(c) 0.9236 - 0.0764 = 0.8472

(d) 0.2389 - 0.0764 = 0.1625

4.S.2 In Table 3, the areas closest to 0.95 are 0.9495, corresponding to z = 1.64, and 0.9505, corresponding to z = 1.65. Using 1.64, the 95th percentile y* satisfies the equation

$$1.64 = \frac{y^* - 25.4}{0.8}$$

which yields y* = (0.8)(1.64) + 25.4 = 26.7 mm.

Alternatively, using 1.65, we find y* = (0.8)(1.65) + 25.4 = 26.72 mm, which rounds to 26.7 mm.

4.S.3 **(a)** = 1; area = 1 – 0.8413 = 0.1587 or 15.87%.

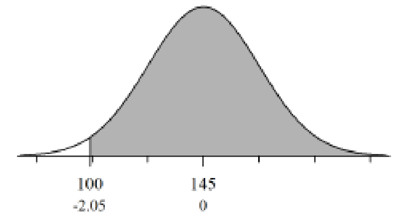
(b) z = -1.75; area = 0.0401 or 4.01%.

• **4.S.4** μ = 145; σ = 22.

(a) For y = 100,

$$z = \frac{y-\mu}{\sigma} = \frac{100-145}{22} = -2.05.$$

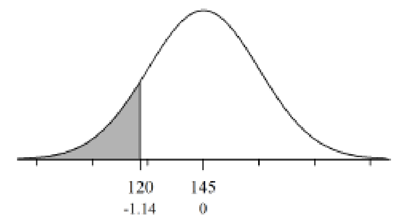
From Table 3, the area below -2.05 is 0.0202.
Thus, Pr{Y $\geq$ 100} = 1 - 0.0202 = 0.9798 or 97.98%.



(b) For y = 120,

$$z = \frac{y-\mu}{\sigma} = \frac{120-145}{22} = -1.14.$$

From Table 3, the area below -1.14 is 0.1271.
Thus, Pr{Y $\leq$ 120} = 0.1271 or 12.71%.



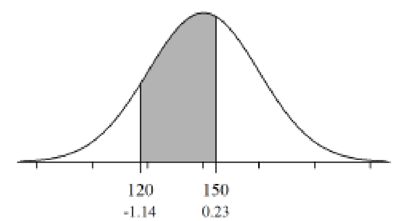
(c) For y = 120,

$$z = \frac{y-\mu}{\sigma} = \frac{120-145}{22} = -1.14.$$

From Table 3, the area below -1.14 is 0.1271.
For y = 150,

$$z = \frac{y-\mu}{\sigma} = \frac{150-145}{22} = 0.23.$$

From Table 3, the area below 0.23 is 0.5910.
Thus, Pr{120 $\leq$ Y $\leq$ 150}= 0.5910 - 0.1271 = 0.4639 or 46.39%.



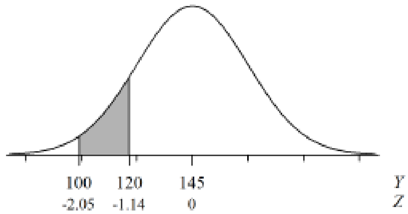
(d) For y = 100,

$z = \frac{y-\mu}{\sigma} = \frac{100-145}{22} = -2.05.$

From Table 3, the area below -2.05 is 0.0202.
For y = 120,

$z = \frac{y-\mu}{\sigma} = \frac{120-145}{22} = -1.14.$

From Table 3, the area below -1.14 is 0.1271.
Thus, Pr{100 ≤ Y ≤ 120}=
0.1271 - 0.0202 = 0.1069 or 10.69%.



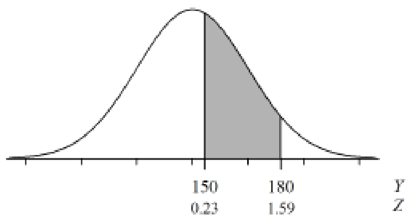
(e) For y = 150,

$z = \frac{y-\mu}{\sigma} = \frac{150-145}{22} = 0.23.$

From Table 3, the area below 0.23 is 0.5910.
For y = 180,

$z = \frac{y-\mu}{\sigma} = \frac{180-145}{22} = 1.59.$

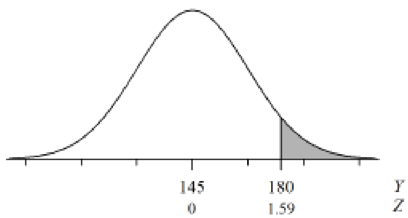
From Table 3, the area below 1.59 is 0.9441.
Thus, Pr{150 ≤ Y ≤ 180}=
0.9441 - 0.5910 = 0.3531 or 35.31%.



(f) For y = 180,

$z = \frac{y-\mu}{\sigma} = \frac{180-145}{22} = 1.59.$

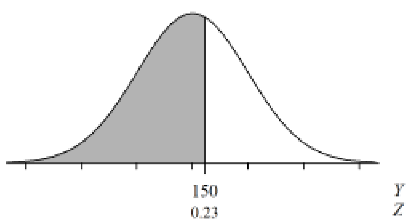
From Table 3, the area below 1.59 is 0.9441.
Thus, Pr{Y ≥ 150}= 1 - 0.9441 = 0.0559 or 5.59%.



(g) For y = 150,

$z = \frac{y-\mu}{\sigma} = \frac{150-145}{22} = 0.23.$

From Table 3, the area below 0.23 is 0.5910.
Thus, Pr{Y ≤ 150}= 0.5910 or 59.10%.



