# Chapitre 3

# 5: Several Useful Discrete Distributions

**5.3** The random variable $x$ is not a binomial random variable since the balls are selected without replacement. For this reason, the probability $p$ of choosing a red ball changes from trial to trial.

**5.4** If the sampling in Exercise 5.3 is conducted with replacement, then $x$ is a binomial random variable with $n = 2$ independent trials, and $p = P[\text{red ball}] = 3/5$, which remains constant from trial to trial.

**5.11** **a** For $n = 10$ and $p = .4$, $P(x = 4) = C_4^{10}(.4)^4(.6)^6 = .251$.

**b** To calculate $P(x \geq 4) = p(4) + p(5) + \cdots + p(10)$ it is easiest to write

$$P(x \geq 4) = 1 - P(x < 4) = 1 - P(x \leq 3).$$

These probabilities can be found individually using the binomial formula, or alternatively using the cumulative binomial tables in Appendix I.

$$P(x = 0) = C_0^{10}(.4)^0(.6)^{10} = .006 \qquad P(x = 1) = C_1^{10}(.4)^1(.6)^9 = .040$$

$$P(x = 2) = C_2^{10}(.4)^2(.6)^8 = .121 \qquad P(x = 3) = C_3^{10}(.4)^3(.6)^7 = .215$$

The sum of these probabilities gives $P(x \leq 3) = .382$ and $P(x \geq 4) = 1 - .382 = .618$.

**c** Use the results of parts **a** and **b**.

$$P(x > 4) = 1 - P(x \leq 4) = 1 - (.382 + .251) = .367$$

**d** From part **c**, $P(x \leq 4) = P(x \leq 3) + P(x = 4) = .382 + .251 = .633$.

**e** $\mu = np = 10(.4) = 4$

**f** $\sigma = \sqrt{npq} = \sqrt{10(.4)(.6)} = \sqrt{2.4} = 1.549$

**5.20** Although there are $n = 30$ days on which it either rains (S) or does not rain (F), the random variable $x$ would not be a binomial random variable because the trials (days) are *not independent*. If there is rain on one day, it will probably affect the probability that there will be rain on the next day.

**5.21** Although there are trials (telephone calls) which result in either a person who will answer (S) or a person who will not (F), the number of trials, $n$, is not fixed in advance. Instead of recording $x$, the number of *successes* in $n$ trials, you record $x$, the number of *trials* until the first success. This is *not* a binomial experiment.

**5.22** There are $n = 100$ students which represent the experimental units.

**a** Each student either took (S) or did not take (F) the SAT. Since the population of students is large, the probability $p = .45$ that a particular student took the SAT will not vary from student to student and the trials will be independent. This is a binomial random variable.

**b** The measurement taken on each student is *score* which can take more than two values. This is not a binomial random variable.

**c** Each student either will (S) or will not (F) score above average. As in part **a**, the trials are independent although the value of $p$, the proportion of students in the population who score above average, is unknown. This is a binomial experiment.

**d** The measurement taken on each student is *amount of time* which can take more than two values. This is not a binomial random variable.

**5.23** Define $x$ to be the number of alarm systems that are triggered. Then $p = P[\text{alarm is triggered}] = .99$ and $n = 9$. Since there is a table available in Appendix I for $n = 9$ and $p = .99$, you should use it rather than the binomial formula to calculate the necessary probabilities.

**a** $P[\text{at least one alarm is triggered}] = P(x \geq 1) = 1 - P(x = 0) = 1 - .000 = 1.000$.

**b** $P[\text{more than seven}] = P(x > 7) = 1 - P(x \leq 7) = 1 - .003 = .997$

**c** $P[\text{eight or fewer}] = P(x \le 8) = .086$

**5.25** Define *x* to be the number of cars that are black. Then $p = P[\text{black}] = .1$ and $n = 25$. Use Table 1 in Appendix I.

**a** $P(x \ge 5) = 1 - P(x \le 4) = 1 - .902 = .098$

**b** $P(x \le 6) = .991$

**c** $P(x > 4) = 1 - P(x \le 4) = 1 - .902 = .098$

**d** $P(x = 4) = P(x \le 4) - P(x \le 3) = .902 - .764 = .138$

**e** $P(3 \le x \le 5) = P(x \le 5) - P(x \le 2) = .967 - .537 = .430$

**f** $P(\text{more than 20 } \textit{not} \text{ black}) = P(\text{less than 5 black}) = P(x \le 4) = .902$

**5.30** Define *x* to be the number of times the mouse chooses the red door. Then, if the mouse actually has no preference for color, $p = P[\text{red door}] = .5$ and $n = 10$. Since $\mu = np = 5$ and $\sigma = \sqrt{npq} = 1.58$, you would expect that, if there is no color preference, the mouse should choose the red door

$$\mu \pm 2\sigma \Rightarrow 5 \pm 3.16 \Rightarrow 1.84 \text{ to } 8.16$$

or between 2 and 8 times. If the mouse chooses the red door more than 8 or less than 2 times, the unusual results might suggest a color preference.

**5.63** Define *x* to be the number supporting the commissioner's claim, with $p = .8$ and $n = 25$.

**a** Using the binomial tables for $n = 25$, $P[x \ge 22] = 1 - P[x \le 21] = 1 - .766 = .234$

**b** $P[x = 22] = P[x \le 22] - P[x \le 21] = .902 - .766 = .136$

**c** The probability of observing an event as extreme as $x = 22$ (or more extreme) is quite high assuming that $p = .8$. Hence, this is not an unlikely event and we would not doubt the claim.

