PART

3

Inferential Statistics

8.1 Practicing the Basics

8.1 **Health care** A study dealing with health care issues plans to take a sample survey of 1500 Americans to estimate the proportion who have health insurance and the mean dollar amount that Americans spent on health care this past year.

a. Identify two population parameters that this study will estimate.

b. Identify two statistics that can be used to estimate these parameters.

8.2 **Video on demand** A recent study from Nielsen (available from nielsen.com; search for "The Cross-Platform Report") took a sample to investigate how video on demand and other subscription-based services (such as Netflix, Amazon Prime, and Hulu) change TV viewing behavior of U.S. adults. Among others, the study recorded how many subscribers primarily watch content time-shifted (i.e., not live) and the weekly number of hours adults spend watching content over the Internet.

a. Mention two population parameters that this survey is trying to estimate.

b. Mention two corresponding statistics that will estimate these population parameters with the help of the survey.

c. The sample mean number of hours watched online is an unbiased estimator for what parameter? Explain what unbiased means.

8.3 **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 $\pm 3\%$.

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.

8.4 **Youth professional football** In a survey of 1009 American adults in 2016, 313 said they would not allow a young son to play competitive football (www.forbes.com). Find the point estimate of the population proportion of these respondents.

8.5 **Government spying** In 2014, news reports worldwide alleged that the U.S. government had hacked German chancellor Angela Merkel's cell phone. A Pew Research Center survey of German citizens at about that time asked whether they find it acceptable or unacceptable for the U.S. government to monitor communications from their country's leaders. Results from the survey show that of 1000 citizens interviewed, 900 found it unacceptable.

(*Source:* Pew Research Center, July 2014, "Global Opposition to U.S. Surveillance and Drones, But Limited Harm to America's Image")

a. Find the point estimate of the population proportion of German citizens who find spying unacceptable.

b. The Pew Research Center reports a margin of error at the 95% confidence level of 4.5% for this survey. Explain what this means.

8.6 **Game apps** The Google Play app store for smartphones offers hundreds of games to download for free or for a small fee. The ones for which a fee is charged are called paid games. For a random sample of five paid games taken in July 2014 on the Google platform, the following fees were charged: $1.09, $4.99, $1.99, $1.99, $2.99.

a. Find a point estimate of the mean fee for paid games available on Google's platform.

b. The margin of error at the 95% confidence level for this point estimate is $1.85. Explain what this means.

8.7 **Nutrient effect on growth rate** Researchers are interested in the effect of a certain nutrient on the growth rate of plant seedlings. Using a hydroponics growth procedure that used water containing the nutrient, they planted six tomato plants and recorded the heights of each plant 14 days after germination. Those heights, measured in millimeters, were as follows: 55.5, 60.3, 60.6, 62.1, 65.5, 69.2.

a. Find a point estimate of the population mean height of this variety of seedling 14 days after germination.

b. A method that we'll study in Section 8.3 provides a margin of error of 4.9 mm for a 95% confidence interval for the population mean height. Construct that interval.

c. Use this example to explain why a point estimate alone is usually insufficient for statistical inference.

8.8 **More youth professional football** In Exercise 8.4, the proportion of American adults who would not allow a young son to play competitive football was 0.31. The estimated standard deviation of this point estimate is 0.015.

a. Find and interpret the margin of error for a 95% confidence interval for the population proportion of adults who would not allow a young son to play competitive football.

b. Construct the 95% confidence interval. Interpret it in context.

8.9 **Feel lonely often?** The GSS has asked "On how many days in the past seven days have you felt lonely?" At sda.berkeley.edu/GSS, enter LONELY as the variable, select Summary Statistics in the menu of table options, and click Run the Table to see the responses.

a. Report the percentage making each response and the mean and standard deviation of the responses. Interpret.

b. The standard deviation of the sample mean can be estimated using this data as 0.06. Interpret the value of 0.06.

8.10 **CI for loneliness** Refer to the previous exercise. The margin of error for a 95% confidence interval for the population mean is 0.12. Construct that confidence interval and interpret it.

8.11 **Barack Obama as president** To answer these questions, refer to the data in a poll conducted by Gallup.com. (*Source:* http://www.gallup.com/poll/113980/gallup-daily-obama-job-approval.aspx)

a. Specify the population parameter, value of the sample statistic (using the chart based on a three-day rolling average for a selected three-day period), the point estimate, and the size of the margin of error.

b. Explain how to interpret the margin of error.

8.2 Practicing the Basics

8.12 Putin A Gallup poll of 2000 Russians taken between April and June 2014 (after the Olympic games in Russia and the annexation of the Crimean peninsula) showed that 83% approved of President Putin's performance. If random sampling were used for this survey, what is the margin of error for this estimate at the 95% confidence level? (In fact, Gallup uses weighted sampling and, based on it, reports a margin of error of 2.7%.) *Source:* gallup.com and search for "Russian Approval."

8.13 Flu shot In a clinical study (the same as mentioned in Example 4), 3900 subjects were vaccinated with a vaccine manufactured by growing cells in fertilized chicken eggs. Over a period of roughly 28 weeks, 24 of these subjects developed the flu.

a. Find the point estimate of the population proportion that were vaccinated with the vaccine but still developed the flu.

b. Find the standard error of this estimate.

c. Find the margin of error for a 95% confidence interval.

d. Construct the 95% confidence interval for the population proportion. Interpret the interval.

e. Can you conclude that fewer than 1% of all people vaccinated with the vaccine will develop the flu? Explain by using results from part d.

8.14 Renewable energy usage in India A survey was conducted by Mercom Communications India in 2014. Out of 1700 respondents, 1479 stated that they support subsidies for solar power over other sources.

a. Estimate the population proportion who supported subsidies for solar power over other sources of energy.

b. Find the margin of error for a 95% confidence interval for this estimate.

c. Find a 95% confidence interval for that proportion and interpret.

d. State and check the assumptions needed for the interval in part c to be valid.

8.15 Make industry help environment? When the 2006 GSS asked subjects whether it should or should not be the government's responsibility to impose strict laws to make industry do less damage to the environment (variable GRNLAWS), 1403 of 1497 subjects said yes.

a. What assumptions are made to construct a 95% confidence interval for the population proportion who would say yes? Do they seem satisfied here?

b. Construct the 95% confidence interval. Interpret in context. Can you conclude whether a majority or minority of the population would answer yes?

8.16 Favor death penalty In the 2012 General Social Survey, respondents were asked whether they favored or opposed the death penalty for people convicted of murder. Software shows results

```
Sample   X     N    Sample p          95% CI
  1    1183  1824  0.648575  (0.626665, 0.670484)
```

Here, X refers to the number of the respondents who were in favor.

a. Show how to obtain the value reported under "Sample p."

b. Interpret the confidence interval reported, in context.

c. Explain what the "95% confidence" refers to, by describing the long-run interpretation.

d. Can you conclude that more than half of all American adults were in favor? Why?

8.17 Oppose death penalty Refer to the previous exercise. Show how you can get a 95% confidence interval for the proportion of American adults who were *opposed* to the death penalty from the confidence interval stated in the previous exercise for the proportion in favor. (*Hint:* The proportion opposed is 1 minus the proportion in favor.)

8.18 Stem cell research A Harris poll of a random sample of 2113 adults in the United States in October 2010 reported that 72% of those polled believe that stem cell research has merit. (*Source:* www.harrisinteractive.com/vault/Harris-Interactive-Poll-HealthDay-2010-10.pdf.) The results, presented using MINITAB software, are

```
  X       N      Sample p          95% CI
1521    2113     0.7198      (0.7007, 0.7390)
```

Here, X denotes the number who believed that stem cell research has merit.

a. Explain how to interpret "Sample p" and "95% CI" on this printout.

b. What is the 95% margin of error associated with the poll?

8.19 z-score and confidence level Which z-score is used in a (a) 90%, (b) 98%, and (c) 99.9% confidence interval for a population proportion?

8.20 Trusting CNN news? USA TODAY conducted a survey of 1000 likely voters in February 2016. In the survey, 134 respondents said that CNN is the TV news or commentary source they trust the most. The interval estimation at the 95% confidence level for the proportion who trust CNN the most is (0.113, 0.155). Explain how to interpret the given confidence interval.

8.21 Budget impact on opportunities for young Canadians A national survey of 1500 Canadian adults conducted by Abacus Data and commissioned by EY to judge reactions to the first federal budget delivered by Finance Minister Bill Morneau, revealed that 33% of the respondents said it would have a positive impact on opportunities for young Canadians. The interval estimation at the 95% confidence level for the proportion who said it would have a positive impact on opportunities for young Canadians is (0.3, 0.36). Specify the population to which this inference applies and explain how to interpret the confidence interval.

8.22 Operations growth in Luxembourg According to a survey conducted by KPMG in 2016, almost 46% of the surveyed companies intend to grow their operations in Luxembourg over the next two years (*Source*: https://www.kpmg.com/LU/en/IssuesAndInsights/Articlespublications/Documents/Management-Company-CEO-Survey-032016.pdf).

a. Can you specify the assumptions made to construct a 95% confidence interval for the population proportion?

b. If the sample size is 5000, verify that the assumptions of part a are satisfied and construct the 95% confidence interval. Determine whether the proportion of companies who intend to grow their operations in Luxembourg is a majority or a minority.

8.23 Chicken breast In a 2014 *Consumer Reports* article titled, "The High Cost of Cheap Chicken," the magazine reported that out of 316 chicken breasts bought in retail stores throughout the United States, 207 contained E. coli bacteria.

a. Find and interpret a 99% confidence interval for the population proportion of chicken breasts that contain E. coli. Can you conclude that the proportion of chicken containing E. coli exceeds 50%? 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.

8.24 Dispute over unlocking iPhone A national survey was conducted by the Pew Research Center (www.people-press.org) between February 18-21, 2016. Among 1002 participating adults, 51% said that Apple Inc. should assist the FBI in their investigations by unlocking the iPhone used by one of the suspects in the San Bernardino terrorist attacks. Based on these data, can we conclude that more than half of Americans support the Department of Justice over Apple Inc. in this dispute over unlocking the concerned iPhone? Explain.

8.25 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.

8.26 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?)

8.27 Simulating confidence intervals Repeat the simulation from Activity 2, but this time simulate 1000 confidence intervals (repeatedly press the Draw Sample(s) button after selecting 100 as the number of samples to generate until you get 1000 simulations) with $p = 0.5$ instead of $p = 0.3$. Do this for a confidence level of 95% and then 99%. What percentage do you expect for each case?

8.28 Simulating confidence intervals with poor coverage Using the Explore Coverage web app, let's check that the large-sample confidence interval for a proportion may work poorly with small samples. In the app, set $p = 0.30$, $n = 10$ and leave the confidence level at 95%. Select to draw 100 random samples of size n and then click on Draw Sample(s).

a. How many of the intervals you generated with the app fail to contain the true value, $p = 0.30$?

b. How many would you expect not to contain the true value? What does this suggest?

c. To see that this is not a fluke, now take 1000 samples and see what percentage of 95% confidence intervals contain 0.30. (*Note:* For every interval formed, the number of successes is smaller than 15, so the large-sample formula is not adequate.)

d. Using the Sampling Distribution for a Sample Proportion web app, generate 10,000 random samples of size 10 when $p = 0.30$. The app will plot the simulated sampling distribution of the sample proportion values. Is it bell shaped? Use this to help you explain why the large-sample confidence interval performs poorly in this case. (This exercise illustrates that assumptions for statistical methods are important, because the methods may perform poorly if we use them when the assumptions are violated.)

8.3 Practicing the Basics

8.29 Average temperature in Florida TRY According to the National Centers for Environmental Information, the mean March monthly average temperature in Florida, for the years 1895 to 2016 (a sample of 122 observations), is 64.264°F with a standard deviation of 3.109°F (*Source*: www.ncdc.noaa.gov).

a. What is the point estimate of the monthly average temperature for March in Florida?

b. Find the standard error of the sample mean.

c. The 95% confidence interval is (63.707, 64.821). Interpret it.

d. Is it plausible that the population mean of the monthly average temperature for March in Florida could be estimated to $\mu = 63$°F? Explain.

8.30 Average temperature in the United States Refer to the previous exercise. For the years 1895 to 2016 (a sample of 122 observations), the mean March monthly average temperature in the United States is 41.699°F with a standard deviation of 2.948°F.

a. Find the point estimate of the monthly average temperature for March in the United States and show that its standard error is 0.267.

b. The 95% confidence interval is 41.170 to 42.227. Explain what "95% confidence" means for this interval.

c. Compared to the interval for Florida, is there enough evidence of a difference between the means? Explain.

8.31 Using *t*-table TECH Using Table B, the web app, software or a calculator, report the *t*-score that you multiply by the standard error to form the margin of error for a

a. 95% confidence interval for a mean with 5 observations.

b. 95% confidence interval for a mean with 15 observations.

c. 99% confidence interval for a mean with 15 observations.

8.32 Anorexia in teenage girls TECH A study[4] compared various therapies for teenage girls suffering from anorexia, an eating disorder. For each girl, weight was measured before and after a fixed period of treatment. The variable measured was the change in weight, $X =$ weight at the end of the study minus weight at the beginning of the study. The therapies were designed to aid weight gain, corresponding to positive values of X. For the sample of 17 girls receiving the family therapy, the changes in weight during the study were 11, 11, 6, 9, 14, −3, 0, 7, 22, −5, −4, 13, 13, 9, 4, 6, 11.

a. Plot these with a dot plot or box plot and summarize.

b. Using the web app, software or a calculator, show that the weight changes have $\bar{x} = 7.29$ and $s = 7.18$ pounds.

c. Using the web app, software or a calculator, show that the standard error of the sample mean was $se = 1.74$.

d. To use the t distribution, explain why the 95% confidence interval uses the t-score equal to 2.120.

---

[4]Data courtesy of Prof. Brian Everitt, Institute of Psychiatry, London.

e. Let μ denote the population mean change in weight for this therapy. Using results from parts b, c, and d, show that the 95% confidence interval for μ is (3.6, 11.0). Explain why this suggests that the true mean change in weight is positive but possibly quite small.

8.33 Talk time on smartphones TECH One feature smartphone manufacturers use in advertising is the amount of time one can talk before recharging the battery. Below are 13 values from a random sample of the talk-time (in minutes) of smartphones running on lithium-ion batteries. The summary statistics are $\bar{x} = 553$, $s = 227$, Q1 = 420, median = 450, Q3 = 650. Talk time: 320, 360, 760, 580, 1050, 900, 360, 500, 420, 420, 650, 420, 450.

a. Construct an appropriate graph (dot plot, stem and leaf plot, histogram, box plot) to describe the shape of the distribution. What assumptions are needed to construct a 95% confidence interval for μ, the mean talk time? Point out any assumptions that seem questionable.

b. Check whether this data set has any potential outliers according to the criterion of (i) 1.5 * IQR below Q1 or above Q3 and (ii) 3 standard deviations from the mean.

c. Using the summary statistics, show that the 95% confidence interval is (416, 690). Interpret it in context.

d. The value 1050 is quite a bit larger than the others. Delete this observation, find the new mean and standard deviation, and construct the 95% confidence interval for μ without this observation. How does it compare to the 95% confidence interval, using all the data?

