Chapter 16 – Sampling Distribution Models and Confidence Intervals for Proportions

Section 16.1

1. Website.

a) Since the sample is drawn at random, and assuming that 200 investors is a small portion of their customers, the sampling distribution for the proportion of 200 investors that use smartphones will be unimodal and symmetric (roughly Normal).

b) The center of the sampling distribution of the sample proportion is 0.36.

c) The standard deviation of the sample proportion is $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.36)(0.64)}{200}} \approx 0.034.$

2. Marketing.

a) The proportion of women in the sample is expected to be 0.51.

b) The standard deviation of the sample proportion is $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.51)(0.49)}{400}} \approx 0.025.$

c) We would expect to find $(0.51)(400) = 204$ women in a sample of 400.

3. Send money.

All of the histograms are centered around $p = 0.05$. As n gets larger, the shape of the histograms gets more unimodal and symmetric, approaching a Normal model, while the variability in the sample proportions decreases.

4. Character recognition.

All of the histograms are centered around $p = 0.85$. As n gets larger, the shapes of the histograms get more unimodal and symmetric, approaching a Normal model, while the variability in the sample proportions decreases.

5. Living online.

a) This means that 56% of the 1060 teens in the sample said they go online several times per day. This is our best estimate of p, the proportion of all U.S. teens who would say they have done so.

b) $SE(\hat{p}) = \sqrt{\frac{(0.56)(0.44)}{1060}} \approx 0.0152$

c) Because we don't know p, we use $\hat{p}$ to estimate the standard deviation of the sampling distribution. The standard error is our estimate of the amount of variation in the sample proportion we expect to see from sample to sample when we ask 1060 teens whether they go online several times per day.

6. How's life?

a) This means that 55.7% of the 105,000 people in the sample considered themselves to be thriving. This is our best estimate of p, the proportion of all U.S. people who would consider themselves as thriving.

b) $SE(\hat{p}) = \sqrt{\frac{(0.557)(0.443)}{105{,}000}} \approx 0.00153$

c) Because we don't know p, we use $\hat{p}$ to estimate the standard deviation of the sampling distribution. The standard error is our estimate of the amount of variation in the sample proportion we expect to see in the proportion of people who say they are thriving from sample to sample when we ask 105,000 people.

Section 16.2

7. Marriage.

The data come from a random sample, so the randomization condition is met. We don't know the exact value of p, we can estimate it with $\hat{p}$. $n\hat{p} = (1500)(0.27) = 405$, and $n\hat{q} = (1500)(0.73) = 1095$. So, there are well over 10 successes and 10 failures, meeting the Success/Failure Condition. Since there are more than $(10)(1500) = 15,000$ adults in the United States, the 10% Condition is met. A Normal model is appropriate for the sampling distribution of the sample proportion.

8. Campus sample.

Stacy plans to use a random sample, so the randomization condition is met. However, $np = (50)(0.10) = 5$, which is less than 10. The Success/Failure condition is not met. It is not appropriate to use a Normal model for the sampling distribution of the sample proportion.

9. Send more money.

a) The histogram for $n = 200$ looks quite unimodal and symmetric. We should be able to use the Normal model.

b) The Success/Failure condition requires np and nq to both be at least 10, which is not satisfied until $n = 200$ for $p = 0.05$. The theory supports the choice in part (a).

10. Character recognition, again.

a) Certainly, the histogram for $n = 100$ is unimodal and symmetric, but the histogram for $n = 75$ looks nearly Normal, too. We should be able to use the Normal model for either.

b) The success/failure condition requires np and nq to both be at least 10, which is satisfied for both $n = 75$ and $n = 100$ when $p = 0.85$. The theory supports the choice in part (a).

11. Sample maximum.

a) A Normal model is not appropriate for the sampling distribution of the sample maximum. The histogram is skewed strongly to the left.

b) No. The 95% rule is based on the Normal model, and the Normal model is not appropriate here.

12. Soup.

a) A Normal model is not appropriate for the sampling distribution of the sample variances. The histogram is skewed to the right.

b) No. The 95% rule is based on the Normal model, and the Normal model is not appropriate here.

Section 16.3

13. Still living online.

a) We are 95% confident that, if we were to ask all U.S. teens whether the go online several times per day, between 53% and 59% of them would say they do.

b) If we were to collect many random samples of 1060 teens, about 95% of the confidence intervals we construct would contain the proportion of all U.S. teens who say they go online several times per day.