# Chapitre 4

# 6: The Normal Probability Distribution

**6.5** **a** $P(-1.43 < z < .68) = A(.68) - A(-1.43) = .7517 - .0764 = .6753$

**b** $P(.58 < z < 1.74) = A(1.74) - A(.58) = .9591 - .7190 = .2401$

**c** $P(-1.55 < z < -.44) = A(-.44) - A(-1.55) = .3300 - .0606 = .2694$

**d** $P(z > 1.34) = 1 - A(1.34) = 1 - .9099 = .0901$

**e** Since the value of $z = -4.32$ is not recorded in Table 3, you can assume that the area to the left of $z = -4.32$ is very close to 0. Then

$$P(z < -4.32) \approx 0$$

**6.6** Similar to Exercise 6.5.

**a** $P(z < 2.33) = A(2.33) = .9901$

**b** As in part **a**, $P(z < 1.645) = A(1.645)$. However, the value for $z = 1.645$ is not given in Table 3, but falls halfway between two tabulated values, $z = 1.64$ and $z = 1.65$. One solution is to choose an area $A(1.645)$ which lies halfway between the two tabulated areas, $A(1.64) = .9495$ and $A(1.65) = .9505$. Then

$$A(1.645) = .9500 \quad \text{and} \quad P(z < 1.645) = A(.9500).$$

This method of evaluation is called "linear interpolation" and is performed as follows:

**1** The difference between two entries in the table is called a "tabular difference". Interpolation is accomplished by taking appropriate portions of this difference.

**2** Let $P_0$ be the probability associated with $z_0$ (i.e. $P_0 = A(z_0)$) and let $P_1$ and $P_2$ be the two tabulated probabilities with corresponding $z$ values, $z_1$ and $z_2$. Consider $\frac{z_0 - z_1}{z_2 - z_1}$ which is the proportion of the distance from $z_1$ to $z_0$.

**3** Multiply

$$\frac{z_0 - z_1}{z_2 - z_1}(P_2 - P_1)$$

to obtain a corresponding proportion for the probabilities and add this value to $P_1$. This value is the desired $P_0 = A(z_0)$. Thus, in this case,

$$\frac{z_0 - z_1}{z_2 - z_1} = \frac{1.645 - 1.64}{1.65 - 1.64} = \frac{.005}{.010} = \frac{1}{2}$$

and

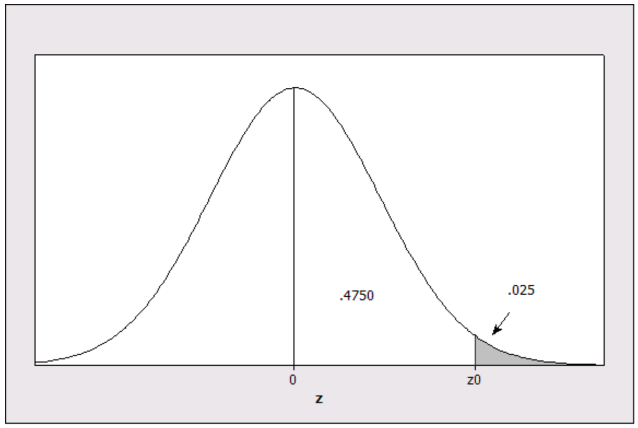
$$P_0 = A(z_0) = P_1 + \frac{z_0 - z_1}{z_2 - z_1}(P_2 - P_1) = .9495 + \left(\frac{1}{2}\right)[.9505 - .9495] = .9495 + .0005 = .9500$$

**c** $P(z > 1.96) = 1 - A(1.96) = 1 - .9750 = .0250$

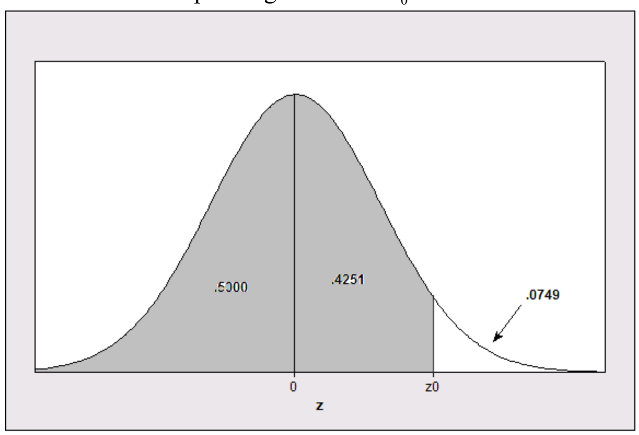
**d** $P(-2.58 < z < 2.58) = A(2.58) - A(-2.58) = .9951 - .0049 = .9902$

**6.7** Now we are asked to find the $z$-value corresponding to a particular area.

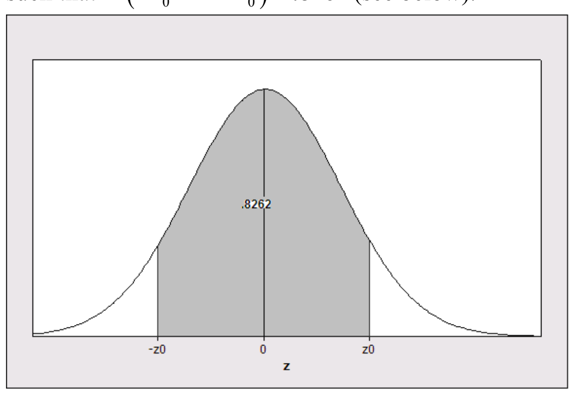
**a** We need to find a $z_0$ such that $P(z > z_0) = .025$. This is equivalent to finding an indexed area of $1 - .025 = .975$. Search the interior of Table 3 until you find the four-digit number **.9750**. The corresponding $z$-value is **1.96**; that is, $A(1.96) = .9750$. Therefore, $z_0 = 1.96$ is the desired $z$-value (see the figure below).




**b** We need to find a $z_0$ such that $P(z < z_0) = .9251$ (see below). Using Table 3, we find a value such that the indexed area is .9251. The corresponding *z*-value is $z_0 = 1.44$.



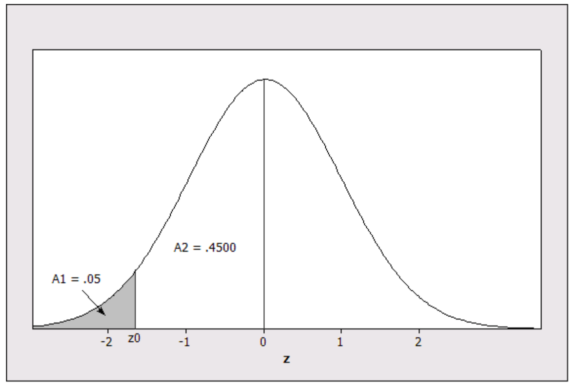

**6.8** We want to find a *z*-value such that $P(-z_0 < z < z_0) = .8262$ (see below).




Since $A(z_0) - A(-z_0) = .8262$, the total area in the two tails of the distribution must be $1 - .8262 = .1738$ so that the lower tail area must be $A(-z_0) = .1738/2 = .0869$. From Table 3, $-z_0 = -1.36$ and $z_0 = 1.36$.

