

4.7 Exercises

4.7.1 Variability in estimates

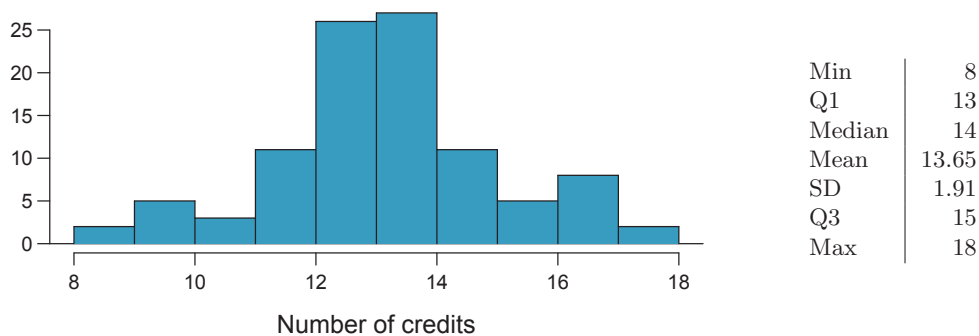
4.1 Identify the parameter, Part I. For each of the following situations, state whether the parameter of interest is a mean or a proportion. It may be helpful to examine whether individual responses are numerical or categorical.

- In a survey, one hundred college students are asked how many hours per week they spend on the Internet.
- In a survey, one hundred college students are asked: “What percentage of the time you spend on the Internet is part of your course work?”
- In a survey, one hundred college students are asked whether or not they cited information from Wikipedia in their papers.
- In a survey, one hundred college students are asked what percentage of their total weekly spending is on alcoholic beverages.
- In a sample of one hundred recent college graduates, it is found that 85 percent expect to get a job within one year of their graduation date.

4.2 Identify the parameter, Part II. For each of the following situations, state whether the parameter of interest is a mean or a proportion.

- A poll shows that 64% of Americans personally worry a great deal about federal spending and the budget deficit.
- A survey reports that local TV news has shown a 17% increase in revenue between 2009 and 2011 while newspaper revenues decreased by 6.4% during this time period.
- In a survey, high school and college students are asked whether or not they use geolocation services on their smart phones.
- In a survey, internet users are asked whether or not they purchased any Groupon coupons.
- In a survey, internet users are asked how many Groupon coupons they purchased over the last year.

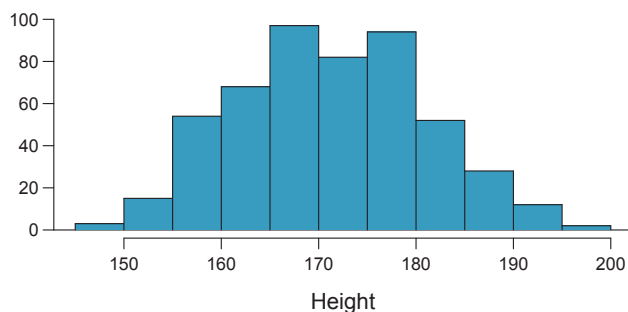
4.3 College credits. A college counselor is interested in estimating how many credits a student typically enrolls in each semester. The counselor decides to randomly sample 100 students by using the registrar’s database of students. The histogram below shows the distribution of the number of credits taken by these students. Sample statistics for this distribution are also provided.



- What is the point estimate for the average number of credits taken per semester by students at this college? What about the median?

- (b) What is the point estimate for the standard deviation of the number of credits taken per semester by students at this college? What about the IQR?
- (c) Is a load of 16 credits unusually high for this college? What about 18 credits? Explain your reasoning. *Hint:* Observations farther than two standard deviations from the mean are usually considered to be unusual.
- (d) The college counselor takes another random sample of 100 students and this time finds a sample mean of 14.02 units. Should she be surprised that this sample statistic is slightly different than the one from the original sample? Explain your reasoning.
- (e) The sample means given above are point estimates for the mean number of credits taken by all students at that college. What measures do we use to quantify the variability of this estimate? Compute this quantity using the data from the original sample.

4.4 Heights of adults. Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters.⁴²



- (a) What is the point estimate for the average height of active individuals? What about the median?
- (b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR?
- (c) Is a person who is 1m 80cm (180 cm) tall considered unusually tall? And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning.
- (d) The researchers take another random sample of physically active individuals. Would you expect the mean and the standard deviation of this new sample to be the ones given above? Explain your reasoning.
- (e) The sample means obtained are point estimates for the mean height of all active individuals, if the sample of individuals is equivalent to a simple random sample. What measure do we use to quantify the variability of such an estimate? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample.

4.5 Wireless routers. John is shopping for wireless routers and is overwhelmed by the number of available options. In order to get a feel for the average price, he takes a random sample of 75 routers and finds that the average price for this sample is \$75 and the standard deviation is \$25.