8.34 Birth weights of elephants TECH The birth weights (in kilograms) of five elephants, selected randomly, are 133, 120, 97, 106, 124 (*Source*: www.elephant.se).

a. Using the web app, software or a calculator, verify that the 95% confidence interval for the population mean is (98.11, 133.89).

b. Name two things you could do to get a narrower interval than the one in part a.

c. Construct a 99% confidence interval. Why is it wider than the 95% interval?

d. On what assumptions is the interval in part a based? Explain how important each assumption is.

8.35 Buy it now TRY Example 6 mentioned closing prices for listings of the iPhone 5s on eBay. If you don't feel comfortable bidding (or can't wait until a listing has ended), you can often purchase the item right away at the indicated Buy It Now price. A random sample of Buy It Now prices for an unlocked iPhone 5s, 16GB and in new condition in July 2014 showed the following prices (in $) with dot plot and box plot shown in the figure: 618, 650, 608, 634, 675, 618, 625, 619, 630.

a. What must we assume to use these data to find a 95% confidence interval for the mean Buy It Now price on eBay? Do these assumptions seem plausible?

b. The table shows the way software reports results. How was the standard error of the mean (SE Mean) obtained?

Variable	N	Mean	StDev	SE Mean	95% CI
BuyItNow	9	630.78	20.47	6.82	(615.04 646.52)

c. Interpret the 95% confidence interval in context.

d. Compared to the confidence interval (569, 599) for prices from auctions obtained in Example 6, is there evidence that the mean price is higher when purchased through Buy It Now? Explain.

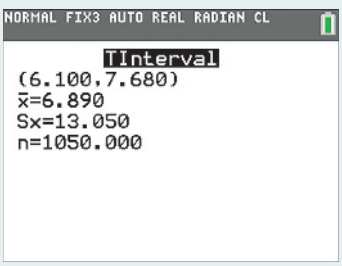
e. Excluding the observation of $675, the confidence interval equals (615, 636). Does this change your answer to part d?

8.36 Time spent on e-mail When the GSS asked $n = 1050$ people in 2012, "About how many hours per week do you spend sending and answering e-mail?" (EMAILHR), the summary statistics were $\bar{x} = 6.89$ and $s = 13.05$. TI output with these data (available on the book's website) is shown in the screen shot.

a. What is the margin of error at the 95% confidence level for the point estimate of the mean number of hours?

b. Interpret the 95% confidence interval shown in context.

c. The distribution of hours spent is right-skewed (e.g., the minimum of zero hours is only 0.5 standard deviations below the mean, but the largest value of 168 hours has a z-score larger than 12). Is this a concern for the validity of the confidence interval?



8.37 Grandmas using e-mail TECH For the question about e-mail in the previous exercise, the 14 females in the GSS of age at least 80 had the responses

0, 0, 0, 0, 1, 1, 1, 2, 2, 6, 6, 7, 7, 10.

a. Using the web app, software or a calculator, find the sample mean and standard deviation and the standard error of the sample mean.

b. Find and interpret a 90% confidence interval for the population mean.

c. Explain why the population distribution may be skewed right. If this is the case, is the interval you obtained in part b useless, or is it still valid? Explain.

8.38 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.

8.39 How often read a newspaper? For the FL Student Survey data file on the book's website, 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     0.388   (3.307, 4.859)
```

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.

8.40 Work hours per week The General Social Survey asked 40 respondents about the number of hours they usually work in a week. A researcher analyzing data from the 2014 GSS obtained the following StatCrunch output:

95% confidence interval results:

μ : Mean of variable

```
          Sample    Std.       Std.          L.        U.
Variable  Mean      Dev.       Err.    DF    Limit     Limit
Number    38.7  12.416284 1.9631868 39 34.72908 42.67092
of work
hours in
a week
```

a. Show how to construct the confidence interval from the other information provided.

b. Can you conclude that the population mean is larger than 43? Explain.

c. Would the confidence interval be wider, or narrower, (i) if you constructed a 99% confidence interval? (ii) if $n = 400$ instead of 40?

8.41 Length of hospital stay for childbirth Data was collected from the records of 2962 patients admitted to a hospital in 2015 to estimate the mean length of stay for childbirth. It was observed that the sample mean was 2.372 days and the margin of error for estimating the population mean is 0.029 at a confidence level of 95%. Explain the meaning of the last sentence, showing what it suggests about the 95% confidence interval. Find the sample standard deviation.

8.42 Effect of *n* Find the margin of error for a 95% confidence interval for estimating the population mean when the sample standard deviation equals 100, with a sample size of (i) 25 and (ii) 100. What is the effect of increasing the sample size? (You can use Table B in the back to find the appropriate *t*-scores.)

8.43 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 25, using confidence levels (i) 95% and (ii) 99%. What is the effect of the choice of confidence level? (You can use Table B in the back to find the appropriate *t*-scores.)

8.44 Catalog mail-order sales A company that sells its products through mail-order catalogs wants information about the success of its most recent catalog. The company decides to estimate the mean dollar amount of items ordered from those who received the catalog. For a random sample of 100 customers from their files, only 5 made an order, so 95 of the response values were \$0. The overall mean of all 100 orders was \$10, with a standard deviation of \$10.

a. Is it plausible that the population distribution is normal? Explain and discuss how much this affects the validity of a confidence interval for the mean.

b. Find a 95% confidence interval for the mean dollar order for the population of all customers who received this catalog. Normally, the mean of their sales per catalog is about \$15. Can we conclude that it declined with this catalog? Explain.

8.45 Number of children For the question, "How many children have you ever had?" use the GSS website sda.berkeley.edu/GSS with the variable CHILDS to find the sample mean and standard deviation for the 2012 survey. (Use YEAR(2012) as filter and select Summary Statistics under Options.) (Treat the 23 respondents who answered "eight or more" as having eight children.)

a. Show how to obtain a standard error of 0.04 for a random sample of 1971 adults.

b. Construct a 95% confidence interval for the population mean. Can you conclude that the population mean is less than 2.0? Explain.

8.46 Simulating the confidence interval Go to the Inference for a Mean web app, accessible on the book's website. Select the Explore Coverage tab and keep the default skewed shape for the population distribution as well as the default confidence level of 95%. Set the sample size to 30 and generate 100 random samples by selecting 100 and clicking on Draw Sample(s). The graph will now show 100 confidence intervals for μ, based on 100 generated sample from the population distribution.

a. How many of the intervals you generated fail to contain the population mean $\mu = 14.3$?

b. How many would you expect not to contain the true value?

c. Now repeat the simulation using at least 1000 random samples of size 30. Why do close to 95% of the intervals contain μ, even though the population distribution is quite skewed?

8.47 Unemployment percentage A statistician conducts a study in order to estimate the proportion of families living in a poor area of a city and having at least one unemployed member. Calculate the sample size needed to estimate it to within 0.05 with 95% confidence.

8.48 Binge drinkers TRY A study at the Harvard School of Public Health found that 44% of 10,000 sampled college students were binge drinkers. A student at the University of Minnesota plans to estimate the proportion of college students at that school who are binge drinkers. How large a random sample would she need to estimate it to within 0.05 with 95% confidence, if before conducting the study she uses the Harvard study results as a guideline?

8.49 Abstainers TRY 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 guess 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.

8.50 How many businesses fail? A study is planned to estimate the proportion of businesses started in the year 2006 that had failed within five years of their start-up. How large a sample size is needed to guarantee estimating this proportion correct to within

a. 0.10 with probability 0.95?

b. 0.05 with probability 0.95?

c. 0.05 with probability 0.99?

d. Compare sample sizes for parts a and b, and b and c, and summarize the effects of decreasing the margin of error and increasing the confidence level.

8.51 Employment percentage in the United States According to the U.S. Bureau of Labor Statistics, 80.3% out of the nation's 81.4 million families had at least one employed member in 2015 (*Source*: http://www.bls.gov/news.release/famee.nr0.htm). What should be the sample size needed to estimate the proportion of families having at least one employed member in 2015 within an accuracy of 3 percentage points at a 95% level of confidence?

8.52 Farm size An estimate is needed of the mean acreage of farms in Ontario, Canada. A 95% confidence interval should have a margin of error of 25 acres. A study 10 years ago in this province had a sample standard deviation of 200 acres for farm size.

a. About how large a sample of farms is needed?

b. A sample is selected of the size found in part a. However, the sample has a standard deviation of 300 acres rather than 200. What is the margin of error for a 95% confidence interval for the mean acreage of farms?

8.53 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.

8.54 Population variability Explain the reasoning behind the following statement: "In studies about a very diverse population, large samples are often necessary, whereas for more homogeneous populations smaller samples are often adequate." Illustrate for the problem of estimating mean income for all medical doctors in the United States compared to estimating mean income for all entry-level employees at McDonald's restaurants in the United States.

8.55 Web survey to get large *n* A newspaper wants to gauge public opinion about legalization of marijuana. The sample size formula indicates that it need a random sample of 875 people to get the desired margin of error. But surveys cost money, and it can only afford to randomly sample 100 people. Here's a tempting alternative: If it places a question about that issue on its website, it will get more than 1000 responses within a day at little cost. Is it better off with the random sample of 100 responses or the website volunteer sample of more than 1000 responses? (*Hint:* Think about the issues discussed in Section 4.2 about proper sampling of populations.)

8.56 Do students like statistics? TRY All respondents out of a random sample of ten students in a college said that they like statistics. Now you want to estimate the proportion of students who like statistics in the whole college.

a. Find the sample proportion of students who like statistics.

b. Find the standard error of the estimate and interpret it.

c. Find a 95% confidence interval using the large-sample formula. Is it appropriate to use the ordinary large-sample confidence interval to obtain an estimation for the population proportion?

d. Why is it not appropriate to use the ordinary large-sample confidence interval in part c? Use a more appropriate approach and interpret the result.

8.57 Movie recommendation In a quick poll at the exit of a movie theater, 8 out of 12 randomly polled viewers said they would recommend the movie to their friends.

a. Construct an appropriate 95% confidence interval for the population proportion.

b. Is it plausible that only half of all the viewers will be willing to recommend the film to their friends? Explain.

8.58 Google Glass TRY Google started selling Google Glass (a type of wearable technology that projects information onto the eye) in the United States in spring 2014 for

$1500. A national poll reveals that out of 500 people sampled, nobody owned Google Glass.

a. Find the sample proportion who don't own Google Glass and its standard error.

b. Find a 95% confidence interval, using the large-sample formula. Is it sensible to conclude that no one in the United States owns Google Glass?

c. Why is it not appropriate to use the ordinary large-sample confidence interval in part b? Use a more appropriate approach and interpret the result.

d. Is it plausible to say that fewer than 1% of the population own Google Glass? Explain.

8.5 Practicing the Basics

8.59 Why bootstrap? Explain the purpose of using the bootstrap method.

8.60 Bootstrap technique for estimation A quality control process consists of crashing cars at various speeds. For a sample of 10 crash tests, the closing speed values (in kilometers per hour) were

61.83, 61.8, 31.82, 61.63, 32.32
32.52, 56.14, 61.67, 56.36, 56.57

a. Explain the steps of how the bootstrap method can be used to get a 95% confidence interval for the standard deviation of the closing speeds of all the crashed cars.

b. Use the Bootstrap web app accessible from the book's website or software to find the 95% bootstrap confidence interval.

8.61 Bootstrap interval for the mean In 2014, the General Social Survey interviewed 1399 randomly selected U.S. residents about how much time per week they spent surfing the web. The responses revealed a very right-skewed distribution, with a sample mean of 12 hours and a standard deviation of 15 hours, indicating that the population distribution is very far from normal. Should we be worried that one of the assumptions for using the t confidence interval from Section 8.3 for the population mean is violated? Let's find the bootstrap confidence interval for the population mean. The data from the survey are available from the book's website under the name "wwwhours".

a. Draw a histogram and find the sample mean and median time people spent surfing the web per week

b. Find a 95% confidence interval for the mean time using the t interval from Section 8.3.

c. Explain the steps of how you would obtain a 95% bootstrap confidence interval for the mean time.

d. Use the Bootstrap web app mentioned in Activity 3 or other software to find the 95% bootstrap confidence interval for the mean. (If you are using the web app, you can open the dataset in a spreadsheet and then copy and paste the values for the variable "hours" into the text field of the web app.)

e. The results from both approaches are virtually identical. Look at the bootstrap distribution of $\bar{x}$ displayed in the web app or other software. Is it surprising that it looks approximately normal? What theorem predicted this shape?

8.62 Bootstrap interval for the proportion We want a 90% confidence interval for the population proportion of students in a high school in Dallas, Texas, who can correctly find Iraq on an unlabeled globe. For a random sample of size 50, 5 get the correct answer.

a. Access the Sampling Distribution of the Sample Proportion web app from the book's website and set the population proportion to $p = 0.1$, pretending that the population proportion is equal to the sample proportion $(10/50 = 0.1)$. Generate one random sample of size $n = 50$ and find the sample proportion of correct answers.

b. Take 1,000 resamples like the one in part a (just click the 1,000 option) and find the 5th and 95th percentiles of the simulated sampling distribution. (*Note:* Click Find Percentile to find the percentiles.) This is the 90% bootstrap confidence interval for p.

c. Explain why the sample proportion does not fall exactly in the middle of the bootstrap confidence interval. (*Hint:* Is the sampling distribution symmetric or skewed?)

9.1 Practicing the Basics

9.1 **H_0 or H_a?** For parts a and b, is the statement a null hypothesis, or an alternative hypothesis?

a. In Canada, the proportion of adults who favor legalized gambling equals 0.50.

b. The proportion of all Canadian college students who are regular smokers is less than 0.24, the value it was 10 years ago.

c. Introducing notation for a parameter, state the hypotheses in parts a and b in terms of the parameter values.

9.2 **H_0 or H_a?** State whether each of the following statements is a null hypothesis or an alternative hypothesis. Why?

a. In 2016, the average price of solar energy in the United States was $3.70 per watt.

b. At least 1 out of every 8 women in the United States will develop breast cancer during her lifetime.

c. Only 48% of all the money donated by telemarketers actually goes to charitable fundraising campaigns.

9.3 **Burden of proof** For a new pesticide, should the Environmental Protection Agency (EPA) bear the burden of proof to show that it is harmful to the environment, or should the producer of the pesticide have to show that it is not harmful to the environment? The pesticide is considered harmful if its toxicity level exceeds a certain threshold and not harmful if its toxicity level is below the threshold. Consider the statements, "The mean toxicity level equals the threshold," "The mean toxicity level exceeds the threshold," and "The mean toxicity level is below the threshold."

a. Which of these statements should be the null and which the alternative hypothesis when the burden of proof is on the EPA to show that the new pesticide is harmful?

b. Which of these statements should be the null and which the alternative hypothesis when the burden of proof is on the producer to show that the new pesticide is not harmful?

9.4 **Electricity prices** According to the U.S. Energy Information Administration, the average monthly household electricity bill in 2014 was $114 before taxes and fees. A consumer association plans to investigate if the average amount has changed this year. Define the population parameter of interest and state the null and alternative hypotheses for this investigation.

9.5 **Low-carbohydrate diet** A study plans to have a sample of obese adults follow a proposed low-carbohydrate diet for three months. The diet imposes limited eating of starches (such as bread and pasta) and sweets, but otherwise no limit on calorie intake. Consider the hypothesis,

The population mean of the values of weight change (= weight at start of study − weight at end of study) is larger than zero.

a. Is this a null or an alternative hypothesis? Explain your reasoning.

b. Define a relevant parameter and express the hypothesis that the diet has no effect in terms of that parameter. Is it a null or an alternative hypothesis?

9.6 **Examples of hypotheses** Give an example of a null hypothesis and an alternative hypothesis about a (a) population proportion and (b) population mean.

9.7 **Proper hypotheses?** Suggest a way to correct each set of null and alternative hypotheses shown such that a proper set of hypotheses can be formed, and then illustrate them through an example.

a. $H_0: \hat{p} = 0.50$, $H_a: \hat{p} > 0.50$

b. $H_0: \mu = 10$, $H_a: \mu = 20$

c. $H_0: p < 0.30$, $H_a: p = 0.10$

9.8 ***z* test statistic** To test $H_0: p = 0.50$ that a population proportion equals 0.50, the test statistic is a z-score that measures the number of standard errors between the sample proportion and the H_0 value of 0.50. If $z = 3.6$, do the data support the null hypothesis, or do they give strong evidence against it? Explain.

9.9 **P-value** Indicate whether each of the following P-values gives strong evidence or not especially strong evidence against the null hypothesis.

a. 0.38

b. 0.001

9.10 Customer satisfaction TRY A customer of a car workshop claimed that majority of customers were not satisfied with the services provided. In order to test this claim, officials in charge of the workshop delegated a third-party statistical company to administrate a satisfaction survey of its current customers. State the parameter of interest and the hypotheses for a significance test for testing this claim, where the alternative hypothesis will reflect the customer's claim.

9.11 Believe in astrology? TRY You plan to apply significance testing to your own experiment for testing astrology, in which astrologers have to guess which of four personality profiles is the correct one for someone who has a particular horoscope. Define the parameter of interest and state the hypotheses, letting one hypothesis reflect the possibility that the astrologers' predictions could be better than random guessing.

9.12 Get P-value from *z* For a test of H_0: $p = 0.50$, the z test statistic equals 1.04.

a. Find the P-value for H_a: $p > 0.50$.

b. Find the P-value for H_a: $p \neq 0.50$.

c. Find the P-value for H_a: $p < 0.50$. (*Hint:* The P-values for the two possible one-sided tests must sum to 1.)

d. Do any of the P-values in part a, part b, or part c give strong evidence against H_0? Explain.

9.13 Get more P-values from *z* Refer to the previous exercise. Suppose $z = 2.50$ instead of 1.04.

a. Find the P-value for (i) H_a: $p > 0.50$, (ii) H_a: $p \neq 0.50$, and (iii) H_a: $p < 0.50$.

b. Do any of the P-values in part a provide strong evidence against H_0? Explain.

9.14 Find test statistic and P-value For a test of H_0: $p = 0.50$, the sample proportion is 0.35 based on a sample size of 100.

a. Show that the test statistic is $z = -3.0$.

b. Find the P-value for H_a: $p < 0.50$.

c. Does the P-value in part b give much evidence against H_0? Explain.

9.15 Dogs and cancer TRY A recent study[4] considered whether dogs could be trained to detect whether a person has lung cancer or breast cancer by smelling the subject's breath. The researchers trained five ordinary household dogs to distinguish, by scent alone, exhaled breath samples of 55 lung and 31 breast cancer patients from those of 83 healthy controls. A dog gave a correct indication of a cancer sample by sitting in front of that sample when it was randomly placed among four control samples. Once trained, the dogs' ability to distinguish cancer patients from controls was tested using breath samples from subjects not previously encountered by the dogs. (The researchers blinded both dog handlers and experimental observers to the identity of breath samples.) Let p denote the probability a dog correctly detects a cancer sample placed among five samples when the other four are controls.

a. Set up the null hypothesis that the dog's predictions correspond to random guessing.

b. Set up the alternative hypothesis to test whether the probability of a correct selection *differs* from random guessing.

c. Set up the alternative hypothesis to test whether the probability of a correct selection is *greater than* with random guessing.

d. In one test with 83 Stage I lung cancer samples, the dogs correctly identified the cancer sample 81 times. The test statistic for the alternative hypothesis in part c was $z = 17.7$. Report the P-value to three decimal

[4] M. McCulloch et al., *Integrative Cancer Therapies*, vol 5, p. 30, 2006.

places and interpret. (The success of dogs in this study made researchers wonder whether dogs can detect cancer at an earlier stage than conventional methods such as MRI scans.)

9.16 Religion important in your life? Americans ages 18 to 29 are considered to be less religious than older Americans. According to recent studies by the Pew Forum on Religion & Public Life, fewer young adults are affiliated with a specific religion than older people today. And, compared with their elders, fewer young people say that religion is very important in their lives. Yet, many young people still believe in traditional religious concepts and practices. Pew Research Center surveys show, for example, that "young adults' beliefs about life after death and the existence of heaven, hell and miracles closely resemble the beliefs of older people today." According to GSS (General Social Survey) results from a random sample of 1,679 subjects, 45% in the 18–29 age group pray daily whereas 55% pray less often.[5] The MINITAB output shows the results for a significance test for which the alternative hypothesis is that the percentage of 18–29-year-olds who pray daily differs from 50%. State and interpret the five steps of a significance test in this context, using information shown in the output to provide the particular values for the hypothesis, test statistic, and P-value.

Test and CI for One Proportion

