Chapter 7 Introduction to Sampling Distributions

Section 7.1

1. Answers vary. Students should identify the individuals (subjects) and variable involved. Answers may include: A population is a set of measurements or counts either existing or conceptual. For example, the population of all ages of all people in Colorado; the population of weights of all students in your school; the population count of all antelope in Wyoming.

2. See Section 1.2. Answer may include:

   A simple random sample of n measurements from a population is a subset of the population selected in a manner such that

   (a) every sample of size n from the population has an equal chance of being selected and

   (b) every member of the population has an equal chance of being included in the sample.

3. A population parameter is a numerical descriptive measure of a population, such as μ, the population mean; σ, the population standard deviation; σ^2, the population variance; p, the population proportion; ρ (rho) the population correlation coefficient for those who have already studied linear regression from Chapter 10.

4. A sample statistic is a numerical descriptive measure of a sample such as $\bar{x}$, the sample mean; s, the sample standard deviation; s^2, the sample variance; $\hat{p}$, the sample proportion; r, the sample correlation coefficient for those who have already studied linear regression from Chapter 10.

5. A statistical inference is a conclusion about the value of a population parameter based on information about the corresponding sample statistic and probability. We will do both estimation and testing.

6. A sampling distribution is a probability distribution for a sample statistic.

7. They help us visualize the sampling distribution by using tables and graphs that approximately represent the sampling distribution.

8. Relative frequencies can be thought of as a measure or estimate of the likelihood of a certain statistic falling within the class bounds.

9. We studied the sampling distribution of mean trout lengths based on samples of size 5. Other such sampling distributions abound. Notice that the sample size remains the same for each sample in a sampling distribution.

Section 7.2

Note: Answers may vary slightly depending on the number of digits carried in the standard deviation.

1. (a) $\mu_{\bar{x}} = \mu = 15$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{14}{\sqrt{49}} = 2.0$$

Because $n = 49 \geq 30$, by the central limit theorem, we can assume that the distribution of $\bar{x}$ is approximately normal.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 15}{2.0}$$

$\bar{x} = 15$ converts to $z = \frac{15 - 15}{2.0} = 0$

$\bar{x} = 17$ converts to $z = \frac{17 - 15}{2.0} = 1$

$$\begin{aligned} P(15 \leq \bar{x} \leq 17) &= P(0 \leq z \leq 1) \\ &= P(z \leq 1) - P(z \leq 0) \\ &= 0.8413 - 0.5000 \\ &= 0.3413 \end{aligned}$$

(b) $\mu_{\bar{x}} = \mu = 15$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{14}{\sqrt{64}} = 1.75$$

Because $n = 64 \geq 30$, by the central limit theorem, we can assume that the distribution of $\bar{x}$ is approximately normal.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 15}{1.75}$$

$\bar{x} = 15$ converts to $z = \frac{15 - 15}{1.75} = 0$

$\bar{x} = 17$ converts to $z = \frac{17 - 15}{1.75} = 1.14$

$$\begin{aligned} P(15 \leq \bar{x} \leq 17) &= P(0 \leq z \leq 1.14) \\ &= P(z \leq 1.14) - P(z \leq 0) \\ &= 0.8729 - 0.5000 \\ &= 0.3729 \end{aligned}$$

(c) The standard deviation of part (b) is smaller because of the larger sample size. Therefore, the distribution about $\mu_{\bar{x}}$ is narrower in part (b).

2. (a) $\mu_{\bar{x}} = \mu = 100$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{48}{\sqrt{81}} = 5.33$$

Because $n = 81 \geq 30$, by the central limit theorem, we can assume that the distribution of $\bar{x}$ is approximately normal.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 100}{5.33}$$

$\bar{x} = 92$ converts to $z = \frac{92 - 100}{5.33} = -1.50$

$\bar{x} = 100$ converts to $z = \frac{100 - 100}{5.33} = 0$

$$\begin{aligned} P(92 \le \bar{x} \le 100) &= P(-1.50 \le z \le 0) \\ &= P(z \le 0) - P(z \le -1.50) \\ &= 0.5000 - 0.0668 \\ &= 0.4332 \end{aligned}$$

(b) $\mu_{\bar{x}} = \mu = 100$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{48}{\sqrt{121}} = 4.36$$

Because $n = 121 \ge 30$, by the central limit theorem, we can assume that the distribution of $\bar{x}$ is approximately normal.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 100}{4.36}$$

$\bar{x} = 92$ converts to $z = \frac{92 - 100}{4.36} = -1.83$

$\bar{x} = 100$ converts to $z = \frac{100 - 100}{4.36} = 0$

$$\begin{aligned} P(92 \le \bar{x} \le 100) &= P(-1.83 \le z \le 0) \\ &= P(z \le 0) - P(z \le -1.83) \\ &= 0.5000 - 0.0336 \\ &= 0.4664 \end{aligned}$$

(c) The probability of part (b) is greater than that of part (a). The standard deviation of part (b) is smaller because of the larger sample size. Therefore, the distribution about $\mu_{\bar{x}}$ is narrower in part (b).

3. (a) No, we cannot say anything about the distribution of sample means because the sample size is only 9 and so it is too small to apply the central limit theorem.

(b) Yes, now we can say that the $\bar{x}$ distribution will also be normal with

$$\mu_{\bar{x}} = \mu = 25 \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{9}} = 1.17.$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 25}{1.17}$$

$$\begin{aligned} P(23 \le \bar{x} \le 26) &= P\left(\frac{23 - 25}{1.17} \le z \le \frac{26 - 25}{1.17}\right) \\ &= P(-1.71 \le z \le 0.86) \\ &= P(z \le 0.86) - P(z \le -1.71) \\ &= 0.8051 - 0.0436 \\ &= 0.7615 \end{aligned}$$

4. (a) No, we cannot say anything about the distribution of sample means because the sample size is only 16 and so it is too small to apply the central limit theorem.

(b) Yes, now we can say that the $\overline{x}$ distribution will also be normal with

$$\mu_{\overline{x}} = \mu = 72 \text{ and } \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{16}} = 2.$$

$$z = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}} = \frac{\overline{x} - 72}{2}$$

$$\begin{aligned} P(68 \le \overline{x} \le 73) &= P\left(\frac{68-72}{2} \le z \le \frac{73-72}{2}\right) \\ &= (-2 \le z \le 0.5) \\ &= P(z \le 0.5) - P(z \le -2) \\ &= 0.6915 - 0.0228 \\ &= 0.6687 \end{aligned}$$

5. (a) $\mu = 75,\ \sigma = 0.8$

$$\begin{aligned} P(x < 74.5) &= P\left(z < \frac{74.5-75}{0.8}\right) \\ &= P(z < -0.63) \\ &= 0.2643 \end{aligned}$$

(b) $\mu_{\overline{x}} = 75,\ \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.8}{\sqrt{20}} = 0.179$

$$\begin{aligned} P(\overline{x} < 74.5) &= P\left(z < \frac{74.5-75}{0.179}\right) \\ &= P(z < -2.79) \\ &= 0.0026 \end{aligned}$$

(c) No. If the weight of only one car were less than 74.5 tons, we cannot conclude that the loader is out of adjustment. If the mean weight for a sample of 20 cars were less than 74.5 tons, we would suspect that the loader is malfunctioning. As we see in part (b), the probability of this happening is very low if the loader is correctly adjusted.

6. (a) $\mu = 68,\ \sigma = 3$

$$\begin{aligned} P(67 \le x \le 69) &= P\left(\frac{67-68}{3} \le z \le \frac{69-68}{3}\right) \\ &= P(-0.33 \le z \le 0.33) \\ &= P(z \le 0.33) - P(z \le -0.33) \\ &= 0.6293 - 0.3707 \\ &= 0.2586 \end{aligned}$$

(b) $\mu_{\bar{x}} = 68,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{9}} = 1$

$$\begin{aligned} P(67 \le \bar{x} \le 69) &= P\left(\frac{67-68}{1} \le z \le \frac{69-68}{1}\right) \\ &= P(-1 \le z \le 1) \\ &= P(z \le 1) - P(z \le -1) \\ &= 0.8413 - 0.1587 \\ &= 0.6826 \end{aligned}$$

(c) The probability in part (b) is much higher because the standard deviation is smaller for the $\bar{x}$ distribution.

7. (a) $\mu = 85,\ \sigma = 25$

$$\begin{aligned} P(x < 40) &= P\left(z < \frac{40-85}{25}\right) \\ &= P(z < -1.8) \\ &= 0.0359 \end{aligned}$$

(b) The probability distribution of $\bar{x}$ is approximately normal with $\mu_{\bar{x}} = 85$; $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{2}} = 17.68$.

$$\begin{aligned} P(\bar{x} < 40) &= P\left(z < \frac{40-85}{17.68}\right) \\ &= P(z < -2.55) \\ &= 0.0054 \end{aligned}$$

(c) $\mu_{\bar{x}} = 85,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{3}} = 14.43$

$$\begin{aligned} P(\bar{x} < 40) &= P\left(z < \frac{40-85}{14.43}\right) \\ &= P(z < -3.12) \\ &= 0.0009 \end{aligned}$$

(d) $\mu_{\bar{x}} = 85,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{5}} = 11.2$

$$\begin{aligned} P(\bar{x} < 40) &= P\left(z < \frac{40-85}{11.2}\right) \\ &= P(z < -4.02) \\ &\approx 0 \end{aligned}$$

(e) Yes; If the average value based on five tests were less than 40, the patient is almost certain to have excess insulin.

8. $\mu = 7500,\ \sigma = 1750$

(a) $P(x < 3500) = P\left(z < \frac{3500 - 7500}{1750}\right)$
$= P(z < -2.29)$
$= 0.0110$

(b) The probability distribution of $\bar{x}$ is approximately normal with

$\mu_{\bar{x}} = 7500;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1750}{\sqrt{2}} = 1237.44.$

$P(\bar{x} < 3500) = P\left(z < \frac{3500 - 7500}{1237.44}\right)$
$= P(z < -3.23)$
$= 0.0006$

(c) $\mu_{\bar{x}} = 7500,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1750}{\sqrt{3}} = 1010.36$

$P(\bar{x} < 3500) = P\left(z < \frac{3500 - 7500}{1010.36}\right)$
$= P(z < -3.96)$
≈ 0

(d) The probabilities decreased as n increased. It would be an extremely rare event for a person to have two or three tests below 3500 purely by chance; the person probably has leukopenia.

9. **(a)** $\mu = 63.0,\ \sigma = 7.1$

$P(x < 54) = P\left(z < \frac{54 - 63.0}{7.1}\right)$
$= P(z < -1.27)$
$= 0.1020$

(b) The expected number undernourished is 2200(0.1020) = 224.4, or about 224.

(c) $\mu_{\bar{x}} = 63.0,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7.1}{\sqrt{50}} = 1.004$

$P(\bar{x} < 60) = P\left(z < \frac{60 - 63.0}{1.004}\right)$
$= P(z < -2.99)$
$= 0.0014$

(d) $\mu_{\bar{x}} = 63.0,\ \sigma_{\bar{x}} = 1.004$

$P(\bar{x} < 64.2) = P\left(z < \frac{64.2 - 63.0}{1.004}\right)$
$= P(z < 1.20)$
$= 0.8849$

Since the sample average is above the mean, it is quite unlikely that the doe population is undernourished.

10. **(a)** From the Central Limit Theorem, we expect the $\bar{x}$ distribution to be approximately normal with the mean $\mu_{\bar{x}} = \mu = 16$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{30}} = 0.3651$.

(b) $\mu_{\bar{x}} = 16,\ \sigma_{\bar{x}} = 0.3651$

$$\begin{aligned} P(16 \le \bar{x} \le 17) &= P\left(\frac{16-16}{0.3651} \le z \le \frac{17-16}{0.3651}\right) \\ &= P(0 \le z \le 2.74) \\ &= P(z \le 2.74) - P(z \le 0) \\ &= 0.9969 - 0.5000 \\ &= 0.4969 \end{aligned}$$

(c) $\mu_{\bar{x}} = 16,\ \sigma_{\bar{x}} = 0.3651$

$$\begin{aligned} P(\bar{x} < 15) &= P\left(z < \frac{15-16}{0.3651}\right) \\ &= P(z < -2.74) \\ &= 0.0031 \end{aligned}$$

11. **(a)** The random variable x is itself an average based on the number of stocks or bonds in the fund. Since x itself represents a sample mean return based on a large (random) sample of stocks or bonds, x has a distribution that is approximately normal (Central Limit Theorem).

(b) $\mu_{\bar{x}} = 1.6\%,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.9\%}{\sqrt{6}} = 0.367\%$

$$\begin{aligned} P(1\% \le \bar{x} \le 2\%) &= P\left(\frac{1\% - 1.6\%}{0.367\%} \le z \le \frac{2\% - 1.6\%}{0.367\%}\right) \\ &= P(-1.63 \le z \le 1.09) \\ &= P(z \le 1.09) - P(z \le -1.63) \\ &= 0.8621 - 0.0516 \\ &= 0.8105 \end{aligned}$$

Note: It does not matter whether you solve the problem using percents or their decimal equivalents as long as you are consistent.

(c) Note: 2 years = 24 months; x is <u>monthly</u> percentage return.

$\mu_{\bar{x}} = 1.6\%,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.9\%}{\sqrt{24}} = 0.1837\%$

$$\begin{aligned} P(1\% \le \bar{x} \le 2\%) &= P\left(\frac{1\% - 1.6\%}{0.1837\%} \le z \le \frac{2\% - 1.6\%}{0.1837\%}\right) \\ &= P(-3.27 \le z \le 2.18) \\ &= P(z \le 2.18) - P(z \le -3.27) \\ &= 0.9854 - 0.0005 \\ &= 0.9849 \end{aligned}$$

(d) Yes. The probability increases as the standard deviation decreases. The standard deviation decreases as the sample size increases.

(e) $\mu_{\bar{x}} = 1.6\%$, $\sigma_{\bar{x}} = 0.1837\%$

$$P(\bar{x} < 1\%) = P\left(z < \frac{1\% - 1.6\%}{0.1837\%}\right) = P(z < -3.27) = 0.0005$$

This is very unlikely if $\mu = 1.6\%$. One would suspect that μ has slipped below 1.6%.

12. (a) The random variable x is itself an average based on the number of stocks in the fund. Since x itself represents a sample mean return based on a large (random) sample of stocks, x has a distribution that is approximately normal (Central Limit Theorem).

(b) $\mu_{\bar{x}} = 1.4\%$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.8\%}{\sqrt{9}} = 0.2667\%$

$$P(1\% \le \bar{x} \le 2\%) = P\left(\frac{1\% - 1.4\%}{0.2667\%} \le z \le \frac{2\% - 1.4\%}{0.2667\%}\right) = P(-1.50 \le z \le 2.25) = P(z \le 2.25) - P(z \le -1.50) = 0.9878 - 0.0668 = 0.9210$$

Note: It does not matter whether you solve the problem using percents or their decimal equivalents as long as you are consistent.

(c) $\mu_{\bar{x}} = 1.4\%$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.8\%}{\sqrt{18}} = 0.1886\%$

$$P(1\% \le \bar{x} \le 2\%) = P\left(\frac{1\% - 1.4\%}{0.1886\%} \le z \le \frac{2\% - 1.4\%}{0.1886\%}\right) = P(-2.12 \le z \le 3.18) = P(z \le 3.18) - P(z \le -2.12) = 0.9993 - 0.0170 = 0.9823$$

(d) Yes. The probability increases as the standard deviation decreases. The standard deviation decreases as the sample size increases.

(e) $\mu_{\bar{x}} = 1.4\%$, $\sigma_{\bar{x}} = 0.1886\%$

$$P(\bar{x} > 2\%) = P\left(z > \frac{2\% - 1.4\%}{0.1886\%}\right) = P(z > 3.18) = 1 - P(z \le 3.18) = 1 - 0.9993 = 0.0007$$

This is very unlikely if $\mu = 1.4\%$. One would suspect that the European stock market may be heating up, i.e., μ is greater than 1.4%.

13. (a) Since x itself represents a sample mean from a large $n \approx 80$ (random) sample of bonds, x is approximately normally distributed according to the Central Limit Theorem.

(b) $\mu_{\bar{x}} = 10.8\%,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4.9\%}{\sqrt{5}} = 2.19\%$

$$\begin{aligned} P(\bar{x} < 6\%) &= P\left(z < \frac{6\% - 10.8\%}{2.19\%}\right) \\ &= P(z < -2.19) \\ &= 0.0143 \end{aligned}$$

Yes. Since this probability is so small, it is very unlikely that $\bar{x}$ would be less than 6% if μ = 10.8%. The junk bond market appears to be weaker, i.e., μ is less than 10.8%.

(c) $\mu_{\bar{x}} = 10.8\%,\ \sigma_{\bar{x}} = 2.19\%$

$$\begin{aligned} P(\bar{x} > 16\%) &= P\left(z > \frac{16\% - 10.8\%}{2.19\%}\right) \\ &= P(z > 2.37) \\ &= 1 - P(z \le 2.37) \\ &= 1 - 0.9911 \\ &= 0.0089 \end{aligned}$$

Yes. Since this probability is so small, it is very unlikely that $\bar{x}$ would be greater than 16% if μ = 10.8%. The junk bond market may be heating up, i.e., μ is greater than 10.8%.

14. (a) $\mu_{\bar{x}} = 6.4,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{40}} = 0.2372$

$$\begin{aligned} P(6 \le \bar{x} \le 7) &= P\left(\frac{6 - 6.4}{0.2372} \le z \le \frac{7 - 6.4}{0.2372}\right) \\ &= P(-1.69 \le z \le 2.53) \\ &= P(z \le 2.53) - P(z \le -1.69) \\ &= 0.9943 - 0.0455 \\ &= 0.9488 \end{aligned}$$

(b) $\mu_{\bar{x}} = 6.4,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{80}} = 0.1677$

$$\begin{aligned} P(6 \le \bar{x} \le 7) &= P\left(\frac{6 - 6.4}{0.1677} \le z \le \frac{7 - 6.4}{0.1677}\right) \\ &= P(-2.39 \le z \le 3.58) \\ &= P(z \le 3.58) - P(z \le -2.39) \\ &\approx 1 - 0.0084 \\ &= 0.9916 \end{aligned}$$

(c) Yes. Since this is such a large probability, the chances of $\bar{x}$ not being in this time interval is extremely unlikely. A second security guard should drop in for a look.

15. (a) The sample size should be 30 or more.

(b) No. If the distribution of x is normal, the distribution of $\bar{x}$ is also normal, regardless of the sample size.

16. (a) By the Central Limit Theorem, the sampling distribution of $\bar{x}$ is approximately normal with mean $\mu_{\bar{x}} = \mu = \$20$ and standard error $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\$7}{\sqrt{100}} = \$0.70$. It is not necessary to make any assumption about the x distribution because n is large.

(b) $\mu_{\bar{x}} = \$20,\ \sigma_{\bar{x}} = \$0.70$

$$\begin{aligned} P(\$18 \le \bar{x} \le \$22) &= P\left(\frac{\$18-\$20}{\$0.70} \le z \le \frac{\$22-\$20}{\$0.70}\right) \\ &= P(-2.86 \le z \le 2.86) \\ &= P(z \le 2.86) - P(z \le -2.86) \\ &= 0.9979 - 0.0021 \\ &= 0.9958 \end{aligned}$$

(c) $\mu_x = \$20,\ \sigma = \$7$

$$\begin{aligned} P(\$18 \le x \le \$22) &= P\left(\frac{\$18-\$20}{\$7} \le z \le \frac{\$22-\$20}{\$7}\right) \\ &= P(-0.29 \le z \le 0.29) \\ &= 0.6141 - 0.3859 \\ &= 0.2282 \end{aligned}$$

(d) We expect the probability in part (b) to be much higher than the probability in part (c) because the standard deviation is smaller for the $\bar{x}$ distribution than it is for the x distribution. By the Central Limit Theorem, the sampling distribution of $\bar{x}$ will be approximately normal as n increases, and its standard deviation, $\sigma/\sqrt{n}$, will decrease as n increases. The standard deviation of $\bar{x}$, a.k.a. the standard error of $\bar{x}$, measures the spread of the $\bar{x}$ values; the smaller $\sigma/\sqrt{n}$ is, the less variability there is in the $\bar{x}$ values. The less variability there is in the values of $\bar{x}$, the more reliable $\bar{x}$ is as an estimate or predictor of μ. For large n, approximately 95% of the possible values of $\bar{x}$ are within $2\sigma/\sqrt{n}$ of μ. The amount x a typical customer spends on impulse buys also estimates μ (recall $\mu_{\bar{x}} = \mu_x = \mu$), but approximately 95% of individual impulse buys x are within 2σ of μ (using either the Empirical Rule for somewhat mound-shaped data, or assuming x has a distribution that is approximately normal). For a fixed interval, such as \$18 to \$22, centered at the mean, \$20 in this case, the proportion of the possible $\bar{x}$ values within the interval will be greater than the proportion of the possible x values within the same interval.