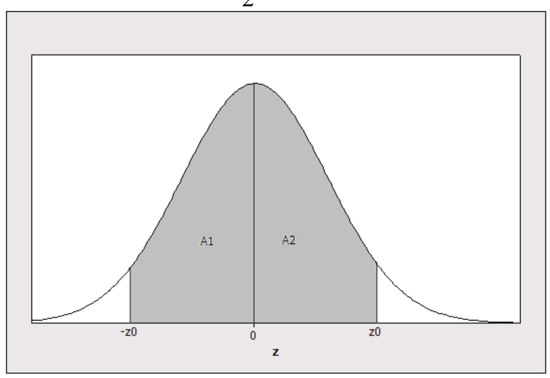
**6.9** **a** Similar to Exercise 6.7b. The value of $z_0$ must be positive and $A(z_0) = .9505$. Hence, $z_0 = 1.65$.

**b** It is given that the area to the left of $z_0$ is .0505, shown as $A_1$ in the figure below. The desired value is not tabulated in Table 3 but falls between two tabulated values, .0505 and .0495. Hence, using linear interpolation (as you did in Exercise 6.6b) $z_0$ will lie halfway between –1.64 and –1.65, or $z_0 = -1.645$.




**6.10** **a** Refer to the figure below. It is given that $P(-z_0 < z < z_0) = .9000$. That is,

$$A(z_0) + A(-z_0) = .9000 \quad \Rightarrow \quad A(-z_0) = \frac{1}{2}(1 - .9000) \quad \Rightarrow \quad A(-z_0) = .0500$$




From Exercise 6.9b, $z_0 = 1.645$.

**b** Refer to the figure above and consider

$$P(-z_0 < z < z_0) = A_1 + A_2 = .9900$$

Then $A(-z_0) = \frac{1}{2}(1 - .9900) = .0050$. Linear interpolation must now be used to determine the value of $-z_0$, which will lie between $z_1 = -2.57$ and $z_2 = -2.58$. Hence, using a method similar to that in Exercise 6.6b, we find

$$z_0 = z_1 + \frac{P_0 - P_1}{P_2 - P_1}(z_2 - z_1) = -2.57 + \left(\frac{.4950 - .4949}{.4951 - .4949}\right)(-2.58 + 2.57)$$

$$= -2.57 - \left(\frac{1}{2}\right)(.01) = -2.575$$

If Table 3 were correct to more than 4 decimal places, you would find that the actual value of $z_0$ is $z_0 = 2.576$; many texts chose to round up and use the value $z_0 = 2.58$.

**6.13** Similar to Exercise 6.12.

**a** Calculate $z_1 = \frac{1.00 - 1.20}{.15} = -1.33$ and $z_2 = \frac{1.10 - 1.20}{.15} = -.67$. Then

$$P(1.00 < x < 1.10) = P(-1.33 < z < -.67) = .2514 - .0918 = .1596$$

**b** Calculate $z = \frac{x - \mu}{\sigma} = \frac{1.38 - 1.20}{.15} = 1.2$. Then

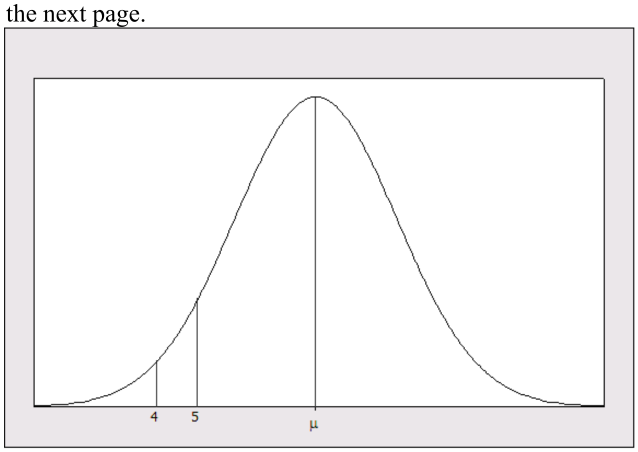
$$P(x > 1.38) = P(z > 1.2) = 1 - .8849 = .1151$$

**c** Calculate $z_1 = \frac{1.35 - 1.20}{.15} = 1$ and $z_2 = \frac{1.50 - 1.20}{.15} = 2$. Then

$$P(1.35 < x < 1.50) = P(1 < z < 2) = .9772 - .8413 = .1359$$

**6.16** The z-value corresponding to $x = 0$ is $z = \frac{x-\mu}{\sigma} = \frac{0-50}{15} = -3.33$. Since the value $x = 0$ lies more than three standard deviations away from the mean, it is considered an unusual observation. The probability of observing a value of z as large or larger than $z = -3.33$ is $A(-3.33) = .0004$.

**6.17** The random variable x is normal with unknown $\mu$ and $\sigma$. However, it is given that $P(x > 4) = P\left(z > \frac{4-\mu}{\sigma}\right) = .9772$ and $P(x > 5) = P\left(z > \frac{5-\mu}{\sigma}\right) = .9332$. These probabilities are shown in the figure on the next page.




The value $\frac{4-\mu}{\sigma}$ is negative, with $A\left(\frac{4-\mu}{\sigma}\right) = 1 - .9772 = .0228$ or $\frac{4-\mu}{\sigma} = -2$ (i)

The value $\frac{5-\mu}{\sigma}$ is also negative, with $A\left(\frac{5-\mu}{\sigma}\right) = 1 - .9332 = .0668$ or $\frac{5-\mu}{\sigma} = -1.5$

(ii)

Equations (i) and (ii) provide two equations in two unknowns which can be solved simultaneously for $\mu$ and $\sigma$. From (i), $\sigma = \frac{\mu - 4}{2}$ which, when substituted into (ii) yields

$$5 - \mu = -1.5\left(\frac{\mu-4}{2}\right)$$
$$10 - 2\mu = -1.5\mu + 6$$
$$\mu = 8$$

and from (i), $\sigma = \frac{8-4}{2} = 2$.