```
Test of p = 0.5 vs p ≠ 0.5
 X    N   Sample p      95% CI      Z-Value  P-Value
756 1679 0.450268  (0.426, 0.474)   −4.08    0.000
Using the normal approximation.
```

9.17 Another test of astrology Examples 1, 3, and 5 referred to a study about astrology. Another part of the study used the following experiment: Professional astrologers prepared horoscopes for 83 adults. Each adult was shown three horoscopes, one of which was the one an astrologer prepared for him or her and the other two were randomly chosen from ones prepared for other subjects in the study. Each adult had to guess which of the three was his or hers. Of the 83 subjects, 28 guessed correctly.

a. Define the parameter of interest and set up the hypotheses to test that the probability of a correct prediction is 1/3 against the astrologers' claim that it exceeds 1/3.

b. Show that the sample proportion $= 0.337$, the standard error of the sample proportion for the test is 0.052, and the test statistic is $z = 0.08$.

c. Find the P-value. Would you conclude that people are more likely to select their horoscope than if they were randomly guessing, or are results consistent with random guessing?

9.18 Opinion on fracking a year earlier TRY The question about the opinion on the increased use of fracking from the November 2014 survey mentioned in Example 6 was also included in an earlier survey in September 2013. Using this earlier survey, let's again focus on those who oppose the increased use of fracking.

a. Define the parameter of interest and set up hypotheses to test that those who oppose fracking in 2013 are in the minority.

b. Of the 1506 respondents in the 2013 survey, 740 indicated that they oppose the increased use of fracking. Find and interpret the test statistic.

c. Report the P-value. Indicate your decision, in the context of this survey, using a 0.05 significance level.

d. Check whether the sample size was large enough to conduct the inference in part c. Indicate what the assumptions are for your inferences to apply to the entire U.S. population.

e. Find the P-value for the two-sided alternative that the proportion opposing is different from 0.50.

9.19 Testing a headache remedy TRY Studies that compare treatments for chronic medical conditions such as headaches can use the same subjects for each treatment. This type of study is commonly referred to as a crossover design. With a crossover design, each person crosses over from using one treatment to another during the study. One such study considered a drug (a pill called Sumatriptan) for treating migraine headaches in a convenience sample of children.[6] The study observed each of 30 children at two times when he or she had a migraine headache. The child received the drug at one time and a placebo at the other time. The order of treatment was randomized and the study was double-blind. For each child, the response was whether the drug or the placebo provided better pain relief. Let p denote the proportion of children having better pain relief with the drug in the population of children who suffer periodically from migraine headaches. Can you conclude that p is different from 0.5, that is, either more or less than half of the population is getting better pain relief with the drug? Of the 30 children, 22 had more pain relief with the drug and 8 had more pain relief with the placebo.

a. For testing H_0: $p = 0.50$ against H_a: $p \neq 0.50$, show that the test statistic $z = 2.56$.

b. Show that the P-value is 0.01. Interpret.

c. Check the assumptions needed for this test and discuss the limitations due to using a convenience sample rather than a random sample.

9.20 Gender bias in selecting managers TRY For a large supermarket chain in Florida, a women's group claimed that female employees were passed over for management training in favor of their male colleagues. The company denied this claim, saying it picked the employees from the eligible pool at random to receive this training. Statewide, the large pool of more than 1000 eligible employees who can be tapped for management training is 40% female and 60% male. Since this program began, 28 of the 40 employees chosen for management training were male and 12 were female.

a. The company claims that it selected employees for training according to their proportion in the pool of eligible employees. Define a parameter of interest and

[5] Accessed from pewforum.org/Age/Religion-Among-the-Millennials.aspx.

[6] Data based on those in a study by M. L. Hamalainen et al., reported in *Neurology,* vol. 48, pp. 1100–1103, 1997.

state this claim as a hypothesis. Explain why this hypothesis is a no-effect hypothesis.

b. State the null and alternative hypotheses for a test to investigate the strength of evidence to support the women's claim of gender bias. (*Hint:* Gender bias means that either males or females are disproportionately selected.)

c. The table shows results of using MINITAB to do a large-sample analysis. Explain why the large-sample analysis is justified and show how software obtained the test statistic value.

```
Test of p = 0.4 vs p ≠ 0.4
X`   N   Sample p      95% CI      Z-Value  P-Value
12  40  0.300000  (0.158, 0.442)   −1.29    0.197
Using the normal approximation.
```

d. To what alternative hypothesis does the P-value in the table refer? Use it to find the P-value for the significance test you specified in part b and interpret it.

e. What decision would you make for a 0.05 significance level? Interpret.

9.21 Gender discrimination Refer to the 95% confidence interval shown in the MINITAB output in the previous exercise. What are the plausible values for the probability of a female to be selected for training? Is this in accordance with our decision of the significance test from part e of the previous exercise? Explain.

9.22 Use of complementary and integrative health (CIH) strategies among nurses in Iran A study (*J Integr Med.* 2016; 14(2): 121–127) was conducted between May 2014 and April 2015 to assess the knowledge, attitude, and use of CIH strategies among nurses in Iran. In this study, 157 nurses from two urban hospitals of Zabol University of Medical Sciences in southeast Iran took part and their responses were analyzed. Most nurses ($n = 95, 60.5\%$) had some knowledge about the strategies. However, a majority ($n = 90, 57.3\%$) of the nurses never applied CIH methods. Does this suggest that nurses who never applied CIH methods would constitute a majority of the population, or are the results consistent with random variation? Answer by:

a. Identifying the relevant variable and parameter. (*Hint:* The variable is categorical with two categories. The parameter is a population proportion for one of the categories.)

b. Stating hypotheses for a large-sample two-sided test and checking that sample size guidelines are satisfied for that test.

c. Finding the test statistic value.

d. Finding and interpreting a P-value and stating the conclusion in context.

9.23 Performance of Egypt's president A poll was conducted between 18 and 20 April, 2016, by the Egyptian Center for Public Opinion Research on the performance of President Abdel Fattah el-Sisi at the end of his 22nd month in office. Out of all the respondents, 51% strongly approved his performance. The poll consisted of 709 responses obtained by randomly sampling citizens aged 18 years and above, covering all governorates. Test that the population proportion of those who approve highly of the president's performance was 0.50 against the alternative that it differed from 0.50. Carry out the five steps of a significance test, at the significance level of 0.05, reporting and interpreting the P-value in context.

9.24 Which cola? The 49 students in a class at the University of Florida made blinded evaluations of pairs of cola drinks. For the 49 comparisons of Coke and Pepsi, Coke was preferred 29 times. In the population that this sample represents, is this strong evidence that a majority prefers one of the drinks? Refer to the following MINITAB printout.

Test and CI for One Proportion

```
Test of p = 0.5 vs. p > 0.5
 X   N  Sample p  95% Lower Bound  Z-Value  P-Value
29  49  0.591837      0.476346       1.29    0.099
Using the normal approximation
```

Test and CI for One Proportion