17. (a) The total checkout time for 30 customers is the sum of the checkout times for each individual customer. Thus, $w = x_1 + x_2 + \cdots + x_{30}$ and the probability that the total checkout time for the next 30 customers is less than 90 is $P(w < 90)$.

(b) If we divide both sides of $w < 90$ by 30, we get $\frac{w}{30} < 3$. However, w is the sum of 30 waiting times, so $\frac{w}{30}$ is $\bar{x}$. Therefore, $P(w < 90) = P(\bar{x} < 3)$.

(c) The probability distribution of $\bar{x}$ is approximately normal with mean $\mu_{\bar{x}} = \mu = 2.7$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.6}{\sqrt{30}} = 0.1095$.

(d) $P(\bar{x} < 3) = P\left(z < \frac{3-2.7}{0.1095}\right)$

$= P(z < 2.74)$

$= 0.9969$

The probability that the total checkout time for the next 30 customers is less than 90 minutes is 0.9969, i.e., $P(w < 90) = 0.9969$.

18. Let $w = x_1 + x_2 + \dots + x_{36}$.

(a) $w < 320$ is equivalent to $\frac{w}{36} < \frac{320}{36}$ or $\bar{x} < 8.889$. $\mu_{\bar{x}} = \mu = 8.5$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{36}} = 0.4167$

$P(w < 320) = P(\bar{x} < 8.889)$

$= P\left(z < \frac{8.889-8.5}{0.4167}\right)$

$= P(z < 0.93)$

$= 0.8238$

(b) $w > 275$ is equivalent to $\frac{w}{36} > \frac{275}{36}$ or $\bar{x} > 7.639$. $\mu_{\bar{x}} = 8.5$, $\sigma_{\bar{x}} = 0.4167$

$P(w > 275) = P(\bar{x} > 7.639)$

$= P\left(z > \frac{7.639-8.5}{0.4167}\right)$

$= P(z > -2.07)$

$= 1 - P(z \le -2.07)$

$= 1 - 0.0192$

≈ 0.9808

(c) $P(275 < w < 320) = P(7.639 < \bar{x} < 8.889)$

$= P(-2.07 < z < 0.93)$

$= P(z < 0.93) - P(z < -2.07)$

$= 0.8238 - 0.0192$

$= 0.8046$

19. Let $w = x_1 + x_2 + \dots + x_{45}$.

(a) $w < 9500$ is equivalent to $\frac{w}{45} < \frac{9500}{45}$ or $\bar{x} < 211.111$. $\mu_{\bar{x}} = 240$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{84}{\sqrt{45}} = 12.522$

$P(w < 9500) = P(\bar{x} < 211.111)$

$= P\left(z < \frac{211.111-240}{12.522}\right)$

$= P(z < -2.31)$

$= 0.0104$

(b) $w < 12{,}000$ is equivalent to $\frac{w}{45} > \frac{12{,}000}{45}$ or $\bar{x} > 266.667$. $\mu_{\bar{x}} = 240$, $\sigma_{\bar{x}} = 12.522$

$$\begin{aligned} P(w > 12{,}000) &= P(\bar{x} > 266.667) \\ &= P\left(z > \frac{266.667 - 240}{12.522}\right) \\ &= P(z > 2.13) \\ &= 1 - P(z \le 2.13) \\ &= 1 - 0.9834 \\ &= 0.0166 \end{aligned}$$

(c)
$$\begin{aligned} P(9500 < w < 12{,}000) &= P(211.111 < \bar{x} < 266.667) \\ &= P(-2.31 < z < 2.13) \\ &= P(z < 2.13) - P(z < -2.31) \\ &= 0.9834 - 0.0104 \\ &= 0.9730 \end{aligned}$$

20. (a) Let $w = x_1 + x_2 + \ldots + x_9$. $\mu_{\bar{x}} = \mu = 6.3$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.2}{\sqrt{9}} = 0.4$

$$\begin{aligned} P(w < 60) &= P\left(\frac{w}{9} < \frac{60}{9}\right) \\ &= P(\bar{x} < 6.667) \\ &= P\left(z < \frac{6.667 - 6.3}{0.4}\right) \\ &= P(z < 0.92) \\ &= 0.8212 \end{aligned}$$

$$\begin{aligned} P(w > 65) &= P\left(\frac{w}{9} > \frac{65}{9}\right) \\ &= P(\bar{x} > 7.222) \\ &= P\left(z > \frac{7.222 - 6.3}{0.4}\right) \\ &= P(z > 2.31) \\ &= 1 - P(z \le 2.31) \\ &= 1 - 0.9896 \\ &= 0.0104 \end{aligned}$$

(b) Let $w = x_1 + x_2 + \ldots + x_{50}$. $\mu_{\bar{x}} = \mu = 6.3$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.2}{\sqrt{50}} = 0.170$

$$P(w < 342) = P\left(\frac{w}{50} < \frac{342}{50}\right)$$

$$\begin{aligned} &P(\bar{x} < 6.84) \\ &= P\left(z < \frac{6.84 - 6.3}{0.170}\right) \\ &= P(z < 3.18) \\ &= 0.9993 \end{aligned}$$

No. By the Central Limit Theorem the sample size is large enough so the sampling distribution of $\bar{x}$ is approximately normal.

Section 7.3

1. (a) Answers vary.

(b) The random variable $\hat{p}$ can be approximated by a normal random variable when both np and nq exceed 5.

$$\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

(c) $np = 33(0.21) = 6.93$, $nq = 33(0.79) = 26.07$

Yes, $\hat{p}$ can be approximated by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.21, \sigma_{\hat{p}} = \sqrt{\frac{0.21(0.79)}{33}} \approx 0.071$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{33} \approx 0.015$$

$$\begin{aligned} P(0.15 \le \hat{p} \le 0.25) &= P(0.15 - 0.015 \le x \le 0.25 + 0.015) \\ &= P(0.135 \le x \le 0.265) \\ &= P\left(\frac{0.135 - 0.21}{0.071} \le z \le \frac{0.265 - 0.21}{0.071}\right) \\ &= P(-1.06 \le z \le 0.77) \\ &= P(z \le 0.77) - P(z \le -1.06) \\ &= 0.7794 - 0.1446 \\ &= 0.6348 \end{aligned}$$

(d) No: $np = 25(0.15) = 3.75$ which does not exceed 5.

(e) $np = 48(0.15) = 7.2,\ nq = 48(0.85) = 40.8$

Yes, $\hat{p}$ can be approximated by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.15,\ \sigma_{\hat{p}} = \sqrt{\frac{0.15(0.85)}{48}} \approx 0.052$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{45} = 0.010$$

$$\begin{aligned} P(\hat{p} \ge 0.22) &= P(x \ge 0.22 - 0.010) \\ &= P(x \ge 0.21) \\ &= P\left(z \ge \frac{0.21 - 0.15}{0.052}\right) \\ &= P(z \ge 1.15) \\ &= 1 - P(z < 1.15) \\ &= 1 - 0.8749 \\ &= 0.1251 \end{aligned}$$

2. (a) $n = 50,\ p = 0.36$

$np = 50(0.36) = 18,\ nq = 50(0.64) = 32$

Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.36,\ \sigma_{\hat{p}} = \sqrt{\frac{0.36(0.64)}{50}} \approx 0.068$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{50} = 0.01$$

$$\begin{aligned} P(0.30 \le \hat{p} \le 0.45) &\approx P(0.30 - 0.01 \le x \le 0.45 + 0.01) \\ &= P(0.29 \le x \le 0.46) \\ &= P\left(\frac{0.29 - 0.36}{0.068} \le z \le \frac{0.46 - 0.36}{0.068}\right) \\ &= P(-1.03 \le z \le 1.47) \\ &= P(z \le 1.47) - P(z \le -1.03) \\ &= 0.9292 - 0.1515 \\ &= 0.7777 \end{aligned}$$

(b) $n = 38$, $p = 0.25$

$np = 38(0.25) = 9.5$, $nq = 38(0.75) = 28.5$

Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.25, \sigma_{\hat{p}} = \sqrt{\frac{0.25(0.75)}{38}} \approx 0.070$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{38} = 0.013$$

$$\begin{aligned} P(\hat{p} > 0.35) &= P(x > 0.35 - 0.013) \\ &= P(x > 0.337) \\ &= P\left(z > \frac{0.337 - 0.25}{0.070}\right) \\ &= P(z > 1.24) \\ &= 1 - P(z \le 1.24) \\ &= 1 - 0.8925 \\ &= 0.1075 \end{aligned}$$

(c) $n = 41$, $p = 0.09$

$np = 41(0.09) = 3.69$

We cannot approximate $\hat{p}$ by a normal random variable since $np < 5$.

3. $n = 30$, $p = 0.60$

$np = 30(0.60) = 18$, $nq = 30(0.40) = 12$

Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.6, \sigma_{\hat{p}} = \sqrt{\frac{0.6(0.4)}{30}} \approx 0.089$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{30} = 0.017$$

(a) $$\begin{aligned} P(\hat{p} \ge 0.5) &\approx P(x \ge 0.5 - 0.017) \\ &= P(x \ge 0.483) \\ &= P\left(z \ge \frac{0.483 - 0.6}{0.089}\right) \\ &= P(z \ge -1.31) \\ &= 0.9049 \end{aligned}$$

(b) $$\begin{aligned} P(\hat{p} \ge 0.667) &\approx P(x \ge 0.667 - 0.017) \\ &= P(x \ge 0.65) \\ &= P\left(z \ge \frac{0.65 - 0.6}{0.089}\right) \\ &= P(z \ge 0.56) \\ &= 0.2877 \end{aligned}$$

(c) $P(\hat{p} \le 0.333) \approx P(x \le 0.333 + 0.017)$
$= P(x \le 0.35)$
$= P\left(z \le \frac{0.35 - 0.6}{0.089}\right)$
$= P(z \le -2.81)$
$= 0.0025$

(d) Yes, both np and nq exceed 5.

4. (a) $n = 38,\ p = 0.73$
$np = 38(0.73) = 27.74,\ nq = 38(0.27) = 10.26$
Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.73,\ \sigma_{\hat{p}} = \sqrt{\frac{0.73(0.27)}{38}} \approx 0.072$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{38} = 0.013$$

$$P(\hat{p} \ge 0.667) \approx P(x \ge 0.667 - 0.013)$$
$$= P(x \ge 0.654)$$
$$= P\left(z \ge \frac{0.654 - 0.73}{0.072}\right)$$
$$= P(z \ge -1.06)$$
$$= 0.8554$$

(b) $n = 45,\ p = 0.86$
$np = 45(0.86) = 38.7,\ nq = 45(0.14) = 6.3$
Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.86,\ \sigma_{\hat{p}} = \sqrt{\frac{0.86(0.14)}{45}} \approx 0.052$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{45} = 0.011$$

$$P(\hat{p} \ge 0.667) \approx P(x \ge 0.667 - 0.011)$$
$$= P(x \ge 0.656)$$
$$= P\left(z \ge \frac{0.656 - 0.86}{0.052}\right)$$
$$= P(z \ge -3.92)$$
$$\approx 1$$

(c) Yes, both np and nq exceed 5 for men and for women.

5. $n = 55,\ p = 0.11$

$np = 55(0.11) = 6.05,\ nq = 55(0.89) = 48.95$

Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.11,\ \sigma_{\hat{p}} = \sqrt{\frac{0.11(0.89)}{55}} \approx 0.042$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{55} = 0.009$$

(a) $$\begin{aligned} P(\hat{p} \le 0.15) &\approx P(x \le 0.15 + 0.009) \\ &= P(x \le 0.159) \\ &= P\left(z \le \frac{0.159 - 0.11}{0.042}\right) \\ &= P(z \le 1.17) \\ &= 0.8790 \end{aligned}$$

(b) $$\begin{aligned} P(0.10 \le \hat{p} \le 0.15) &\approx P(0.10 - 0.009 \le x \le 0.15 + 0.009) \\ &= P(0.091 \le x \le 0.159) \\ &= P\left(\frac{0.091 - 0.11}{0.042} \le z \le \frac{0.159 - 0.11}{0.042}\right) \\ &= P(-0.45 \le z \le 1.17) \\ &= P(z \le 1.17) - P(z \le -0.45) \\ &= 0.8790 - 0.3264 \\ &= 0.5526 \end{aligned}$$

(c) Yes, both np and nq exceed 5.

6. $n = 28,\ p = 0.31$

$np = 28(0.31) = 8.68,\ nq = 28(0.69) = 19.32$

Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.31,\ \sigma_{\hat{p}} = \sqrt{\frac{0.31(0.69)}{28}} \approx 0.087$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{28} = 0.018$$

(a) $$\begin{aligned} P(\hat{p} \ge 0.25) &\approx P(x \ge 0.25 - 0.018) \\ &= P(x \ge 0.232) \\ &= P\left(z \ge \frac{0.232 - 0.31}{0.087}\right) \\ &= P(z \ge -0.90) \\ &= 0.8159 \end{aligned}$$

(b) $P(0.25 \le \hat{p} \le 0.50) \approx P(0.25 - 0.018 \le x \le 0.50 + 0.018)$
$= P(0.232 \le x \le 0.518)$
$= P\left(\frac{0.232 - 0.31}{0.087} \le z \le \frac{0.518 - 0.31}{0.087}\right)$
$= P(-0.90 \le z \le 2.39)$
$= P(z \le 2.39) - P(z \le -0.90)$
$= 0.9916 - 0.1841$
$= 0.8075$

(c) Yes, both np and nq exceed 5.

7. (a) $n = 100$, $p = 0.06$
$np = 100(0.06) = 6$, $nq = 100(0.94) = 94$
$\hat{p}$ can be approximated by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.06,\ \sigma_{\hat{p}} = \sqrt{\frac{0.06(0.94)}{100}} \approx 0.024$$

$$\text{continuity correction} = \frac{0.5}{100} = 0.005$$

(b) $P(\hat{p} \ge 0.07) \approx P(x \ge 0.07 - 0.005)$
$= P(x \ge 0.065)$
$= P\left(z \ge \frac{0.065 - 0.06}{0.024}\right)$
$= P(z \ge 0.21)$
$= 0.4168$

(c) $P(\hat{p} \ge 0.11) \approx P(x \ge 0.11 - 0.005)$
$= P(x \ge 0.105)$
$= P\left(z \ge \frac{0.105 - 0.06}{0.024}\right)$
$= P(z \ge 1.88)$
$= 0.0301$

Yes, since this probability is so small, it should rarely occur. The machine might need an adjustment.

8. (a) $n = 50$, $p = 0.565$
$np = 50(0.565) = 28.25$, $nq = 50(0.435) = 21.75$
$\hat{p}$ can be approximated by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.565,\ \sigma_{\hat{p}} = \sqrt{\frac{0.565(0.435)}{50}} \approx 0.070$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{50} = 0.01$$

(b) $$\begin{aligned} P(\hat{p} \le 0.53) &\approx P(x \le 0.53 + 0.01) \\ &= P(x \le 0.54) \\ &= P\left(z \le \frac{0.54 - 0.565}{0.070}\right) \\ &= P(z \le -0.36) \\ &= 0.3594 \end{aligned}$$

(c) $$\begin{aligned} P(\hat{p} \le 0.41) &\approx P(x \le 0.41 + 0.01) \\ &= P(x \le 0.42) \\ &= P\left(z \le \frac{0.42 - 0.565}{0.070}\right) \\ &= P(z \le -2.07) \\ &= 0.0192 \end{aligned}$$

(d) Meredith has the more serious case because the probability of having such a low reading in a healthy person is less than 2%.

9. $$\begin{aligned} \bar{p} &= \frac{\text{total number of successes from all 12 quarters}}{\text{total number of families from all 12 quarters}} \\ &= \frac{11 + 14 + \ldots + 19}{12(92)} \\ &= \frac{206}{1104} \\ &= 0.1866 \end{aligned}$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.1866 = 0.8134$$

$$\mu_{\hat{p}} = p = \bar{p} = 0.1866$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{0.1866(0.8134)}{92}} \approx 0.0406$$

Check: $n\bar{p} = 92(0.1866) = 17.2$, $n\bar{q} = 92(0.8134) = 74.8$

Since both $n\bar{p}$ and $n\bar{q}$ exceed 5. the normal approximation should be reasonably good.

Center line $= \bar{p} = 0.1866$

Control limits at $\bar{p} \pm 2\sqrt{\frac{\bar{p}\bar{q}}{n}}$

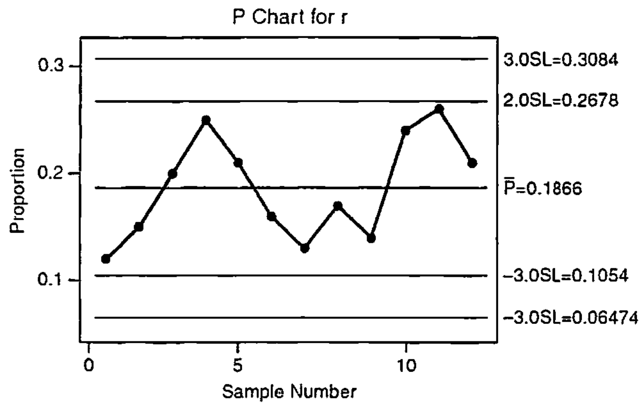
$$\begin{aligned} &= 0.1866 \pm 2(0.0406) \\ &= 0.1866 \pm 0.0812 \end{aligned}$$

or 0.1054 and 0.2678

Control limits at $\bar{p} \pm 3\sqrt{\frac{\bar{p}\bar{q}}{n}}$

$$\begin{aligned} &= 0.1866 \pm 3(0.0406) \\ &= 0.1866 \pm 0.1218 \end{aligned}$$

or 0.0648 and 0.3084



There are no out-of-control signals.

10. $$\bar{p} = \frac{\text{total number of defective cans}}{\text{total number of cans}} = \frac{8+11+\ldots+10}{110(15)} = \frac{133}{1650} = 0.08061$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.08061 = 0.91939$$

$$\mu_{\hat{p}} = p \approx \bar{p} = 0.08061$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{(0.08061)(0.91939)}{110}} \approx 0.02596$$

Check: $n\bar{p} = 110(0.08061) = 8.9$, $n\bar{q} = 110(0.91939) = 101.1$

Since both $n\bar{p}$ and $n\bar{q}$ exceed 5, the normal approximation should be reasonably good.