14. Spanking.

a) We are 95% confident that, if we were to ask all parents about spanking their children, between 50.6% and 55.4% would say they never do.

b) If we were to collect many random samples of 1807 parents, about 95% of the confidence intervals we construct would contain the proportion of all parents who would say they never spank their children.

Section 16.4

15. Wrong direction.

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645 \times \sqrt{\frac{(0.54)(0.46)}{2214}} \approx 0.017$ or 1.7%

b) We are 90% confident that the observed proportion responding "Wrong Track" is within 0.017 of the population proportion.

16. Smoking.

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 2.576 \times \sqrt{\frac{(0.18)(0.82)}{105{,}000}} \approx 0.0031$ or 0.31%

b) We are 99% confident that the observed proportion of smokers is within 0.0031 of the population proportion.

Section 16.5

17. Wrong direction again.

a) The sample is a simple random sample. Both $n\hat{p} = (2214)(0.54) = 1196$ and $n\hat{q} = (2214)(0.46) = 1018$ are at least ten. The sample size of 2214 is less than 10% of the population of all US voters. Conditions for the confidence interval are met.

b) The margin of error for 95% confidence would be larger. The critical value of 1.96 is greater than the 1.645 needed for 90% confidence.

18. More spanking.

a) The sample is a simple random sample. Both $n\hat{p} = (1807)(0.53) = 958$ and $n\hat{p} = (1807)(0.96) = 849$ are at least ten. The sample size of 1807 is less than 10% of the population of parents. Conditions for the confidence interval are met.

b) The margin of error for 95% confidence would be smaller. The critical value of 1.96 is less than the 2.576 needed for 99% confidence.

Section 16.6

19. Graduation.

a)
$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.06 = 1.645\sqrt{\frac{(0.25)(0.75)}{n}}$$
$$n = \frac{(1.645)^2(0.25)(0.75)}{(0.06)^2}$$
$$n \approx 141 \text{ people}$$

In order to estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 6% with 90% confidence, we would need a sample of at least 141 people. All decimals in the final answer must be rounded up, to the next person.

(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 188 people.)

19. (continued)

b) $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$0.04 = 1.645\sqrt{\frac{(0.25)(0.75)}{n}}$$

$$n = \frac{(1.645)^2(0.25)(0.75)}{(0.04)^2}$$

$n \approx 318$ people

In order to estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 4% with 90% confidence, we would need a sample of at least 318 people. All decimals in the final answer must be rounded up, to the next person.

(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 423 people.)

Alternatively, the margin of error is now 2/3 of the original, so the sample size must be increased by a factor of 9/4. 141(9/4) ≈ 318 people.

c) $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$0.03 = 1.645\sqrt{\frac{(0.25)(0.75)}{n}}$$

$$n = \frac{(1.645)^2(0.25)(0.75)}{(0.03)^2}$$

$n \approx 564$ people

In order estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 3% with 90% confidence, we would need a sample of at least 564 people. All decimals in the final answer must be rounded up, to the next person. (For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 752 people.)

Alternatively, the margin of error is now half that of the original, so the sample size must be increased by a factor of 4. 141(4) ≈ 564 people.

20. Hiring.

a) $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$0.05 = 2.326\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$n = \frac{(2.326)^2(0.5)(0.5)}{(0.05)^2}$$

$n \approx 542$ businesses

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 5% with 98% confidence, we would need a sample of at least 542 businesses. All decimals in the final answer must be rounded up, to the next business.

20. (continued)

b) $ME = z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

$$0.03 = 2.326\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$n = \frac{(2.326)^2(0.5)(0.5)}{(0.03)^2}$$

$$n \approx 1504 \text{ businesses}$$

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 3% with 98% confidence, we would need a sample of at least 1503 businesses. All decimals in the final answer must be rounded up, to the next business.

(Alternatively, the margin of error is being decreased to 3/5 of its original size, so the sample size must increase by a factor of 25/9. 542(25/9) ≈ 1506 businesses. A bit off, because 542 was rounded, but close enough!

c) $ME = z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

$$0.01 = 2.326\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$n = \frac{(2.326)^2(0.5)(0.5)}{(0.01)^2}$$

$$n \approx 13{,}526 \text{ businesses}$$

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 1%with 98% confidence, we would need a sample of at least 13,526 businesses.