```
Test of p = 0.5 vs. p ≠ 0.5
 X  N Sample p        95% CI        Z-Value P-Value
29 49 0.591837 (0.454221, 0.729452)   1.29    0.199
Using the normal approximation
```

a. Explain how to get the test statistic value that MINITAB reports.

b. Explain how to get the P-value. Interpret it.

c. Based on the result in part b, does it make sense to accept H_0? Explain.

d. What does the 95% confidence interval tell you that the test does not?

9.25 How to sell a burger A fast-food chain wants to compare two ways of promoting a new burger (a turkey burger). One way uses a coupon available in the store. The other way uses a poster display outside the store. Before the promotion, its marketing research group matches 50 pairs of stores. Each pair has two stores with similar sales volume and customer demographics. The store in a pair that uses coupons is randomly chosen, and after a month-long promotion, the increases in sales of the turkey burger are compared for the two stores. The increase was higher for 28 stores using coupons and higher for 22 stores using the poster. Is this strong evidence to support the coupon approach, or could this outcome be explained by chance? Answer by performing all five steps of a two-sided significance test about the population proportion of times the sales would be higher with the coupon promotion.

9.26 ◆◆ **A binomial headache** A null hypothesis states that the population proportion p of headache sufferers who have better pain relief with aspirin than with another pain reliever equals 0.50. For a crossover study with 10 subjects, all 10 have better relief with aspirin. If the null hypothesis were true, by the binomial distribution the probability of this sample result (which is the most extreme) equals $(0.50)^{10} = 0.001$. In fact, this is the small-sample P-value for testing H_0: $p = 0.50$ against H_a: $p > 0.50$. Does this P-value give (a) strong evidence in favor of H_0 or (b) strong evidence against H_0? Explain why.

9.27 ◆◆ **P-value for small samples** Example 4, on whether dogs can detect bladder cancer by selecting the correct urine specimen (out of seven), used the normal sampling distribution to find the P-value. The normal distribution P-value approximates a P-value using the binomial distribution. That binomial P-value is more appropriate when either expected count is less than 15. In Example 4, n was 54, and 22 of the 54 selections were correct.

a. If H_0: $p = 1/7$ is true, $X =$ number of correct selections has the binomial distribution with $n = 54$ and $p = 1/7$. Why?

b. For H_a: $p > 1/7$, with $x = 22$, the small sample P-value using the binomial is $P(22)+P(23)+\cdots+P(54)$, where $P(x)$ denotes the binomial probability of outcome x with $p = 1/7$. (This equals 0.0000019.) Why would the P-value be this sum rather than just P(22)?

9.3 Practicing the Basics

9.28 Which *t* has P-value = 0.05? A *t* test for a mean uses a sample of 15 observations. Find the *t* test statistic value that has a P-value of 0.05 when the alternative hypothesis is (a) H_a: $\mu \neq 0$, (b) H_a: $\mu > 0$, and (c) H_a: $\mu < 0$. (Among others, you can use the *t* Distribution web app to find the answer.)

9.29 Practice mechanics of a *t* test A study has a random sample of 20 subjects. The test statistic for testing H_0: $\mu=100$ is $t = 2.40$. Find the approximate P-value for the alternative, (a) H_a: $\mu \neq 100$, (b) H_a: $\mu > 100$, and (c) H_a: $\mu < 100$. (Among others, you can use the *t* Distribution web app to find the answer.)

9.30 Effect of *n* Refer to the previous exercise. If the same sample mean and standard deviation had been based on $n = 5$ instead of $n = 20$, the test statistic would have been $t = 1.20$. Would the P-value for H_a: $\mu \neq 100$ be larger, or smaller, than when $t = 2.40$? Why?

9.31 Photovoltaic solar energy in Europe According to Eurostat, the yearly growth rate of renewable energy production from photovoltaic cells is, on average, 1.65 in Europe and 1.57 in Germany. However, the rate is likely to decrease with time. The StatCrunch output shows the results for a significance

test for which the alternative hypothesis is: "the average of the yearly growth rate of renewable energy production from photovoltaic cells in Germany in the last 5 years differs from 1.57". (*Source*: http://ec.europa.eu/eurostat/statistics-explained/index.php/Renewable_energy_statistics)

One sample T hypothesis test:

μ: Mean of variable

H_0: μ = 1.57

H_A: μ ≠ 1.57

Hypothesis test results:

Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
X	1.3388108	0.11834682	3	−1.9534892	0.1458

a. Identify the sample mean of the yearly growth rate of renewable energy production from photovoltaic cells in the sample.

b. Identify the P-value for this test.

c. What conclusion about H_0 can be drawn from this statistical output?

9.32 **Female work week** When the 583 female workers in the 2012 GSS were asked how many hours they worked in the previous week, the mean was 37.0 hours, with a standard deviation of 15.1 hours. Does this suggest that the population mean work week for females is significantly different from 40 hours? Answer by:

a. Identifying the relevant variable and parameter.

b. Stating null and alternative hypotheses.

c. Reporting and interpreting the P-value for the test statistic value.

d. Explaining how to make a decision for the significance level of 0.01.

9.33 **StatCrunch for statistics** For effective learning, an instructor advised his students to practice solving statistical problems on StatCrunch, a web-based software, at least 7 hours per week. In an in-class activity, a student surveyed 15 of her 35 classmates to assess the number of hours they spend per week practicing statistics on StatCrunch. The following data was collected:

6.5, 4, 3.5, 0, 12, 5, 12, 15, 12, 0, 1, 8, 0.5, 2, 7

Is there strong evidence that the mean number of hours of StatCrunch usage in the entire class is larger than 7? Answer by:

a. Identifying the relevant variable and parameter.

b. Stating null and alternative hypotheses.

c. Finding and interpreting the test statistic value.

d. Reporting and interpreting the P-value and stating the conclusion in context.

9.34 **Lake pollution** An industrial plant claims to discharge no more than 1000 gallons of wastewater per hour, on the average, into a neighboring lake. An environmental action group decides to monitor the plant in case this limit is being exceeded. Doing so is expensive, and only a small sample is possible. A random sample of four hours is selected over a period of a week. The observations (gallons of wastewater discharged per hour) are

2000, 1000, 3000, 2000.

a. Show that $\bar{x} = 2000$, $s = 816.5$, and standard error = 408.25.

b. To test H_0: $\mu = 1000$ vs. H_a: $\mu > 1000$, show that the test statistic equals 2.45.

c. Using Table B or software, show that the P-value is less than 0.05, so there is enough evidence to reject the null hypothesis at the 0.05 significance level.

d. Explain how your one-sided analysis in part b implicitly tests the broader null hypothesis that $\mu \leq 1000$.

9.35 **Weight change for controls** A disadvantage of the experimental design in Example 8 on weight change in anorexic girls is that girls could change weight merely from participating in a study. In fact, girls were randomly assigned to receive a therapy or to serve in a control group, so it was possible to compare weight change for the therapy group to the control group. For the 26 girls in the control group, the weight change had $\bar{x} = -0.5$ and $s = 8.0$. Repeat all five steps of the test of H_0: $\mu = 0$ against H_a: $\mu \neq 0$ for this group and interpret the P-value.

9.36 **Water fluoridation** Fluoridated water has fluoride at a level that is effective for preventing cavities. A study determined the number of cavity-free children (NCF) per 100 children in 16 North American cities BEFORE and AFTER public water fluoridation projects.

AFTER	BEFORE	AFTER	BEFORE	AFTER	BEFORE	AFTER	BEFORE
49.2	18.2	3.4	2.8	23	25	46.8	25.6
30	21.9	16.8	21	17	13	84.9	50.4
16	5.2	10.7	11.3	79	76	65.2	41.2
47.8	20.4	5.7	6.1	66	59	52	21

Let μ denote the population mean of the difference between the NCF values BEFORE and AFTER public water fluoridation projects. Use a calculator or software for the following analyses:

a. Form the 16 difference scores, for instance 49.2 − 18.2 = 31 for city 1 and 52 − 21 = 31 for city 16, always taking AFTER–BEFORE. Construct a dot plot or a box plot. Describe the sample data distribution.

b. Carry out the five steps of the significance test for a mean of the difference scores, using H_0: $\mu = 0$ and H_a: $\mu \neq 0$.

c. Discuss whether the assumptions seem valid for this example. What is the impact of using a convenience sample?

9.37 **Too little or too much wine?** Wine-pouring vending machines, previously available in Europe and international airports, have become popular in the past few years in the United States. They are even approved to dispense wine in some Walmart stores. The available pouring options are a 5-ounce glass, a 2.5-ounce half-glass, and a 1-ounce taste. When the machine is in statistical control (see Exercise 7.43), the amount dispensed for a full glass is 5.1 ounces. Four observations are taken each day to plot a daily mean over time on a control chart to check for irregularities. The most recent day's observations were 5.05, 5.15, 4.95, and 5.11. Could the difference between the sample mean and the target value be due to random

variation, or can you conclude that the true mean is now different from 5.1? Answer by showing the five steps of a significance test, making a decision using a 0.05 significance level.

9.38 Selling a burger In Exercise 9.25, a fast-food chain compared two ways of promoting a turkey burger. In a separate experiment with 10 pairs of stores, the difference in the month's increased sales between the store that used coupons and the store with the outside poster had a mean of $3000. Does this indicate a true difference between mean sales for the two advertising approaches? Answer by using the output shown to test that the population mean difference is 0, carrying out the five steps of a significance test. Make a decision using a 0.05 significance level.

One-Sample T