Center line = $\bar{p} = 0.08061$

Control limits at $\bar{p} \pm 2\sqrt{\frac{\bar{p}\bar{q}}{n}}$

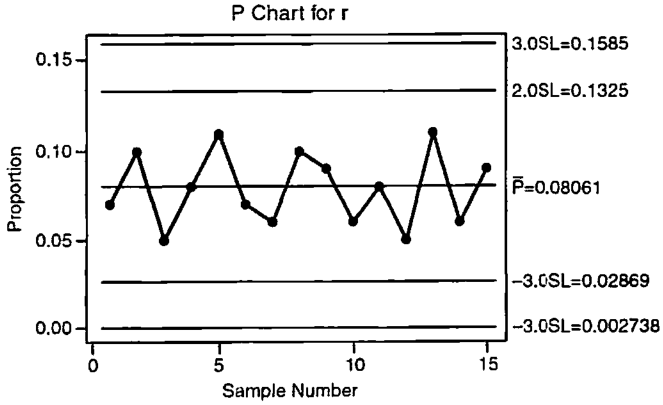
$$= 0.08061 \pm 2(0.02596)$$
$$= 0.08061 \pm 0.05192$$

or 0.02869 and 0.1325

Control limits at $\bar{p} \pm 3\sqrt{\frac{\bar{p}\bar{q}}{n}}$

$$= 0.08061 \pm 3(0.02596)$$
$$= 0.08061 \pm 0.07788$$

or 0.00273 and 0.1585



There are no out-of-control signals. It appears that the production process is in reasonable control.

11. $$\bar{p} = \frac{\text{total number who got jobs}}{\text{total number of people}}$$
$$= \frac{60 + 53 + \ldots + 58}{75(15)}$$
$$= \frac{872}{1125}$$
$$= 0.7751$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.7751 = 0.2249$$
$$\mu_{\hat{p}} = p \approx \bar{p} = 0.7751$$
$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{(0.7751)(0.2249)}{75}} \approx 0.0482$$

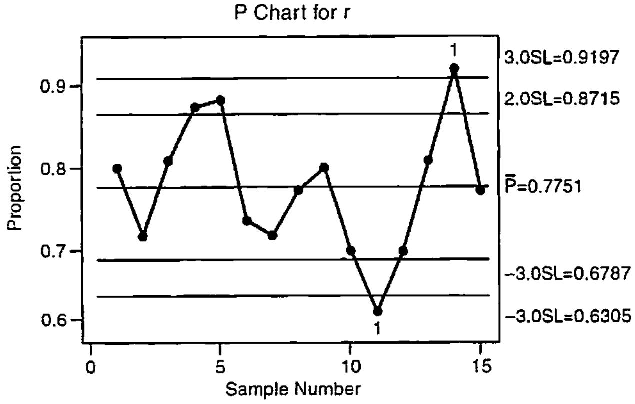
Check: $n\bar{p} = 75(0.7751) = 58.1$, $n\bar{q} = 75(0.2249) = 16.9$

Since both $n\bar{p}$ and $n\bar{q}$ exceed 5, the normal approximation should be reasonably good.

Center line = $\bar{p} = 0.7751$

Control limits at $\bar{p} \pm 2\sqrt{\frac{\bar{p}\bar{q}}{n}}$
$$= 0.7751 \pm 2(0.0482)$$
$$= 0.7751 \pm 0.0964$$
or 0.6787 to 0.8715

Control limits at $\bar{p} \pm 3\sqrt{\frac{\bar{p}\bar{q}}{n}}$
$$= 0.7751 \pm 3(0.0482)$$
$$= 0.7751 \pm 0.1446$$
or 0.6305 to 0.9197



Out-of-control signal III occurs on days 4 and 5. Out-of-control signal I occurs on day 11 on the low side and day 14 on the high side. Out-of-control signals on the low side are of most concern for the homeless seeking work. The foundation should look to see what happened on that day. The foundation might take a look at the out of control periods on the high side to see if there is a possibility of cultivating more jobs.

Chapter 7 Review

1. **(a)** The $\bar{x}$ distribution approaches a normal distribution.

(b) The mean $\mu_{\bar{x}}$ of the $\bar{x}$ distribution equals the mean μ of the x distribution, regardless of the sample size.

(c) The standard deviation $\sigma_{\bar{x}}$ of the sampling distribution equals $\frac{\sigma}{\sqrt{n}}$, where σ is the standard deviation of the x distribution and n is the sample size.

(d) They will both be approximately normal with the same mean, but the standard deviations will be $\frac{\sigma}{\sqrt{50}}$ and $\frac{\sigma}{\sqrt{100}}$ respectively.

2. All the $\bar{x}$ distributions will be normal with mean $\mu_{\bar{x}} = \mu = 15$. The standard deviations will be:

$$n = 4:\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{4}} = \frac{3}{2}$$

$$n = 16:\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{16}} = \frac{3}{4}$$

$$n = 100:\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{100}} = \frac{3}{10}$$

3. **(a)** $\mu = 35,\ \sigma = 7$

$$\begin{aligned} P(x \geq 40) &= P\left(z \geq \frac{40-35}{7}\right) \\ &= P(z \geq 0.71) \\ &= 0.2389 \end{aligned}$$

(b) $\mu_{\bar{x}} = \mu = 35,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7}{\sqrt{9}} = \frac{7}{3}$

$$\begin{aligned} P(\bar{x} \geq 40) &= P\left(z \geq \frac{40-35}{\frac{7}{3}}\right) \\ &= P(z \geq 2.14) \\ &= 0.0162 \end{aligned}$$

4. (a) $\mu = 38,\ \sigma = 5$

$$\begin{aligned} P(x \leq 35) &= P\left(z \leq \frac{35-38}{5}\right) \\ &= P(z \leq -0.6) \\ &= 0.2743 \end{aligned}$$

(b) $\mu_{\bar{x}} = \mu = 38,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{10}} = 1.58$

$$\begin{aligned} P(\bar{x} \leq 35) &= P\left(z \leq \frac{35-38}{1.58}\right) \\ &= P(z \leq -1.90) \\ &= 0.0287 \end{aligned}$$

(c) The probability in part (b) is much smaller because the standard deviation is smaller for the $\bar{x}$ distribution.

5. $\mu_{\bar{x}} = \mu = 100,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5$

$$\begin{aligned} P(100 - 2 \leq \bar{x} \leq 100 + 2) &= P(98 \leq \bar{x} \leq 102) \\ &= P\left(\frac{98-100}{1.5} \leq z \leq \frac{102-100}{1.5}\right) \\ &= P(-1.33 \leq z \leq 1.33) \\ &= P(z \leq 1.33) - P(z \leq -1.33) \\ &= 0.9082 - 0.0918 \\ &= 0.8164 \end{aligned}$$

6. $\mu_{\bar{x}} = \mu = 15,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{36}} = 0.333$

$$\begin{aligned} P(15 - 0.5 \leq \bar{x} \leq 15 + 0.5) &= P(14.5 \leq \bar{x} \leq 15.5) \\ &= P\left(\frac{14.5-15}{0.333} \leq z \leq \frac{15.5-15}{0.333}\right) \\ &= P(-1.5 \leq z \leq 1.5) \\ &= P(z \leq 1.5) - P(z \leq -1.5) \\ &= 0.9332 - 0.0668 \\ &= 0.8664 \end{aligned}$$

7. $\mu_{\bar{x}} = \mu = 750,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{64}} = 2.5$

(a)
$$\begin{aligned} P(\bar{x} \ge 750) &= P\left(z \ge \frac{750-750}{2.5}\right) \\ &= P(z \ge 0) \\ &= 0.5000 \end{aligned}$$

(b)
$$\begin{aligned} P(745 \le \bar{x} \le 755) &= P\left(\frac{745-750}{2.5} \le z \le \frac{755-750}{2.5}\right) \\ &= P(-2 \le z \le 2) \\ &= P(z \le 2) - P(z \le -2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544 \end{aligned}$$

8. (a) Miami: $\mu = 76,\ \sigma = 1.9$

$$\begin{aligned} P(x < 77) &= P\left(z < \frac{77-76}{1.9}\right) \\ &= P(z < 0.53) \\ &= 0.7019 \end{aligned}$$

Fairbanks: $\mu = 0,\ \sigma = 5.3$

$$\begin{aligned} P(x < 3) &= P\left(z < \frac{3-0}{5.3}\right) \\ &= P(z < 0.57) \\ &= 0.7157 \end{aligned}$$

(b) Since x has a normal distribution, the sampling distribution of $\bar{x}$ is also normal regardless of the sample size.

Miami: $\mu_{\bar{x}} = \mu = 76,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.9}{\sqrt{7}} = 0.718$

$$\begin{aligned} P(\bar{x} < 77) &= P\left(z < \frac{77-76}{0.718}\right) \\ &= P(z < 1.39) \\ &= 0.9177 \end{aligned}$$

Fairbanks: $\mu_{\bar{x}} = \mu = 0,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5.3}{\sqrt{7}} = 2.003$

$$\begin{aligned} P(\bar{x} < 3) &= P\left(z < \frac{3-0}{2.003}\right) \\ &= P(z < 1.50) \\ &= 0.9332 \end{aligned}$$

(c) We cannot say anything about the probability distribution of $\bar{x}$, because the sample size is not 30 or greater. Consider using all 31 days.

Miami: $\mu_{\bar{x}} = \mu = 76,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.9}{\sqrt{31}} = 0.341$

$$\begin{aligned} P(\bar{x} < 77) &= P\left(z < \frac{77-76}{0.341}\right) \\ &= P(z < 2.93) \\ &= 0.9983 \end{aligned}$$

Fairbanks: $\mu_{\bar{x}} = \mu = 0,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5.3}{\sqrt{31}} = 0.952$

$$\begin{aligned} P(\bar{x} < 3) &= P\left(z < \frac{3-0}{0.952}\right) \\ &= P(z < 3.15) \\ &= 0.9992 \end{aligned}$$

9. (a) $n = 50,\ p = 0.22$

$np = 50(0.22) = 11,\ nq = 50(0.78) = 39$

Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.22,\ \sigma_{\hat{p}} = \sqrt{\frac{0.22(0.78)}{50}} \approx 0.0586$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{50} = 0.01$$

$$\begin{aligned} P(0.20 \le \hat{p} \le 0.25) &\approx P(0.20 - 0.01 \le x \le 0.25 + 0.01) \\ &= P(0.19 \le x \le 0.26) \\ &= P\left(\frac{0.19-0.22}{0.0586} \le z \le \frac{0.26-0.22}{0.0586}\right) \\ &= P(-0.51 \le z \le 0.68) \\ &= P(z \le 0.68) - P(z \le -0.51) \\ &= 0.7517 - 0.3050 \\ &= 0.4467 \end{aligned}$$

(b) $n = 38$, $p = 0.27$

$np = 38(0.27) = 10.26$, $nq = 38(0.73) = 27.74$

Approximate $\hat{p}$ by a normal random variable since both np and nq exceed 5.

$$\mu_{\hat{p}} = p = 0.27, \sigma_{\hat{p}} = \sqrt{\frac{(0.27)(0.73)}{38}} \approx 0.0720$$

$$\text{continuity correction} = \frac{0.5}{n} = \frac{0.5}{38} = 0.013$$

$$\begin{aligned} P(\hat{p} \geq 0.35) &\approx P(x \geq 0.35 - 0.013) \\ &= P(x \geq 0.337) \\ &= P\left(z \geq \frac{0.337 - 0.27}{0.0720}\right) \\ &= P(z \geq 0.93) \\ &= 0.1762 \end{aligned}$$

(c) $n = 51$, $p = 0.05$

$np = 51(0.05) = 2.55$

No, we cannot approximate $\hat{p}$ by a normal random variable since $np < 5$.

Chapter 8 Estimation

Section 8.1

Answers may vary slightly due to rounding.

1. (a) $\bar{x} = \frac{\sum x_i}{n} = \frac{5128}{35} = 146.5143 \approx 146.5,$

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{756{,}820 - \frac{(5128)^2}{35}}{34} = 161.6101$$

$$s = \sqrt{s^2} = \sqrt{161.6101} = 12.7126 \approx 12.7 \text{ as stated}$$

Since $n = 35 \geq 30$, we can use s to approximate σ.

(b) $c = 80\%$ so $z_c = 1.28$

$$\left(\bar{x} - \frac{z_c s}{\sqrt{n}}\right) < \mu < \left(\bar{x} + \frac{z_c s}{\sqrt{n}}\right),\ \left(146.5 - \frac{1.28(12.7)}{\sqrt{35}}\right) < \mu < \left(146.5 + \frac{1.28(12.7)}{\sqrt{35}}\right)$$

$(146.5 - 2.7) < \mu < (146.5 + 2.7)$

143.8 calories $< \mu <$ 149.2 calories

(c) $c = 90\%$ so $z_c = 1.645$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(12.7)}{\sqrt{35}} \approx 3.5$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(146.5 - 3.5) < \mu < (146.5 + 3.5)$

143.0 calories $< \mu <$ 150.0 calories

(d) $c = 99\%$ so $t_c = 2.58$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{2.58(12.7)}{\sqrt{35}} \approx 5.5;\ (\bar{x} - E) < \mu < (\bar{x} + E)$$

$(146.5 - 5.5) < \mu < (146.5 + 5.5)$

141.0 calories $< \mu <$ 152.0 calories

(e)

c	z_c	Length of confidence interval
80%	1.28	149.2 − 143.8 = 5.4
90%	1.645	150.0 − 143.0 = 7.0
99%	2.58	152.0 − 141.0 = 11.0

As the confidence level, c, increases, so does z_c; therefore, all else being the same, the length of the confidence interval increases, too. We can be more confident the interval captures μ if the interval is longer.

2. (a) $n = 30, \bar{x} = 15.71$ inches, $s = 4.63$ inches, since $n = 30 \geq 30$, we can use s to approximate σ

$c = 95\%$ so $z_c = 1.96$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(4.63)}{\sqrt{30}} = 1.66$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(15.71 - 1.66) < \mu < (15.71 + 1.66)$

14.05 inches $< \mu < 17.37$ inches

(b) $n = 90$, $\bar{x} = 15.58$ inches, $s = 4.61$ inches

Since $n = 90 \geq 30$, we can use s to approximate σ.

$c = 95\%$ so $z_c = 1.96$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(4.61)}{\sqrt{90}} = 0.95$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(15.58 - 0.95) < \mu < (15.58 + 0.95)$

14.63 inches $< \mu < 16.53$ inches

(c) $n = 300$, $\bar{x} = 15.59$ inches, $s = 4.62$ inches

Since $n = 300 \geq 30$, we can use s to approximate σ.

$c = 95\%$ so $z_c = 1.96$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(4.62)}{\sqrt{300}} = 0.52$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(15.59 - 0.52) < \mu < (15.59 + 0.52)$

15.07 inches $< \mu < 16.11$ inches

(d)

Sample size, n	Length of confidence interval
30	17.37 − 14.05 = 3.32
90	16.53 − 14.63 = 1.90
300	16.11 − 15.07 = 1.04

As n increases, so does $\sqrt{n}$, which appears in the denominator of E. All else being the same, the length of the confidence interval decreases as n increases.

3. (a) $n = 42$, $\bar{x} = \frac{\sum x_i}{n} = \frac{1511.8}{42} = 35.9952 \approx 36.0$, as stated

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{58,714.96 - 42(36.0)^2}{41} = 104.4624$$

$s^2 = \sqrt{104.4624} = 10.2207 \approx 10.2$, as stated

Since $n = 42 \geq 30$, we can use s to approximate σ.

(b) $c = 75\%$, $z_c = 1.15$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.15(10.2)}{\sqrt{42}} = 1.81$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(36.0 - 1.81) < \mu < (36.0 + 1.81)$

$34.19 < \mu < 37.81$ thousand dollars per employee profit

(c) Since \$30 thousand per employee profit is less than the lower limit of the confidence interval (34.19), your bank profits are low, compared to other similar financial institutions.

(d) Since \$40 thousand per employee profit exceeds the upper limit of the confidence interval (37.81). your bank profit is higher than other similar financial institutions.

(e) $c = 90\%,\ z_c = 1.645$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(10.2)}{\sqrt{42}} = 2.59$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(36.0 - 2.59) < \mu < (36.0 + 2.59)$

$33.41 < \mu < 38.59$ thousand dollars per employee profit

\$30 thousand is less than the lower limit of the confidence interval (33.44), so your bank's profit is less than that of other financial institutions.

\$40 thousand is more than the upper limit of the confidence interval (38.59), so your bank is doing better (profit-wise) than other financial institutions.

4. (a) $n = 35,\ \bar{x} = 5.1029 \approx 5.1$, as stated

$s = 3.7698 \approx 3.8$, as stated

Since $n = 35 \geq 30$, we can use s to estimate σ.

(b) $c = 80\%,\ z_c = 1.28$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.28(3.8)}{\sqrt{35}} = 0.82$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(5.1 - 0.82) < \mu < (5.1 + 0.82)$

$4.28 < \mu < 5.92$ thousand dollars per employee profit

(c) Yes. \$3 thousand per employee profit is less than the lower limit of the confidence interval (4.28)

(d) Yes. \$6.5 thousand dollars profit per employee is larger than the upper limit of the confidence interval (5.92).

(e) $c = 95\%,\ z_c = 1.96$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(3.8)}{\sqrt{35}} = 1.26$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(5.1 - 1.26) < \mu < (5.1 + 1.26)$

$3.84 < \mu < 6.36$ thousand dollars per employee profit

Yes. \$3 thousand per employee profit is less than the lower limit of the confidence interval (3.84)

Yes. \$6.5 thousand dollars profit per employee is larger than the upper limit of the confidence interval (6.36).

5. (a) $n = 40,\ \bar{x} = 51.1575 \approx 51.16$, as stated

$s = 3.0404 \approx 3.04$, as stated

Since $n = 40 \geq 30$, we can use s to estimate σ.

(b) $c = 90\%,\ z_c = 1.645$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(3.04)}{\sqrt{40}} = 0.79$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(51.16 - 0.79) < \mu < (51.16 + 0.79)$

$50.37 < \mu < 51.95^\circ F$

(c) $c = 99\%,\ z_c = 2.58$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{2.58(3.04)}{\sqrt{40}} = 1.24$$
$$(\bar{x} - E) < \mu < (\bar{x} + E)$$
$$(51.16 - 1.24) < \mu < (51.16 + 1.24)$$
$$49.92^\circ\text{F} < \mu < 52.40^\circ\text{F}$$

(d) Since 53°F > 52.4°F, the upper confidence interval limit, it is unlikely the average January temperature is 53°F. Plotting the data by year to see if there is an upward trend (lately or overall), and/or adding several more years of observation might provide evidence to support or refute the claim.

6. $n = 115,\ \bar{x} = \$9.74,\ s = \$2.93$

Since $n = 115 \geq 30$, we can use s to estimate σ.

(a) $c = 95\%,\ z_c = 1.96$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(2.93)}{\sqrt{115}} = 0.54$$
$$(\bar{x} - E) < \mu < (\bar{x} + E)$$
$$(9.74 - 0.54) < \mu < (9.74 + 0.54)$$
$$\$9.20 < \mu < \$10.28$$

(b) Multiply the endpoints of the confidence interval for the average tab per customer by the number of customers. Let μ^* be the total lunch income.

$$(115)(\$9.20) < \mu^* < (115)(\$10.28)$$
$$\$1058.00 < \mu^* < \$1182.20$$

7. (a) $n = 102,\ \bar{x} = 1.2,\ s = 0.4,\ c = 99\%,\ z_c = 2.58$

Since $n = 102 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{2.58(0.4)}{\sqrt{102}} = 0.10$$
$$(\bar{x} - E) < \mu < (\bar{x} + E)$$
$$(1.2 - 0.10) < \mu < (1.2 + 0.10)$$
$$1.10 \text{ seconds} < \mu < 1.30 \text{ seconds}$$

(b) $\bar{x} = 609,\ s = 248,\ c = 95\%,\ z_c = 1.96$, n still 102

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(248)}{\sqrt{102}} = 48.13$$
$$(\bar{x} - E) < \mu < (\bar{x} + E)$$
$$(609 - 48.13) < \mu < (609 + 48.13)$$
$$560.87 \text{ Hz} < \mu < 657.13 \text{ Hz}$$

8. $n = 56,\ \bar{x} = 97^\circ\text{C},\ s = 17^\circ\text{C}$

Since $n = 56 \geq 30$, we can use s to estimate σ.