(Alternatively, the margin of error has been decreased to 1/5 of its original size, so a sample 25 times as large would be needed. 25(542) = 13,550. Close enough!

It would probably be very expensive and time consuming to sample that many businesses.

Chapter Exercises

21. Margin of error.

The newscaster believes the true proportion of voters with a certain opinion is within 4% of the estimate, with some degree of confidence, perhaps 95% confidence.

22. Another margin of error.

He believes the true percentage of children who are exposed to lead-based paint is within 3% of his estimate, with some degree of confidence, perhaps 95% confidence.

23. Conditions.

a) *Population* – all cars; *sample* – 134 cars actually stopped at the checkpoint;
p – proportion of all cars with safety problems; $\hat{p}$ – proportion of cars in the sample that actually have safety problems (10.4%).
Randomization condition: This sample is not random, so hopefully the cars stopped are representative of cars in the area.
10% condition: The 134 cars stopped represent a small fraction of all cars, certainly less than 10%.
Success/Failure condition: $n\hat{p} = 14$ and $n\hat{q} = 120$ are both greater than 10, so the sample is large enough.

A one-proportion z-interval can be created for the proportion of all cars in the area with safety problems.

23. (continued)

b) *Population* – the general public; *sample* – 602 viewers that logged on to the Web site; p –proportion of the general public that support prayer in school; $\hat{p}$ - proportion of viewers that logged on to the Web site and voted that support prayer in schools (81.1%).
Randomization condition: This sample is not random, but biased by voluntary response.
It would be very unwise to attempt to use this sample to infer anything about the opinion of the general public related to school prayer.

c) *Population* – parents at the school; *sample* – 380 parents who returned surveys; p – proportion of all parents in favor of uniforms; $\hat{p}$ – proportion of those who responded that are in favor of uniforms (60%).
Randomization condition: This sample is not random, but rather biased by nonresponse. There may be lurking variables that affect the opinions of parents who return surveys (and the children who deliver them!).
It would be very unwise to attempt to use this sample to infer anything about the opinion of the parents about uniforms.

d) *Population* – all freshmen enrollees at the college (not just one year); *sample* – 1632 freshmen during the specified year; p – proportion of all students who will graduate on time; $\hat{p}$ – proportion of on time graduate that year (85.05%).
Randomization condition: This sample is not random, but this year's freshmen class is probably representative of freshman classes in other years.
10% condition: The 1632 students in that years freshmen class represent less than 10% of all possible students.
Success/Failure condition: $n\hat{p} = 1388$ and $n\hat{q} = 244$ are both greater than 10, so the sample is large enough.
A one-proportion z-interval can be created for the proportion of freshmen that graduate on time from this college.

24. More conditions.

a) *Population* – all customers who recently bought new cars; *sample* – 167 people surveyed about their experience; p – proportion of all new car buyers who are dissatisfied with the salesperson; $\hat{p}$ – proportion of new car buyers surveyed who are dissatisfied with the salesperson (3%).
Success/Failure condition: $n\hat{p} = 167(0.03) = 5$ and $n\hat{q} = 162$. Since only 5 people were dissatisfied, the sample is **not** large enough to use a confidence interval to estimate the proportion of dissatisfied car buyers.

b) *Population* – all college students; *sample* – 2883 who were asked about their cell phones at the football stadium; p – proportion of all college students with cell phones; $\hat{p}$ – proportion of college students at the football stadium with cell phones (84.3%).
Randomization condition: This sample is not random. The best we can hope for is that the students at the football stadium are representative of all college students.
10% condition: The 2883 students at the football stadium represent less than 10% of all college students.
Success/Failure condition: $n\hat{p} = 2430$ and $n\hat{q} = 453$ are both greater than 10, so the sample is large enough.
Extreme caution should be used when using a one-proportion z-interval to estimate the proportion of college students with cell phones. The students at the football stadium may not be representative of all students.