```
Test of μ = 0 vs ≠ 0
 N   Mean   StDev  SE Mean     95% CI      T      P
10   3000    4000     1265   (139, 5861)  2.37  0.042
```

9.39 Assumptions important? Refer to the previous exercise.

a. Explain how the result of the 95% confidence interval shown in the table agrees with the test decision using the 0.05 significance level.

b. Suppose you instead wanted to perform a one-sided test because the study predicted that the increase in sales would be higher with coupons. Explain why the normal population assumption might be problematic.

9.40 Anorexia in teenage girls Example 8 described a study about various therapies for teenage girls suffering from anorexia. For each of 17 girls who received the family therapy, the changes in weight were

11, 11, 6, 9, 14, −3, 0, 7, 22, −5, −4, 13, 13, 9, 4, 6, 11.

a. Plot these data with a dot plot or box plot and summarize.

b. Verify that the weight changes have $\bar{x} = 7.29$, $s = 7.18$, and $se = 1.74$ pounds.

c. Give all steps of a significance test about whether the population mean was 0 against an alternative designed to see whether there is any effect.

9.41 Sensitivity study Ideally, results of a statistical analysis should not depend greatly on a single observation. To check this, it's a good idea to conduct a **sensitivity study.** This entails redoing the analysis after deleting an outlier from the data set or changing its value to a more typical value and checking whether results change much. If results change little, this gives us more faith in the conclusions that the statistical analysis reports. For the weight changes in Table 9.4 from the anorexia study (shown again here and available on the book's website), the greatest reported value of 20.9 pounds was a severe outlier. Suppose this observation was actually 2.9 pounds but was incorrectly recorded. Redo the two-sided test of that example and summarize how the results differ. Does the ultimate conclusion depend on that single observation?

Weight Changes in Anorexic Girls		
1.7	11.7	−1.4
0.7	6.1	−0.8
−0.1	1.1	2.4
−0.7	−4.0	12.6
−3.5	20.9	1.9
14.9	−9.3	3.9
3.5	2.1	0.1
17.1	1.4	15.4
−7.6	−0.3	−0.7
1.6	−3.7	

9.42 Test and CI Results of 99% confidence intervals are consistent with results of two-sided tests with which significance level? Explain the connection.

9.4 Practicing the Basics

9.43 Dr. Dog In the experiment in Example 4, we got a P-value < 0.001 for testing H_0: $p = 1/7$ about dogs' ability to diagnose urine from bladder cancer patients.

a. For the significance level 0.05, what decision would you make?

b. If you made an error in part a, what type of error was it? Explain what the error means in context of the dog experiment.

9.44 Error probability A significance test about a mean is conducted using a significance level of 0.05. The test statistic equals 10.52. The P-value is 0.003.

a. If H_0 was true, for what probability of a Type I error was the test designed?

b. If the P-value was 0.3 and the test resulted in a decision error, what type of error was it?

9.45 Fracking errors Example 6, in testing H_0: $p = 0.5$ against H_a: $p < 0.5$, analyzed whether those opposing increased use of fracking are in the minority. In the words of that example, what would be (a) a Type I error and (b) a Type II error?

9.46 Anorexia errors Example 8 tested a therapy for anorexia, using hypotheses H_0: $\mu = 0$ and H_a: $\mu \neq 0$ about the population mean weight change μ. In the words of that example, what would be (a) a Type I error and (b) a Type II error?

9.47 Anorexia decision Refer to the previous exercise. When we test H_0: $\mu = 0$ against H_a: $\mu > 0$, we get a P-value of 0.02.

a. What would the decision be for a significance level of 0.05? Interpret in context.

b. If the decision in part a is in error, what type of error is it?

c. Suppose the significance level were instead 0.01. What decision would you make, and if it is in error, what type of error is it?

9.48 Errors in the courtroom Consider the test of H_0: The defendant is not guilty against H_a: The defendant is guilty.

a. Explain, in context, the conclusion of the test if H_0 is rejected.

b. Describe, in context, a Type I error.

c. Explain, in context, the conclusion of the test if you fail to reject H_0.

d. Describe, in context, a Type II error.

9.49 Errors in medicine Consider the test of H_0: The new drug is safe against H_a: the new drug is not safe.

a. Explain, in context, the conclusion of the test if H_0 is rejected.

b. Describe, in context, a Type I error.

c. Explain, in context, the conclusion of the test if you fail to reject H_0.

d. Describe, in context, a Type II error.

9.50 Decision errors in prostate cancer detection In the year 2016, approximately 181,000 new prostate cancer cases were expected to be diagnosed in the United States and approximately 26,100 deaths were linked to prostate cancer (www.uptodate.com/contents/screening-for-prostate-cancer). The prostate-specific antigen (PSA) test is used to detect the presence of prostate cancer. Define the null hypothesis as the patient does not have prostate cancer. See the table for a summary of the possible outcomes:

Prostate Cancer Detection		
	Cancer Detection	
Cancer	**Negative**	**Positive**
No (H_0)	Correct	Type I error
Yes (H_a)	Type II error	Correct

a. When an oncologist interprets a PSA test result, explain why a Type I error is a false positive, predicting that a man has prostate cancer when actually he does not.

b. A Type II error is a false negative. What does this mean, and what is the consequence of such an error to the man?

c. An oncologist wants to reduce the chance of telling a man that he may have prostate cancer when actually he does not. Consequently, a positive test result will be reported only when there is *extremely* strong evidence that prostate cancer is present. What is the disadvantage of this approach?

9.51 Detecting pregnancy Home pregnancy tests claim an accuracy rate of over 99%. A positive test does turn out later to be a false positive. There are several reasons why a woman may obtain a positive result in a home pregnancy test when she is not actually pregnant.

a. For the home pregnancy test, explain what a Type I error is and explain the consequence to a woman of this type of error.

b. For the home pregnancy test, what is a Type II error? What is the consequence to a woman of this type of error?

c. To which diagnostic does the probability of 99% refer?

d. Can you state that the probability that a woman who receives a positive test result is not actually pregnant is 0.01? Explain your answer.

9.52 Which error is worse? Which error, Type I or Type II, would usually be considered more serious for decisions in the following tests? Explain why.

a. A trial to test a murder defendant's claimed innocence, when conviction results in the death penalty.

b. A medical diagnostic procedure, such as a mammogram.

9.5 Practicing the Basics

9.53 Misleading summaries? Two researchers conduct separate studies to test H_0: $p = 0.50$ against H_a: $p \neq 0.50$, each with $n = 400$.

a. Researcher A gets 220 observations in the category of interest, and $\hat{p} = 220/400 = 0.550$ and test statistic $z = 2.00$. Show that the P-value $= 0.046$ for Researcher A's analysis.

b. Researcher B gets 219 in the category of interest, and $\hat{p} = 219/400 = 0.5475$ and test statistic $z = 1.90$. Show that the P-value $= 0.057$ for Researcher B's analysis.

c. Using $\alpha = 0.05$, indicate in each case from part a and part b whether the result is "statistically significant." Interpret.

d. From part a, part b, and part c, explain why important information is lost by reporting the result of a test as "P-value ≤ 0.05" versus "P-value > 0.05," or as "reject H_0" versus "do not reject H_0," instead of reporting the actual P-value.

e. Show that the 95% confidence interval for p is (0.501, 0.599) for Researcher A and (0.499, 0.596) for Researcher B. Explain how this method shows that, in practical terms, the two studies had very similar results.

9.54 Practical significance TRY A study considers whether the mean score on a college entrance exam for students in 2010 is any different from the mean score of 500 for students who took the same exam in 1985. Let μ represent the mean score for all students who took the exam in 2010. For a random sample of 25,000 students who took the exam in 2010, $\bar{x} = 498$ and $s = 100$.

a. Show that the test statistic is $t = -3.16$.

b. Find the P-value for testing H_0: $\mu = 500$ against H_a: $\mu \neq 500$.

c. Explain why the test result is statistically significant but not practically significant.

9.55 Effect of *n* Example 10 analyzed political conservatism and liberalism in the United States. Suppose that the sample mean of 4.09 and sample standard deviation of 1.43 were from a sample size of only 25, rather than 2449.

a. Find the test statistic.

b. Find the P-value for testing H_0: $\mu = 4.0$ against H_a: $\mu \neq 4.0$. Interpret.

c. Show that a 95% confidence interval for μ is (3.5, 4.7).

d. Together with the results of Example 10, explain what this illustrates about the effect of sample size on (i) the size of the P-value (for a given mean and standard deviation) and (ii) the width of the confidence interval.

9.56 Fishing for significance A marketing study conducts 60 significance tests about means and proportions for several groups. Of them, 3 tests are statistically significant at the 0.05 level. The study's final report stresses only the tests with significant results, not mentioning the other 57 tests. What is misleading about this?

9.57 Selective reporting In 2004, New York Attorney General Eliot Spitzer filed a lawsuit against GlaxoSmithKline pharmaceutical company, claiming that the company failed to publish results of one of its studies that showed that an antidepressant drug (Paxil) may make adolescents more likely to commit suicide. Partly as a consequence, editors of 11 medical journals agreed to a new policy to make researchers and companies register all clinical trials when they begin, so that negative results cannot later be covered up. The *International Journal of Medical Journal Editors* wrote, "Unfortunately, selective reporting of trials does occur, and it distorts the body of evidence available for clinical decision-making." Explain why this controversy relates to the argument that it is misleading to report results only if they are "statistically significant." (*Hint:* See the subsection of this chapter on misinterpretations of significance tests.)

9.58 How many medical discoveries are Type I errors? Refer to Example 11. Using a tree diagram, given that H_0 is rejected, approximate P(Type I error) under the assumption that a true effect exists 20% of the time and that there's a 30% chance of a Type II error.

9.59 Selective reporting and p-hacking In the webcomic on the link http://xkcd.com/882/, a girl claims that jelly beans cause acne. Scientists investigate and find no link between the two ($p > 0.05$). They are asked to check if jelly beans of a particular color cause acne. They test 20 different colors each at a significance level of 5% and find a link between green jelly beans and acne. This leads to a newspaper headline, "Green Jellybeans Cause Acne" where the 5% chance of the link is mentioned as 95% confidence. When the scientists repeat the same experiment, they are unable to find any link between acne and color of jelly beans. They conclude that the earlier result might be coincidental. Using this example, explain why you need to have some skepticism when research suggests that some therapy or drug has an impact in treating a disease.

9.6 Practicing the Basics

9.60 Find P(Type II error) A study is designed to test H_0: $p = 0.50$ against H_a: $p > 0.50$, taking a random sample of size $n = 100$, using significance level 0.05.

a. Show that the rejection region consists of values of $\hat{p} > 0.582$.

b. Sketch a single picture that shows (i) the sampling distribution of $\hat{p}$ when H_0 is true and (ii) the sampling distribution of $\hat{p}$ when $p = 0.60$. Label each sampling distribution with its mean and standard error and highlight the rejection region.

c. Find P(Type II error) when $p = 0.60$.

9.61 Gender bias in selecting managers Exercise 9.20 tested the claim that female employees were passed over for management training in favor of their male colleagues.

Statewide, the large pool of more than 1000 eligible employees who can be tapped for management training is 40% female and 60% male. Let p be the probability of selecting a female for any given selection. For testing H_0: $p = 0.40$ against H_a: $p < 0.40$ based on a random sample of 50 selections, using the 0.05 significance level, verify that:

a. A Type II error occurs if the sample proportion falls less than 1.645 standard errors below the null hypothesis value, which means that $\hat{p} > 0.286$.

b. When $p = 0.20$, a Type II error has probability 0.06.

9.62 Balancing Type I and Type II errors Recall that for the same sample size, the smaller the probability of a Type I error, α, the larger the P(Type II error). Let's check this for Example 12. There we found P(Type II error) for testing H_0: $p = 1/3$ (astrologers randomly guessing) against H_a: $p > 1/3$ when actually $p = 0.50$, with $n = 116$. If we use $\alpha = 0.01$, verify that:

a. A Type II error occurs if the sample proportion falls less than 2.326 standard errors above the null hypothesis value, which means $\hat{p} < 0.435$.

b. When $p = 0.50$, a Type II error has probability 0.08. (By comparison, Example 12 found P(Type II error) $= 0.02$ when $\alpha = 0.05$, so we see that P(Type II error) increased when P(Type I error) decreased.)

9.63 P (Type II error) increase Consider the hypothesis H_0: $p = 0.3$ against H_a: $p > 0.3$ for testing at a 5% significance level.

a. Assume $n = 200$, find P(Type II error) when (i) $p = 0.4$ and (ii) $p = 0.35$

b. Assume $n = 100$, find P(Type II error) when (i) $p = 0.4$ and (ii) $p = 0.35$

c. Explain intuitively why P (Type II error) is larger when the parameter value is closer to the value in H_0 and when the sample size decreases.

9.64 Type II error with two-sided H_a In Example 12 for testing H_0: $p = 1/3$ (astrologers randomly guessing) with $n = 116$ when actually $p = 0.50$, suppose we used H_a: $p \neq 1/3$. Then show that:

a. A Type II error occurs if $0.248 < \hat{p} < 0.419$.

b. The probability is 0.00 that $\hat{p} < 0.248$ and 0.96 that $\hat{p} > 0.419$.

c. P(Type II error) $= 0.04$.

9.65 Power for knee osteoarthritis treatment An ankle-foot orthosis (AFO) is a specially designed brace to support and improve the function of the foot and ankle. A 2016 study on the treatment of knee osteoarthritis investigated the biomechanical effects of the Agilium Freestep AFO on the lever arm of the ground reaction force (GRF) in a gait analysis lab. Results show that the lever arm of the GRF was significantly reduced by 14% with the Agilium Freestep AFO (www.oandp.org). Statistical analyses were conducted using the Student's t-test with a power of 80%.

a. What should be the null and the alternative hypotheses in this study?

b. How should the power of 80% be interpreted?

c. In context, what is a Type II error for this test?

9.66 Exploring Type II errors Refer to the web app from Activity 2 at the end of this section, now assuming that we are using the two-sided test H_0: $p = 0.33$ against H_a: $p \neq 0.33$. (Select "not equal" for the type of the alternative hypothesis in the web app.)

a. Explain the effect of increasing the sample size on the probability of a Type II error when the true $p = 0.50$.

b. Use the app to find the sample size needed to achieve a power of at least 90% when truly $p = 0.50$.

c. For a fixed sample size, do you think the probability of a Type II error will increase or decrease when the true p is 0.40 instead of 0.50? Check your answer with the app.

d. Does the power increase or decrease when the significance level is 0.10 instead of 0.05? Check your answer with the app.

10.1 Practicing the Basics

10.1 Unemployment rate According to the Bureau of Labor Statistics, the official unemployment rate was 10.4% among blacks and 4.7% among whites as of February 2015. (www.bls.gov/).

a. Identify the response variable and the explanatory variable.

b. Identify the two groups that are the categories of the explanatory variable.

c. The unemployment statistics are based on a sample of individuals. Were the samples of white individuals and black individuals independent samples or dependent samples? Explain.

10.2 Sampling sleep The 2011 Bedroom Sleep poll of a random sample of 1500 adults reported that respondents slept an average of 6.5 hours on weekdays and 7.2 hours on weekends, and that 21% of respondents got eight or more hours of sleep on weekdays, whereas 44% got eight or more hours of sleep on weekends (www.sleepfoundation.org).

a. To compare the means or the percentages using inferential methods, should you treat the samples on weekdays and weekends as independent samples or as dependent samples? Explain.

b. To compare these results to polls of other people taken in previous years, should you treat the samples in the two years as independent samples or as dependent samples? Explain.

10.3 Basic life support knowledge In 2015, a survey of first-year university students in Brazil was conducted to determine if they knew how to activate the Mobile Emergency Attendance Service (MEAS). Of the 1038 respondents (59.5% studying biological sciences, 11.6% physical sciences, and 28.6% humanities), 54.3% students of nonbiological subjects ($n = 564$) knew how to activate the MEAS as compared to 61.4% students of biological sciences ($n = 637$). (*Source*: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4661033/)

a. Estimate the difference between the proportions of students of biological sciences and nonbiological subjects who know how to activate the MEAS and interpret.

b. Find the standard error for this difference. Interpret it.

c. Define the two relevant population parameters for comparison in the context of this exercise.

d. Construct and interpret a 95% confidence interval for the difference in proportions, explaining how your interpretation reflects whether the interval contains 0.

e. State and check the assumptions for the confidence interval in part d to be valid.

10.4 Smoking and lung obstruction A National Center for Health Statistics data brief published in 2015 (Nr. 181) looked at the association between lung obstruction and smoking status in adults 40 to 79 years old. In a random sample of 6927 adults without any lung obstruction, 54.1% never smoked. In a random sample of 1146 adults with lung obstruction (such as asthma or COPD), 23.8% never smoked.

a. Find and interpret a point estimate of the difference between the proportion of adults without and with lung obstruction who never smoked.

b. A 99% confidence interval for the true difference is (0.267, 0.339). Interpret.

c. What assumptions must you make for the interval in part b to be valid?

10.5 Risky behaviors among HIV positive female sex workers In 2014, questionnaire surveys were administrated among 181 female sex workers in the Yunnan province of China who confirmed themselves to be HIV positive (www.ncbi.nlm.gov/pubmed/26833008). The participants were divided into two age groups—76 cases were below 35 years and 105 cases were 35 years old and above. 26 females below 35 years and 54 females of ages 35 years and above reported using drugs. Let p_1 and p_2 denote the population proportions of females below 35 years and of females 35 years or above who took drugs, respectively.

a. Report point estimates of p_1 and p_2.

b. Construct a 95% confidence interval for $(p_1 - p_2)$, specifying the assumptions you made to use this method. Interpret.

c. Based on the interval in part b, explain why the proportion using drugs may have been quite a bit larger for females of ages 35 years or above, or it might have been only moderately larger.

10.6 Aspirin and heart attacks in Sweden A Swedish study used 1360 patients who had suffered a stroke. The study randomly assigned each subject to an aspirin treatment or a placebo treatment.[3] The table shows MINITAB output,

[3]Based on results described in *Lancet*, vol. 338, pp. 1345–1349 (1991).

where X is the number of deaths due to heart attack during a follow-up period of about 3 years. Sample 1 received the placebo and sample 2 received aspirin.

a. Explain how to obtain the values labeled "Sample p."

b. Explain how to interpret the value given for "estimate for difference."

c. Explain how to interpret the confidence interval, indicating the relevance of 0 falling in the interval.

d. If we instead let sample 1 refer to the aspirin treatment and sample 2 the placebo treatment, explain how the estimate of the difference and the 95% confidence interval would change. Explain how then to interpret the confidence interval.

Deaths due to heart attacks in Swedish study

```
Sample     X      N     Sample p
  1       28    684    0.040936
  2       18    676    0.026627
Difference = p(1) − p(2)
Estimate for difference: 0.0143085
95% CI for difference: (−0.00486898, 0.0334859)
Test for differnce = 0 (vs ≠ 0):
z = 1.46 P-Value = 0.144
```

10.7 Swedish study test Refer to the previous exercise.

a. State the hypotheses that were tested.

b. Explain how to interpret the P-value for the test.

c. Even though the difference between the sample proportions was larger than in the Physicians Health Study (Examples 2–4), the P-value is larger (i.e., the test results are less statistically significant). Explain how this could be. (*Hint:* How do the sample sizes compare for the two studies? How does this affect the standard error and thus the test statistic and P-value?)

d. Report the P-value for the one-sided alternative hypothesis that the chance of death due to heart attack is lower for the aspirin group.

10.8 Significance test for aspirin and cancer deaths study TRY In the study for cancer death rates, consider the null hypothesis that the population proportion of cancer deaths p_1 for placebo is the same as the population proportion p_2 for aspirin. The sample proportions were $\hat{p}_1 = 347/11535 = 0.0301$ and $\hat{p}_2 = 327/14035 = 0.0233$.

a. For testing H_0: $p_1 = p_2$ against H_a: $p_1 \neq p_2$, show that the pooled estimate of the common value p under H_0 is $\hat{p} = 0.026$ and the standard error is 0.002.

b. Show that the test statistic is $z = 3.4$.

c. Find and interpret the P-value in context.

10.9 Basic life support knowledge and willingness to enroll in a first-aid course. In the study of basic life support knowledge mentioned in Exercise 10.3, 51.1% students from non-biological subjects said they would enroll in a first-aid course versus 74.7% of students from biological subjects. Is this change statistically significant at the 0.05 significance level?

a. Specify assumptions, notation, and hypotheses for a two-sided test.

b. Show how to find the pooled estimate of the proportion to use in a test. (*Hint:* You need the count in each year rather than the proportion.) Interpret this estimate.

c. Find the test statistic and P-value. Make a decision using a significance level of 0.05.

10.10 Comparing marketing commercials TECH Two TV commercials are developed for marketing a new product. A volunteer test sample of 200 people is randomly split into two groups of 100 each. In a controlled setting, Group A watches commercial A and Group B watches commercial B. In Group A, 25 say they would buy the product. In group B, 20 say they would buy the product. The marketing manager who devised this experiment concludes that commercial A is better. Is this conclusion justified? Analyze the data. If you prefer, use software (such as MINITAB or a web app) for which you can enter summary counts.

a. Show all steps of your analysis (perhaps including an appropriate graph) and check assumptions.

b. Comment on the manager's conclusion and indicate limitations of the experiment.

10.11 Hormone therapy for menopause TECH The Women's Health Initiative conducted a randomized experiment to see whether hormone therapy was helpful for postmenopausal women. The women were randomly assigned to receive the estrogen plus progestin hormone therapy or a placebo. After five years, 107 of the 8506 on the hormone therapy developed cancer and 88 of the 8102 in the placebo group developed cancer. Is this a significant difference?

a. Set up notation and state assumptions and hypotheses.

b. Find the test statistic and P-value and interpret. (If you prefer, use software, such as MINITAB or a web app, for which you can conduct the analysis by entering summary counts.)

c. What is your conclusion for a significance level of 0.05? (The study was planned to be eight years long but was stopped after five years because of increased heart and cancer problems for the therapy group. This shows a benefit of doing two-sided tests, because results sometimes have the opposite direction from the expected one.)

10.12 Obama A/B testing To increase Barack Obama's visibility and to raise money for the campaign leading up to the 2008 presidential election, Obama's analytics team conducted an A/B test with his website. In the original version, the button to join the campaign read "Sign Up". In an alternative version, it read "Learn More". Of 77,858 visitors to the original version, 5851 clicked the button. Of 77,729 visitors to the alternative version, 6927 clicked the button. Is there evidence that one version was more successful than the other in recruiting campaign members?