**6.18** The random variable x, the weight of a package of ground beef, has a normal distribution with $\mu = 1$ and $\sigma = .15$.

**a** $P(x > 1) = P\left(z > \frac{1-1}{.15}\right) = P(z > 0) = 1 - .5 = .5$

**b** $P(.95 < x < 1.05) = P\left(\frac{.95-1}{.15} < z < \frac{1.05-1}{.15}\right) = P(-.33 < z < .33) = .6293 - .3707 = .2586$

**c** $P(x < .80) = P\left(z < \frac{.80-1}{.15}\right) = P(z < -1.33) = .0918$

**d** The z-value corresponding to $x = 1.45$ is $z = \frac{x-\mu}{\sigma} = \frac{1.45-1}{.15} = 3$, which would be considered an unusual observation. Perhaps the setting on the scale was accidentally changed to 1.5 pounds!

**6.19** The random variable *x*, the height of a male human, has a normal distribution with $\mu = 69$ and $\sigma = 3.5$.

**a** A height of 6'0" represents $6(12) = 72$ inches, so that

$$P(x > 72) = P\left(z > \frac{72-69}{3.5}\right) = P(z > .86) = 1 - .8051 = .1949$$

**b** Heights of 5'8" and 6'1" represent $5(12)+8 = 68$ and $6(12)+1 = 73$ inches, respectively. Then

$$P(68 < x < 73) = P\left(\frac{68-69}{3.5} < z < \frac{73-69}{3.5}\right) = P(-.29 < z < 1.14) = .8729 - .3859 = .4870$$

**c** A height of 6'0" represents $6(12) = 72$ inches, which has a *z*-value of

$$z = \frac{72-69}{3.5} = .86$$

This would not be considered an unusually large value, since it is less than two standard deviations from the mean.

**d** The probability that a man is 6'0" or taller was found in part **a** to be .1949, which is not an unusual occurrence. However, if you define *y* to be the number of men in a random sample of size $n = 42$ who are 6'0" or taller, then *y* has a binomial distribution with mean $\mu = np = 42(.1949) = 8.1858$ and standard deviation $\sigma = \sqrt{npq} = \sqrt{42(.1949)(.8051)} = 2.567$. The value $y = 18$ lies

$$\frac{y-\mu}{\sigma} = \frac{18-8.1858}{2.567} = 3.82$$

standard deviations from the mean, and would be considered an unusual occurrence for the general population of male humans. Perhaps our presidents do not represent a *random* sample from this population.

**6.21** The random variable *x*, cerebral blood flow, has a normal distribution with $\mu = 74$ and $\sigma = 16$.

**a** $P(60 < x < 80) = P\left(\frac{60-74}{16} < z < \frac{80-74}{16}\right) = P(-.88 < z < .38) = .6480 - .1894 = .4586$

**b** $P(x > 100) = P\left(z > \frac{100-74}{16}\right) = P(z > 1.62) = 1 - .9474 = .0526$

**c** $P(x < 40) = P\left(z < \frac{40-74}{16}\right) = P(z < -2.12) = .0170$

**6.29** It is given that the counts of the number of bacteria are normally distributed with $\mu = 85$ and $\sigma = 9$. The *z*-value corresponding to $x = 100$ is $z = \frac{x-\mu}{\sigma} = \frac{100-85}{9} = 1.67$ and

$$P(x > 100) = P(z > 1.67) = 1 - .9525 = .0475$$

**6.31** Let *w* be the number of words specified in the contract. Then *x*, the number of words in the manuscript, is normally distributed with $\mu = w + 20,000$ and $\sigma = 10,000$. The publisher would like to specify *w* so that

$$P(x < 100,000) = .95.$$

As in Exercise 6.30, calculate

$$z = \frac{100,0000 - (w + 20,000)}{10,000} = \frac{80,000 - w}{10,000}.$$

Then $P(x < 100,000) = P\left(z < \frac{80,000 - w}{10,000}\right) = .95$. It is necessary that $z_0 = (80,000 - w)/10,000$ be such that

$$P(z < z_0) = .95 \quad \Rightarrow \quad A(z_0) = .9500 \quad \text{or} \quad z_0 = 1.645.$$

Hence,

$$\frac{80,000-w}{10,000}=1.645 \quad \text{or} \quad w=63,550 .$$

**6.45** **a** The approximating probability will be $P(x<29.5)$ where *x* has a normal distribution with $\mu=50(.7)=35$ and $\sigma=\sqrt{50(.7)(.3)}=3.24$. Then

$$P(x<29.5)=P\left(z<\frac{29.5-35}{3.24}\right)=P(z<-1.70)=.0446$$

**b** The approximating probability is

$$P(x>42.5)=P\left(z>\frac{42.5-35}{3.24}\right)=P(z>2.31)=1-.9896=.0104$$

**c** If fewer than 10 individuals *did not* watch a movie at home this week, then more than $50-10=40$ did watch a movie at home. The approximating probability is

$$P(x>40.5)=P\left(z>\frac{40.5-35}{3.24}\right)=P(z>1.70)=1-.9554=.0446$$

**6.48** Define *x* to be the number of workers with identifiable lung cancer. If the rate of lung cancer in the population of workers in the air-polluted environment is the same as the population in general, then $p=1/40=.025$ and the random variable *x* has a binomial distribution with $n=400$ and $p=.025$. Calculate

$$\mu=np=400(.025)=10 \text{ and } \sigma=\sqrt{400(.025)(.975)}=\sqrt{9.75}=3.122$$

If we want to show that the rate of lung cancer in the polluted environment is greater than the general population rate, we need to show that the number of cancer victims in the sample from the polluted environment is unusually high. Hence, we calculate the *z*-score associated with $x=19$ as

$$z=\frac{x-\mu}{\sigma}=\frac{19-10}{3.122}=2.882$$

which is quite large. This would imply that the value $x=19$ is unusually large, and would indicate that in fact *p* is greater than 1/40 for the workers in the polluted environment.

