8.2 Projecting winning candidate News coverage during a recent election projected that a certain candidate would receive 54.8% of all votes cast; the projection had a margin of error of ±3%

a. Give a point estimate for the proportion of all votes the candidate will receive.

b. Give an interval estimate for the proportion of all votes the candidate will receive.

c. In your own words, state the difference between a point estimate and an interval estimate.

a. 54.8%

b. 54.8% ±3% = (51.8%, 57.8%)

c. Point estimate uses one specific value while interval estimate uses a range of values.

8.22 Wife doesn't want kids TRY The 1996 GSS asked, "If the husband in a family wants children, but the wife decides that she does not want any children, is it all right for the wife to refuse to have children?" Of 699 respondents, 576 said yes.

a. Find a 99% confidence interval for the population proportion who would say yes. Can you conclude that the population proportion exceeds 75%? Why?

b. Without doing any calculation, explain whether the interval in part a would be wider or narrower than a 95% confidence interval for the population proportion who would say yes.

a. 99% CI for p is $\hat{p} \pm 2.58(se)$

$$= 576/699 \pm 2.58\sqrt{(576/699)(1-576/699)/699}$$

= 0.824 ± 2.58 (0.0144) = 0.824 ± 0.0372 = (0.787, 0.861)

Yes, we can conclude that the population proportion exceeds 75% because 75% is below the lowest believable value.

b. 99% CI is wider than a 95% CI because the margin of error is larger for a higher confidence level.

8.23 Exit poll predictions A national television network takes an exit poll of 1400 voters after each has cast a vote in a state gubernatorial election. Of them, 660 say they voted for the Democratic candidate and 740 say they voted for the Republican candidate.

a. Treating the sample as a random sample from the population of all voters, would you predict the winner? Base your decision on a 95% confidence interval.

b. Base your decision on a 99% confidence interval. Explain why you need stronger evidence to make a prediction when you want greater confidence.

a. Let p= proportion of voters who vote for the Democratic candidate.

95% CI for p is $\hat{p} \pm 1.96(se)$

$$= \frac{660}{1400} \pm 1.96\sqrt{\left(\frac{660}{1400}\right)\left(1-\frac{660}{1400}\right)/1400}$$ **=0.471±0.025=(0.446, 0.496)**

We could predict the winner because 0.50 falls outside of the CI.

b. 99% CI for p is (0.437, 0.505). We can not predict a winner since 0.50 falls within the CI. The more confident we are, the wider the CI.

8.24 Exit poll with smaller sample In the previous exercise, suppose the same proportions resulted from $n = 140$ (instead of 1400), with counts 66 and 74.

a. Now does a 95% confidence interval allow you to predict the winner? Explain.

b. Explain why the same proportions but with smaller samples provide less information. (*Hint:* What effect does n have on the standard error?)

a. The sample proportion is the same, but se= $\sqrt{\left(\frac{66}{140}\right)\left(1-\frac{66}{140}\right)/140}$ **=0.042. The 95% CI is (0.389, 0.553). We cannot predict the winner.**

b. The standard error becomes smaller when n is larger.

8.36 Wage discrimination? According to a union agreement, the mean income for all senior-level assembly-line workers in a large company equals \$500 per week. A representative of a women's group decides to analyze whether the mean income for female employees matches this norm. For a random sample of nine female employees, using software she obtains a 95% confidence interval of (371, 509). Explain what is wrong with each of the following interpretations of this interval.

a. We infer that 95% of the women in the population have income between \$371 and \$509 per week.

b. If random samples of nine women were repeatedly selected, then 95% of the time the sample mean income would be between \$371 and \$509.

c. We can be 95% confident that $\bar{x}$ is between \$371 and \$509.

d. If we repeatedly sampled the entire population, then 95% of the time the population mean would be between \$371 and \$509.

Your explanation to 8.36 should be in two sentences or less.

a. 95% is the confidence level, not the proportion of females.
b. Should be 'If random samples ..., then 95% of the time the CI would contain the population mean'.
c. The CI is for population mean, not sample mean.
d. The population mean is known if the entire population is sampled.