```
Sample          X       N     Sample p
SignUp       77858    5851    0.075150
LearnMore    77729    6927    0.089117
Difference = p(SignUp) − p(LearnMore)
Estimate for difference: −0.013968
95% CI for difference: (−0.016695, −0.011239)
Test for difference = 0(vs ≠ 0):
Z = −10.03 P-Value = 0.000
```

a. Sketch an appropriate graph to compare the sample proportions visually.

b. Show all steps of a significance test, using the computer output. Define any parameters you are using when

specifying the hypotheses. Mention whether there is a significant difference at the 0.05 significance level.

c. Interpret the confidence interval shown in the output. Why is this interval more informative than just reporting the P-value?

10.13 Prevalence of allergen-specific IgE antibodies in school children A study was conducted in Japan to estimate the prevalence of allergen-specific IgE antibodies in children. A group of school children between the ages of 9 and 15 years were surveyed in 2001 and in 1996 (retrospectively). The authors of the study concluded that the percentage of positive cases for allergen-specific IgE antibodies in 2001 had increased as compared to 1996. Did the authors use the inferential methods (confidence interval, significance test) in this section to compare the two proportions? Explain. (*Source*: http://www.ncbi.nlm.nih.gov/pubmed/16883099)

10.2 Practicing the Basics

10.14 Energy drinks: health risks and toxicity A study was carried out in Saudi Arabia in which 31 male university students (18 overweight/obese and 13 having normal weight) were enrolled from December 2013 to December 2014 (www.annsaudimed.net). The heart rate variability was significantly less in obese subjects as compared to subjects with normal weight at 60 minutes after consuming an energy drink as indicated by the mean heart rate range MHRR (P-value = 0.012).

a. The conclusion was based on a significance test comparing means. Define notation in context, identify the groups and the population means and state the null hypothesis for the test.

b. What information you are not able to obtain from the P-value approach which you could learn if the confidence interval comparing the means was provided?

10.15 Address global warming You would like to determine what students at your school would be willing to do to help address global warming and the development of alternatively fueled vehicles. To do this, you take a random sample of 100 students. One question you ask them is, "How high of a tax would you be willing to add to gasoline (per gallon) to encourage drivers to drive less or to drive more fuel-efficient cars?" You also ask, "Do you believe (yes or no) that global warming is a serious issue that requires immediate action such as the development of alternatively fueled vehicles?" In your statistical analysis, use inferential methods to compare the mean response on gasoline taxes (the first question) for those who answer yes and for those who answer no to the second question. For this analysis,

a. Identify the response variable and the explanatory variable.

b. Are the two groups being compared independent samples or dependent samples? Why?

c. Identify a confidence interval you could form to compare the groups, specifying the parameters used in the comparison.

10.16 Housework for women and men Do women tend to spend more time on housework than men? If so, how much more? Based on data from the National Survey of Families and Households, one study reported the results in the table for the number of hours spent in housework per week. (*Source*: Data from A. Lincoln, *Journal of Marriage and Family,* vol. 70, 2008, pp. 806–814.)

	Housework Hours		
Gender	**Sample Size**	**Mean**	**Standard Deviation**
Women	476	33.0	21.9
Men	496	19.9	14.6

a. Based on this study, calculate how many more hours, on the average, women spend on housework than men.

b. Find the standard error for comparing the means. What factor causes the standard error to be small compared to the sample standard deviations for the two groups?

c. Calculate the 95% confidence interval comparing the population means for women and men. Interpret the result including the relevance of 0 being within the interval or not. (*Hint:* For such large sample sizes, the *t*-score is practically identical to a *z*-score.)

d. State the assumptions upon which the interval in part c is based.

10.17 More confident about housework Refer to part c in the previous exercise.

a. Show that a 99% confidence interval is (10.0, 16.2).

b. Explain why this interval is wider than the 95% confidence interval.

10.18 Employment by gender The study described in Exercise 10.16 also evaluated the weekly time spent in employment. This sample comprises men and women with a high level of labor force attachment. Software shows the results.

```
Gender      N      Mean     StDev    SE Mean
Men       496    47.54      9.92       0.45
Women     476    42.01      6.53       0.30
Difference = mu(Men) - mu(Women)
95% CI for difference: (4.477, 6.583)
T-Test of difference = 0 (vs ≠):
T-Value = 10.30  P-Value = 0.000
```

a. Does it seem plausible that employment has a normal distribution for each gender? Explain.

b. What effect does the answer to part a have on inference comparing population means? What assumptions are made for the inferences in this table?

c. Explain how to interpret the confidence interval.

d. Refer to part c. Do you think that the population means are equal? Explain.

10.19 Ideal number of children In 2014, the GSS asked, "What is the ideal number of children for a family to have?" For those giving a numerical response (and treating the response "7 or more" as 7), the following table shows summary statistics. (You can reproduce these data by typing CHLDIDEL(0–7) as the row variable, SEX as the column variable, and YEAR(2014) as the selection filter on the GSS website sda.berkeley.edu/GSS.)

	Ideal Number of Children		
Gender	**Sample Size**	**Mean**	**Standard Deviation**
Women	921	2.54	0.89
Men	754	2.50	0.85

a. Find the standard error for the difference in the sample means.

b. The sample size is large, resulting in degrees of freedom in excess of 100, so the *t*-score is essentially the same as the *z*-score. Using this, find and interpret the 95% confidence interval for the difference between the population means for females and males.

10.20 Annual income of CEOs A study analyzes the total annual pay of CEOs (in pounds) for a sample of UK companies over the period 2003–2006 categorized according to the number of compensation consultants employed. The sample included 311 firms having one consultant and 203 firms having two consultants. Software output shows the following results:

Two sample T hypothesis test:

μ_1 : Mean of CEO total pay in firms with one compensation consultant

μ_2 : Mean of CEO total pay in firms with two compensation consultants

$\mu_1 - \mu_2$: Difference between two means

$H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 \neq 0$

Sample Statistics:

Sample	N	Mean	Std. Dev.
Population 1	311	1658	1314
Population 2	203	1779	1461

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	−121	126.75	399.12	−0.95	0.34

(*Source*: http://www.globalequity.org/geo/sites/default/files/SSRN-id1646926.pdf)

a. Does it seem plausible that income has a normal distribution for each firms' category? Explain.

b. What effect does the answer to part a have on inference comparing population means? What assumptions are made for the inferences in this table?

c. A 95% confidence interval for the difference in the population means for CEOs (men and women) is (£-370.19, £128.19). Interpret, indicating the relevance of £0 falling in the interval.

10.21 Bulimia CI A study of bulimia among college women (J. Kern and T. Hastings, *Journal of Clinical Psychology*, vol. 51, 1995, p. 499) studied the connection between childhood sexual abuse and a measure of family cohesion (the higher the score, the greater the cohesion). The sample mean on the family cohesion scale was 2.0 for 13 sexually abused students ($s = 2.1$) and 4.8 for 17 nonabused students ($s = 3.2$).

a. Find the standard error for comparing the means.

b. Construct a 95% confidence interval for the difference between the mean family cohesion for sexually abused students and non-abused students. Interpret.

10.22 Empagliflozin and renal function over time A study published in June 2016 in *New England Journal of Medicine* wanted to determine the long-term renal effects (measured by eGFR: estimated glomerular filtration rate) of empagliflozin in patients with type 2 Diabetes. 7020 patients with type 2 Diabetes at 590 sites in 42 countries received at least one dose of a study drug. Patients were randomly assigned to receive either empagliflozin (at a dose level of either 10 mg or 25 mg) or a placebo once daily in addition to standard care. The difference between the study groups in the average rate of change in eGFR was estimated after a duration of 4 weeks.

a. The authors stated that there was a short-term decrease in the eGFR in the empagliflozin groups, with 95% confidence interval of weekly decreases of 0.62 ± 0.04 ml per minute per 1.73 m^2 of body-surface area in the 10-mg group. Interpret the confidence interval.

b. The authors also provided a p-value that is <0.001 for the comparisons in eGFR means of the 10 mg dose empagliflozin group with the placebo group. Specify the hypotheses for this test of comparison of means, which was two-sided. Interpret the p-value obtained.

10.23 Nicotine dependence A study on nicotine dependence for teenage smokers obtained a random sample of seventh graders. The response variable was constructed from a questionnaire called the Hooked on Nicotine Checklist (HONC). This is a list of ten questions such as, "Have you ever tried to quit but couldn't?" and "Is it hard to keep from smoking in places where it is banned, like school?" The HONC score is the total number of questions to which a student answered yes, so each student's HONC score falls between 0 and 10. The higher the score, the more hooked that student is on nicotine. One explanatory variable considered in the study was whether a subject reported inhaling when smoking. The following table reports descriptive statistics.

	HONC Score		
Group	**Students**	**Mean**	**Standard Deviation**
Inhalers	237	2.9	3.6
Noninhalers	95	0.1	0.5

a. Explain why (i) the overwhelming majority of noninhalers must have had HONC scores of 0 and (ii) on average, those who reported inhaling answered yes to nearly three more questions than those who denied inhaling.

b. Might the HONC scores have been approximately normal for each group? Why or why not?

c. Find the standard error for the estimate $(\bar{x}_1 - \bar{x}_2) = 2.8$. Interpret.

d. The 95% confidence interval for $(\mu_1 - \mu_2)$ is (2.3, 3.3). What can you conclude about the population means for inhalers and noninhalers?

10.24 Inhaling affect HONC? Refer to the previous exercise.

a. Show that the test statistic for H_0: $\mu_1 = \mu_2$ equals $t = 11.7$. If the population means were equal, explain why it would be nearly impossible by random variation to observe this large a test statistic.

b. What decision would you make about H_0, at common significance levels? Can you conclude which group had higher mean nicotine dependence? How?

c. State the assumptions for the inference in this exercise.

10.25 Females or males more nicotine dependent? Refer to Exercise 10.23 about studying nicotine dependence using a random sample of teenagers. Of those seventh graders in the study who had tried tobacco, the mean HONC score was 2.8 ($s = 3.6$) for the 150 females and 1.6 ($s = 2.9$) for the 182 males.

a. Find a standard error for comparing the sample means. Interpret.

b. Find the test statistic and P-value for H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 \neq \mu_2$. Interpret and explain what (if any) effect gender has on the mean HONC score.

c. Do you think that the HONC scores were approximately normal for each gender? Why or why not? How does this affect the validity of the analysis in part b?

10.26 Female and male monthly smokers Refer to the previous exercise. A subject was called a monthly smoker if he or she had smoked cigarettes over an extended period of time. The 74 female monthly smokers had a mean HONC score of 5.4 ($s = 3.5$), and the 71 male monthly smokers had a mean HONC score of 3.9 ($s = 3.6$). Using software (such as the Comparing Means web app or MINITAB) that can conduct analyses using summary statistics, repeat parts b and c of the previous exercise.

10.27 Kuwaiti men versus Swedish men The following descriptive statistics were obtained from a study (Saud al-Obaidi et al., *Journal of Rehabilitation Research and Development,* vol. 40, 2003) that aimed to compare the weight of Kuwaiti men with Swedish men between the ages of 20 to 29 years.

	Group size	**Mean weight (kg)**	**Standard deviation**
Kuwaiti men	15	81.57	26.26
Swedish men	15	70.73	12.56

a. Using software (such as StatCrunch) which can conduct analyses using summary statistics, find the test statistic and P-value for H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 \neq \mu_2$. Interpret and explain what (if any) effect a country has on the mean weight of its men.

b. Do you think that the weights were approximately normal for each country? Why or why not? How does this affect the validity of the analysis in part a?

10.28 Kidnapping in southern and eastern European countries The following data on kidnapping offences in countries of east and south Europe in 2014 were obtained from https://data.unodc.org.

(Crime and Criminal Justice –> Crime –> Kidnapping –> Filter by Region and Sub Region as appropriate)

Eastern Europe: 31, 95, 12, 3, 292, 88, 369, 10

Southern Europe: 2, 1, 3, 1, 58, 297, 22, 376, 11, 5, 99, 8

Using statistical software,

a. Construct and interpret a plot comparing responses by region.

b. Construct and interpret a 95% confidence interval comparing population means for kidnapping counts in Eastern and Southern Europe in 2014.

c. Show all five steps of a significance test comparing the population means.

d. State and check the assumptions for part b and c.

10.29 Study time A graduate teaching assistant for Introduction to Statistics (STA 2023) at the University of Florida collected data from students in one of her classes in spring 2007 to investigate whether study time per week (average number of hours) differed between students in the class who planned to go to graduate school and those who did not. The data were as follows:

Graduate school: 15, 7, 15, 10, 5, 5, 2, 3, 12, 16, 15, 37, 8, 14, 10, 18, 3, 25, 15, 5, 5

No graduate school: 6, 8, 15, 6, 5, 14, 10, 10, 12, 5

Using software or a calculator,

a. Find the sample mean and standard deviation for each group. Interpret.

b. Find the standard error for the difference between the sample means. Interpret.

c. Find a 95% confidence interval comparing the population means. Interpret.

10.30 Gum flavor longevity In a test to determine the flavor longevity of a chewing gum, clients entering a store were asked to participate in an activity. The activity consisted of chewing a certain brand of gum and recording how long the gum flavor lasted in minutes. Records from groups of males and females were as follows:

Females: 15, 21, 29, 22, 19, 25, 35, 23

Males: 22, 24, 23, 30, 12, 17, 28

Use a statistical software (e.g., StatCrunch) to perform a two-sided significance test of the null hypothesis that the population mean is equal for the two groups. Show the software output and all five steps of a significance test comparing the population means. Interpret results in context.

10.31 Time spent on social networks As part of a class exercise, an instructor at a major university asks her students how many hours per week they spend on social networks. She wants to investigate whether time spent on social networks differs for male and female students at this university. The results for those age 21 or under were:

Males: 5, 7, 9, 10, 12, 12, 12, 13, 13, 15, 15, 20

Females: 5, 7, 7, 8, 10, 10, 11, 12, 12, 14, 14, 14, 16, 18, 20, 20, 20, 22, 23, 25, 40

a. Using software or a calculator, find the sample mean and standard deviation for each group. Interpret.

b. Find the standard error for the difference between the sample means.

c. Find and interpret a 90% confidence interval comparing the population means.

10.32 More time on social networks In the previous exercise, plot the data. Do you see any outliers that could influence the results? Remove the most extreme observation from each group and redo the analyses. Compare results and summarize the influence of the extreme observations.

10.33 Normal assumption The methods of this section make the assumption of a normal population distribution. Why do you think this is more relevant for small samples than for large samples? (*Hint:* What shape does the sampling distribution of $\bar{x}_1 - \bar{x}_2$ have for large samples, regardless of the actual shape of the population distributions?)

10.34 Vital capacity One of the authors of this book has his lung function checked every other year. At each checkup, his lung volume (called the forced vital capacity, or FVC) is measured before and after using an inhaler that contains medication against asthma. The last five checkups provided the following results (in liters):

	Checkup				
	1	**2**	**3**	**4**	**5**
Before using inhaler:	5.08	5.99	5.32	6.03	5.44
After using inhaler:	5.36	5.98	5.62	6.26	5.68

Can we use the inferential methods (confidence interval, significance test) developed in this section to compare the mean FVC before and after using the inhaler? If yes, use software to find a 95% confidence interval for the difference in the population means before and after using the inhaler and interpret it. If no, explain why.

10.35 Body dissatisfaction Female college student participation in athletics has increased dramatically over the past few decades. Sports medicine providers are aware of some unique health concerns of athletic women, including disordered eating. A study (M. Reinking and L. Alexander, *Journal of Athletic Training,* vol. 40, 2005, pp. 47–51) compared disordered-eating symptoms and their causes for collegiate female athletes (in lean and nonlean sports) and nonathletes. The sample mean of the body dissatisfaction assessment score was 13.2 ($s = 8.0$) for 16 lean sport athletes (those sports that place value on leanness, including distance running, swimming, and gymnastics) and 7.3 ($s = 6.0$) for the 68 nonlean sport athletes. Assuming equal population standard deviations,

a. Find the standard error for comparing the means.

b. Construct a 95% confidence interval for the difference between the mean body dissatisfaction for lean sport athletes and nonlean sport athletes. Interpret.

10.36 Body dissatisfaction test Refer to the previous exercise.

a. Find the P-value for testing whether the population means are equal. Use a two-sided alternative.

b. Summarize assumptions for the analysis in part a. Do you think the normality assumption is justified? If not, what is the consequence of violating it?

10.37 Surgery versus placebo for knee pain Refer to Example 10, "Arthroscopic Surgery." Here we show MINITAB output comparing mean knee pain scores for the placebo (Group 1) to lavage arthroscopic surgery (Group 2) assuming equal population standard deviations.

a. State and interpret the result of the confidence interval.

b. Is it reasonable to assume equal population standard deviations?

c. State all steps and interpret the result of the significance test.