**6.51** Define *x* to be the number of consumers who preferred a *Pepsi* product. Then the random variable *x* has a binomial distribution with $n=500$ and $p=.26$, if *Pepsi's* market share is indeed 26%. Calculate

$$\mu=np=500(.26)=130 \text{ and } \sigma=\sqrt{500(.26)(.74)}=\sqrt{96.2}=9.808$$

**a** Using the normal approximation with correction for continuity, we find the area between $x=149.5$ and $x=150.5$:

$$P(149.5<x<150.5)=P\left(\frac{149.5-130}{9.808}<z<\frac{150.5-130}{9.808}\right)=P(1.99<z<2.09)=.9817-.9767=.0050$$

**b** Find the area between $x=119.5$ and $x=150.5$:

$$P(119.5<x<150.5)=P\left(\frac{119.5-130}{9.808}<z<\frac{150.5-130}{9.808}\right)=P(-1.07<z<2.09)=.9817-.1423=.8394$$

**c** Find the area to the left of $x=149.5$:

$$P(x<149.5)=P\left(z<\frac{149.5-130}{9.808}\right)=P(z<1.99)=.9767$$

**d** The value $x=232$ lies $z=\frac{232-130}{9.808}=10.40$ standard deviations above the mean, if *Pepsi's* market share is indeed 26%. This is such an unusual occurrence that we would conclude that *Pepsi's* market share is higher than claimed.

**6.71** It is given that the random variable *x* (ounces of fill) is normally distributed with mean $\mu$ and standard deviation $\sigma=.3$. It is necessary to find a value of $\mu$ so that $P(x>8)=.01$. That is, an 8-ounce cup will overflow when $x>8$, and this should happen only 1% of the time. Then

$$P(x > 8) = P\left(z > \frac{8-\mu}{.3}\right) = .01 .$$

From Table 3, the value of *z* corresponding to an area (in the upper tail of the distribution) of .01 is $z_0 = 2.33$. Hence, the value of $\mu$ can be obtained by solving for $\mu$ in the following equation:

$$2.33 = \frac{8-\mu}{.3} \quad or \quad \mu = 7.301$$

**6.72** The 3000 light bulbs utilized by the manufacturing plant comprise the entire population (that is, this is not a sample from the population) whose length of life is normally distributed with mean $\mu = 500$ and standard deviation $\sigma = 50$. The objective is to find a particular value, $x_0$, so that

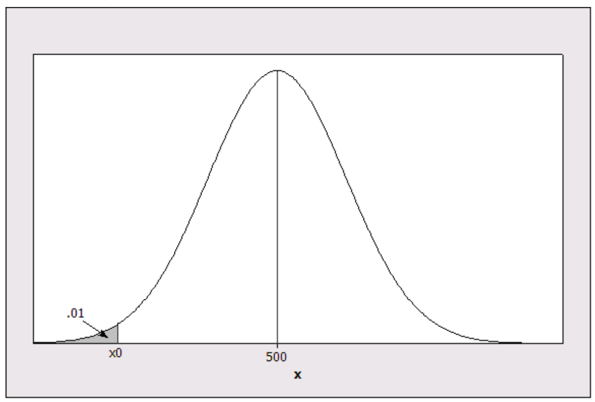
$$P(x \le x_0) = .01 .$$

That is, only 1% of the bulbs will burn out before they are replaced at time $x_0$. Then

$$P(x \le x_0) = P(z \le z_0) = .01 \quad \text{where} \quad z_0 = \frac{x_0 - 500}{50} .$$

From Table 3, the value of *z* corresponding to an area (in the left tail of the distribution) of .01 is $z_0 = -2.33$. Solving for $x_0$ corresponding to $z_0 = -2.33$,

$$-2.33 = \frac{x_0 - 500}{50} \quad \Rightarrow \quad -116.5 = x_0 - 500 \quad \Rightarrow \quad x_0 = 383.5$$



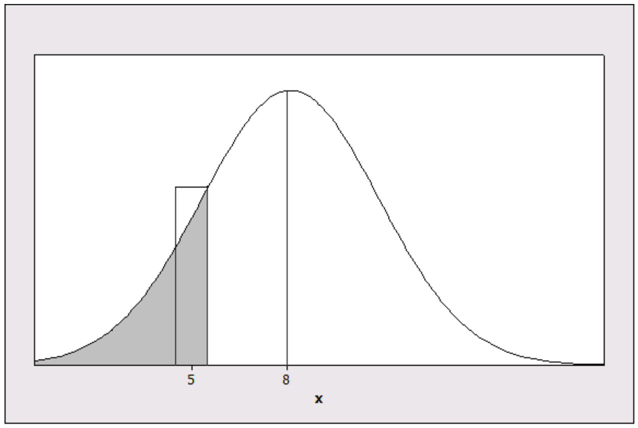

**6.74** The random variable of interest is *x*, the number of persons not showing up for a given flight. This is a binomial random variable with $n = 160$ and $p = P[\text{person does not show up}] = .05$. If there is to be a seat available for every person planning to fly, then there must be at least five persons not showing up. Hence, the probability of interest is $P(x \ge 5)$. Calculate

$$\mu = np = 160(.05) = 8 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{160(.05)(.95)} = \sqrt{7.6} = 2.7$$

Referring to the figure on the next page, a correction for continuity is made to include the entire area under the rectangle associated with the value $x = 5$, and the approximation becomes $P(x \ge 4.5)$. The z-value corresponding to $x = 4.5$ is

$$z = \frac{x-\mu}{\sigma} = \frac{4.5-8}{\sqrt{7.6}} = -1.27$$

so that $$P(x \ge 4.5) = P(z \ge -1.27) = 1 - .1020 = .8980$$


5
8
x

# 7: Sampling Distributions

**7.19** Regardless of the shape of the population from which we are sampling, the sampling distribution of the sample mean will have a mean $\mu$ equal to the mean of the population from which we are sampling, and a standard deviation equal to $\sigma/\sqrt{n}$ .

**a** $\mu = 10;\quad \sigma/\sqrt{n} = 3/\sqrt{36} = .5$

**b** $\mu = 5;\quad \sigma/\sqrt{n} = 2/\sqrt{100} = .2$

**c** $\mu = 120;\quad \sigma/\sqrt{n} = 1/\sqrt{8} = .3536$

**7.20** **a** If the sampled populations are normal, the distribution of $\overline{x}$ is also normal *for all values of n*.
**b** The Central Limit Theorem states that for sample sizes as small as $n = 25$ , the sampling distribution of $\overline{x}$ will be approximately normal. Hence, we can be relatively certain that the sampling distribution of $\overline{x}$ for parts **a** and **b** will be approximately normal. However, the sample size is part **c**, $n = 8$ , is too small to assume that the distribution of $\overline{x}$ is approximately normal.