- (a) Based on this information, how much variability should he expect to see in the mean prices of repeated samples, each containing 75 randomly selected wireless routers?
- (b) A consumer website claims that the average price of routers is \$80. Is a true average of \$80 consistent with John's sample?

⁴²G. Heinz et al. "Exploring relationships in body dimensions". In: *Journal of Statistics Education* 11.2 (2003).

4.6 Chocolate chip cookies. Students are asked to count the number of chocolate chips in 22 cookies for a class activity. They found that the cookies on average had 14.77 chocolate chips with a standard deviation of 4.37 chocolate chips.

- Based on this information, about how much variability should they expect to see in the mean number of chocolate chips in random samples of 22 chocolate chip cookies?
- The packaging for these cookies claims that there are at least 20 chocolate chips per cookie. One student thinks this number is unreasonably high since the average they found is much lower. Another student claims the difference might be due to chance. What do you think?

4.7.2 Confidence intervals

4.7 Relaxing after work. The General Social Survey (GSS) is a sociological survey used to collect data on demographic characteristics and attitudes of residents of the United States. In 2010, the survey collected responses from 1,154 US residents. The survey is conducted face-to-face with an in-person interview of a randomly-selected sample of adults. One of the questions on the survey is “After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?” A 95% confidence interval from the 2010 GSS survey is 3.53 to 3.83 hours.⁴³

- Interpret this interval in the context of the data.
- What does a 95% confidence level mean in this context?
- Suppose the researchers think a 90% confidence level would be more appropriate for this interval. Will this new interval be smaller or larger than the 95% confidence interval? Assume the standard deviation has remained constant since 2010.

4.8 Mental health. Another question on the General Social Survey introduced in Exercise 4.7 is “For how many days during the past 30 days was your mental health, which includes stress, depression, and problems with emotions, not good?” Based on responses from 1,151 US residents, the survey reported a 95% confidence interval of 3.40 to 4.24 days in 2010.

- Interpret this interval in context of the data.
- What does a 95% confidence level mean in this context?
- Suppose the researchers think a 99% confidence level would be more appropriate for this interval. Will this new interval be smaller or larger than the 95% confidence interval?
- If a new survey asking the same questions was to be done with 500 Americans, would the standard error of the estimate be larger, smaller, or about the same. Assume the standard deviation has remained constant since 2010.

4.9 Width of a confidence interval. Earlier in Chapter 4, we calculated the 99% confidence interval for the average age of runners in the 2012 Cherry Blossom Run as (32.7, 37.4) based on a sample of 100 runners. How could we decrease the width of this interval without losing confidence?

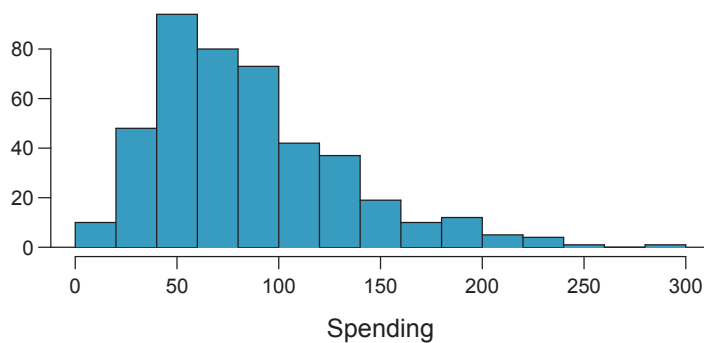
4.10 Confidence levels. If a higher confidence level means that we are more confident about the number we are reporting, why don't we always report a confidence interval with the highest possible confidence level?

⁴³National Opinion Research Center, General Social Survey, 2010.

4.11 Waiting at an ER, Part I. A hospital administrator hoping to improve wait times decides to estimate the average emergency room waiting time at her hospital. She collects a simple random sample of 64 patients and determines the time (in minutes) between when they checked in to the ER until they were first seen by a doctor. A 95% confidence interval based on this sample is (128 minutes, 147 minutes), which is based on the normal model for the mean. Determine whether the following statements are true or false, and explain your reasoning for those statements you identify as false.