```
Sample     N     Mean     StDev     SE Mean
  1       60     51.6      23.7        3.1
  2       61     53.7      23.7        3.0
Difference = μ(1) − μ(2)
Estimate for difference: −2.10
95% CI for difference: (−10.63, 6.43)
T-Test of difference = 0(vs ≠):
T-Value = −0.49 P-Value = 0.627 DF = 119
Both use Pooled StDev = 23.7000
```

10.38 Comparing clinical therapies A clinical psychologist wants to choose between two therapies for treating severe cases of mental depression. She selects six patients who are similar in their depressive symptoms and in their overall quality of health. She randomly selects three of the patients to receive Therapy 1, and the other three receive Therapy 2. She selects small samples for ethical reasons—if her experiment indicates that one therapy is superior, she will use that therapy on all her other depression patients. After one month of treatment, the improvement in each patient is measured by the change in a score for measuring severity of mental depression. The higher the score, the better. The improvement scores are

Therapy 1: 30, 45, 45
Therapy 2: 10, 20, 30

Analyze these data (you can use software if you wish), assuming equal population standard deviations.

a. Show that $\bar{x}_1 = 40, \bar{x}_2 = 20, s = 9.35, se = 7.64$, $df = 4$, and a 95% confidence interval comparing the means is $(-1.2, 41.2)$.

b. Explain how to interpret what the confidence interval tells you about the therapies. Why do you think that it is so wide?

c. When the sample sizes are very small, it may be worth sacrificing some confidence to achieve more precision. Show that a 90% confidence interval is (3.7, 36.3). At this confidence level, can you conclude that Therapy 1 is better?

10.39 Clinical therapies 2 Refer to the previous exercise.

a. For the null hypothesis, H_0: $\mu_1 = \mu_2$, show that $t = 2.62$ and the two-sided P-value $= 0.059$. Interpret.

b. What decision would you make in the test, using a (i) 0.05 and (ii) 0.10 significance level? Explain what this means in the context of the study.

c. Suppose the researcher had predicted ahead of time that Therapy 1 would be better. To which H_a does this correspond? Report the P-value for it, and make a decision with significance level 0.05.

10.40 Vegetarians more liberal? When a sample of social science graduate students at the University of Florida gave their responses on political ideology (ranging from 1 = very liberal to 7 = very conservative), the mean was

3.18 ($s = 1.72$) for the 51 nonvegetarian students and 2.22 ($s = 0.67$) for the 9 vegetarian students. Software for comparing the means provides the printout, which shows results first for inferences that assume equal population standard deviations and then for inferences that allow them to be unequal.

```
Sample    N     Mean     StDev    SE Mean
  1      51    3.18     1.72       0.24
  2       9    2.220    0.670      0.22
Difference = μ(1) − μ(2)
Estimate for difference: 0.960
95% CI for difference: (−0.210, 2.130)
T-Test of difference = 0(vs ≠): T-Value = 1.64
P-Value = 0.106 DF = 58
Both use Pooled StDev = 1.6162
95% CI for difference: (0.289, 1.631)
T-Test of difference = 0(vs ≠):
T-Value = 2.92 P-Value = 0.007 DF = 30
```

a. Explain why the results of the two approaches differ so much. Which do you think is more reliable?

b. State your conclusion about whether the true means are plausibly equal.

10.41 Teeth whitening results One scientific "test of whiteness" tested the effect of a self applied tooth-whitening peroxide gel system in a randomized, controlled clinical trial.[9] The 58 adults assigned to the gel whitening group applied the gel after normal brushing according to the manufacturer's instructions. The fluoride toothpaste group was instructed to brush twice a day. The procedure was repeated for both groups twice a day for 14 days. An experienced examiner determined the tooth shades comparing each tooth to the shade tabs from an accepted shade scale (Vita shade guide) at the start of the experiment to create a baseline and then after one and two weeks of product application. Changes between the baseline score and the one- and two-week assessments were expressed as a difference of the respective Vita score, with a positive difference indicating an improvement in tooth whiteness. The results of the study are shown in the following table.

Mean Vita Shade Score Recorded at Two Weeks and Change from Baseline

Group	n	Two Weeks Mean Vita Shade (s.d.)	Change From Baseline (s.d.)	Treatment Difference
Xtra White whitening gel	58	6.80 (2.48)	1.02 (1.32)	0.67 ($p < 0.05$)
Toothpaste only	59	7.01 (2.19)	0.35 (1.29)	

a. State the hypotheses that were tested for the change from baseline means.

b. The P-value is reported as <0.05 for the test comparing the means. Explain how to interpret this value.

c. Calculate the pooled standard error, the t statistic, and the resulting P-value.

d. The ratio of the change from baseline sample means was 2.91. Interpret this ratio.

[9]*Source*: Data from L. Z. Collins et. al., *Journal of Dentistry*, vol. 32, 2004, pp. 13–17.

10.42 Permuting therapies Refer to Exercise 10.38, which compared two therapies for depression patients. Suppose that in a different experiment, only four patients took part; two were randomly assigned to the group that received therapy 1 and the remaining two to the group that received therapy 2. The following table shows the improvement scores for these 4 patients.

Patient:	1	2	3	4
Therapy:	1	1	2	2
Score:	30	60	20	30

a. List all six possible ways these four patients could have been assigned to the two therapies. That is, give all six possible ways of dividing the four patients into two groups of two patients each.

b. Using the observed improvement scores shown in the table, find the sample means in each group and their difference under each possible assignment.

10.43 Permutations equally likely Refer to the previous exercise comparing improvement scores under two therapies for depression patients.

a. State the null hypothesis of equal population distributions in the context of this experiment.

b. Argue that if the null hypothesis is true, the sampling distribution for the difference in sample means is $P(-20) = 2/6, P(-10) = 1/6, P(10) = 1/6, P(20) = 2/6$.

c. Find the permutation P-value for testing the null hypothesis in part a (which implies the population mean scores are the same) against the alternative that the population mean improvement score is larger under therapy 1, using the difference in the sample means as a test statistic. Interpret it.

10.44 Two-sided permutation P-value Refer to the previous exercise and the sampling distribution mentioned there. What is the permutation P-value for the two-sided test with alternative hypothesis that the mean improvement scores are different under therapies 1 and 2? (*Hint:* As for any regular test, the two-sided P-value is the the sum of the tail probabilities of test statistic values as extreme as the observed one. Extreme differences are those with an absolute value as large as or larger than the absolute value of the observed difference.)

10.45 Time spent on social networks revisited Exercise 10.31 considered the following data on the number of hours students spent on social network sites per week:

Males: 5, 7, 9, 10, 12, 12, 12, 13, 13, 15, 15, 20

Females: 5, 7, 7, 8, 10, 10, 11, 12, 12, 14, 14, 14, 16, 18, 20, 20, 20, 22, 23, 25, 40

Enter the observations (separated by spaces) into the Permutation Test web app. Let males be Group 1 and females be Group 2. We are interested in testing whether the population mean number of hours spent on social networks differs between males and females, using as test statistic the difference in sample means.

a. What are the observed group means and their difference? (The subtitle of the dot plot shows this information.)

b. Press the Generate Random Permutation(s) button once to generate one permutation of the original data. What are the two group means and their difference under this permutation? (The subtitle of the dot plot shows this information.)

c. Did this one permutation lead to a difference that is less extreme, as extreme, or more extreme than the observed difference?

d. Select to generate 10,000 random permutations. How many of them resulted in a test statistic as or more extreme than the observed difference? (Remember, the alternative hypothesis is H_a: $\mu_1 \neq \mu_2$.)

e. Find and interpret the permutation P-value.

10.46 Compare permutation test to *t* test Refer to the previous exercise.

a. Run a *t* test and report the P-value for the two-sided alternative hypothesis.

b. Using a significance level of 0.10, are the decisions based on the *t* test and based on the permutation test comparable? Which one would you trust more?

c. Remove the largest observations (which could be outliers) in each of the two groups, as was done in Exercise 10.32. Does the conclusion of the permutation test change?

10.47 Dominance of politicians For a rating experiment, researchers translated short video clips of randomly selected speeches by 30 male and 30 female politicians into animated, gender-neutral stick-figures. These animated sequences were rated (on a scale from −100 to 100) by a panel of students on the perceived dominance of the speaker. A higher rating indicates higher dominance. The following data show the average rating each speech received. (The data are also in the stickfigures.csv file.)

TECH

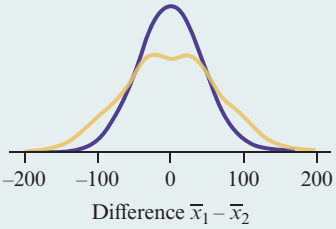
Male politicians: −11, 37, 40, −14, 41, −43, 19, 35, 42, 18, 2, −1, −6, 49, −6, 20, 27, −24, 19, −33, 48, 21, 26, −21, 28, −11, −7, 15, 39, −29

Female politicians: 23, 32, 12, −45, 19, 21, −25, 2, −27, −32, −26, −14, −6, 22, −22, −18, 66, 22, 3, −9, −12, −43, −1, −36, 23, 44, 9, 32, −40, −7

a. State the null and alternative hypothesis and run a permutation test (e.g., using the web app) to see whether there is a difference in the ratings on competence between female and male speakers. What is the conclusion when the significance level is 0.05?

b. Run a *t* test. Are results comparable to the ones from the permutation test? Why or why not?

10.48 Sampling distribution of $\bar{x}_1 - \bar{x}_2$ Refer to Example 12, which compared two groups of seven dogs each in terms of their time interacting with their owners. The following graph shows a smoothed version of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ derived from the permutation approach. Superimposed (in blue) is the theoretical sampling distribution of $\bar{x}_1 - \bar{x}_2$ if the sample sizes were large.



a. What would the mean, standard error, and approximate shape of the theoretical sampling distribution of $\bar{x}_1 - \bar{x}_2$ be if the sample sizes were large?

b. Comment on how the permutation sampling distribution of $\bar{x}_1 - \bar{x}_2$ differs from the theoretical, large-sample one.

c. Consider the one-sided alternative H_a: $\mu_1 > \mu_2$. If the actually observed difference in sample means were 100, would the P-values computed from each sampling distribution be the same or different? Explain.

10.4 Practicing the Basics

10.49 Does exercise help blood pressure? Several recent studies have suggested that people who suffer from abnormally high blood pressure can benefit from regular exercise. A medical researcher decides to test her belief that walking briskly for at least half an hour a day has the effect of lowering blood pressure. She conducts a small pilot study. If results from it are supportive, she will apply for funding for a larger study. She randomly samples three of her patients who have high blood pressure. She measures their systolic blood pressure initially and then

again a month later after they participate in her exercise program. The table shows the results.

Subject	Before	After
1	150	130
2	165	140
3	135	120

a. Explain why the three before observations and the three after observations are dependent samples.

b. Find the sample mean of the before scores, the sample mean of the after scores, and the sample mean of the difference scores d = before − after. How are they related?

c. Find a 95% confidence interval for the difference between the population means of subjects before and after going through such a study. Interpret.

10.50 Test for blood pressure Refer to the previous exercise. The output shows some results of using software to analyze the data with a significance test.