**7.26** **a** Since the sample size is large, the sampling distribution of $\overline{x}$ will be approximately normal with mean $\mu = 64,571$ and standard deviation $\sigma/\sqrt{n} = 4000/\sqrt{60} = 516.3978$ .
**b** From the Empirical Rule (and the general properties of the normal distribution), approximately 95% of the measurements will lie within 2 standard deviations of the mean:

$$\mu \pm 2SE \;\Rightarrow\; 64,571 \pm 2(516.3978)$$
$$64,571 \pm 1032.80 \text{ or } 63,538.20 \text{ to } 65,603.80$$

**c** Use the mean and standard deviation for the distribution of $\overline{x}$ given in part **a**.

$$P(\overline{x} > 66,000) = P\left(z > \frac{66,000 - 64,571}{516.3978}\right)$$
$$= P(z > 2.77) = 1 - .9972 = .0028$$

**d** Refer to part **c**. You have observed a very unlikely occurrence, assuming that $\mu = 64,571$ . Perhaps your sample was not a random sample, or perhaps the average salary of \$64,571 is no longer correct.

**7.32** **a** Since the total daily sales is the sum of the sales made by a fixed number of customers on a given day, it is a sum of random variables, which, according to the Central Limit Theorem, will have an approximate normal distribution.
**b** Let $x_i$ be the total sales for a single customer, with i = 1, 2, …., 30. Then $x_i$ has a probability distribution with $\mu = 8.50$ and $\sigma = 2.5$ . The total daily sales can now be written as $x = \sum x_i$ . If $n = 30$ , the mean and standard deviation of the sampling distribution of $x$ are given as

$$n\mu = 30(8.5) = 255 \text{ and } \sigma\sqrt{n} = 2.5\sqrt{30} = 13.693$$

**7.34** The sampled population has a mean of 5.97 with a standard deviation of 1.95.

**a** With $n = 31$, calculate $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{6.5 - 5.97}{1.95/\sqrt{31}} = 1.51$, so that

$P(\overline{x} \le 6.5) = P(z \le 1.51) = .9345$

**b** Calculate $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{9.80 - 5.97}{1.95/\sqrt{31}} = 10.94$ , so that

$P(\overline{x} \ge 9.80) = P(z \ge 10.94) \approx 1 - 1 = 0$

**c** The probability of observing an average diameter of 9.80 or higher is extremely unlikely, if indeed the average diameter in the population of affected tendons was no different from that of unaffected tendons (5.97). We would conclude that the average diameter in the population of patients affected with AT is higher than 5.97.

**7.43** **a** For $n = 100$ and $p = .19$, $np = 19$ and $nq = 81$ are both greater than 5. Therefore, the normal approximation will be appropriate, with mean $p = .19$ and $SE = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.19(.81)}{100}} = .03923$.

**b** $P(\hat{p} > .25) = P\left(z > \frac{.25 - .19}{.03923}\right) = P(z > 1.53) = 1 - .9370 = .0360$

**c** $P(.25 < \hat{p} < .30) = P\left(\frac{.25 - .19}{.03923} < z < \frac{.30 - .19}{.03923}\right) = P(1.53 < z < 2.80) = .9974 - .9370 = .0604$

**d** The value $\hat{p} = .30$ lies

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{.30 - .19}{.03923} = 2.80$$

standard deviations from the mean. Also, $P(\hat{p} \geq .30) = P(z \geq 2.80) = 1 - .9974 = .0026$. This is an unlikely occurrence, assuming that $p = .19$. Perhaps the sampling was not random, or the 19% figure is not correct.

**7.45** **a** The random variable $\hat{p}$, the sample proportion of brown M&Ms in a package of $n = 55$, has a binomial distribution with $n = 55$ and $p = .13$. Since $np = 7.15$ and $nq = 47.85$ are both greater than 5, this binomial distribution can be approximated by a normal distribution with mean $p = .13$ and

$SE = \sqrt{\frac{.13(.87)}{55}} = .04535.$

**b** $P(\hat{p} < .2) = P\left(z < \frac{.2 - .13}{.04535}\right) = P(z < 1.54) = .9382$

**c** $P(\hat{p} > .35) = P\left(z > \frac{.35 - .13}{.04535}\right) = P(z > 4.85) \approx 1 - 1 = 0$

**d** From the Empirical Rule (and the general properties of the normal distribution), approximately 95% of the measurements will lie within 2 (or 1.96) standard deviations of the mean:

$$p \pm 2SE \quad \Rightarrow \quad .13 \pm 2(.04535)$$

$$.13 \pm .09 \quad \text{or} \quad .04 \text{ to } .22$$

# Chapitre 6

# 8: Large-Sample Estimation

**8.1** The margin of error in estimation provides a practical upper bound to the difference between a particular estimate and the parameter which it estimates. In this chapter, the margin of error is $1.96\times$ (standard error of the estimator).

**8.3** For the estimate of $\mu$ given as $\bar{x}$, the margin of error is $1.96\ SE = 1.96\frac{\sigma}{\sqrt{n}}$.

**a** $1.96\sqrt{\frac{0.2}{30}} = .160$ **b** $1.96\sqrt{\frac{0.9}{30}} = .339$ **c** $1.96\sqrt{\frac{1.5}{30}} = .438$

**8.4** Refer to Exercise 8.3. As the population variance $\sigma^2$ increases, the margin of error also increases.

**8.5** The margin of error is $1.96\ SE = 1.96\frac{\sigma}{\sqrt{n}}$, where $\sigma$ can be estimated by the sample standard deviation *s* for large values of *n*.