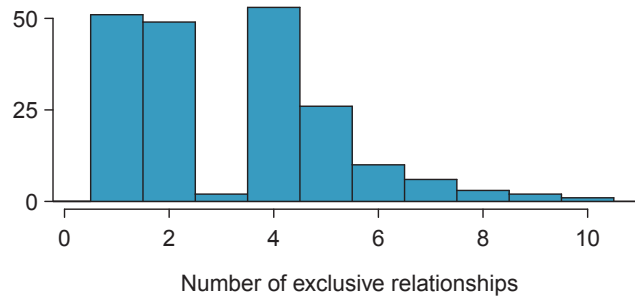
- This confidence interval is not valid since we do not know if the population distribution of the ER wait times is nearly normal.
- We are 95% confident that the average waiting time of these 64 emergency room patients is between 128 and 147 minutes.
- We are 95% confident that the average waiting time of all patients at this hospital's emergency room is between 128 and 147 minutes.
- 95% of such random samples would have a sample mean between 128 and 147 minutes.
- A 99% confidence interval would be narrower than the 95% confidence interval since we need to be more sure of our estimate.
- The margin of error is 9.5 and the sample mean is 137.5.
- In order to decrease the margin of error of a 95% confidence interval to half of what it is now, we would need to double the sample size.

4.12 Thanksgiving spending, Part I. The 2009 holiday retail season, which kicked off on November 27, 2009 (the day after Thanksgiving), had been marked by somewhat lower self-reported consumer spending than was seen during the comparable period in 2008. To get an estimate of consumer spending, 436 randomly sampled American adults were surveyed. Daily consumer spending for the six-day period after Thanksgiving, spanning the Black Friday weekend and Cyber Monday, averaged \$84.71. A 95% confidence interval based on this sample is (\$80.31, \$89.11). Determine whether the following statements are true or false, and explain your reasoning.



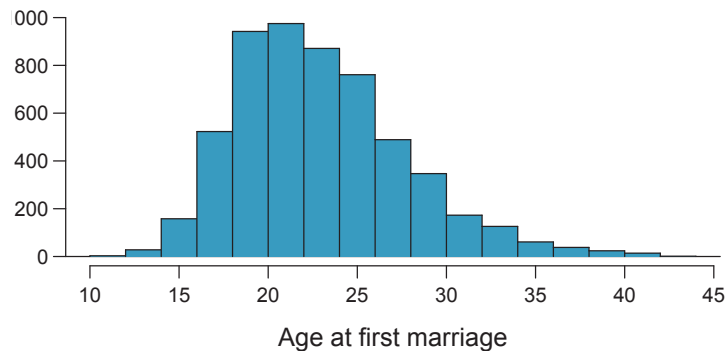
- We are 95% confident that the average spending of these 436 American adults is between \$80.31 and \$89.11.
- This confidence interval is not valid since the distribution of spending in the sample is right skewed.
- 95% of such random samples would have a sample mean between \$80.31 and \$89.11.
- We are 95% confident that the average spending of all American adults is between \$80.31 and \$89.11.
- A 90% confidence interval would be narrower than the 95% confidence interval.
- In order to decrease the margin of error of a 95% confidence interval to a third of what it is now, we would need to use a sample 3 times larger.
- The margin of error for the reported interval is 4.4.

4.13 Exclusive relationships. A survey was conducted on 203 undergraduates from Duke University who took an introductory statistics course in Spring 2012. Among many other questions, this survey asked them about the number of exclusive relationships they have been in. The histogram below shows the distribution of the data from this sample. The sample average is 3.2 with a standard deviation of 1.97.



Estimate the average number of exclusive relationships Duke students have been in using a 90% confidence interval and interpret this interval in context. Check any conditions required for inference, and note any assumptions you must make as you proceed with your calculations and conclusions.

4.14 Age at first marriage, Part I. The National Survey of Family Growth conducted by the Centers for Disease Control gathers information on family life, marriage and divorce, pregnancy, infertility, use of contraception, and men's and women's health. One of the variables collected on this survey is the age at first marriage. The histogram below shows the distribution of ages at first marriage of 5,534 randomly sampled women between 2006 and 2010. The average age at first marriage among these women is 23.44 with a standard deviation of 4.72.⁴⁴



Estimate the average age at first marriage of women using a 95% confidence interval, and interpret this interval in context. Discuss any relevant assumptions.

⁴⁴National Survey of Family Growth, 2006-2010 Cycle.

4.7.3 Hypothesis testing

4.15 Identify hypotheses, Part I. Write the null and alternative hypotheses in words and then symbols for each of the following situations.

- New York is known as “the city that never sleeps”. A random sample of 25 New Yorkers were asked how much sleep they get per night. Do these data provide convincing evidence that New Yorkers on average sleep less than 8 hours a night?
- Employers at a firm are worried about the effect of March Madness, a basketball championship held each spring in the US, on employee productivity. They estimate that on a regular business day employees spend on average 15 minutes of company time checking personal email, making personal phone calls, etc. They also collect data on how much company time employees spend on such non-business activities during March Madness. They want to determine if these data provide convincing evidence that employee productivity decreases during March Madness.

4.16 Identify hypotheses, Part II. Write the null and alternative hypotheses in words and using symbols for each of the following situations.