24. (continued)

c) *Population* – potato plants in the U.S.; *sample* – 240 potato plants in a field in Maine; p – proportion of all potato plants in the U.S. that show signs of blight;
$\hat{p}$ – proportion of potato plants in the sample that show signs of blight (2.9%).
Randomization condition: Although potato plants are randomly selected from the field in Maine, it doesn't seem reasonable to assume that these potato plants are representative of all potato plants in the U.S.
Success/Failure condition: $n\hat{p} = 7$ and $n\hat{q} = 233$. There are only 7 (less than 10!) plants with signs of blight. The sample is not large enough.
Three conditions are not met! Don't use a confidence interval to attempt to estimate the percentage of potato plants in the U.S. that show signs of blight.

d) *Population* – all employees at the company; *sample* – all employees during the specified year; p – proportion of all employees who will have an injury on the job in a year; $\hat{p}$ – proportion of employees who had an injury on the job during the specified year.
Randomization condition: This sample is not random, but this year's employees are probably representative of employees in other years, with regards to injury on the job (3.9%).

10% condition: The 309 employees represent less than 10% of all possible employees over many years.
Success/Failure condition: $n\hat{p} = 12$ and $n\hat{q} = 297$ are both greater than 10, so the sample is large enough.

A one-proportion *z*-interval can be created for the proportion of employees who are expected to suffer an injury on the job in future years, provided that this year is representative of future years.

25. Conclusions.

a) Not correct. This statement implies certainty. There is no level of confidence in the statement.

b) Not correct. Different samples will give different results. Many fewer than 95% of samples are expected to have *exactly* 88% on-time orders.

c) Not correct. A confidence interval should say something about the unknown population proportion, not the sample proportion in different samples.

d) Not correct. We *know* that 88% of the orders arrived on time. There is no need to make an interval for the sample proportion.

e) Not correct. The interval should be about the proportion of on-time orders, not the days.

26. More conclusions.

a) Not correct. This statement implies certainty. There is no level of confidence in the statement.

b) Not correct. We *know* that 56% of the spins in this experiment landed heads. There is no need to make an interval for the sample proportion.

c) Correct.

d) Not correct. The interval should be about the proportion of heads, not the spins.

e) Not correct. The interval should be about the proportion of heads, not the percentage of euros.

27. Confidence intervals.

a) False. For a given sample size, higher confidence means a *larger* margin of error.

b) True. Larger samples lead to smaller standard errors, which lead to smaller margins of error.

c) True. Larger samples are less variable, which makes us more confident that a given confidence interval succeeds in catching the population proportion.

d) False. The margin of error decreases as the square root of the sample size increases. Halving the margin of error requires a sample four times as large as the original.

28. Confidence intervals, again.

a) True. The smaller the margin of error is, the less confidence we have in the ability of our interval to catch the population proportion.

b) True. Larger samples are less variable, which translates to a smaller margin of error. We can be more precise at the same level of confidence.

c) True. Smaller samples are more variable, leading us to be less confident in the ability of our interval to catch the true population proportion.

d) True. The margin of error decreases as the square root of the sample size increases.

29. Cars.

We are 90% confident that between 29.9% and 47.0% of cars are made in Japan.

30. Parole.

We are 95% confident that between 56.1% and 62.5% of paroles are granted by the Nebraska Board of Parole.

31. Mislabeled seafood.

a) $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.33) \pm 1.960\sqrt{\frac{(0.33)(0.67)}{1215}} = (0.304, 0.356)$

b) We are 95% confident that between 30.4% and 35.6% of all seafood packages purchased in the United States are mislabeled.

c) The size of the population is irrelevant. If *Consumer Reports* had a random sample, 95% of intervals generated by studies like this are expected to capture the true proportion of seafood packages that are mislabeled.

32. Mislabeled seafood, second course.

a) **Randomization condition:** It's not clear how the sample was chosen, but we will assume that the seafood packages are representative of all seafood packages.
10% condition: 22 is far less than 10% of all packages of "red snapper"..
Success/Failure condition: $n\hat{p} = 12$ and $n\hat{q} = 10$ are at least 10, so the sample is large enough.

b) Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of "red snapper" packages that are mislabeled.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{12}{22}\right) \pm 1.960\sqrt{\frac{\left(\frac{12}{22}\right)\left(\frac{10}{22}\right)}{22}} = (0.34, 0.75)$$

c) We are 95% confident that between 34% and 75% of all "red snapper" packages are mislabeled.