**a** $1.96\sqrt{\frac{4}{50}} = .554$ **b** $1.96\sqrt{\frac{4}{500}} = .175$ **c** $1.96\sqrt{\frac{4}{5000}} = .055$

**8.6** Refer to Exercise 8.5. As the sample size *n* increases, the margin of error decreases.

**8.13** The point estimate of $\mu$ is $\bar{x} = 39.8^{\circ}$ and the margin of error with $s = 17.2$ and $n = 50$ is

$$1.96\ SE = 1.96\frac{\sigma}{\sqrt{n}} \approx 1.96\frac{s}{\sqrt{n}} = 1.96\frac{17.2}{\sqrt{50}} = 4.768$$

**8.17** **a** The point estimate for *p* is given as $\hat{p} = \frac{x}{n} = .51$ and the margin of error is approximately

$$1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{.51(.49)}{900}} = .0327$$

**8.24** **a** $\bar{x} \pm z_{.005}\frac{\sigma}{\sqrt{n}} = \bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}} \approx 34 \pm 2.58\sqrt{\frac{12}{38}} = 34 \pm 1.450$ or $32.550 < \mu < 35.450$.

**b** $\bar{x} \pm z_{.05}\frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.645\frac{\sigma}{\sqrt{n}} \approx 1049 \pm 1.645\sqrt{\frac{51}{65}} = 1049 \pm 1.457$ or $1047.543 < \mu < 1050.457$.

**c** $\bar{x} \pm z_{.025}\frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} \approx 66.3 \pm 1.96\sqrt{\frac{2.48}{89}} = 66.3 \pm .327$ or $65.973 < \mu < 66.627$.

**8.27** The width of a 95% confidence interval for $\mu$ is given as $1.96\frac{\sigma}{\sqrt{n}}$. Hence,

**a** When $n = 100$, the width is $2\left(1.96\frac{10}{\sqrt{100}}\right) = 2(1.96) = 3.92$.

**b** When $n = 200$, the width is $2\left(1.96\frac{10}{\sqrt{200}}\right) = 2(1.386) = 2.772$.

**c** When $n = 400$, the width is $2\left(1.96\frac{10}{\sqrt{400}}\right) = 2(.98) = 1.96$.

**8.28** Refer to Exercise 8.27.

**a** When the sample size is doubled, the width is decreased by $1/\sqrt{2}$.

**b** When the sample size is quadrupled, the width is decreased by $1/\sqrt{4} = 1/2$.

**8.29** **a** A 90% confidence interval for $\mu$ is $\bar{x} \pm 1.645\frac{\sigma}{\sqrt{n}}$. Hence, its width is

$$2\left(1.645\frac{\sigma}{\sqrt{n}}\right) = 2\left(1.645\frac{10}{\sqrt{100}}\right) = 2(1.645) = 3.29$$

**b** A 99% confidence interval for $\mu$ is $\bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$. Hence, its width is

$$2\left(2.58\frac{\sigma}{\sqrt{n}}\right) = 2\left(2.58\frac{10}{\sqrt{100}}\right) = 2(2.58) = 5.16$$

**c** Notice that as the confidence coefficient increases, so does the width of the confidence interval. If we wish to be more confident of enclosing the unknown parameter, we must make the interval wider.

**8.32** **a** An approximate 95% confidence interval for *p* is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = .54 \pm 1.96\sqrt{\frac{.54(.46)}{400}} = .54 \pm .049$$

or $.491 < p < .589$.

**b** An approximate 95% confidence interval for *p* is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = .30 \pm 1.96\sqrt{\frac{.30(.70)}{350}} = .30 \pm .048$$

or $.252 < p < .348$.

**8.34** **a** The 90% confidence interval for *p* is

$$\hat{p} \pm 1.645\sqrt{\frac{\hat{p}\hat{q}}{n}} = .39 \pm 1.645\sqrt{\frac{.39(.61)}{1002}} = .39 \pm .025$$

or $.365 < p < .415$.

**b** The 90% confidence interval for *p* is

$$\hat{p} \pm 1.645\sqrt{\frac{\hat{p}\hat{q}}{n}} = .53 \pm 1.645\sqrt{\frac{.53(.47)}{1002}} = .53 \pm .026$$

or $.504 < p < .556$.

**8.37** **a** The 99% confidence interval for $\mu$ is

$$\bar{x} \pm 2.58\frac{s}{\sqrt{n}} = 98.25 \pm 2.58\frac{0.73}{\sqrt{130}} = 98.25 \pm .165 \text{ or } 98.085 < \mu < 98.415$$

**b** Since the possible values for μ given in the confidence interval does not include the value μ = 98.6, it is not likely that the true average body temperature for healthy humans is 98.6, the usual average temperature cited by physicians and others.

**8.42** Similar to previous exercises. The 90% confidence interval for $\mu_1 - \mu_2$ is approximately

$$(\bar{x}_1 - \bar{x}_2) \pm 1.645\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(2.4 - 3.1) \pm 1.645\sqrt{\frac{1.44}{100} + \frac{2.64}{100}}$$

$$-0.7 \pm .332 \quad \text{or} \quad -1.032 < (\mu_1 - \mu_2) < -0.368$$

Intervals constructed in this manner will enclose $\mu_1 - \mu_2$ 90% of the time. Hence, we are fairly certain that this particular interval encloses $(\mu_1 - \mu_2)$.

**8.44** Similar to previous exercises. The 95% confidence interval for $\mu_1 - \mu_2$ is approximately

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(21.3 - 13.4) \pm 1.96\sqrt{\frac{(2.6)^2}{30} + \frac{(1.9)^2}{30}}$$

$$7.9 \pm 1.152 \quad \text{or} \quad 6.748 < (\mu_1 - \mu_2) < 9.052$$

Intervals constructed in this manner will enclose $(\mu_1 - \mu_2)$ 95% of the time in repeated sampling. Hence, we are fairly certain that this particular interval encloses $(\mu_1 - \mu_2)$.

**8.45** **a** The point estimate of the difference $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 = 53,659 - 51,042 = 2617$$

and the margin of error is

$$1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx 1.96\sqrt{\frac{2225^2}{50} + \frac{2375^2}{50}} = 902.08$$

**b** Since the margin of error does not allow the estimate of the difference $\mu_1 - \mu_2$ to be negative—the lower limit is $2617 - 902.08 = 1714.92$ —it is likely that the mean for chemical engineering majors is larger than the mean for computer science majors.