8.38 How often read a newspaper? For the FL Student Survey data file on the text CD, software reports the results for responses on the number of times a week the subject reads a newspaper:

Variable	N	Mean	Std Dev	SE Mean	95.0% CI
News	60	4.1	3.0	0.387	(3.325, 4.875)

a. Is it plausible that $\mu = 7$, where μ is the population mean for all Florida students? Explain.

b. Suppose that the sample size had been 240, with $\bar{x} = 4.1$ and $s = 3.0$. Find a 95% confidence interval, and compare it to the one reported. Describe the effect of sample size on the margin of error.

c. Does it seem plausible that the population distribution of this variable is normal? Why?

d. Explain the implications of the term *robust* regarding the normality assumption made to conduct this analysis.

a. No, 7 is not in the 95% CI.
b. 95% CI = x-bar ± 1.96 (se) = 4.1 ± 1.98 (0.1936) = 4.1 ± 0.38 = (3.72, 4.48). This CI is narrower than the CI above using a smaller n. When the sample size is increased, the margin of error would decrease.
c. No, the variable (number of times a week the subject reads a newspaper) is non-negative and the standard deviation (3.0) is relatively large to the mean value (4.1). *X* is non-negative. The minimum value, 0 has a z-score of 1.37, thus *X* is not normally distributed. Due to large *sd*, X is possibly bimodal (most values are far from the mean).
d. Even if the normality assumption is not met, the analysis is still likely to produce valid results.

For b, *df*=239, you may use *t* table with *df*=100 or *df*=infinity to approximate the exact *t* value.

8.42 Effect of confidence level Find the margin of error for estimating the population mean when the sample standard deviation equals 100 for a sample size of 400, using confidence level (i) 95% and (ii) 99%. What is the effect of the choice of confidence level?

s = 100, n = 400, ME = t(SE)

(i) For 95% CI, the ME = 1.98 (100/√400) = 9.9

(ii) For 99% CI, the ME = 2.617(100/√400) = 13.085

The margin of error increases when the confidence level is higher.

For (i) and (ii), *df*=399, you may use *t* table with *df*=100 or *df*=infinity

8.48 Abstainers The Harvard study mentioned in the previous exercise estimated that 19% of college students abstain from drinking alcohol. To estimate this proportion in your school, how large a random sample would you need to estimate it to within 0.05 with probability 0.95, if before conducting the study

a. You are unwilling to predict the proportion value at your school.

b. You use the Harvard study as a guideline.

c. Use the results from parts a and b to explain why strategy (a) is inefficient if you are quite sure you'll get a sample proportion that is far from 0.50.

ME or m =0.05, z=1.96

a. p=0.5 $n = \frac{z^2\hat{p}(1-\hat{p})}{m^2} = \frac{1.96^2 * 0.5 * (1-0.5)}{0.05^2} = 384.16 = 385$

b. P=0.19 $n = \frac{z^2\hat{p}(1-\hat{p})}{m^2} = \frac{1.96^2 * 0.19 * (1-0.19)}{0.05^2} = 236.49 = 237$

For sample size, always round it to the next integer.

c. It overestimated the sample size by quite a lot. It would be more costly than needed.

8.52 Income of Native Americans TRY How large a sample size do we need to estimate the mean annual income of Native Americans in Onondaga County, New York, correct to within $1000 with probability 0.99? No information is available to us about the standard deviation of their annual income. We guess that nearly all of the incomes fall between $0 and $120,000 and that this distribution is approximately bell shaped.

ME=$1000, z=2.58,

For a roughly normal distribution,

$$\sigma \approx \text{range}/6 = 120{,}000/6 = 20{,}000$$

n=(Zσ/ME)^2=(2.58*20,000/1,000)^2=2662.56=2663

round *n* up to the next integer.

❖ How to find the *t* value to construct a CI for the population mean? *df*=*n*-1.

If the exact *t* value is provided, use the exact value;

If the exact value is unkown,

use *t* table with **the largest *df* found on the table** that is smaller than the exact *df*=*n*-1 (conservative approach).

Alternatively, you may use the *Z* value to estimate *t* (when *n* is not too small).

For example,

if *df*=39; Conservative approach→ use *t* table with *df*=30;

Use Z to estimate→ use *Z or t* table with *df*=infinity.