33. Baseball fans.

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645 \times \sqrt{\frac{(0.48)(0.52)}{1006}} \approx 0.026$

b) We're 90% confident that this poll's estimate is within 2.6% of the true proportion of people who are baseball fans.

c) The margin of error for 99% confidence would be larger. To be more certain, we must be less precise.

d) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 2.576 \times \sqrt{\frac{(0.48)(0.52)}{1006}} \approx 0.040$

e) Smaller margins of error involve less confidence. The narrower the confidence interval, the less likely we are to believe that we have succeeded in capturing the true proportion of people who are baseball fans.

34. Still living online.

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.960 \times \sqrt{\frac{(0.56)(0.44)}{1060}} \approx 0.030$

b) The pollsters are 95% confident that the true proportion of teens who say they go online several times per day is within 3.0% of the estimated 56%.

c) A 90% confidence interval results in a smaller margin of error. If confidence is decreased, a smaller interval is constructed.

d) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645 \times \sqrt{\frac{(0.56)(0.44)}{1060}} \approx 0.025$

e) Smaller samples generally produce larger intervals. Smaller samples are more variable, which increases the margin of error.

35. Contributions please.

a) **Randomization condition:** Letters were sent to a random sample of 100,000 potential donors.
10% condition: We assume that the potential donor list has more than 1,000,000 names.
Success/Failure condition: $n\hat{p} = 4781$ and $n\hat{q} = 95{,}219$ are both much greater than 10, so the sample is large enough.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{4781}{100{,}000}\right) \pm 1.960 \sqrt{\frac{\left(\frac{4781}{100{,}000}\right)\left(\frac{95{,}219}{100{,}000}\right)}{100{,}000}} = (0.0465, 0.0491)$$

We are 95% confident that the between 4.65% and 4.91% of potential donors would donate.

b) The confidence interval gives the set of plausible values with 95% confidence. Since 5% is above the interval, it seems to be a bit optimistic.

36. Take the offer.

a) **Randomization condition:** Offers were sent to a random sample of 50,000 cardholders.
10% condition: We assume that there are more than 500,000 cardholders.

Success/Failure condition: $n\hat{p} = 1184$ and $n\hat{q} = 48{,}816$ are both much greater than 10, so the sample is large enough.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{1184}{50{,}000}\right) \pm 1.960 \sqrt{\frac{\left(\frac{1184}{50{,}000}\right)\left(\frac{48{,}816}{50{,}000}\right)}{50{,}000}} = (0.0223, 0.0250)$$

We are 95% confident that the between 2.23% and 2.5% of all cardholders would register for double miles.

b) The confidence interval gives the set of plausible values with 95% confidence. Since 2% is below the interval, there is evidence that the true proportion is above 2%. The campaign should be worth the expense.

37. Teenage drivers.

a) **Randomization condition:** The insurance company randomly selected 582 accidents.
10% condition: 582 accidents represent less than 10% of all accidents.
Success/Failure condition: $n\hat{p} = 91$ and $n\hat{q} = 491$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of accidents involving teenagers.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{91}{582}\right) \pm 1.960 \sqrt{\frac{\left(\frac{91}{582}\right)\left(\frac{491}{582}\right)}{582}} = (12.7\%, 18.6\%)$$

37. (continued)

b) We are 95% confident that between 12.7% and 18.6% of all accidents involve teenagers.

c) About 95% of random samples of size 582 will produce intervals that contain the true proportion of accidents involving teenagers.

d) Our confidence interval contradicts the assertion of the politician. The figure quoted by the politician, 1 out of every 5, or 20%, is above the interval.

38. Junk mail.

a) **Independence assumption:** There is no reason to believe that one randomly selected person's response will affect another's.
Randomization condition: The company randomly selected 1000 recipients.
10% condition: 1000 recipients is less than 10% of the population of 200,000 people.
Success/Failure condition: $n\hat{p} = 123$ and $n\hat{q} = 877$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of people who will respond to the new flyer.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{123}{1000}\right) \pm 1.645 \sqrt{\frac{\left(\frac{123}{1000}\right)\left(\frac{877}{1000}\right)}{1000}} = (10.6\%, 14.0\%)$$

b) We are 90% confident that between 10.6% and 14.0% of people will respond to the new flyer.

c) About 90% of random samples of size 1000 will produce intervals that contain the true proportion of people who will respond to the new flyer.

d) Our confidence interval suggests that the company should do the mass mailing. The entire interval is well above the cutoff of 5%.