```
Paired T for Before—After
              N     Mean     StDev    SE Mean
Before        3    150.0     15.0      8.660
After         3    130.0     10.0      5.774
Difference    3     20.0      5.0      2.887
T-Test of mean difference = 0 (vs ≠ 0):
T-Value = 6.93  P-Value = 0.020
```

a. State the hypotheses to which the reported P-value refers.

b. Explain how to interpret the P-value. Does the exercise program seem beneficial to lowering blood pressure?

c. What are the assumptions on which this analysis is based?

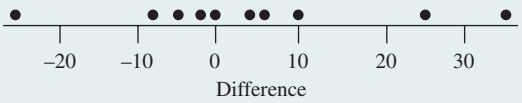
10.51 Social activities for students As part of her class project, a student at the University of Florida randomly sampled 10 fellow students to investigate their most common social activities. As part of the study, she asked the students to state how many times they had done each of the following activities during the previous year: going to a movie, going to a sporting event, or going to a party. The table shows the data.

Frequency of Attending Movies, Sports Events, and Parties

	Activity		
Student	Movies	Sports	Parties
1	10	5	25
2	4	0	10
3	12	20	6
4	2	6	52
5	12	2	12
6	7	8	30
7	45	12	52
8	1	25	2
9	25	0	25
10	12	12	4

a. To compare the mean movie attendance and mean sports attendance using statistical inference, should we treat the samples as independent or dependent? Why?

b. The figure is a dot plot of the $n = 10$ difference scores for movies and sports. Does this show any irregularities that would make statistical inference unreliable?



Dot plot of difference scores.

c. Using the MINITAB output shown for these data, show how the 95% confidence interval was obtained from the other information given in the printout. Interpret the interval.

d. Show how the test statistic shown on the printout was obtained from the other information given. Report the P-value and interpret in context.

MINITAB Output for Inferential Analyses:

```
Paired T for Movies - Sports
              N     Mean     StDev    SE Mean
Movies       10    13.00     13.17      4.17
Sports       10     9.00      8.38      2.65
Difference   10     4.00     16.17      5.11
95% CI for mean difference: (−7.56, 15.56)
T-Test of mean difference = 0 (vs ≠ 0):
T-Value = 0.78  P-Value = 0.454
```

10.52 More social activities Refer to the previous exercise. The output shows the result of comparing the mean responses on parties and sports.

```
Paired T for Parties - Sports
              N     Mean     StDev    SE Mean
Parties      10    21.80     18.58      5.87
Sports       10     9.00      8.38      2.65
Difference   10    12.80     22.55      7.13
95% CI for mean difference: (−3.33, 28.93)
T-Test of mean difference = 0 (vs ≠ 0):
T-Value = 1.80  P-Value = 0.106
```

a. Explain how to interpret the reported 95% confidence interval.

b. State the hypotheses to which the P-value refers and interpret its value.

c. Explain the connection between the results of the test and the confidence interval.

d. What assumptions are necessary for these inferences to be appropriate?

10.53 Movies versus parties Refer to the previous two exercises. Using software, compare the responses on movies and parties using (a) all steps of a significance test and (b) a 95% confidence interval. Interpret results in context.

10.54 Mileage of midsized cars The following table lists the fuel economy of 15 midsized cars when driven in the city and when driven on the highway, along with their overall mileage. Their performance in miles per gallon (mpg) in each test was recorded and the means and standard deviations of their gas mileage were obtained.

	Std. Dev.	Mean
Overall (mpg)	4.58	30.4
City (mpg)	6.03	21.47
Highway (mpg)	2.50	40.87

a. Estimate the change in the mean overall performance as compared to the mean value of performance in the city.

b. Is this information sufficient to find a confidence interval or conduct a test about the change in the mean? If not, what else do you need to know?

c. What assumptions are necessary for the inference in part b?

10.55 Midsized cars' gas mileage change Refer to the previous exercise. Statistics of the change in the overall performance and the city performance are summarized in the following table:

	n	Mean	Std. Dev.
Difference (Overall - City)	15	8.93	1.79

a. Explain how this standard deviation could be so much less than the standard deviations for the miles per gallon (mpg) performance for each test alone.

b. Is 10 mpg a plausible mean change in the population of midsized cars? Answer by constructing a 95% confidence interval for the population mean change in mpg or conducting a significance test of the hypothesis that the mean mpg change in the population equals 10. Interpret.

10.56 Internet book prices Anna's project for her introductory statistics course was to compare the selling prices of textbooks at two Internet bookstores. She first took a random sample of 10 textbooks used that term in courses at her college, based on the list of texts compiled by the college bookstore. The prices of those textbooks at the two Internet sites were

Site A: $115, $79, $43, $140, $99, $30, $80, $99, $119, $69
Site B: $110, $79, $40, $129, $99, $30, $69, $99, $109, $66

a. Are these independent samples or dependent samples? Justify your answer.

b. Find the mean for each sample. Find the mean of the difference scores. Compare and interpret.

c. Using software or a calculator, construct a 90% confidence interval comparing the population mean prices of all textbooks used that term at her college. Interpret.

10.57 Comparing book prices 2 For the data in the previous exercise, use software or a calculator to perform a significance test comparing the population mean prices. Show all steps of the test and indicate whether you would conclude that the mean price is lower at one of the two Internet bookstores.

10.58 Lung capacity revisited Refer to Exercise 10.34 about measuring the lung function (called the forced vital capacity, or FVC, measured in liters) before and after using an inhaler. The data are shown again in the following table.

	Checkup				
	1	**2**	**3**	**4**	**5**
Before using inhaler:	5.08	5.99	5.32	6.03	5.44
After using inhaler:	5.36	5.98	5.62	6.26	5.68

a. By how much, on average, did the FVC improve after using the inhaler?

b. Find a 95% confidence interval for the mean improvement and interpret.

c. Suppose you wrongly analyzed the data as two independent samples. Find the confidence interval for the difference in the population means and show how the conclusions would change substantially.

10.59 Comparing speech recognition systems Table 10.20 in Example 17, repeated here, showed results of an experiment comparing the results of two speech recognition systems, GMDS and CDHMM.

	CDHMM	
GMDS	**Correct**	**Incorrect**
Correct	1921	58
Incorrect	16	5

a. Estimate the population proportion p_1 of correct results for GMDS and p_2 of correct results for CDHMM.

b. Software reports a 95% interval for $p_1 - p_2$ of (0.013, 0.029). Interpret.

10.60 Treat juveniles as adults? The table that follows refers to a sample of juveniles convicted of a felony in Florida. Matched pairs were formed using criteria such as age and the number of prior offenses. For each pair, one subject was handled in the juvenile court and the other was transferred to the adult court. The response of interest was whether the juvenile was rearrested within a year.

a. Are the outcomes for the courts independent samples or dependent samples? Explain.

b. Estimate the population proportions of rearrest for the adult and juvenile courts.

c. Test the hypothesis that the population proportions rearrested were identical for the adult and juvenile court assignments. Use a two-sided alternative, and interpret the P-value.

	Juvenile Court	
Adult Court	**Rearrest**	**No Rearrest**
Rearrest	158	515
No Rearrest	290	1134

Source: Data provided by Larry Winner.

10.61 Change coffee brand? A study was conducted to see if an advertisement campaign would increase market share for Sanka instant decaffeinated coffee (R. Grover and V. Srinivasan, *J. Marketing Research*, vol. 24, 1987, pp. 139–153). Subjects who use instant decaffeinated coffee were asked which brand they bought last. They were asked this before the campaign and after it. The results are shown in the table, with computer output based on it below.

a. Estimate the population proportion choosing Sanka for the (i) first purchase and (ii) second purchase. Find the difference and interpret it.

b. Explain how to interpret the 95% confidence interval.

c. The software output also shows a P-value. State the hypotheses that this P-value refers to in the context of this exercise and give the conclusion of the test when using a significance level of 0.05.

Two Purchases of Coffee

	Second Purchase	
First Purchase	**Sanka**	**Other Brand**
Sanka	155	49
Other brand	76	261

```
McNemar's Test
Estimated
Difference          95% CI              P
 −0.0499      (−0.0920, −0.0078)     0.020
Difference = p(First Purchase) − p(Second
Purchase Sanka)
```

10.62 President's popularity Last month a random sample of 1000 subjects was interviewed and asked whether they thought the president was doing a good job. This month the same subjects were asked this again. The results are: 450 said yes each time, 450 said no each time, 60 said yes on the first survey and no on the second survey, and 40 said no on the first survey and yes on the second survey.

a. Form a contingency table showing these results.

b. Estimate the proportion giving a favorable rating (i) last month and (ii) this month.

c. Find the difference in the proportions and interpret.

d. Find the test statistic and P-value for applying McNemar's test that the population proportion was the same each month. Interpret.

10.63 Marital status and life insurance Adult males participating in a poll were asked whether they were married and whether they had subscribed to a life insurance policy. Of all the respondents, 42% males said that they were married and 32.6% said that they had a life insurance policy.

a. Estimate the difference between the population proportions who were married and who had a life insurance policy.

b. In this survey, 75 respondents said they were married and did not have a life insurance policy whereas 47 respondents stated having a life insurance policy but were unmarried. Report the (i) assumptions, (ii) hypotheses, (iii) test statistic, (iv) P-value, and (v) conclusion for testing that the probability of being married and the probability of having a life insurance policy are the same.

10.64 Marital status and life insurance by age Refer to the previous exercise. Results in this poll also depended strongly on the age of the respondents. For instance, the percentages of respondents (married, having a life insurance policy) were (10%, 5%) between the ages of 18 to 24 years, (45%, 31%) between the ages of 35 to 44 years, (39%, 42%) between the ages of 55 to 64 years, and (28%, 39%) for respondents of 65 years of age or more. If the sample size of respondents between the ages of 35 to 44 years was 120, is this information enough to compare the two proportions inferentially for that range of age? Explain.

10.5 Practicing the Basics

10.65 Benefits of drinking A *USA Today* story (May 22, 2010) about the medical benefits of moderate drinking of alcohol stated that a major French study links those who drink moderately to a lower risk for cardiovascular disease but challenges the idea that moderate drinking is the cause. "Instead, the researchers say, people who drink moderately tend to have a higher social status, exercise more, suffer less depression and enjoy superior health overall compared to heavy drinkers and lifetime abstainers. A causal relationship between cardiovascular risk and moderate drinking is not at all established." The study looked at the health status and drinking habits of 149,773 French adults.

a. Explain how this story refers to an analysis of three types of variables. Identify those variables.

b. Suppose socioeconomic status is treated as a control variable when we compare moderate drinkers to abstainers in their heart attack rates. Explain how this analysis shows that an effect of an explanatory variable on a response variable can change at different values of a control variable.

10.66 Death penalty in Kentucky A study of the death penalty in Kentucky reported the results shown in the table. (*Source*: Data from T. Keil and G. Vito, *Amer. J. Criminal Justice*, vol. 20, 1995, pp. 17–36.)

a. Find and compare the percentage of white defendants with the percentage of black defendants who received the death penalty, when the victim was (i) white and (ii) black.

b. In the analysis in part a, identify the response variable, explanatory variable, and control variable.

c. Construct the summary 2×2 table that ignores, rather than controls, victim's race. Compare the overall percentages of white defendants and black defendants who got the death penalty (ignoring, rather than controlling, victim's race). Compare to part a.

d. Do these data satisfy Simpson's paradox? If not, explain why not. If so, explain what is responsible for Simpson's paradox occurring.

e. Explain, without doing the calculations, how you could inferentially compare the proportions of white and black defendants who get the death penalty (i) ignoring victim's race and (ii) controlling for victim's race.

Victim's Race	Defendant's Race	Death Penalty		Total
		Yes	No	
White	White	31	360	391
	Black	7	50	57
Black	White	0	18	18
	Black	2	106	108

10.67 Stress at work A study performed in Austria in 2015 addressed the problem of workplace stress by the gender of the worker. Researchers also used the type of working area as a control variable to shed more light on the association between the variables. The data shown in the following table was obtained.

				Stress at Work		
				No	Yes	Total
				Count	Count	Count
Type of working area	Intermediate	Gender	Female	3	13	16
			Male	11	29	40
	Rural	Gender	Female	6	5	11
			Male	14	26	40
	Urban	Gender	Female	8	14	22
			Male	19	25	44

a. Treating the type of working area as a control variable, whether the respondents stated having too much stress in their current job as the response variable, and the gender of the respondent as the explanatory variable, explain whether these results illustrate Simpson's paradox or not.

b. Which working area is the most appropriate for female workers and which working area is most unsuitable for both genders?

10.68 Teacher salary, gender, and academic rank The American Association of University Professors (AAUP) reports yearly on faculty salaries for all types of higher education institutions across the United States. The following table lists the mean salary, in thousands of dollars, of full-time instructional faculty on nine-month contracts at four-year public institutions of higher education in 2010, by gender and academic rank. Regard salary as the response variable, gender as the explanatory variable, and academic rank as the control variable.

Mean Salary (Thousands of Dollars) for Men and Women Faculty Members

	Academic Rank				
Gender	Professor	Associate	Assistant	Instructor	Overall
Men	109.2	77.8	66.1	46.0	84.4
Women	96.2	72.7	61.8	46.9	68.8

a. Find the difference between men and women faculty members on their mean salary (i) overall and (ii) after controlling for academic rank.

b. The overall difference between the mean salary of men and women faculty members was larger than the difference at each academic rank. What could be a reason for this? (Simpson's paradox does not hold here, but the gender effect does weaken when we control for academic rank.)

10.69 Family size in Canada The table shows the mean number of children in Canadian families, classified by whether the family was English speaking or French speaking and by whether the family lived in Quebec or in another province.

Mean Number of Children in Canada

Province	English Speaking	French Speaking
Quebec	1.64	1.80
Other	1.97	2.14
Overall	1.95	1.85

a. Overall, compare the mean number of children for English-speaking and French-speaking families.

b. Compare the means, controlling for province (Quebec, Other).

c. How is it possible that for each level of province the mean is higher for French-speaking families, yet overall the mean is higher for English-speaking families? Which paradox does this illustrate?

10.70 Heart disease and age In the United States, the median age of residents is lowest in Utah. At each age level, the death rate from heart disease is higher in Utah than in Colorado. Overall, the death rate from heart disease is lower in Utah than Colorado. Are there any contradictions here, or is this possible? Explain.

10.71 Breast cancer over time The percentage of women who get breast cancer sometime during their lifetime is higher now than in 1900. Suppose that breast cancer incidence tends to increase with age, and suppose that women tend to live longer now than in 1900. Explain why a comparison of breast cancer rates now with the rate in 1900 could show different results if we control for the age of the woman.