(a) $c = 95\%,\ z_c = 1.96$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(17)}{\sqrt{56}} = 4.5$$
$$(\bar{x} - E) < \mu < (\bar{x} + E)$$
$$(97 - 4.5) < \mu < (97 + 4.5)$$
$$92.5^\circ\text{C} < \mu < 101.5^\circ\text{C}$$

(b) If the temperature rises, the hot air in the balloon rises, so the balloon would go up. The upper limit of the confidence interval (101.5°C) is an estimate of the (maximum) temperature at which the balloon is at equilibrium (neither going up nor down).

9. (a) $n = 38,\ \bar{x} = 2.5,\ s = 0.7,\ c = 90\%,\ z_c = 1.645$

Since $n = 38 \geq 30$, we can use s to estimate σ.

$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(0.7)}{\sqrt{38}} = 0.2$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(2.5 - 0.2) < \mu < (2.5 + 0.2)$

2.3 minutes $< \mu <$ 2.7 minutes; length = 2.7 – 2.3 = 0.4 minute

(b) $\bar{x} = 15.2,\ s = 4.8,\ n = 38,\ c = 90\%,\ z_c = 1.645$

$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(4.8)}{\sqrt{38}} = 1.3$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(15.2 - 1.3) < \mu < (15.2 + 1.3)$

13.9 min $< \mu <$ 16.5 min; length = 16.5 – 13.9 = 2.6 min

(c) $\bar{x} = 25.7,\ s = 8.3,\ n = 38,\ c = 90\%,\ z_c = 1.645$

$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(8.3)}{\sqrt{38}} = 2.2$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(25.7 - 2.2) < \mu < (25.7 + 2.2)$

23.5 min $< \mu <$ 27.9 min, length = 27.9 – 23.5 = 4.4 min

(d)

length	s
0.4	0.7
2.6	4.8
4.4	8.3

As s increases, so does the length of the interval. This is because s is in the numerator of E and the length of any confidence interval is $(\bar{x} + E) - (\bar{x} - E) = 2E$.

10. $n = 50 \geq 30$, so we can use s to estimate σ.

(a) $\bar{x} = 5.55,\ s = 0.57,\ c = 85\%,\ z_c = 1.44$

$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.44(0.57)}{\sqrt{50}} = 0.12$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(5.55 - 0.12) < \mu < (5.55 + 0.12)$

5.43 cm $< \mu <$ 5.67 cm

(b) $\bar{x} = 2.03,\ s = 0.27,\ c = 90\%,\ z_c = 1.645$

$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(0.27)}{\sqrt{50}} = 0.06$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(2.03 - 0.06) < \mu < (2.03 + 0.06)$

1.97 cm $< \mu <$ 2.09 cm

11. $n = 56$, $\bar{x} = 3.4$, $s = 1.2$, $c = 85\%$, $z_c = 1.44$

Since $n = 56 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.44(1.2)}{\sqrt{56}} = 0.23$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(3.4 - 0.23) < \mu < (3.4 + 0.23)$

3.17 years $< \mu <$ 3.63 years

12. **(a)** $n = 75$, $\bar{x} = 19.5$, $s = 2.25$, $c = 95\%$, $z_c = 1.96$

Since $n = 75 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(2.25)}{\sqrt{75}} = 0.51$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(19.5 - 0.51) < \mu < (19.5 + 0.51)$

18.99 years $< \mu <$ 20.01 years

(b) $n = 89$, $\bar{x} = 22.8$, $s = 2.79$, $c = 99\%$, $z_c = 2.58$

Since $n = 89 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{2.58(2.79)}{\sqrt{89}} = 0.76$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(22.8 - 0.76) < \mu < (22.8 + 0.76)$

22.04 years $< \mu <$ 23.56 years

13. $n = 36$, $\bar{x} = 16{,}000$, $s = 2400$, $c = 90\%$, $z_c = 1.645$

Since $n = 36 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.645(2400)}{\sqrt{36}} = 658$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(16{,}000 - 658) < \mu < (16{,}000 + 658)$

15,342 cars $< \mu <$ 16,658 cars

14. **(a)** $n = 50$, $\bar{x} = 98.75$, $s = 15.91$, $c = 95\%$, $z_c = 1.96$

Since $n = 50 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(15.91)}{\sqrt{50}} = 4.41$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(98.75 - 4.41) < \mu < (98.75 + 4.41)$

\$94.34 $< \mu <$ \$103.16

(b) The WSJ's figure of \$111 is not in the confidence interval; it is higher than the upper limit of \$103.16. The WSJ may have surveyed only upper echelon hotels, where as your survey, based on Yellow Page listings, covered a wider variety of accommodations.

15. $n = 196$, $\bar{x} = 11.9$, $s = 4.30$, $c = 95\%$, $z_c = 1.96$

Since $n = 196 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(4.30)}{\sqrt{196}} = 0.6$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(11.9 - 0.6) < \mu < (11.9 + 0.6)$

11.3 mg/liter $< \mu <$ 12.5 mg/liter

16. $n = 99$, $\bar{x} = 10.5$, $s = 3.2$, $c = 95\%$, $z_c = 1.96$

Since $n = 99 \geq 30$, we can use s to estimate σ.

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(3.2)}{\sqrt{99}} = 0.63$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(10.5 - 0.63) < \mu < (10.5 + 0.63)$

9.87 mg/liter $< \mu <$ 11.13 mg/liter

17. (a) $n = 40$, $\bar{x} = 287.4$, $s = 15.543 \approx 15.5$, as stated

(b) Since $n = 40 \geq 30$, we can use s to estimate σ.

$c = 85\%$, $z_c = 1.44$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.44(15.5)}{\sqrt{40}} = 3.5$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(287.4 - 3.5) < \mu < (287.4 + 3.5)$

283.9 pounds $< \mu <$ 290.9 pounds

(c) Yes; 200 pounds is considerably below the lower limit of the confidence interval (283.9).

(d) Yes; 291 pounds is just barely above the upper limit of the confidence interval (290.0), which rounds to 291. His weight is appropriate for such positions.

(e) $c = 95\%$, $z_c = 1.96$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(15.5)}{\sqrt{40}} = 4.8$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(287.4 - 4.8) < \mu < (287.4 + 4.8)$

282.6 pounds $< \mu <$ 292.2 pounds

Yes; 200 pounds is still much smaller than the lower limit.

Yes; 291 pounds is now within the confidence interval.

18. (a) $n = 40$, $\bar{x} = 27.775 \approx 27.8$ as stated, $s = 3.7994 \approx 3.8$, as stated.

(b) $c = 80\%$, $z_c = 1.28$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.28(3.8)}{\sqrt{40}} = 0.77$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(27.8 - 0.77) < \mu < (27.8 + 0.77)$

27.03 years $< \mu <$ 28.57 years

(c) Yes, he would be somewhat old for such a position because 33 is higher than the upper limit of the confidence interval (28.57).

(d) $c = 99\%$. $z_c = 2.58$

$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{2.58(3.8)}{\sqrt{40}} = 1.55$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(27.8 - 1.55) < \mu < (27.8 + 1.55)$

26.25 years $< \mu <$ 29.35 years

33 years is still quite "old," it is above the upper limit of 29.35 years.

Section 8.2

Answers may vary slightly due to rounding.

1. $n = 18$ so $d.f. = n - 1 = 18 - 1 = 17$, $c = 0.95$

$t_c = t_{0.95} = 2.110$

2. $n = 4$ so $d.f. = n - 1 = 4 - 1 = 3$, $c = 0.99$

$t_c = t_{0.99} = 5.841$

3. $n = 22$ so $d.f. = n - 1 = 22 - 1 = 21$, $c = 0.90$

$t_c = t_{0.90} = 1.721$

4. $n = 12$ so $d.f. = n - 1 = 12 - 1 = 11$, $c = 0.95$

$t_c = t_{0.95} = 2.201$

5. $n = 9$ so $d.f. = n - 1 = 9 - 1 = 8$

(a) $\bar{x} = \frac{\sum x}{n} = \frac{11,450}{9} \approx 1272$, as stated

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{14,577,854 - \frac{(11,450)^2}{9}}{8} = 1363.6944$$

$s = \sqrt{1363.6944} = 36.9282 \approx 37$, as stated

(b) $c = 90\%$, $t_c = t_{0.90}$ with $8\ d.f. = 1.860$

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.86(37)}{\sqrt{9}} = 22.94 \approx 23$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(1272 - 23) < \mu < (1272 + 23)$

1249 A.D. $< \mu <$ 1295 A.D.

6. $n = 16$ so $d.f. = n - 1 = 15$

(a) $\bar{x} = \frac{\sum x_i}{n} = \frac{57.2}{16} = 3.575 \approx 3.58$, as stated

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{256.02 - 16(3.575)^2}{15} = 3.4353$$

$s = \sqrt{3.4353} = 1.853 \approx 1.85$, as stated

(b) $c = 95\%$, so $t_{0.95}$ with 15 $d.f. = 2.131$

$$E = \frac{t_c s}{\sqrt{n}} = \frac{2.131(1.85)}{\sqrt{16}} \approx 0.99$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(3.58 - 0.99) < \mu < (3.58 + 0.99)$

2.59 hours $< \mu <$ 4.57 hours

7. $n = 12$ so $d.f. = n - 1 = 11$

(a) $\bar{x} = 148.3333 \approx 148.33$, as stated
$s = 53.0151 \approx 53.02$, as stated

(b) $c = 90\%$, so $t_{0.90}$ with 11 $d.f. = 1.796$

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.796(53.02)}{\sqrt{12}} \approx 27.49$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(148.33 - 27.49) < \mu < (148.33 + 27.49)$

$\$120.84 < \mu < \$175.82$

8. $n = 20$ so $d.f. = n - 1 = 19$

(a) $\bar{x} = 83.75$, as stated
$s = 28.9662 \approx 28.97$, as stated

(b) $c = 90\%$, so $t_{0.90}$ with 19 $d.f. = 1.729$

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.729(28.97)}{\sqrt{20}} \approx 11.20$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(83.75 - 11.20) < \mu < (83.75 + 11.20)$

$\$72.55 < \mu < \$94.95$

9. $n = 6$ so $d.f. = n - 1 = 5$

(a) $\bar{x} = 91.0$, as stated
$s = 30.7181 \approx 30.7$, as stated

(b) $c = 75\%$, so $t_{0.75}$ with 5 $d.f. = 1.301$

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.301(30.7)}{\sqrt{6}} \approx 16.3$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(91.0 - 16.3) < \mu < (91.0 + 16.3)$

74.7 pounds $< \mu <$ 107.3 pounds

10. $n = 16$ so $d.f. = n - 1 = 15$

(a) $\bar{x} = 5.625 \approx 5.63$, as stated
$s = 1.7842 \approx 1.78$, as stated

(b) $c = 85\%$, so $t_{0.85}$ with 15 $d.f. = 1.517$

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.517(1.78)}{\sqrt{16}} \approx 0.68$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(5.63 - 0.68) < \mu < (5.63 + 0.68)$

4.95 pups $< \mu <$ 6.31 pups

11. $n = 6$ so $d.f. = n - 1 = 5$
$\bar{x} = 79.25,\ s = 5.33$
$c = 80\%$, so $t_{0.80}$ with 5 $d.f. = 1.476$
$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.476(5.33)}{\sqrt{6}} \approx 3.21$$
$(\bar{x} - E) < \mu < (\bar{x} + E)$
$(79.25 - 3.21) < \mu < (79.25 + 3.21)$
76.04 cm $< \mu < 82.46$ cm

12. $n = 8$, so $d.f. = n - 1 = 7$
$\bar{x} = 244.5,\ s = 21.73,\ c = 99\%$
$t_{0.99}$ with 7 $d.f. = 3.499$
$$E = \frac{t_c s}{\sqrt{n}} = \frac{3.499(21.73)}{\sqrt{8}} = 26.8818 \approx 26.9$$
$(\bar{x} - E) < \mu < (\bar{x} + E)$
$(244.5 - 26.9) < \mu < (244.5 + 26.9)$
217.6 calories $< \mu < 271.4$ calories

13. $n = 8$, so $d.f. = n - 1 = 7$

(a) $\bar{x} = 12.3475 \approx 12.35$, as stated
$s = 2.2487 \approx 2.25$, as stated

(b) $c = 90\%$, so $t_{0.90}$ with 7 $d.f. = 1.895$
$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.895(2.25)}{\sqrt{8}} = 1.5075 \approx 1.51$$
$(\bar{x} - E) < \mu < (\bar{x} + E)$
$(12.35 - 1.51) < \mu < (12.35 + 1.51)$
$\$10.84 < \mu < \$13.86$

14. $n = 9,\ d.f. = n - 1 = 8$
$\bar{x} = 106.8889 \approx 106.9$, as stated
$s = 29.4425 \approx 29.4$, as stated
$c = 90\%$, so $t_{0.90}$ with 8 $d.f. = 1.860$
$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.860(29.4)}{\sqrt{9}} = 18.228 \approx 18.2$$
$(\bar{x} - E) < \mu < (\bar{x} + E)$
$(106.9 - 18.2) < \mu < (106.9 + 18.2)$
$\$88.7$ thousand $< \mu < \$125.1$ thousand

15. **(a)** $n = 19,\ d.f. = n - 1 = 18,\ c = 90\%,\ t_{0.90}$ with 18 $d.f. = 1.734$
$\bar{x} = 9.8421 \approx 9.8$, as stated
$s = 3.3001 \approx 3.3$, as stated
$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.734(3.3)}{\sqrt{19}} \approx 1.3$$
$(\bar{x} - E) < \mu < (\bar{x} + E)$
$(9.8 - 1.3) < \mu < (9.8 + 1.3)$
8.5 inches $< \mu < 11.1$ inches

(b) $n = 7$, $d.f. = n - 1 = 6$. $c = 80\%$, $t_{0.80}$ with 6 $d.f. = 1.440$

$\bar{x} = 17.0714 \approx 17.1$, as stated

$s = 2.4568 \approx 2.5$, as stated

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.440(2.5)}{\sqrt{7}} \approx 1.4$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(17.1 - 1.4) < \mu < (17.1 + 1.4)$

15.7 inches $< \mu <$ 18.5 inches

16. (a) $n = 12$, $d.f. = n - 1 = 6$, $c = 90\%$, $t_{0.90}$ with 11 $d.f. = 1.796$

$\bar{x} = 4.70$, as stated

$s = 1.9781 \approx 1.98$, as stated

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.796(1.98)}{\sqrt{12}} \approx 1.03$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(4.70 - 1.03) < \mu < (4.70 + 1.03)$

$3.67\% < \mu < 5.73\%$

(b) $n = 14$. $d.f. = n - 1 = 13$, $c = 90\%$, $t_{0.90}$ with 13 $d.f. = 1.771$

$\bar{x} = 3.1786 \approx 3.18$, as stated

$s = 1.3429 \approx 1.34$, as stated

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.771(1.34)}{\sqrt{12}} \approx 0.63$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(3.18 - 0.63) < \mu < (3.18 + 0.63)$

$2.55\% < \mu < 3.81\%$

17. (a) $n = 8$, $d.f. = n - 1 = 7$. $c = 85\%$, $t_{0.85}$ with 7 $d.f. = 1.617$

$\bar{x} = 33.125 \approx 33.1$, as stated

$s = 6.3852 \approx 6.4$, as stated

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.617(6.4)}{\sqrt{8}} = 3.65885 \approx 3.7$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(33.1 - 3.7) < \mu < (33.1 + 3.7)$

\$29.4 thousand $< \mu <$ \$36.8 thousand

(b) $n = 9$, $d.f. = n - 1 = 8$, $c = 99\%$, $t_{0.99}$ with 8 $d.f. = 3.355$

$\bar{x} = 20.8222 \approx 20.8$, as stated

$s = 2.5228 \approx 2.5$, as stated

$$E = \frac{t_c s}{\sqrt{n}} = \frac{3.355(2.5)}{\sqrt{9}} \approx 2.8$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(20.8 - 2.8) < \mu < (20.8 + 2.8)$

\$18.0 thousand $< \mu <$ \$23.6 thousand

18. (a) $n = 15$, $d.f. = n - 1 = 14$, $c = 99\%$, $t_{0.99}$ with 14 $d.f. = 2.977$

$\bar{x} = 7.2867 \approx 7.3$, as stated

$s = 0.7954 \approx 0.8$, as stated

$$E = \frac{t_c s}{\sqrt{n}} = \frac{2.977(0.8)}{\sqrt{15}} \approx 0.6$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(7.3 - 0.6) < \mu < (7.3 + 0.6)$

$6.7 \text{ days} < \mu < 7.9 \text{ days}$

(b) $n = 10$, $d.f. = n - 1 = 9$, $c = 90\%$, $t_{0.90}$ with 9 $d.f. = 1.833$

$\bar{x} = 62.26 \approx 62.3$, as stated

$s = 8.0185 \approx 8.0$, as stated

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.833(8.0)}{\sqrt{10}} \approx 4.6$$

$(\bar{x} - E) < \mu < (\bar{x} + E)$

$(62.3 - 4.6) < \mu < (62.3 + 4.6)$

$57.7\% < \mu < 66.9\%$

19. Notice that the four figures are drawn on different scales.

(a) The box plots differ in range (distance between whisker ends), in interquartile range (distance between box ends), in medians (line through boxes), in symmetry (indicated by the placement of the median within the box, and by the placement of the median relative to the whisker endpoints), in whisker lengths, and in the presence/absence of outliers. These differences are to be expected since each box plot represents a different sample of size 20. (Although the data sets were all selected as samples of size $n = 20$ from a normal distribution with $\mu = 68$ and $\sigma = 3$, it is very interesting that Sample 2, figure (b), shows 2 outliers.)

(b)

Sample	Confidence interval width	Includes $\mu = 68$?
1	69.407 – 66.692 = 2.715	Yes
2	69.426 – 66.490 = 2.936	Yes
3	69.211 – 66.741 = 2.470	Yes
4	68.050 – 65.766 = 2.284	Yes (barely)

The intervals differ in length; all 4 cover/capture/enclose $\mu = 68$. If many additional samples of size 20 were generated from this distribution, we would expect about 95% of the confidence intervals created from these samples to cover/capture/enclose the number 68; in approximately 5% of the intervals, 68 would be outside the confidence interval, i.e., 68 would be less than the lower limit or greater than the upper limit. Drawing all these samples (at least conceptually) and checking whether or not the confidence intervals include the number 68, keeping track of the percentage that do, is an illustration of the definition of (95%) confidence intervals.

Section 8.3

Answers may vary slightly due to rounding.

1. $r = 39$, $n = 62$, $\hat{p} = \frac{r}{n} = \frac{39}{62}$, $\hat{q} = 1 - \hat{p} = \frac{23}{62}$

(a) $\hat{p} = \frac{39}{62} = 0.6290$

(b) $c = 95\%, z_c = z_{0.95} = 1.96$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.96\sqrt{(0.6290)(1-0.6290)/62} = 0.1202$
$(\hat{p}-E) < p < (\hat{p}+E), (0.6290-0.1202) < p < (0.6290+0.1202)$
$0.5088 < p < 0.7492$ or approximately 0.51 to 0.75.
We are 95% confident that the true proportion of actors who are extroverts is between 0.51 and 0.75, approximately. In repeated sampling from the same population, approximately 95% of the samples would generate confidence intervals that would cover/capture/enclose the true value of p.

(c) $np \approx n\hat{p} = r = 39$
$nq \approx n\hat{q} = n - r = 62 - 39 = 23$
It is quite likely that np and $nq > 5$, since their estimates are much larger than 5. If np and $nq > 5$ then $\hat{p}$ is approximately normal with $\mu = p$ and $\sigma = \sqrt{pq/n}$. This forms the basis for the large sample confidence interval derivation.