39. Safe food.

The grocer can conclude nothing about the opinions of all his customers from this survey. Those customers who bothered to fill out the survey represent a voluntary response sample, consisting of people who felt strongly one way or another about irradiated food. The random condition was not met.

40. Local news.

The city council can conclude nothing about general public support for the mayor's initiative. Those who showed up for the meeting are probably a biased group. In addition, a show of hands vote may influence people, affecting the independence of the votes.

41. Death penalty, again.

a) There may be response bias based on the wording of the question.

b) $$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.585) \pm 1.960 \sqrt{\frac{(0.585)(0.415)}{1020}} = (55.4\%, 61.5\%)$$

c) The margin of error based on the pooled sample is smaller, since the sample size is larger.

42. Gambling.

a) The interval based on the survey conducted by the college Statistics class will have the larger margin of error, since the sample size is smaller.

b) **Independence assumption:** There is no reason to believe that one randomly selected voter's response will influence another.
Randomization condition: Both samples were random.
10% condition: Both samples are probably less than 10% of the city's voters, provided the city has more than 12,000 voters.

42. (continued)

Success/Failure condition:
For the newspaper, $n_1\hat{p}_1 = (1200)(0.53) = 636$ and $n_1\hat{q}_1 = (1200)(0.47) = 564$
For the Statistics class, $n_2\hat{p}_2 = (450)(0.54) = 243$ and $n_2\hat{q}_2 = (450)(0.46) = 207$
All the expected successes and failures are greater than 10, so the samples are large enough.

Since the conditions are met, we can use one-proportion *z*-intervals to estimate the proportion of the city's voters that support the gambling initiative.

Newspaper poll: $\hat{p}_1 \pm z^*\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1}} = (0.53) \pm 1.960\sqrt{\dfrac{(0.53)(0.47)}{1200}} = (50.2\%, 55.8\%)$

Statistics class poll: $\hat{p}_2 \pm z^*\sqrt{\dfrac{\hat{p}_2\hat{q}_2}{n_2}} = (0.54) \pm 1.960\sqrt{\dfrac{(0.54)(0.46)}{450}} = (49.4\%, 58.6\%)$

c) The Statistics class should conclude that the outcome is too close to call, because 50% is in their interval.

43. Rickets.

a) **Randomization condition:** The 2700 children were chosen at random.
10% condition: 2700 children are less than 10% of all English children.
Success/Failure condition: $n\hat{p} = (2700)(0.20) = 540$ and $n\hat{q} = (2700)(0.80) = 2160$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion *z*-interval to estimate the proportion of the English children with vitamin D deficiency.

$$\hat{p} \pm z^*\sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.20) \pm 2.326\sqrt{\frac{(0.20)(0.80)}{2700}} = (18.2\%, 21.8\%)$$

b) We are 98% confident that between 18.2% and 21.8% of English children are deficient in vitamin D.

c) About 98% of random samples of size 2700 will produce confidence intervals that contain the true proportion of English children that are deficient in vitamin D.

44. Teachers.

a) **Randomization condition:** The poll was conducted from a random sample.
10% condition: 1002 people is less than 10% of all Americans.
Success/Failure condition: $n\hat{p} = 762$ and $n\hat{q} = 240$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion *z*-interval to estimate the proportion of all Americans who believe that high-achieving high school students should be recruited to become teachers.

$$\hat{p} \pm z^*\sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{762}{1002}\right) \pm 1.645\sqrt{\frac{\left(\frac{762}{1002}\right)\left(\frac{240}{1002}\right)}{1002}} = (73.8\%, 78.3\%)$$

b) We are 90% confident that between 73.8% and 78.3% of all Americans believe that high-achieving high school students should be recruited to become teachers.

c) About 90% of random samples of size 1002 will produce confidence intervals that contain the true proportion of all Americans who believe that high-achieving high school students should be recruited to become teachers.

d) These data refute the pundit's claim of that 2/3 of Americans believe this statement, since 66.7% is not in the interval.

45. Privacy or Security?

a) The confidence interval will be wider. The sample size is probably about one-sixth (17%) of the sample size of for all adults, so we would expect the confidence interval to be about two and a half times as wide.

b) The second poll's margin of error should be slightly wider. There are fewer "young" people (13%) in the sample than seniors (17%).