**8.49** The 95% confidence interval for $\mu_1 - \mu_2$ is approximately

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(98.11 - 98.39) \pm 1.96\sqrt{\frac{.7^2}{65} + \frac{.74^2}{65}}$$

$$-.28 \pm .248 \quad \text{or} \quad -.528 < (\mu_1 - \mu_2) < -.032$$

**b** Since the confidence interval in part **a** has two negative endpoints, it does not contain the value $\mu_1 - \mu_2 = 0$. Hence, it is not likely that the means are equal. It appears that there is a real difference in the mean temperatures for males and females.

**8.54** Calculate $\hat{p}_1 = \frac{x_1}{995} = .41$ and $\hat{p}_2 = \frac{x_2}{1094} = .44$. The approximate 95% confidence interval is

$$(\hat{p}_1 - \hat{p}_2) \pm 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.41 - .44) \pm 1.96\sqrt{\frac{.41(.59)}{995} + \frac{.44(.56)}{1094}}$$

$$-.03 \pm .042 \quad \text{or} \quad -.072 < (p_1 - p_2) < .012$$

Since the value $p_1 - p_2 = 0$ is in the confidence interval, it is possible that $p_1 = p_2$. You should not conclude that there is a difference in the proportion of Republicans and Democrats who favor mentioned the economy as an important issue in the elections.

**8.56** **a** Calculate $\hat{p}_1 = \frac{410}{451} = .909$ and $\hat{p}_2 = \frac{505}{550} = .918$. The approximate 95% confidence interval is

$$(\hat{p}_1 - \hat{p}_2) \pm 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.909 - .918) \pm 1.96\sqrt{\frac{.909(.091)}{451} + \frac{.918(.082)}{550}}$$

$$-.009 \pm .035 \quad \text{or} \quad -.044 < (p_1 - p_2) < .026$$

Since the value $p_1 - p_2 = 0$ is in the confidence interval, it is possible that $p_1 = p_2$. You should not conclude that there is a difference in the proportion of fans versus non-fans who favor mandatory drug testing.

**8.62** **a** The point estimate for $p$ is given as $\hat{p} = \frac{x}{n} = \frac{23}{41} = .561$ and the margin of error is approximately

$$1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{.56(.44)}{41}} = .152$$

**b** Calculate $\hat{p}_1 = \frac{10}{32} = .3125$ and $\hat{p}_2 = \frac{23}{41} = .561$. The approximate 95% confidence interval is

$$(\hat{p}_1 - \hat{p}_2) \pm 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.3125 - .561) \pm 1.96\sqrt{\frac{.3125(.6875)}{32} + \frac{.561(.439)}{41}}$$

$$-.2485 \pm .2211 \quad \text{or} \quad -.4696 < (p_1 - p_2) < -.0274$$

**8.68** It is necessary to find the sample size required to estimate a certain parameter to within a given bound with confidence $(1-\alpha)$. Recall from Section 8.5 that we may estimate a parameter with $(1-\alpha)$ confidence within the interval (estimator) $\pm z_{\alpha/2} \times$ (std error of estimator). Thus, $z_{\alpha/2} \times$ (std error of estimator) provides the margin of error with $(1-\alpha)$ confidence. The experimenter will specify a given bound B. If we let $z_{\alpha/2} \times$ (std error of estimator) $\le$ B, we will be $(1-\alpha)$ confident that the estimator will lie within B units of the parameter of interest.

For this exercise, the parameter of interest is $\mu$, $B = 1.6$ and $1-\alpha = .95$. Hence, we must have

$$1.96\frac{\sigma}{\sqrt{n}} \le 1.6 \Rightarrow \quad 1.96\frac{12.7}{\sqrt{n}} \le 1.6$$

$$\sqrt{n} \ge \frac{1.96(12.7)}{1.6} = 15.5575$$

$$n \ge 242.04 \quad \text{or} \quad n \ge 243$$

**8.69** For this exercise, $B = .04$ for the binomial estimator $\hat{p}$, where $SE(\hat{p}) = \sqrt{\frac{pq}{n}}$. Assuming maximum variation, which occurs if $p = .3$ (since we suspect that $.1 < p < .3$) and $z_{.025} = 1.96$, we have

$$1.96\sigma_{\hat{p}} \le B \Rightarrow \quad 1.96\sqrt{\frac{pq}{n}} \le B$$

$$1.96\sqrt{\frac{.3(.7)}{n}} \le .04 \Rightarrow \quad \sqrt{n} \ge \frac{1.96\sqrt{.3(.7)}}{.04} \Rightarrow n \ge 504.21 \quad \text{or} \quad n \ge 505$$

**8.70** In this exercise, the parameter of interest is $\mu_1 - \mu_2$, $n_1 = n_2 = n$, and $\sigma_1^2 \approx \sigma_2^2 \approx 27.8$. Then we must have

$$z_{\alpha/2} \times (\text{std error of } \bar{x}_1 - \bar{x}_2) \le B$$

$$1.645\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le .17 \quad \Rightarrow \quad 1.645\sqrt{\frac{27.8}{n} + \frac{27.8}{n}} \le .17$$

$$\sqrt{n} \ge \frac{1.645\sqrt{55.6}}{.17} \quad \Rightarrow \quad n \ge 5206.06 \quad \text{or} \quad n_1 = n_2 = 5207$$

**8.74** Similar to Exercise 8.71.

$$z_{.025}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}} \le .03 \quad \Rightarrow \quad 1.96\sqrt{\frac{(.5)(.5)}{n}+\frac{(.5)(.5)}{n}} \le .03$$

$$\sqrt{n} \ge \frac{1.96\sqrt{.5}}{.03} \quad \Rightarrow \quad n \ge 2134.2 \quad \text{or} \quad n_1 = n_2 = 2135$$

**8.80** The parameter of interest is $\mu_1 - \mu_2$, the difference in grade-point averages for the two populations of students. Assume that $n_1 = n_2 = n$, and $\sigma_1^2 \approx \sigma_2^2 \approx (.6)^2 = .36$ and that the desired bound is .2. Then

$$1.96\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le .2 \quad \Rightarrow \quad 1.96\sqrt{\frac{.36}{n}+\frac{.36}{n}} \le .2$$

$$\sqrt{n} \ge \frac{1.96\sqrt{.72}}{.2} \quad \Rightarrow \quad n \ge 69.149$$

or $n_1 = n_2 = 70$ students should be included in each group.