2. $n = 519,\ r = 285$

(a) $\hat{p} = \frac{r}{n} = \frac{285}{519} = 0.5491$

(b) $c = 99\%, z_c = 2.58, \hat{q} = 1 - \hat{p} = 0.4509$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 2.58\sqrt{(0.5491)(0.4509)/519} = 0.0564$
$(\hat{p}-E) < p < (\hat{p}+E), (0.5491-0.0564) < p < (0.5491+0.0564)$
$0.4927 < p < 0.6055$ or approximately 0.49 to 0.61.
In repeated sampling, approximately 99% of the intervals generated from the samples would include p, the proportion of judges who are introverts.

(c) $np \approx n\hat{p} = n\left(\frac{r}{n}\right) = r = 285 > 5$
$nq, n\hat{q} = n(1-\hat{p}) = n\left(1 - \frac{r}{n}\right) = n - r = 234 > 5$
Since the estimates of np and nq are substantially greater than 5, it is quite likely $np, nq > 5$. This allows us to use the normal approximation to the distribution of $\hat{p}$, with $\mu = p$ and $\sigma = \sqrt{pq/n}$.

3. $n = 5222, r = 1619$

(a) $\hat{p} = \frac{r}{n} = \frac{1619}{5222} = 0.3100$
so $\hat{q} = 1 - \hat{p} = 0.6900$

(b) $c = 99\%$, so $z_c = 2.58$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 2.58\sqrt{(0.3100)(0.6900)/5222} = 0.0165$
$(\hat{p}-E) < p < (\hat{p}+E), (0.3100-0.0165) < p < (0.3100+0.0165)$
$0.2935 < p < 0.3265$ or approximately 0.29 to 0.33.
In repeated sampling, approximately 99% of the confidence intervals generated from the samples would include p, the proportion of judges who are hogans.

(c) $np \approx n\hat{p} = r = 1619, nq \approx n\hat{q} = n(1-\hat{p}) = n - r = 3603$
Since the estimates of np and nq are much greater than 5, it is reasonable to assume $np, nq > 5$. Then we can use the normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$ to approximate the distribution of $\hat{p}$.

4. $n = 592, r = 360$

(a) $\hat{p} = \frac{r}{n} = \frac{360}{592} = 0.6081$

so $\hat{q} = 1 - \hat{p} = 0.3919$

(b) $c = 95\%, z_c = 1.96$

$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.96\sqrt{(0.6081)(0.3919)/592} = 0.0393$

$(\hat{p} - E) < p < (\hat{p} + E)$

$(0.6081 - 0.0393) < p < (0.6081 + 0.0393)$

$0.5688 < p < 0.6474$, or approximately 0.57 to 0.65.

In repeated sampling from this population, approximately 95% of the samples would generate confidence intervals capturing p, the population proportion of Santa Fe black on white potsherds at the excavation site.

(c) $np \approx n\hat{p} = r = 360 > 5$

$nq \approx n\hat{q} = n(1 - \hat{p}) = n - r = 592 - 360 = 232 > 5$

Since the estimates of np and nq are both greater than 5, it is reasonable to assume that np, nq are also greater than 5. We can then approximate the sampling distribution of $\hat{p}$ with a normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$.

5. $n = 5792, r = 3139$

(a) $\hat{p} = \frac{r}{n} = \frac{3139}{5792} = 0.5420$

so $\hat{q} = 1 - \hat{p} = 0.4580$

(b) $c = 99\%$, so $z_c = 2.58$

$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 2.58\sqrt{(0.5420)(0.4580)/5792} = 0.0169$

$(\hat{p} - E) < p < (\hat{p} + E)$

$(0.5420 - 0.0169) < p < (0.5420 + 0.0169)$

$0.5251 < p < 0.5589$, or approximately 0.53 to 0.56.

If we drew many samples of size 5792 physicians from those in Colorado, and generated a confidence interval from each sample, we would expect approximately 99% of the intervals to include the true proportion of Colorado physicians providing at least some charity care.

(c) $np \approx n\hat{p} = r = 3139 > 5; nq \approx n\hat{q} = n - r = 2653 > 5.$

Since the estimates of np and nq are much larger than 5, it is reasonable to assume np and nq are both greater than 5. Under the circumstances, it is appropriate to approximate the distribution of $\hat{p}$ with a normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$.

6. $n = 250, r = 40$

(a) $\hat{p} = \frac{r}{n} = \frac{40}{250} = 0.1600$

so $\hat{q} = 1 - \hat{p} = 0.8400$

(b) $c = 90\%$, so $z_c = 1.645$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.645\sqrt{(0.1600)(0.8400)/250} = 0.0381$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.1600 - 0.0381) < p < (0.1600 + 0.0381)$
$0.1219 < p < 0.1981$, or approximately 0.12 to 0.20.
In repeated sampling, approximately 90% of the confidence intervals generated from the samples would include p, the true proportion of egg cartons with at least one broken egg.

(c) $np \approx n\hat{p} = r = 40 > 5; nq \approx n\hat{q} = n - r = 250 - 40 = 210 > 5.$
Since the estimates of np and nq are larger than 5, it is reasonable to assume np and $nq > 5$. When np and $nq > 5$, the normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$ provides a good approximation to the distribution of $\hat{p}$.

7. $n = 99$, $r = 17$

(a) $\hat{p} = \frac{r}{n} = \frac{17}{99} = 0.1717$
so $\hat{q} = 0.8283$

(b) $c = 85\%$, so $z_c = 1.44$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.44\sqrt{(0.1717)(0.8283)/99} = 0.0546$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.1717 - 0.0546) < p < (0.1717 + 0.0546)$
$0.1171 < p < 0.2263$, or approximately 0.12 to 0.23.
In repeated sampling from this population about 85% of the intervals generated from these samples would include p, the proportion of fugitives arrested after their photographs appeared in the newspaper.

(c) $np \approx n\hat{p} = r = 17 > 5; nq \approx n\hat{q} = n - r = 82 > 5.$
Since the estimates of np and $nq > 5$, it is reasonable to assume np, $nq > 5$. If np and $nq > 5$, the normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$ provides a good approximation to the distribution of $\hat{p}$.

8. $n = 10{,}351$, $r = 7867$

(a) $\hat{p} = \frac{r}{n} = \frac{7867}{10{,}351} = 0.7600$
so $\hat{q} = 1 - \hat{p} = 0.2400$

(b) $c = 99\%$, so $z_c = 2.58$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 2.58\sqrt{(0.7600)(0.2400)/10{,}351} = 0.0108$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.7600 - 0.0108) < p < (0.7600 + 0.0108)$
$0.7492 < p < 0.7708$, or approximately 0.75 to 0.77.
In repeated sampling from the population of convicts who escaped from U.S. prisons, approximately 99% of the confidence intervals created from those samples would include p, the proportion of recaptured escaped convicts.

(c) $np \approx n\hat{p} = r = 7867 > 5; nq \approx n\hat{q} = n(1 - \hat{p}) = n - r = 2484 > 5.$
Since the estimates of np and nq are each considerably greater than 5, it is reasonable to assume that np and $nq > 5$. When np and $nq > 5$, the distribution of $\hat{p}$ can be quite accurately approximated by a normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$.

9. $n = 855, r = 26$

(a) $\hat{p} = \frac{r}{n} = \frac{26}{855} = 0.0304$
so $\hat{q} = 1 - \hat{p} = 0.9696$

(b) $c = 99\%$, so $z_c = 2.58$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 2.58\sqrt{(0.0304)(0.9696)/855} = 0.0151$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.0304 - 0.0151) < p < (0.0304 + 0.0151)$
$0.0153 < p < 0.0455$, or approximately 0.02 to 0.05.

If many additional samples of size n = 855 were drawn from this fish population, and a confidence interval was created from each such sample, approximately 99% of those confidence intervals would contain p, the catch-and-release mortality rate (barbless hooks removed).

(c) $np \approx n\hat{p} = r = 26 > 5; nq \approx n\hat{q} = n - r = 829 > 5.$

Based on the estimates of np and nq, it is safe to assume both np and $nq > 5$. When np and $nq > 5$, the distribution of $\hat{p}$ can be accurately approximated by a normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$.

10. $n = 200, r = 58$

(a) $\hat{p} = \frac{r}{n} = \frac{58}{200} = 0.2900$
so $\hat{q} = 1 - \hat{p} = 0.7100$

(b) $c = 95\%$, so $z_c = 1.96$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.96\sqrt{(0.2900)(0.7100)/200} = 0.0629$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.2900 - 0.0629) < p < (0.2900 + 0.0629)$
$0.2271 < p < 0.3529$, or approximately 0.23 to 0.35.

If many additional samples of size $n = 200$ were drawn from this fish population, and if a confidence interval was created from each of these samples, approximately 95% of the intervals would include the true value of p, the catch-and-release mortality rate (barbed hooks not removed).

(c) $np \approx n\hat{p} = r = 58 > 5; nq \approx n\hat{q} = n - r = 200 - 58 = 142 > 5.$

Based on these estimates, it is reasonable to assume np, $nq > 5$, and when np and $nq > 5$, the normal distribution with $\mu = p$ and $\sigma = \sqrt{pq/n}$ closely approximates the actual distribution of $\hat{p}$, and can be used instead of the actual distribution of $\hat{p}$.

11. $n = 900$, $r = 54$, $n - r = 846$; both np and $nq > 5$.

(a) $\hat{p} = \frac{r}{n} = \frac{54}{900} = 0.0600$
so $\hat{q} = 1 - \hat{p} = 0.9400$

(b) $c = 99\%$, so $z_c = 2.58$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 2.58\sqrt{(0.0600)(0.9400)/900} = 0.0204$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.0600 - 0.0204) < p < (0.0600 + 0.0204)$
$0.0396 < p < 0.0804$, or about 0.04 to 0.08.

12. $n = 1000$, $r = 590$, $n - r = 410$; both np and $nq > 5$.

(a) $\hat{p} = \frac{r}{n} = \frac{590}{1000} = 0.5900$
so $\hat{q} = 0.4100$

(b) $c = 99\%$, so $z_c = 2.58$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 2.58\sqrt{(0.5900)(0.4100)/1000} = 0.0401$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.5900 - 0.0401) < p < (0.5900 + 0.0401)$
$0.5499 < p < 0.6301$, or about 0.55 to 0.63.

13. $n = 2000$, $r = 382$, $n - r = 1618$; both np and $nq > 5$.

(a) $\hat{p} = \frac{r}{n} = \frac{382}{2000} = 0.1910$, so $\hat{q} = 0.8090$

(b) $c = 80\%$, so $z_c = 1.28$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.28\sqrt{(0.1910)(0.8090)/2000} = 0.0112$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.1910 - 0.0112) < p < (0.1910 + 0.0112)$
$0.1798 < p < 0.2022$, or about 0.18 to 0.20.

14. $n = 328$, $r = 171$, $n - r = 157$; both np and $nq > 5$.

(a) $\hat{p} = \frac{r}{n} = \frac{171}{328} = 0.5213$, so $\hat{q} = 0.4787$

(b) $c = 95\%$, so $z_c = 1.96$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.96\sqrt{(0.5213)(0.4787)/328} = 0.0541$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.5213 - 0.0541) < p < (0.5213 + 0.0541)$
$0.4672 < p < 0.5754$, or about 0.47 to 0.58.
In repeated sampling, approximately 95% of the intervals created from these samples would include p, the proportion of M.D.s with solo practices.

(c) Margin of error $= E = 0.0541 \approx 5.4\%$
A recent study published by the American Medical Association showed that approximately 52% of M.D.s have solo practices. The study has a margin of error of 5.4 percentage points.

15. $n = 730$, $r = 628$, $n - r = 102$; both np and $nq > 5$.

(a) $\hat{p} = \frac{r}{n} = \frac{628}{730} = 0.8603$, so $\hat{q} = 0.1397$

(b) $c = 95\%$, so $z_c = 1.96$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.96\sqrt{(0.8603)(0.1397)/730} = 0.0251$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.8603 - 0.0251) < p < (0.8603 + 0.0251)$
$0.8352 < p < 0.8854$, or about 0.84 to 0.89.
In repeated sampling, approximately 95% of the intervals created from the samples would include p, the proportion of loyal women shoppers.

(c) Margin of error $= E \approx 2.5\%$
A recent study by the Food Marketing Institute showed that about 86% of women shoppers remained loyal to their favorite supermarket last year. The study's margin of error is 2.5 percentage points.

16. $n = 1001, r = 273, n - r = 728$: both np and $nq > 5$.

(a) $\hat{p} = \frac{r}{n} = \frac{273}{1001} = 0.2727$, so $\hat{q} = 0.7273$

(b) $c = 95\%$, so $z_c = 1.96$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.96\sqrt{(0.2727)(0.7273)/1001} = 0.0276$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.2727 - 0.0276) < p < (0.2727 + 0.0276)$
$0.2451 < p < 0.3003$, or about 0.25 to 0.30.

If many additional samples of size $n = 1001$ were drawn from this population, about 95% of the confidence intervals created from these samples would include p, the proportion of shoppers who stock up on bargains.

(c) Margin of error = $E \approx 2.8\%$
The Food Marketing Institute reported that, based on a recent study, 27.3% of shoppers stock up on an item when it is a real bargain. The study had a margin of error of 2.8 percentage points.

17. $n = 1000, r = 250, n - r = 750$; both np and $nq > 5$.

(a) $\hat{p} = \frac{r}{n} = \frac{250}{1000} = 0.2500$, so $\hat{q} = 0.7500$

(b) $c = 95\%$, so $z_c = 1.96$
$E \approx z_c\sqrt{\hat{p}\hat{q}/n} = 1.96\sqrt{(0.2500)(0.7500)/1000} = 0.0268$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.2500 - 0.0268) < p < (0.2500 + 0.0268)$
$0.2232 < p < 0.2768$, or about 0.22 to 0.28.

(c) Margin of error = $E \approx 2.7\%$
In a survey reported in *USA Today*, 25% of large corporations interviewed admitted that, given a choice between equally qualified applicants, they would offer the job to the nonsmoker. The survey's margin of error was 2.7 percentage points.

18. $\hat{p} = 19\% = 0.19$, margin of sampling error $E = 3\% = 0.03$ confidence interval is $(\hat{p} - E) < p < (\hat{p} + E)$ so $(0.19 - 0.03) < p < (0.19 + 0.03)$, or 0.16 to 0.22.

Section 8.4

1. The goal is to estimate μ, the population mean number of new lodgepole pine saplings in a 50 square meter plot in Yellowstone National Park. Use $n = (z_c\sigma / E)^2$, where $c = 95\%$, $z_c = 1.96$, $\sigma = 44$, and $E = 10$.

$$n \approx \left[\frac{1.96(44)}{10}\right]^2 = 74.37, \text{ “round up” to 75 plots.}$$

2. The goal is to estimate the mean root depth, μ, in glacial outwash soil.
Use $n = (z_c\sigma / E)^2$ with $c = 90\%$, $z_c = 1.645, E = 0.5$, and $\sigma = 8.94$.
$n \approx [1.645(8.94)/0.5]^2 = 865.10$; “roundup” to 866 plants.

3. The goal is to estimate the proportion, p. of people in your neighborhood who go to McDonald's. Use $n = pq(z_c/E)^2$.

(a) $c = 85\%, z_c = 1.44, E = 0.05$

With no preliminary estimate of p, we will use $p = q = 0.5$, which maximizes the value of $pq = p(1-p)$, and, therefore, maximizes the value of n; no other choice of p would give a larger n. The "worst case" scenario, when pq and n are the largest values possible, occurs when $p = q = \frac{1}{2} = 0.5$.

Then, $n = pq(z_c/E)^2 \approx 0.5 \cdot 0.5(z_c/E)^2 = 0.25(z_c/E)^2 = 0.25(1.44/0.05)^2 = 207.36$, which is rounded up to 208 people.

(b) $\hat{p} = 1/20 = 0.05$ is a preliminary estimate of $q = 1-p$, so $\hat{q} = 1-\hat{p} = 0.95$ is a preliminary estimate of q.

Then $n = pq(z_c/E)^2 \approx (0.05)(0.95)(1.44/0.05)^2 = 39.4$, or, rounded up, 40 people.

(Note that this sample size is much, much smaller than the sample size of $n = 208$ derived in part (a) where, in the absence of a preliminary estimate of p, we used the worst case estimate of p. $\hat{p} = 1/2$.)

4. The goal is to estimate μ, the mean height of NBA players; use $n = (z_c\sigma/E)^2$. Preliminary, $n = 41$. $s = 3.32, c = 95\%, z_c = 1.96, E = 0.75$.

Then $n \approx [1.96(3.32)/0.75]^2 = 75.3$, or 76 players. However, the preliminary sample had 41 players in it, so we need to sample only $76 - 41 = 35$ additional players.

5. The goal is to find μ, the mean player weight. Preliminary, $n = 56$.
Use $n = (z_c\sigma/E)^2$. $s = 26.58, c = 90\%, z_c = 1.645, E = 4$.

Then $n \approx [1.645(26.58)/4]^2 = 119.5$, or 120 players. However, since 56 players have already been drawn to estimate σ, we need only $120 - 56 = 64$ additional players.

6. The goal is to estimate p, the proportion of callers who reach a business person by phone on the first call, so use $n = pq(z_c/E)^2$.

(a) $c = 80\%, z_c = 1.28, E = 0.03$

Since there is no preliminary estimate of p, we will use $\hat{p} = \hat{q} = 1/2$, the "worst case" estimate; no other choice of $\hat{p}$ would give us a larger sample size.

When $n = pq(z_c/E)^2$ becomes $n \approx 0.5 \cdot 0.5(z_c/E)^2$ or $n \approx 0.25(z_c/E)^2 = 0.25(1.28/0.03)^2 = 455.1$, or 456. We need to have a sample size of 456 business phone calls.

(b) $\hat{p} = 17\% = 0.17$, so $\hat{q} = 1-\hat{p} = 1-0.17 = 0.83$
$n = pq(z_c/E)^2 \approx (0.17)(0.83)(1.28/0.03)^2 = 256.9$, or 257 business phone calls.

7. The goal is to estimate the proportion of women students; use $n = pq(z_c/E)^2$.

(a) $c = 99\%, z_c = 2.58, E = 0.05$

Since there is no preliminary estimate of p, we'll use the "worst case" estimate, $\hat{p} = \frac{1}{2}, \hat{q} = 1-\hat{p} = \frac{1}{2}$.

$n \approx 0.5 \cdot 0.5(z_c/E)^2 = 0.25(z_c/E)^2 = 0.25(2.58/0.05)^2 = 665.6$, or 666 students.

(b) Preliminary estimate of p: $\hat{p} = 54\% = 0.54, \hat{q} = 1-\hat{p} = 0.46$.

$n \approx (0.54)(0.46)(2.58/0.05)^2 = 661.4 \approx 662$ students.

(There is very little difference between (a) and (b) because (a) uses $\hat{p} = \frac{1}{2} = 0.50$ and (b) uses $\hat{p} = 0.54$, and the $\hat{p}s$ are approximately the same.)

8. The goal is to estimate p, the proportion of self-service gas customers; use $n = pq(z_c / E)^2$.

(a) $c = 90\%, z_c = 1.645, E = 0.08$

Since there is no preliminary estimate of p to use, we'll use the "worst case" $\hat{p} = \hat{q} = \frac{1}{2}$.

$n = pq(z_c / E)^2 \approx 0.5 \cdot 0.5(z_c / E)^2 = 0.25(z_c / E)^2 = 0.25(1.645 / 0.08)^2 = 105.7 \approx 106$ customers.

(b) $\hat{p} = 81\% = 0.81$, so $\hat{q} = 1 - \hat{p} = 0.19$.

$n \approx \hat{p}\hat{q}(z_c / E)^2 = (0.81)(0.19)(1.645 / 0.08)^2 = 65.07 \approx 66$ customers.

9. The goal is to estimate μ, the mean reconstructed clay vessel diameter, use $n = (z_c\sigma / E)^2$.
Preliminary $n = 83, s = 5.5, c = 95\%, \; z_c = 1.96, E = 1.0$

$n \approx [1.96(5.5)/1.0]^2 = 116.2 \approx 117$ clay pots.