46. Back to campus.

a) The confidence interval for the retention rate in private colleges will be narrower than the confidence interval for the retention rate in public colleges, since it is based on a larger sample.

b) Since the overall sample size is larger, the margin of error in retention rate is expected to be smaller.

47. Deer ticks.

a) **Independence assumption:** Deer ticks are parasites. A deer carrying the parasite may spread it to others. Ticks may not be distributed evenly throughout the population.
Randomization condition: The sample is not random and may not represent all deer.
10% condition: 153 deer are less than 10% of all deer.
Success/Failure condition: $n\hat{p} = 32$ and $n\hat{q} = 121$ are both greater than 10, so the sample is large enough.

The conditions are not satisfied, so we should use caution when a one-proportion *z*-interval is used to estimate the proportion of deer carrying ticks.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{32}{153}\right) \pm 1.645 \sqrt{\frac{\left(\frac{32}{153}\right)\left(\frac{121}{153}\right)}{153}} = (15.5\%, 26.3\%)$$

We are 90% confident that between 15.5% and 26.3% of deer have ticks.

b) In order to cut the margin of error in half, they must sample 4 times as many deer.
$4(153) = 612$ deer.

c) The incidence of deer ticks is not plausibly independent, and the sample may not be representative of all deer, since females and young deer are usually not hunted.

48. Back to campus II.

a) In order to cut the margin of error in half, they must sample 4 times as many college freshmen.
$4(1644) = 6576$.

b) A sample this large may be more than 10% of the population of all potential students.

49. Graduation, again.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.02 = 1.960\sqrt{\frac{(0.25)(0.75)}{n}}$$

$$n = \frac{(1.960)^2(0.25)(0.75)}{(0.02)^2}$$

$$n \approx 1801 \text{ people}$$

In order to estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 2% with 95% confidence, we would need a sample of at least 1801 people. All decimals in the final answer must be rounded up, to the next person.

(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 2401 people.)

50. Better hiring info.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 2.576\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$n = \frac{(2.576)^2(0.5)(0.5)}{(0.04)^2}$$

$n \approx 1037$ businesses

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 4% with 99% confidence, we would need a sample of at least 1037 businesses. All decimals in the final answer must be rounded up, to the next business.

51. Pilot study.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.03 = 1.645\sqrt{\frac{(0.15)(0.85)}{n}}$$

$$n = \frac{(1.645)^2(0.15)(0.85)}{(0.03)^2}$$

$n \approx 384$ cars

Use $\hat{p} = \frac{9}{60} = 0.15$ from the pilot study as an estimate.

In order to estimate the percentage of cars with faulty emissions systems to within 3% with 90% confidence, the state's environmental agency will need a sample of at least 384 cars. All decimals in the final answer must be rounded up, to the next car.

52. Another pilot study.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 2.326\sqrt{\frac{(0.22)(0.78)}{n}}$$

$$n = \frac{(2.326)^2(0.22)(0.78)}{(0.04)^2}$$

$n \approx 581$ adults

Use $\hat{p} = 0.22$ from the pilot study as an estimate.

In order to estimate the percentage of adults with higher than normal levels of glucose in their blood to within 4% with 98% confidence, the researchers will need a sample of at least 581 adults. All decimals in the final answer must be rounded up, to the next adult.

53. Approval rating.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.025 = z^* \sqrt{\frac{(0.65)(0.35)}{972}}$$

$$z^* = \frac{0.025}{\sqrt{\frac{(0.65)(0.35)}{972}}}$$

$$z^* \approx 1.634$$

Since $z^* \approx 1.634$, which is close to 1.645, the pollsters were probably using 90% confidence. The slight difference in the z^* values is due to rounding of the governor's approval rating.

54. Amendment.

a) This poll is inconclusive because the confidence interval, 52% ± 3% contains 50%. The true proportion of voters in favor of the constitutional amendment is estimated to be between 49% (minority) to 55% (majority). We can't be sure whether or not the majority of voters support the amendment or not.

b)

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.03 = z^* \sqrt{\frac{(0.52)(0.48)}{1505}}$$

$$z^* = \frac{0.03}{\sqrt{\frac{(0.52)(0.48)}{1505}}}$$

$$z^* \approx 2.3295$$

Since $z^* \approx 2.3295$, which is close to 2.326, the pollsters were probably using 98% confidence. The slight difference in the z^* values is due to rounding of the amendment's approval rating.