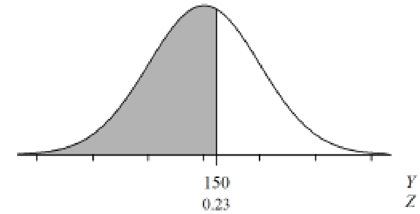
• **4.S.5** μ = 145; σ = 22.

If none of the plants is more than 150 cm tall, then all of the plants are less than or equal to 150 cm tall.

For y = 150,

$z = \frac{y-\mu}{\sigma} = \frac{150-145}{22} = 0.23.$

From Table 3, the area below 0.23 is 0.5910.
Thus, Pr{Y ≤ 150}= 0.5910 or 59.10%.



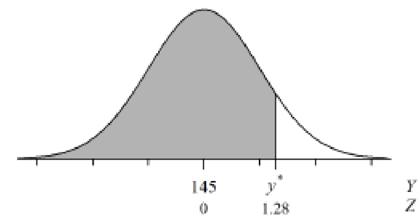
We can apply the binomial formula with n = 4, p = 0.5910, and j = 4.
Thus, Pr{none more than 150 cm tall}
= Pr{all less than or equal to 150 cm tall}
$= {}_4C_4 0.5910^4 = 0.122.$

• **4.S.6** μ = 145; σ = 22.

The 90th percentile is the value that is larger than 90% of the distribution. In Table 3, the area closest to .9 is .8997, corresponding to z = 1.28. Thus, the 90th percentile y* satisfies the equation

$1.28 = \frac{y^*-145}{22}.$

The solution of this equation is
$y^* = (22)(1.28) + 145 = 173.2.$
Thus, the 90th percentile of the distribution is 173.2 cm.



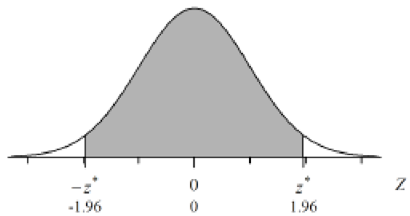
4.S.7 In Table 3, the area closest to 0.25 is 0.2486, which corresponds to z = -0.67. Thus, the first quartile y* satisfies

$-0.67 = \frac{y^*-145}{22}$

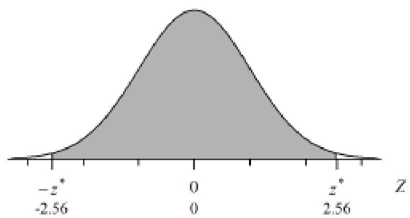
which yields y* = (-0.67)(22) + 145 = 130.26 cm. Similarly, the third quartile is (0.67)(22) + 145 = 159.74 cm. The interquartile range is 159.74 – 130.26 = 29.48 cm/

• **4.S.8** μ = 145; σ = 22.

(a) We wish to find z* such that the shaded area is 0.95. This means that the area in the left tail is 0.025. In Table 3, the area 0.025 corresponds to z = -1.96, which is -z*. Likewise, the area 0.975 corresponds to z = 1.96. Thus, z* = 1.96.



(b) We wish to find z* such that the shaded area is 0.99. This means that the area in the left tail is 0.005. In Table 3, the area 0.0005 corresponds to z = -2.58, which is -z*. Likewise, the area 0.995 corresponds to z = 2.56. Thus, z* = 2.58.



4.S.9 **(a)** 1 - 0.0668 = 0.9332

(b) 1 - 0.9878 - 0.0122

(c) 0.9878 - 0.0668 = 0.9210

(d) 0.4013 - 0.0668 = 0.3345

4.S.10 In Table 3, the area closest to 0.25 is 0.2486, which corresponds to z = -0.67. Thus, the first quartile y* satisfies

$$-0.67 = \frac{y^* - 15.6}{0.4}$$

which yields y* = (-0.67)(0.4) + 15.6 = 15.33 ms. Similarly, the third quartile is (0.67)(0.4) + 15.6 = 15.87 ms. The interquartile range is 15.87 - 15.33 = 0.54 ms.

4.S.11 By symmetry of the normal curve,

$$\mu = \frac{1}{2}(61.2 + 67.4) = 64.3 \text{ inches.}$$

In Table 3, the area closest to 0.9 is 0.8997, which corresponds to z = 1.28. Therefore, we have

$$1.28 = \frac{67.4 - 64.3}{\sigma}$$

which yields σ = (67.4 - 64.3)/1.28 = 2.4 inches.

4.S.12 **(a)** 1 - 0.9938 = 0.0062 or 0.62%

(b) 0.1056 or 10.56%

(c) 0.8944 - 0.1056 = 0.7888 or 78.88%

(d) 0.9938 - 0.1056 = 0.8882 or 88.82%

(e) 0.9938 - 0.8944 = 0.0994 or 9.94%

4.S.13 0.9938 - 0.1056 = 0.8882

4.S.14 From Table 3, the probability that a single child has a score of 80 or less is 0.1056. Using the binomial formula,

Pr{one 80 or less, four higher than 80} = (5)(0.1056)(0.8944^4) = 0.3379.

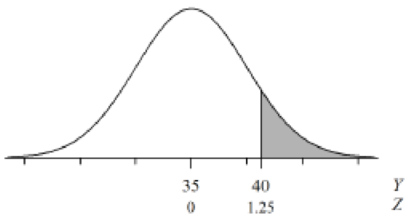
• **4.S.15** The distribution of readings is a normal distribution with mean μ (the true concentration) and standard deviation σ. A reading of 40 or more is considered "unusually high." Suppose that μ = 35 and σ = 4.

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).