Since 83 pots were already measured, we need 117 – 83 = 34 additional reconstructed clay pots.

10. The goal is to estimate the mean weight, μ, of bighorn sheep; use $n = (z_c\sigma / E)^2$.
Preliminary $n = 37, s = 15.8, c = 90\%, z_c = 1.645, E = 2.5$.

$n \approx [1.645(15.8)/2.5]^2 = 108.1 \approx 109$ sheep.

Since 37 bighorn sheep have already been weighed, we need 109 – 37 = 72 additional bighorn sheep for the sample.

11. The goal is to estimate the proportion, p, of pine beetle-infested trees; use $n = pq(z_c / E)^2$.

(a) $c = 85\%, z_c = 1.44, E = 0.06$

No preliminary $\hat{p}$ available, so use "worst case" estimate of $p, \hat{p} = \frac{1}{2}; \hat{q} = 1 - \hat{p} = \frac{1}{2}$.

$n \approx 0.5 \cdot 0.5(z_c / E)^2 = 0.25(z_c / E)^2 = 0.25(1.44 / 0.06)^2 = 144$ trees.

(b) Preliminary study showed $r = 19$ infested trees in a sample of $n = 58$ trees, so
$\hat{p} = r / n = 19/58 = 0.3276; \hat{q} = 1 - \hat{p} = 0.6724$.

$n = pq(z_c / E)^2 \approx (0.3276)(0.6724)(1.44 / 0.06)^2 = 126.88 \approx 127$ trees.

Since 58 trees have already been checked for infestation, we need to check 127 – 58 = 69 additional trees.

12. The goal is to estimate the mean daily net income, μ; use $n = (z_c\sigma / E)^2$.
Preliminary sample of $n = 40, s = 57.19, E = 10, c = 85\%. \; z_c = 1.44$.

$n \approx [1.44(57.19)/10]^2 = 67.8 \approx 68$ business days. Since 40 days' records have already been examined, we need to check the net income on an additional 68 – 40 = 28 days.

13. The goal is to estimate p, the proportion of voters favoring capital punishment; use $n = pq(z_c / E)^2$.

(a) $c = 99\%, z_c = 2.58, E = 0.01$

Since there is no preliminary estimate of p, we will use the "worst case" estimate,
$\hat{p} = \frac{1}{2}; \hat{q} = 1 - \hat{p} = \frac{1}{2}$.

$n \approx 0.5 \cdot 0.5(z_c / E)^2 = 0.25(z_c / E)^2 = 0.25(2.58 / 0.01)^2 = 16{,}641$ people.

(b) Preliminary estimate $\hat{p} = 67\% = 0.67; \hat{q} = 1 - \hat{p} = 0.33$.

$n \approx (0.67)(0.33)(2.58/0.01)^2 = 14,717.3 \approx 14,718$ people.

(Comment: this is an exceptionally large sample size and it would cost David quite a large sum of money to survey this many people. Good surveys are often done with sample sizes of 300 to 1000. David needs to reconsider the confidence level, c, and the margin of error, E, he wants for his project. If David opted for $c = 95\%$, $z_c = 1.96$, and $E = 0.05$, he could use a sample of about 340.)

14. The goal is to estimate μ, the average of women at their first marriage: use $n = (z_c\sigma / E)^2$.

Preliminary sample of $n = 35, s = 2.3, c = 95\%, z_c = 1.96, E = 0.25$.

$n \approx [1.96(2.3)/0.25]^2 = 325.15 \approx 326$ women.

Since 35 women have already been interviewed, the sociologist needs to interview $326 - 35 = 291$ additional women.

15. The goal is to estimate μ, the average time phone customers are on hold; use $n = (z_c\sigma / E)^2$.

Preliminary sample of size $n = 167, s = 3.8$ minutes, $c = 99\%, z_c = 2.58, E = 30$ seconds $= 0.5$ minute (all time figures must be in the same units).

$n \approx [2.58(3.8)/0.5]^2 = 384.5 \approx 385$ phone calls.

Since the airline already measured the time on hold for 167 calls, it needs to measure the time on hold for $385 - 167 = 218$ more phone calls.

16. The goal is to estimate p, the proportion of small businesses declaring bankruptcy; use $n = pq(z_c / E)^2$.

(a) $c = 95\%, z_c = 1.96, E = 0.10$

Since there is no preliminary estimate of p available, use the worst case estimate. $\hat{p} = \frac{1}{2}; \hat{q} = 1 - \hat{p} = \frac{1}{2}$.

$n \approx 0.5 \cdot 0.5(z_c / E)^2 = 0.25(z_c / E)^2 = 0.25(1.96/0.10)^2 = 96.04 \approx 97$ small businesses.

(b) Preliminary sample of $n = 38$ found $r = 6$ small businesses had declared bankruptcy, so $\hat{p} = r/n = 6/38 = 0.1579, \hat{q} = 1 - \hat{p} = 0.8421$.

$n \approx (0.1579)(0.8421)(1.96/0.10)^2 = 51.08 \approx 52$ small businesses, total.

Since 38 have already been surveyed, the National Council of Small Businesses needs to survey an additional $52 - 38 = 14$ small businesses.

17. The goal is to estimate p, the proportion of pickup truck owners who are women; use $n = pq(z_c / E)^2$.

(a) $c = 90\%, z_c = 1.645, E = 0.1$

Since no preliminary estimate is available, use the worst case estimate. $\hat{p} = \frac{1}{2}; \hat{q} = 1 - \hat{p} = \frac{1}{2}$.

$n \approx 0.5 \cdot 0.5(z_c / E)^2 = 0.25(z_c / E)^2 = 0.25(1.645/0.1)^2 = 67.65 \approx 68$ owners.

(b) $\hat{p} = 24\% = 0.24; \hat{q} = 1 - \hat{p} = 0.76$.

$n \approx (0.24)(0.76)(1.645/0.1)^2 = 49.36 \approx 50$ owners.

18. The goal is to estimate p, the proportion of voters who favor the bond issue; use $n = pq(z_c / E)^2$.

(a) $c = 90\%, z_c = 1.645, E = 5\% = 0.05$

Since no preliminary estimate of p is available, Linda Silbers should use the worst case estimate. $\hat{p} = \frac{1}{2}; \hat{q} = 1 - \hat{p} = \frac{1}{2}$.

$n \approx 0.5 \cdot 0.5(z_c / E)^2 = 0.25(z_c / E)^2 = 0.25(1.645/0.05)^2 = 270.6 \approx 271$ voters.

(b) $\hat{p} = 73\% = 0.73; \hat{q} = 1 - \hat{p} = 0.27.$

$n \approx (0.73)(0.27)(1.645/0.05)^2 = 213.3 \approx 214$ voters.

19. **(a)** Hint: it is usually easier to go from the more complicated version to the easier version in showing the equality of two things. So, start with:

$$\frac{1}{4}-\left(p-\frac{1}{2}\right)^2 = \frac{1}{4}-\left[p^2-2p\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)^2\right] \quad \text{Recall: } (a-b)^2 = a^2-2ab+b^2$$
$$= \frac{1}{4}-\left[p^2-p+\frac{1}{4}\right]$$
$$= -p^2+p$$
$$= p-p^2 = p(1-p)$$

as was to be shown.

(b) For any number $a, a^2 \geq 0$, so

$$\left(p-\frac{1}{2}\right)^2 \geq 0$$
$$(-1)\left(p-\frac{1}{2}\right)^2 \leq (-1)(0) = 0 \quad \text{Multiply both sides by } -1\text{, remembering to reverse the order of inequality.}$$
$$-\left(p-\frac{1}{2}\right)^2 \leq 0$$
$$0 \geq -\left(p-\frac{1}{2}\right)^2$$
$$\frac{1}{4} \geq \frac{1}{4}-\left(p-\frac{1}{2}\right)^2 \quad \text{Add } \tfrac{1}{4} \text{ to both sides.}$$

but $\frac{1}{4}-\left(p-\frac{1}{2}\right)^2 = p(1-p)$ from part (a), so $\frac{1}{4} \geq p(1-p)$

[$p(1-p)$ is never greater than $1/4$ because it is always less than or equal to $1/4$.]

20. Note: all samples size calculations, after "rounding up," give the <u>minimum</u> sample size required to meet the c, E, etc., criteria.

Recall from text page 407 that the margin of error is the maximal error E of a <u>95%</u> confidence interval for p.

The problem, therefore, implies $E = 3\% = 0.03$, $c = 95\%$, $z_c = 1.96$, and since no preliminary estimate or p is given, use $\hat{p} = \frac{1}{2}; \hat{q} = 1 - \hat{p} = \frac{1}{2}$.

Because the goal is to estimate p, the proportion of registered voters who favor using lottery proceeds for park improvements, use

$$n = pq(z_c/E)^2 \approx 0.5 \cdot 0.5(z_c/E)^2 = 0.25(z_c/E)^2 = 0.25(1.96/0.03)^2 = 1067.1 \approx 1068 \text{ registered voters.}$$

21. The goal is to estimate the proportion, p, of votes for the Democratic presidential candidate. Use $n = pq(z_c / E)^2$.

(a) $E = 0.001, c = 99\%, z_c = 2.58$

Since there is no preliminary estimate of p, use $\hat{p} = \hat{q} = \frac{1}{2}$.

$$n = 0.5 \cdot 0.5(z_c / E)^2 = 0.25(z_c / E)^2 = 0.25(2.58 / 0.001)^2 = 1{,}664{,}100 \text{ votes.}$$

(b) No. If the preliminary estimate of p was $\hat{p} = 0.5$, the same as the no information, worst case estimate of p, the sample size would be exactly the same as in (a). The general formula $n = pq(z_c / E)^2$ with preliminary estimates of p (and q) = 0.5 is the same as the no-information-about-p formula, $n = \frac{1}{4}(z_c / E)^2$.

[The above solutions have repeatedly used this fact, demonstrating that using $\hat{p} = \hat{q} = 0.5$ (the worst case estimates of p and q, giving the largest possible sample size that meets the stated criteria) in the special no-information about-p formula, $n = \frac{1}{4}(z_c / E)^2$, directly.]

Section 8.5

Answers may vary slightly due to rounding. For clarity, pooled estimates will be subscripted with "*p*" or the word "pooled."

1.

	Sample 1	Sample 2
n	45	40
s	0.37	0.31
$\bar{x}$	6.18	6.45

The confidence interval is for the difference in mean heights; since both n_1 and $n_2 > 30$ use large-sample technique.

(a) $c = 95\%, z_c = 1.96$

$$E = z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx z_c\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.96\sqrt{\frac{0.37^2}{45} + \frac{0.31^2}{40}} = 0.1446$$

$$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$$
$$[(6.18 - 6.45) - 0.14] < (\mu_1 - \mu_2) < [(6.18 - 6.45) + 0.14]$$
$$-0.41 < (\mu_1 - \mu_2) < -0.13$$

(b) Since the interval values are all negative, we are 95% confident that $\mu_1 - \mu_2 < 0$, i.e., that $\mu_1 < \mu_2$. The football players appear to be between 0.13 and 0.41 feet shorter on average than the basketball players.

(c) $c = 99\%, z_c = 2.58$

$$E = 2.58\sqrt{\frac{0.37^2}{45} + \frac{0.31^2}{40}} = 0.1904$$

$$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$$
$$[(6.18 - 6.45) - 0.19] < (\mu_1 - \mu_2) < [(6.18 - 6.45) + 0.19]$$
$$-0.46 < (\mu_1 - \mu_2) < -0.08$$

No; Since all the interval values are all negative, we are 99% confidence that football players are, on average, between 0.08 and 0.46 feet shorter than basketball players.

2.

	Sample 1	Sample 2
n	38	41
$\bar{x}$	31.5	30.3
s	4.4	3.8

The confidence interval is for the difference between mean ages; since n_1 and $n_2 \geq 30$, use large-sample technique.

(a) $c = 75\%, z_c = 1.15$

$$E = z_c\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \approx 1.15\sqrt{\frac{4.4^2}{38}+\frac{3.8^2}{41}} = 1.0675$$

$$[(\bar{x}_1-\bar{x}_2)-E] < (\mu_1-\mu_2) < [(\bar{x}_1-\bar{x}_2)+E]$$
$$[(31.5-30.3)-1.07] < (\mu_1-\mu_2) < [(31.5-30.3)+1.07]$$
$$0.13 < (\mu_1-\mu_2) < 2.27$$

(b) Because all values in the interval are positive we are 75% confident that baseball players are, on average between 0.13 and 2.27 years older than basketball players.

(c) $c = 90\%, z_c = 1.645$

$$E = 1.645\sqrt{\frac{4.4^2}{38}+\frac{3.8^2}{41}} = 1.5270$$

$$[(\bar{x}_1-\bar{x}_2)-E] < (\mu_1-\mu_2) < [(\bar{x}_1-\bar{x}_2)+E]$$
$$[(31.5-30.3)-1.53] < (\mu_1-\mu_2) < [(31.5-30.3)+1.53]$$
$$-0.33 < (\mu_1-\mu_2) < 2.73$$

Yes. Because the interval includes 0, at the 90% confidence level, there appears to be no difference in mean ages.

3. (a) $n_1 = 15, \bar{x}_1 = 5.593 \approx 5.6, s_1 = 0.9505 \approx 1.0$
$n_2 = 12, \bar{x}_2 = 2.533 \approx 2.5, s_2 = 0.8998 \approx 0.9$

(b) Since we want a confidence interval for $\mu_1 - \mu_2$, and n_1 and n_2 are both < 30, use small sample technique; $c = 99\%$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{14(1.0)^2+11(0.9)^2}{15+12-2}} = \sqrt{0.9164} = 0.9573$$

$t_c = t_{0.99} = 2.787$ with $n_1 + n_2 - 2 = 15 + 12 - 2 = 25\ d.f.$

$$E \approx t_c s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (2.787)(0.9573)\sqrt{\frac{1}{15}+\frac{1}{12}} = 1.0333 \approx 1.03$$

$$[(\bar{x}_1-\bar{x}_2)-E] < (\mu_1-\mu_2) < [(\bar{x}_1-\bar{x}_2)+E]$$
$$[(5.6-2.5)-1.03] < (\mu_1-\mu_2) < [(5.6-2.5)+1.03]$$
$$2.07 < (\mu_1-\mu_2) < 4.13$$

(c) Since all numbers in the interval are positive, we are 99% confident that the profit as a percentage of revenue is between 2.07% and 4.13% higher for the manufacturers than for the food and drugstores.

4. (a) $n_1 = 12, \bar{x}_1 = 4.425 \approx 4.4, s_1 = 1.5829 \approx 1.6$
$n_2 = 10, \bar{x}_2 = 4.62 \approx 4.6, s_2 = 1.4235 \approx 1.4$

since $n_i < 30$, use small sample technique to estimate $\mu_1 - \mu_2$.

(b) $c = 90\%. n_1 + n_2 - 2 = 12 + 10 - 2 = 20, t_{0.90}$ with 20 $d.f. = 1.725$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{11(1.6)^2 + 9(1.4)^2}{12 + 10 - 2}} = \sqrt{2.29} = 1.5133$$

$$E \approx t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (1.725)(1.5133)\sqrt{\frac{1}{12} + \frac{1}{10}} = 1.1177 \approx 1.1$$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$
$[(4.4 - 4.6) - 1.1] < (\mu_1 - \mu_2) < [(4.4 - 4.6) + 1.1]$
$-1.3 < (\mu_1 - \mu_2) < 0.9$

(c) No; both positive and negative values are in the interval, so we are 90% confident that there is no difference in mean profit as a percentage of revenue between insurance companies and health care organizations.

5.

	Sample 1	Sample 2
n	375	571
r	289	23
$\hat{p}$	289/375 = 0.7707	23/571 = 0.0403

$n_1\hat{p}_1 = 289, n_1\hat{q}_1 = 86, n_2\hat{p}_2 = 23, n_2\hat{q}_2 = 548$
Since all four of these estimates are > 5, use the large sample technique to estimate $p_1 - p_2$.

(a) $c = 99\%. z_c = 2.58$

$$E \approx z_c \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 2.58\sqrt{\frac{(0.7707)(0.2293)}{375} + \frac{(0.0403)(0.9597)}{571}} = 2.58(0.0232) = 0.0599 \approx 0.06$$

$[(\hat{p}_1 - \hat{p}_2) - E] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E]$
$[(0.7707 - 0.0403) - 0.0599] < (p_1 - p_2) < [(0.7707 - 0.0403) + 0.0599]$
$0.6705 < (p_1 - p_2) < 0.7903$, or approximately 0.67 to 0.79

(b) Because the confidence interval contains only positive values, $p_1 > p_2$ and we can be 99% confident that $p_1 - p_2$ is between 0.67 and 0.79, inclusive.

6.

	Sample 1	Sample 2
n	375	571
r	132	217
$\hat{p}$	132/375 = 0.3520	217/571 = 0.3800

Since $n_1\hat{p}_1 = 132, n_1\hat{q}_1 = 243. n\hat{p}_2 = 217, n\hat{q}_2 = 354$ are all greater than 5. we can use large sample methods to estimate $p_1 - p_2$.

(a) $c = 90\%. z_c = 1.645$

$$E \approx z_c \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.645\sqrt{\frac{(0.3520)(0.6480)}{375} + \frac{(0.3800)(0.6200)}{571}} = 1.645(0.0320) = 0.0526$$

$[(\hat{p}_1 - \hat{p}_2) - E] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E]$
$[(0.3520 - 0.3800) - 0.0526] < (p_1 - p_2) < [(0.3520 - 0.3800) + 0.0526]$
$-0.0806 < (p_1 - p_2) < 0.0246$, or about -0.08 to 0.02.

(b) The confidence interval contains both positive and negative values. With 90% confidence, we can conclude there is no difference between p_1 and p_2.

7.

	Sample 1	Sample 2
n	9340	25,111
$\bar{x}$	63.3	72.1
s	9.17	12.67

Since $n_1, n_2 > 30$, we can use large sample methods to estimate $\mu_1 - \mu_2$.

(a) $c = 99\%, z_c = 2.58$

$$E \approx z_c\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 2.58\sqrt{\frac{9.17^2}{9340}+\frac{12.67^2}{25{,}111}} = 0.3201$$

$$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$$
$$[(63.3 - 72.1) - 0.3201] < (\mu_1 - \mu_2) < [(63.3 - 72.1) + 0.3201]$$
$$-9.1201 < \mu < -8.4799, \text{ or about } -9.12 \text{ to } -8.48$$

(b) The interval includes only negative numbers, leading us to believe $\mu_1 < \mu_2$. We are 99% confident that the mean interval between eruptions during the period 1983 to 1987 is between 8.48 and 9.12 minutes longer than the mean interval between Old Faithful eruptions during the period 1948 to 1952. [Comment: it is highly unlikely the data in this problem constitute the required two independent random samples. First, the data are time series observations and are probably highly correlated. It is possible the 30 year gap between time periods would be sufficient to wipe out the effects of serial correlation so that the two samples could be considered independent. However, the times within each sample are still correlated, and random samples consist of data that are independent (and identically distributed). Second, the large sample sizes, much larger than needed, might indicate a census rather than a sample of data was used.)

8.

	Sample 1	Sample 2
n	32	32
$\bar{x}$	69.44	59.00
s	11.69	11.60

Since $n_1, n_2 > 30$, we can use large sample methods to estimate $\mu_1 - \mu_2$.

(a) $c = 99\%, z_c = 2.58$

$$E \approx z_c\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 2.58\sqrt{\frac{11.69^2}{32}+\frac{11.60^2}{32}} = 2.58(2.9113) = 7.5111$$

$$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$$
$$[(69.44 - 59.00) - 7.51] < (\mu_1 - \mu_2) < [(69.44 - 59.00) + 7.51]$$
$$2.93 < (\mu_1 - \mu_2) < 17.95$$

(b) All values in the interval are positive, indicating $\mu_1 > \mu_2$. We are 99% confident that mothers score 2.93 to 17.95 points higher than fathers on the empathy scale, i.e., mothers are more sensitive to baby temperament.

9. (a) $n_1 = 14, \bar{x}_1 = 4.8571 \approx 4.9, s = 2.7695 \approx 2.8$
$n_2 = 16, \bar{x}_2 = 4.1875 \approx 4.2, s = 2.4824 \approx 2.5.$

(b) Since both sample sizes are less than 30, we will use the small sample technique to estimate $\mu_1 - \mu_2$.
$c = 95\%, n_1 + n_2 - 2 = 14 + 16 - 2 = 28, t_{0.95}$ with 28 $d.f. = 2.048$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{13(2.8)^2 + 15(2.5)^2}{14+16-2}} = \sqrt{6.9882} = 2.6435$$

$$E \approx t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.048(2.6435)\sqrt{\frac{1}{14} + \frac{1}{16}} = 1.9813$$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$
$[(4.9 - 4.2) - 1.98] < (\mu_1 - \mu_2) < [(4.9 - 4.2) + 1.98]$
$-1.28 < (\mu_1 - \mu_2) < 2.68$

(c) The interval includes both positive and negative numbers. With 95% confidence, there sums to be no difference in mean number of children by income group.

10. (a) $n_1 = 10, \bar{x}_1 = 75.80, s = 8.3240 \approx 8.32$
$n_2 = 18, \bar{x}_2 = 66.8333 \approx 66.83, s_2 = 8.8667 \approx 8.87$

(b) Since both n_1 and $n_2 < 30$, we will use small sample methods to estimate $\mu_1 - \mu_2$.
$c = 85\%, n_1 + n_2 - 2 = 10 + 18 - 2 = 26, t_{0.85}$ with 26 $d.f. = 1.483$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{9(8.32^2) + 17(8.87^2)}{10+18-2}} = 8.6836$$

$$E \approx t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.483(8.6836)\sqrt{\frac{1}{10} + \frac{1}{18}} = 5.0791$$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$
$[(75.80 - 66.83) - 5.0791] < (\mu_1 - \mu_2) < [(75.80 - 66.83) + 5.0791]$
$3.89 < (\mu_1 - \mu_2) < 14.05$

(c) Because the interval contains only positive numbers, we can claim $\mu_1 > \mu_2$. We are 85% confident that grey wolves in the Chihuahua Region weight 3.89 to 14.05 pounds more, on average, than grey wolves in the Durango Region.

11. $n_1 = 210, r_1 = 65, \hat{p}_1 = 65/210 = 0.3095, \hat{q}_1 = \frac{145}{210} = 0.6905$
$n_2 = 152, r_2 = 18, \hat{p}_2 = 18/152 = 0.1184, \hat{q}_2 = \frac{134}{152} = 0.8816.$
Since $n_1\hat{p}_1 = 65, n_1\hat{q}_1 = 145, n_2\hat{p}_2 = 18$, and $n_2\hat{q}_2 = 134$ are all > 5, we will use large sample methods to estimate $p_1 - p_2$.

(a) $c = 99\%, z_c = 2.58$

$$E \approx z_c \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 2.58\sqrt{\frac{0.3095(0.6905)}{210} + \frac{(0.1184)(0.8816)}{152}} = 2.58(0.0413) = 0.1065$$

$[(\hat{p}_1 - \hat{p}_2) - E] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E]$
$[(0.3095 - 0.1184) - 0.1065] < (p_1 - p_2) < [(0.3095 - 0.1184) + 0.1065]$
$0.0846 < (p_1 - p_2) < 0.2976$, or about 0.085 to 0.298.

(b) The interval consists only of positive values, indicating $p_1 > p_2$. We are 99% confident that the difference in the percentage of traditional Navajo hogans is between 0.085 and 0.298, i.e., there are between 8.5% and 29.8% more hogans in the Fort Defiance Region than in the Indian Wells Region. If it is true that traditional Navajo tend to live in hogans, then, percentage-wise, there are more traditional Navajo at Fort Defiance than at Indian Hills.

12. $\hat{p}_1 = 69/112 = 0.6161, \hat{q}_1 = 0.3839$
$\hat{p}_2 = 26/140 = 0.1857, q_2 = 0.8143$
$n_1\hat{p}_1 = 69, n_1\hat{q}_1 = 43, n_2\hat{p}_2 = 26, n_2\hat{q}_2 = 114$
Since all four of the above estimates exceed 5, we can use large sample methods to estimate $p_1 - p_2$.

(a) $c = 99\%, z_c = 2.58$

$$E \approx z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} = 2.58\sqrt{\frac{0.6161(0.3839)}{112}+\frac{0.1857(0.8143)}{140}} = 2.58(0.0565) = 0.1458$$

$[(\hat{p}_1 - \hat{p}_2) - E] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E]$
$[(0.6161 - 0.1857) - 0.1458] < (p_1 - p_2) < [(0.6161 - 0.1857) + 0.1458]$
$0.2846 < (p_1 - p_2) < 0.5762$, or about 0.28 to 0.58

(b) The interval contains only positive values implying $p_1 > p_2$. We are 99% confident that the difference in percentage unidentified is between 28% and 58%. The higher the altitude, the greater the percentage of unidentified artifacts, which supports the hypothesis.

13. $n_1 = 51, \bar{x}_1 = 74.04, s_1 = 17.19$
$n_2 = 36, \bar{x}_2 = 94.53, s_2 = 19.66$.

Since $n_1, n_2 \geq 30$ we will use large sample methods to estimate $\mu_1 - \mu_2$.

(a) $c = 95\%, z_c = 1.96$

$$E \approx z_c\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 1.96\sqrt{\frac{17.19^2}{51}+\frac{19.66^2}{36}} = 7.9689$$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$
$[(74.04 - 94.53) - 7.9689] < (\mu_1 - \mu_2) < [(74.04 - 94.53) + 7.9689]$
$-28.4589 < (\mu_1 - \mu_2) < -12.5211$, or about -28.46 to -12.52

(b) The confidence interval covers only negative values, indicating $\mu_2 > \mu_1$. We are 95% that the mean buck weight in the Mesa Verde Region is between 12.52 and 28.46 kg more than the mean buck weight in the Cache la Poudre River. More older deer and more abundant browse may help explain the size advantage in the Mesa Verde Region.

14.

	Sample 1	Sample 2
n	316	419
r	259	94
$\hat{p} = \frac{r}{n}$	259/316 = 0.8196	94/419 = 0.2243
$\hat{q} = 1 - \hat{p}$	57/316 = 0.1804	325/419 = 0.7757

Note: Sample 2, which received no plasma compress treatment, could be considered the control group.
$n_1\hat{p}_1 = 259, n_1\hat{q}_1 = 57, n_2\hat{p}_2 = 94, n_2\hat{q}_2 = 325$
Since each of these four estimates ≥ 30, we can use large sample methods to estimate $p_1 - p_2$.

(a) $c = 95\%, z_c = 1.96$

$$E \approx z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.96\sqrt{\frac{0.8196(0.1804)}{316}+\frac{0.2243(0.7757)}{419}} = 1.96(0.0297) = 0.0582$$

$[(\hat{p}_1-\hat{p}_2)-E] < (p_1-p_2) < [(\hat{p}_1-\hat{p}_2)+E]$

$[(0.8196-0.2243)-0.0582] < (p_1-p_2) < [(0.8196-0.2243)+0.0582]$

$0.5371 < (p_1-p_2) < 0.6535$, or about 0.54 to 0.65.

(b) The numbers in the confidence interval are all positive, indicating $p_1 > p_2$, i.e., the proportion of patients with no visible scars is greater among those who received the plasma compress treatment than among those without this treatment. With 95% confidence we can say that the plasma compress treatment increased the proportion of patients with no visible scars by between 54 and 65 percentage points. The treatment seems to be quite effective in reducing scars, based on this data.

15. Because the original group of 45 subjects was randomly split into 3 subgroups of 15, each subgrouping can be considered a random sample, and it is independent of the other subgroups/samples. Since the sample sizes (15) are all < 30, we will use small sample procedures (based on t) to find estimates of $\mu_i-\mu_j$. Because the t distribution requires the data to be normal but is robust against some departures from normality, the authors of this study would have to make a case for their self esteem scores' distributions being at least mound-shaped (unimodal) and symmetric.

$c = 85\%, n_i + n_j - 2 = 15+15-2 = 28. t_{0.85}$ with $28\ d.f. = 1.480$

Preliminary calculations:

$\mu_i-\mu_j$	$\bar{x}_i-\bar{x}_j$	$s_{\text{pooled } ij} = \sqrt{\frac{(n_i-1)s_i^2+(n_j-1)s_j^2}{n_i+n_j-2}}$
1 *versus* 2	19.84 − 19.32 = 0.52	$\sqrt{\frac{14(3.07)^2+14(3.62)^2}{15+15-2}} = 3.3563$
1 *versus* 3	19.84 − 17.88 = 1.96	$\sqrt{\frac{14(3.07)^2+14(3.74)^2}{15+15-2}} = 3.4214$
2 *versus* 3	19.32 − 17.88 = 1.44	$\sqrt{\frac{14(3.62)^2+14(3.74)^2}{15+15-2}} = 3.6805$

$$E_{ij} = t_c s_{\text{spooled } ij}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}} = t_c s_{\text{pooled } ij}\sqrt{\frac{1}{15}+\frac{1}{15}}$$

$$= (1.480)s_{\text{pooled } ij}(0.3651) = s_{\text{pooled } ij}(0.5403)$$

$[(\bar{x}_i-\bar{x}_j)-E_{ij}] < (\mu_i-\mu_j) < [(\bar{x}_i-\bar{x}_j)+E_{ij}]$

(a) $i = 1, j = 2$:

$[0.52-(3.3563)(0.5403)] < (\mu_1-\mu_2) < [0.52+(3.3563)(0.5403)]$

$-1.2934 < (\mu_1-\mu_2) < 2.3334$, or about -1.29 to 2.33

(b) $i = 1, j = 3$:

$[1.96-(3.4214)(0.5403)] < (\mu_1-\mu_3) < [1.96+(3.4214)(0.5403)]$

$0.1114 < (\mu_1-\mu_3) < 3.8086$, or about 0.11 to 3.81

(c) $i = 2, j = 3$:

$[1.44-3.6805(0.5403)] < (\mu_2-\mu_3) < [1.44+3.6805(0.5403)]$

$-0.5486 < (\mu_2-\mu_3) < 3.4286$ or about -0.55 to 3.43

(d) With 85% confidence we can say there is no significant difference between the self-esteem scores on competence and social acceptance, and no significant difference between self-esteem scores on social acceptance and physical attractiveness. However, the interval estimate for $\mu_1 - \mu_3$ contains only positive numbers, indicating that $\mu_1 > \mu_3$. With 85% confidence we can say that the self-esteem score for competence was between 0.11 and 3.81 points higher than that for physical attractiveness.
Notes. (1) There are better ways to study these 3 self-esteem scores than the method used here; however, that technique is beyond the level of the text. (2) The confidence interval formula used in (a)-(c) is designed to capture the true difference, $\mu_i - \mu_j$, 85% of the time. However, the change that all the intervals will simultaneously include the true value, for there is and *j*s is less than 85%. (For further details, consult a more advanced textbook and look up multiple comparisons, comparison wise error rate, and family-wise or experiment wise error rate, or comparison wise, family-wise, and experiment wise confidence coefficients.) (3) Although (b) showed a statistically significant difference (at 85%) between μ_1, and μ_3, the paper's authors would have to argue whether a difference of 0.11 to 3.81 points has any practical significance, especially since 3.81 is very close to s_{pooled}. Statistical significance does not necessarily mean the results have practical significance.

16. Refer to the focus problem at the beginning of this chapter.

Group I: $n_1 = 474, r_1 = 270, \hat{p}_1 = \frac{270}{474} = 0.5696$

Group II: $n_2 = 805, r_2 = 270, \hat{p} = \frac{270}{805} = 0.3354$

(a) $\hat{p}_1 = 0.5696, \hat{q}_1 = 1 - \hat{p}_1 = 0.4304, c = 95\%, z_c = 1.96$
$n_1\hat{p}_1 = 270, n_1\hat{q}_1 = 204$; since the estimate for n_1p_1 and $n_1q_1 > 5$, we can use the large sample approach to find an estimate of p_1.

$$E_1 \approx z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = 1.96\sqrt{\frac{(0.5696)(0.4304)}{474}} = 0.0446$$

$(\hat{p}_1 - E_1) < p_1 < (\hat{p}_1 + E_1), (0.5696 - 0.0446) < p_1 < (0.5696 + 0.0446)$
$0.5250 < p_1 < 0.6055$, or about 0.53 to 0.61

(b) $\hat{p}_2 = 0.3354, \hat{q}_2 = 1 - \hat{p}_2 = 1 - 0.3354 = 0.6646$
$n_2\hat{p}_2 = 270, n_2\hat{q}_2 = 535$; since estimates of n_2p_2 and $n_2q_2 > 5$, we will use large sample methods to find an estimate of p_2.

$$E_2 \approx z_c\sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.96\sqrt{\frac{(0.3354)(0.6646)}{805}} = 0.0326$$

$(\hat{p}_2 - E_2) < p_2 < (\hat{p}_2 + E_2), (0.3354 - 0.0326) < p_2 < (0.3354 + 0.0326)$
$0.3028 < p_2 < 0.3680$, or about 0.30 to 0.37

(c) Since all $n_i\hat{p}_i$ and $n_i\hat{q}_i > 5$, we will use large sample methods to estimate $p_1 - p_2$.

$$E_{12} \approx z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.96\sqrt{\frac{(0.5696)(0.4304)}{474}+\frac{0.3354(0.6646)}{805}} = 1.96(0.0282) = 0.0553$$

$$[(\hat{p}_1 - \hat{p}_2) - E_{12}] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E_{12}]$$

$$[(0.5696 - 0.3354) - 0.0553] < (p_1 - p_2) < [0.5696 - 0.3354) + 0.0553]$$

$0.1789 < (p_1 - p_2) < 0.2895$, or about 0.18 to 0.29.

Since the interval contains only positive numbers, we can say with 95% confidence that $p_1 > p_2$, i.e., the proportion of eggs hatched in well separated and well hidden nesting boxes is greater than the proportion of eggs hatches in highly visible, closely grouped nesting boxes; in fact, with type I nesting boxes, the percentage hatched is roughly 20 to 30 percentage points higher than that of type II nesting boxes.

(d) A greater proportion of wood duck eggs hatch if the eggs are laid in well separated, well hidden, nesting boxes.

17. (a) $[(\bar{x}_1 - \bar{x}_2) - E] < \mu_1 - \mu_2 < [(\bar{x}_1 - \bar{x}_2) + E]$

where $E \approx z_c\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

$z_{0.90} = 1.645$, $z_{0.95} = 1.96$, $z_{0.99} = 2.58$

As we change from one c confidence interval to another, the only number that changes is z_c. The interval width is $2E$, and it depends on z_c; if c increases, z_c increases, and the interval gets wider; if c decreases, z_c decreases, and the interval gets narrower. The larger interval (larger z_c) always includes all the points that were in a smaller interval created from a smaller z_c. All intervals are centerd at $(\bar{x}_1 - \bar{x}_2)$.

Therefore, if a 95% confidence interval includes both positive and negative numbers, the 99% confidence interval must include both positive and negative numbers. However, in going from 95% to 90%, the width of the interval decreases evenly from both sides. If $(\bar{x}_1 - \bar{x}_2)$ is near 0, the 90% confidence interval could have both positive and negative values, just like the 95% confidence interval did, and all values in the 90% confidence interval would have also been in the 95% confidence interval. However, the farther $(\bar{x}_1 - \bar{x}_2)$ is from zero, the more likely it is that the narrower 90% confidence interval will contain only positive or only negative numbers.

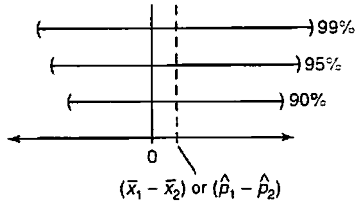
(b) $[(\hat{p}_1 - \hat{p}_2) - E] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E]$

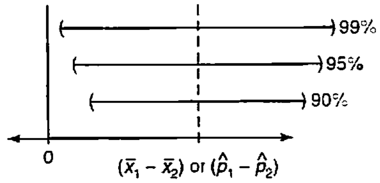
where $E \approx z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$

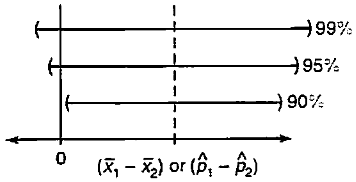
As in (a), the only number that changes (and, therefore, affects the width) is z_c, and all the intervals are centered at $(\hat{p}_1 - \hat{p}_2)$.

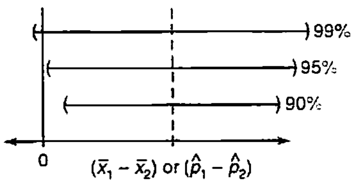
If a 95% confidence interval contains only positive values, a 99% confidence interval could contain all positive numbers, or both positive and negative numbers, depending on how close $(\hat{p}_1 - \hat{p}_2)$ is to zero. However, with a narrower 90% confidence interval, if the 95% confidence interval had only positive values, the 90% interval will, too, and every value in the 90% interval would also have been in the 95% confidence interval.

Note: the principle is the same whether it is a confidence interval for $\mu, \mu_1 - \mu_2, p, p_1 - p_2$, etc. One can make statements paralleling those of (b) for the 95% confidence interval having only negative values.









18. (a) $E = z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ Let $n = n_1 = n_2$; then $E = z_c\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}} = \frac{z_c}{\sqrt{n}}\sqrt{\sigma_1^2 + \sigma_2^2}$

Solve for n:

$\sqrt{n}E = z_c\sqrt{\sigma_1^2 + \sigma_2^2}$ Multiply both sides by $\sqrt{n}$.

$\sqrt{n} = \frac{z_c}{E}\sqrt{\sigma_1^2 + \sigma_2^2}$ Divide both sides by E.

$n = \left(\frac{z_c}{E}\right)^2 (\sigma_1^2 + \sigma_2^2)$ Square both sides.

Note that this is $n_1 = n$ *and* $n_2 = n$; $2n$ units total.

(b) $c = 95\%, z_c = 1.96, E = 0.05$

$s_1 = 0.37, s_2 = 0.31$ from Problem 1 above

$$n \approx \left(\frac{z_c}{E}\right)^2 (s_1^2 + s_2^2) = \left(\frac{1.96}{0.05}\right)^2 (0.37^2 + 0.31^2) = 358.037 \approx 359$$

so n_1 and n_2 should each be 359.

(c) $c = 90\%$ so $z_c = 1.645, E = 0.5$

$s_1 = 4.4, s_2 = 3.8$ from Problem 2 above

$$n \approx \left(\frac{z_c}{E}\right)^2 (s_1^2 + s_2^2) = \left(\frac{1.645}{0.5}\right)^2 (4.4^2 + 3.8^2) = 365.85 \approx 366$$

so each n_1 and n_2 should be 366.

19. $E = z_c\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} = \frac{z_c}{\sqrt{n}}\sqrt{p_1q_1 + p_2q_2}$ if $n = n_1 = n_2$

So $\sqrt{n}E = z_c\sqrt{p_1q_1 + p_2q_2}$ Multiply both sides by $\sqrt{n}$.

$\sqrt{n} = \frac{z_c}{E}\sqrt{p_1q_1 + p_2q_2}$ Divide both sides by E.

$n = \left(\frac{z_c}{E}\right)^2 (p_1q_1 + p_2q_2)$ Square both sides.

so $n \approx \left(\frac{z_c}{E}\right)^2 (\hat{p}_1\hat{q}_1 + \hat{p}_2\hat{q}_2)$

If we have no estimate for p_i, we would use the "worst case" estimate, i.e., the conservative approach would be to make n as large as possible, by using $\hat{p}_1 = 0.5, \hat{q}_1 = 1 - \hat{p}_1 = 0.5$

So, in this case,

$$n \approx \left(\frac{z_c}{E}\right)^2 [(0.5)(0.5) + (0.5)(0.5)] = 0.5\left(\frac{z_c}{E}\right)^2 \text{ or } \left(\frac{1}{2}\right)\left(\frac{z_c}{E}\right)^2$$

Again, all sample size estimates are the minimum number meeting the stated criteria.

(a) $c = 99\%, z_c = 2.58, E = 0.04.$

$\hat{p}_1 = \frac{289}{375} = 0.7707, \hat{q}_1 = 1 - \hat{p}_1 = 0.2293.$

$\hat{p}_2 = \frac{23}{571} = 0.0403, \hat{q}_2 = 0.9597$

(Recall the n_ip_i and n_iq_i conditions were checked in Problem 5 above.)

$$n \approx \left(\frac{z_c}{E}\right)^2 (\hat{p}_1\hat{q}_1 + \hat{p}_2\hat{q}_2) = \left(\frac{2.58}{0.04}\right)^2 [0.7707(0.2293) + 0.0403(0.9597)] = 896.107 \approx 897$$

n_1 and n_2 should each be 897 (married couples)

(b) Let $\hat{p}_i = \hat{q}_i = 0.5; c = 95\%, z_c = 1.96, E = 0.05$

$$n \approx \left(\frac{1}{2}\right)\left(\frac{z_c}{E}\right)^2 = \left(\frac{1}{2}\right)\left(\frac{1.96}{0.05}\right)^2 = 768.32 \approx 769$$

n_1 and n_2 should each be 769 (married couples)

Chapter 8 Review

1. point estimate: a single number used to estimate a population parameter
critical value: the x-axis values (arguments) of a probability density function (such as the standard normal or Student's t) which cut off an area of $c, 0 \le c \le 1$, under the curve between them. Examples: the area under the standard normal curve between $-z_c$ and $+z_c$ is c; the area under the curve of a Student's t distribution between $-t_c$ and $+t_c$ as c. The area is symmetric about the curve's mean, μ.
maximal error of estimate, E: the largest distance ("error") between the point estimate and the parameter it estimates that can be tolerated under certain circumstances; E is the half-width of a confidence interval.
confidence level, c: A measure of the reliability of an (interval) estimate: c denotes the proportion of all possible confidence interval estimates of a parameter (or difference between 2 parameters) that will cover/capture/enclose the true value being estimated. It is a statement about the probability the procedure being used has of capturing the value of interest; it cannot be considered a measure of the reliability of a specific interval, because any specific interval is either right or wrong-either it captures the parameter value, or it does not, period.
confidence interval: a procedure designed to give a range of values as an (interval) estimate of an unknown parameter value; compare point estimate. What separates confidence interval estimates from any other interval estimate (such as 4, give or take 2.8) is that the reliability of the procedure can be determined: if $c = 0.90 = 90\%$, for example, a 90% confidence interval (estimate) for μ says that it all possible samples of size n were drawn, and a 90% confidence interval for μ was created for each such sample using the prescribed method (such as $\bar{x} \pm z_c\sigma/\sqrt{n}$), then if the true value of μ became known, 90% of the confidence intervals so created would cover/capture/enclose the value of μ.
large/small samples: a large sample in our context is one that is of sufficient size to warrant using a normal approximation to the exact method, i.e., large enough that the central limit theorem can reasonably be applied and that the approximation results in sufficiently accurate estimates of the exact method results. We have said that $n \ge 30$ is large enough, in all but the most extreme cases, to say that the Student's t-distribution can be approximated by the normal and that s^2 can be used to estimate σ^2. Similarly, if np_i and nq_i are both greater than 5, the normal distribution can be used to approximate the exact binomial calculations of the probability of r successes.
Small samples are those of a size where using a normal approximation instead of the exact method would give unreliable results; the difference between the exact and appropriate answers is too large to be tolerated. For Student's t distribution, if $n < 30$, the normal approximation results are considered too crude to be useful. For the central limit theorem to be applied when estimating p or $p_1 - p_2$, or, specifically, for the normal approximation to the binomial to be applied we have said np_i and nq_i must both be >5.* Here the criteria are the products np_i and nq_i, not just the size of n. A sample of size $n = 300$ seems large enough for just about any purpose, but if $p = 0.01$, $np_i = 300(0.01) = 3 \le 5$, and $n = 300$ is not large enough for a normal approximation to be accurate enough. In general, if p (or q) is near 0 or near 1, n must be quite large before the normal approximation can be used.
*Some textbooks use other criteria or rules of thumb, such as np_i and $nq_i > 10$.

2. $n = 370 \ge 30$ is sufficiently large to use large sample procedures to estimate μ.
$\bar{x} = 750, s = 150, c = 0.90$ so $z_c = 1.645$, or $c = 0.99$ so $z_c = 2.58$
$$E \approx \frac{z_c s}{\sqrt{n}} = z_c \frac{150}{\sqrt{370}} = 7.7981 z_c$$
$$(\bar{x} - E) < \mu < (\bar{x} + E)$$
For $c = 0.90$: $E \approx 7.7981(1.645) = 12.8279 \approx 13$
$$(750 - 13) < \mu < (750 + 13)$$
$$\$737 < \mu < \$763$$
For $c = 0.99$: $E \approx 7.7981(2.58) = 20.1191 \approx 20$
$$(750 - 20) < \mu < (750 + 20)$$
$$\$730 < \mu < \$770$$

3. $n = 73 \geq 30$; use a large sample procedure to estimate μ.
$\bar{x} = 178.70, s = 7.81, c = 95\%, z_c = 1.96$
$$E \approx \frac{z_c s}{\sqrt{n}} = \frac{1.96(7.81)}{\sqrt{73}} = 1.7916 \approx 1.79$$
$(\bar{x} - E) < \mu < (\bar{x} + E)$
$(178.70 - 1.79) < \mu < (178.70 + 1.79)$
$176.91 < \mu < 180.49$

4. $c = 99\%, z_c = 2.58, E = 2, s = 7.81$ from Problem 3 above
$$n = \left(\frac{z_c s}{E}\right)^2 = \left[\frac{2.58(7.81)}{2}\right]^2 = 101.5036 \approx 102$$

5. (a) $\bar{x} = 74.2, s = 18.2530 \approx 18.3$, as indicated
   (b) $c = 95\%, n = 15$ so use small sample procedure to estimate μ
   $t_{0.95}$ with $n - 1 = 14$ $d.f. = 2.145$
   $$E \approx \frac{t_c s}{\sqrt{n}} = \frac{2.145(18.3)}{\sqrt{15}} = 10.1352 \approx 10.1$$
   $(\bar{x} - E) < \mu < (\bar{x} + E)$
   $(74.2 - 10.1) < \mu < (74.2 + 10.1)$
   64.1 centimeters $< \mu <$ 84.3 centimeters

6. (a) $n = 10, \bar{x} = 15.78 \approx 15.8, s = 3.4608 \approx 3.5$, as indicated
   (b) estimate μ with small sample procedure because $n = 10 < 30$
   $c = 80\%, n - 1 = 9$ $d.f., t_{0.80}$ with 9 $d.f. = 1.383$
   $$E \approx \frac{t_c s}{\sqrt{n}} = \frac{1.383(3.5)}{\sqrt{10}} = 1.5307 \approx 1.53$$
   $(\bar{x} - E) < \mu < (\bar{x} + E)$
   $(15.8 - 1.53) < \mu < (15.8 + 1.53)$
   14.27 centimeters $< \mu <$ 17.33 centimeters

7. $n = 2958, r = 1538, \hat{p} = \frac{r}{n} = \frac{1538}{2958} = 0.5199 \approx 0.52$
$\hat{q} = 1 - \hat{p} = 0.4801, n - r = 1420, c = 90\%, z_c = 1.645$
$np \approx n\hat{p} = r = 1538 > 5, nq \approx n\hat{q} = n - r = 1420 > 5$, so use large sample method to estimate p
$$E \approx z_c\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645\sqrt{\frac{(0.5199)(0.4801)}{2958}} = 0.0151 \approx 0.02$$
$(\hat{p} - E) < p < (\hat{p} + E)$
$(0.52 - 0.02) < p < (0.52 + 0.02)$
$0.50 < p < 0.54$

8. $95\%, z_c = 1.96$, preliminary estimate $\hat{p} = 0.52$
$E = 0.01, \hat{q} = 1 - \hat{p} = 0.48$
$$n = \left(\frac{z_c}{E}\right)^2 \hat{p}\hat{q} = \left(\frac{1.96}{0.01}\right)^2 (0.52)(0.48) = 9{,}588.6336$$
sample size of 9589

9. $n = 167,\ r = 68$

(a) $\hat{p} = \frac{r}{n} = \frac{68}{167} = 0.4072, \hat{q} = 0.5928, n - r = 99$

(b) $n\hat{p} = r = 68 > 5$ and $n\hat{q} = n - r = 99 > 5$, so use large sample method to estimate p

$c = 95\%, z_c = 1.96$

$E \approx z_c\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(0.4072)(0.5928)}{167}} = 0.0745$

$(\hat{p} - E) < p < (\hat{p} + E)$

$(0.4072 - 0.0745) < p < (0.4072 + 0.0745)$

$0.3327 < p < 0.4817$, or about 0.333 to 0.482

10. $c = 95\%, z_c = 1.96, E = 0.06, \hat{p} = 0.4072, \hat{q} = 0.5928$ from Problem 9

$n \approx \left(\frac{z_c}{E}\right)^2 \hat{p}\hat{q} = \left(\frac{1.96}{0.06}\right)^2 (0.4072)(0.5928) = 257.5880 \approx 258$ potshards

Since Problem 9 says 167 potshards have already been collected, we need 258 – 167 = 91 additional potshards to be collected.

11. $n_1 = 43, \bar{x}_1 = 3.6, s_1 = 1.8$
$n_2 = 40, \bar{x}_2 = 3.3, s_2 = 1.7$

(a) Since $n_i \geq 30$, we can use a large sample approach to estimating $\mu_1 - \mu_2$

$c = 90\%,\ z_c = 1.645$

$E \approx z_c\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.645\sqrt{\frac{1.8^2}{43} + \frac{1.7^2}{40}} = 0.6320$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$

$[(3.6 - 3.3) + 0.632] < (\mu_1 - \mu_2) < [(3.6 - 3.3) + 0.632]$

-0.332 percent $< (\mu_1 - \mu_2) < 0.932$ percent

(negative value means a salary decrease)

(b) Since the interval contains both positive and negative values, with confidence 90%, there appears to be no significant difference in percent salary increases between western and eastern colleges.

12. $n_1 = 32, \bar{x}_1 = 13.7, s_1 = 4.1$
$n_2 = 34, \bar{x}_2 = 10.1, s_2 = 2.7$

(a) Since $n_i \geq 30$, we can use large sample procedures to estimating $\mu_1 - \mu_2$.

$c = 95\%,\ z_c = 1.96$

$E \approx z_c\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.96\sqrt{\frac{4.1^2}{32} + \frac{2.7^2}{34}} = 1.6857 \approx 1.69$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$

$[(13.7 - 10.1) - 1.69] < (\mu_1 - \mu_2) < [(13.7 - 10.1) + 1.69]$

1.91 percent of stockholder equity $< (\mu_1 - \mu_2) < 5.29$ percent of stockholder equity

(b) Since all interval values are positive it appears $\mu_1 > \mu_2$, i.e., in terms of profit as a percentage of stockholder equity, retail stores do better than utilities. The difference is estimated to be between 1.9 and 5.3 percentage points, with 95% confidence.

13. $n_1 = 18, \bar{x}_1 = 98, s_1 = 6.5$
$n_2 = 24, \bar{x}_2 = 90, s_2 = 7.3$

(a) Since $n_i < 30$, use small sample procedures to estimate $\mu_1 - \mu_2$.
$c = 75\%$, $n_1 + n_2 - 2 = 40, t_{0.75}$ with 40 $d.f. = 1.167$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{17(6.5)^2 + 23(7.3)^2}{18+24-2}} = 6.9712$$

$$E \approx t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.167(6.9712)\sqrt{\frac{1}{18} + \frac{1}{24}} = 2.5367 \approx 2.54$$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$
$[(98-90) - 2.54] < (\mu_1 - \mu_2) < [(98-90) + 2.54]$
5.46 pounds $< (\mu_1 - \mu_2) <$ 10.54 pounds

(b) Since the interval contains only positive values, we can say with 75% confidence, that $\mu_1 > \mu_2$, i.e., that Canadian wolves weight more than Alaska wolves, and that the difference is approximately 5.5 to 10.5 pounds.

14. $n_1 = 17, \bar{x}_1 = 4.9, s_1 = 1.0$
$n_2 = 6, \bar{x}_2 = 2.8, s_2 = 1.2$
Since $n_i < 30$, use small sample procedures to estimate $\mu_1 - \mu_2$.

(a) $c = 85\%$. $n_1 + n_2 - 2 = 17 + 6 - 2 = 21, t_{0.85}$ with 21 $d.f. = 1.494$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{16(1.0)^2 + 5(1.2)^2}{17+6-2}} = 1.0511$$

$$E \approx t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.494(1.0511)\sqrt{\frac{1}{17} + \frac{1}{6}} = 0.7457 \approx 0.75$$

$[(\bar{x}_1 - \bar{x}_2) - E] < (\mu_1 - \mu_2) < [(\bar{x}_1 - \bar{x}_2) + E]$
$[(4.9-2.8) - 0.75] < (\mu_1 - \mu_2) < [(4.9-2.8) + 0.75]$
$1.35 < (\mu_1 - \mu_2) < 2.85$ wolf pups per litter

(b) The interval includes only positive values, so we can say $\mu_1 > \mu_2$, i.e., with 85% confidence, the average litter size in Canada is larger than that in Finland by 1.35 to 2.85 wolf pups.

15. $n_1 = 93, r_1 = 79, \hat{p}_1 = \frac{79}{93} = 0.8495, \hat{q}_1 = 0.1505, n_1 - r_1 = 14$
$n_2 = 83, r_2 = 74, \hat{p}_2 = \frac{74}{83} = 0.8916, \hat{q}_2 = 0.1084, n_2 - r_2 = 9$
Since $n_i\hat{p}_i$ and $n_i\hat{q}_i$ are all > 5, we can use large sample procedures to estimate $p_1 - p_2$.

(a) $c = 95\%, z_c = 1.96$

$$E \approx z_c \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.96\sqrt{\frac{(0.8495)(0.1505)}{93} + \frac{(0.8916)(0.1084)}{83}} = 0.0988$$

$[(\hat{p}_1 - \hat{p}_2) - E] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E]$
$[(0.8495 - 0.8916) - 0.0988] < (p_1 - p_2) < [(0.8495 - 0.8916) + 0.0988]$
$-0.1409 < (p_1 - p_2) < 0.0567$

(b) Since the interval contains positive and negative values, we can say, with 95% confidence, that there is no significant differences between the proportion of accurate responses for face-to-face interviews and that for telephone interviews.

16. $n_1 = 30, r_1 = 16, \hat{p}_1 = \frac{16}{30} = 0.5333, \hat{q}_1 = 0.4667, n_1 - r_1 = 14$

$n_2 = 46, r_2 = 25, \hat{p}_2 = \frac{25}{46} = 0.5435, \hat{q}_2 = 0.4565, n_2 - r_2 = 21$

Since $n_i\hat{p}_i$ and $n_i\hat{q}_i$ are all > 5, we can use large sample procedure to estimate $p_1 - p_2$.

(a) $c = 90\%, z_c = 1.645$

$$E \approx z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.645\sqrt{\frac{0.5333(0.4667)}{30} + \frac{(0.5435)(0.4565)}{46}} = 0.1925$$

$[(\hat{p}_1 - \hat{p}_2) - E] < (p_1 - p_2) < [(\hat{p}_1 - \hat{p}_2) + E]$

$[(0.5333 - 0.5435) - 0.1925] < (p_1 - p_2) < [(0.5333 - 0.5435) + 0.1925]$

$-0.2027 < (p_1 - p_2) < 0.1823$

(b) Since the interval contains both positive and negative values, we can conclude, with 90% confidence, that there is no significant difference between the proportion of accurate responses in face-to-face interviews and that in telephone interviews.

17. (a) $P(A_1 < \mu_1 < B_1) = 0.80$

$P(A_2 < \mu_2 < B_2) = 0.80$

i.e., the two intervals are designed so that the confidence interval procedure produces intervals A_i to B_i that capture μ_i 80% of the time.

$P(A_1 < \mu_1 < B_1 \text{ and } A_2 < \mu_2 < B_2) = P(A_1 < \mu_1 < B_1) \cdot P(A_2 < \mu_2 < B_2, \text{ given } A_1 < \mu_1 < B_1)$

but the intervals were created using independent samples, so the intervals themselves are independent, so

$= P(A_1 < \mu_1 < B_1) \cdot P(A_2 < \mu_2 < B_2)$

by the definition of independent events: if *C* and *D* are independent, *P*(*C*, given *D*) = *P*(*C*).

$= (0.80)(0.80)$

$= 0.64$

The probability that both intervals are simultaneously correct, i.e., that both intervals capture their μ_i, is 0.64.

$P(\text{at least one interval fails to capture its } \mu_i) = 1 - P(\text{both intervals capture their } \mu_i)$

$= 1 - 0.64$

$= 0.36$

[There are 4 possible outcomes. Using the (*x*, *y*) interval notation, they are

(1) $\mu_1 \in (A_1, B_1), \mu_2 \in (A_2, B_2)$ both capture μ_i

(2) $\mu_1 \in (A_1, B_1), \mu_2 \notin (A_2, B_2)$ only μ_1, is captured

(3) $\mu_1 \notin (A_1, B_1), \mu_2 \in (A_2, B_2)$ only μ_2 is captured

(4) $\mu_1 \notin (A_1, B_1), \mu_2 \notin (A_2, B_2)$ neither μ_i is captured

By the complimentary event rule, *P*(at least 1 fails) = *P*[case (2), (3), or (4)] = 1 − *P*[case (1)]

(b) $P(A_1 < \mu_1 < B_1) = c$
$P(A_2 < \mu_2 < B_2) = c$

(Both confidence intervals are at level c.)

$$\begin{aligned} 0.90 &= P(A_1 < \mu_1 < B_1 \text{ and } A_2 < \mu_2 < B_2) \\ &= P(A_1 < \mu_1 < B_1) \cdot P(A_2 < \mu_2 < B_2) \quad \text{since the intervals are independent} \\ &= c \cdot c \\ &= c^2 \end{aligned}$$

If $0.90 = c^2$, then $\sqrt{0.90} = c = 0.9487$, or about 0.95

(c) Answers vary.
In large, complex engineering designs, each component must be within design specifications or the project will fail.
Consider the hundreds (if not thousands) of components which must function properly to launch the space shuttle, keep it orbiting, and return it safely to earth. For example, nuts, bolts, rivets, wiring, and the like must be a certain size, give or take some tiny amount. Tiles and the glue securing them must be able to withstand a huge range of temperatures, from the ambient air temperature at launch time to the extreme heat of re-entry.
Each of the design specifications can be thought of as a confidence interval. Manufacturers and suppliers want to be very, very confident their parts are well within the specifications, or they might lose their contracts to competitors. Similarly, NASA wants to be very, very, confident all the parts, as a group, meet specifications; otherwise, costly delays or catastrophic failures may occur. (Recall that much of the challenger disaster was due to o-ring failure-because NASA decided to go ahead with the launch even though they had been warned by the o-ring manufacturer that the temperature at the launch site was below the lowest temperature at which the o-rings had been tested, and that the o-rings might not completely seat at that temperature.)
If NASA will tolerate only a 1 in 1,000 or 1 in 1,000,000 chance of failure, i.e., $c = 0.999$ or $c = 0.999999$, the individual components' confidence levels, c must be (much) higher than NASA's.