- Since 2008, chain restaurants in California have been required to display calorie counts of each menu item. Prior to menus displaying calorie counts, the average calorie intake of diners at a restaurant was 1100 calories. After calorie counts started to be displayed on menus, a nutritionist collected data on the number of calories consumed at this restaurant from a random sample of diners. Do these data provide convincing evidence of a difference in the average calorie intake of a diners at this restaurant?
- Based on the performance of those who took the GRE exam between July 1, 2004 and June 30, 2007, the average Verbal Reasoning score was calculated to be 462. In 2011 the average verbal score was slightly higher. Do these data provide convincing evidence that the average GRE Verbal Reasoning score has changed since 2004?⁴⁵

4.17 Online communication. A study suggests that the average college student spends 2 hours per week communicating with others online. You believe that this is an underestimate and decide to collect your own sample for a hypothesis test. You randomly sample 60 students from your dorm and find that on average they spent 3.5 hours a week communicating with others online. A friend of yours, who offers to help you with the hypothesis test, comes up with the following set of hypotheses. Indicate any errors you see.

$$H_0 : \bar{x} < 2 \text{ hours}$$

$$H_A : \bar{x} > 3.5 \text{ hours}$$

4.18 Age at first marriage, Part II. Exercise 4.14 presents the results of a 2006 - 2010 survey showing that the average age of women at first marriage is 23.44. Suppose a researcher believes that this value has increased in 2012, but he would also be interested if he found a decrease. Below is how he set up his hypotheses. Indicate any errors you see.

$$H_0 : \bar{x} = 23.44 \text{ years old}$$

$$H_A : \bar{x} > 23.44 \text{ years old}$$

4.19 Waiting at an ER, Part II. Exercise 4.11 provides a 95% confidence interval for the mean waiting time at an emergency room (ER) of (128 minutes, 147 minutes).

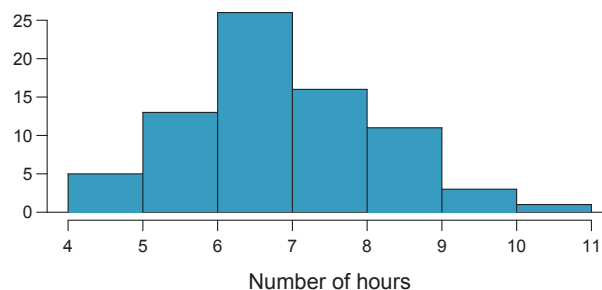
- A local newspaper claims that the average waiting time at this ER exceeds 3 hours. What do you think of this claim?
- The Dean of Medicine at this hospital claims the average wait time is 2.2 hours. What do you think of this claim?
- Without actually calculating the interval, determine if the claim of the Dean from part (b) would be considered reasonable based on a 99% confidence interval?

⁴⁵ETS, Interpreting your GRE Scores.

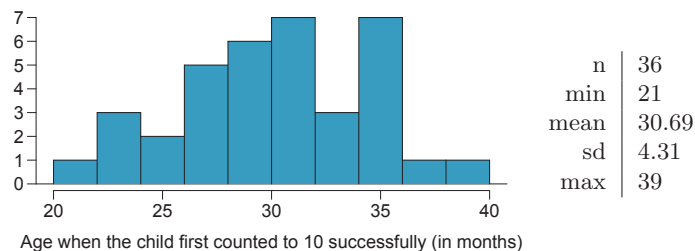
4.20 Thanksgiving spending, Part II. Exercise 4.12 provides a 95% confidence interval for the average spending by American adults during the six-day period after Thanksgiving 2009: (\$80.31, \$89.11).

- A local news anchor claims that the average spending during this period in 2009 was \$100. What do you think of this claim?
- Would the news anchor's claim be considered reasonable based on a 90% confidence interval? Why or why not?

4.21 Ball bearings. A manufacturer claims that bearings produced by their machine last 7 hours on average under harsh conditions. A factory worker randomly samples 75 ball bearings, and records their lifespans under harsh conditions. He calculates a sample mean of 6.85 hours, and the standard deviation of the data is 1.25 working hours. The following histogram shows the distribution of the lifespans of the ball bearings in this sample. Conduct a formal hypothesis test of this claim. Make sure to check that relevant conditions are satisfied.



4.22 Gifted children, Part I. Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.⁴⁶



- Are conditions for inference satisfied?
- Suppose you read on a parenting website that children first count to 10 successfully when they are 32 months old, on average. Perform a hypothesis test to evaluate if these data provide convincing evidence that the average age at which gifted children first count to 10 successfully is different than the general average of 32 months. Use a significance level of 0.10.
- Interpret the p-value in context of the hypothesis test and the data.
- Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully.
- Do your results from the hypothesis test and the confidence interval agree? Explain.

⁴⁶F.A. Graybill and H.K. Iyer. *Regression Analysis: Concepts and Applications*. Duxbury Press, 1994, pp. 511–516.

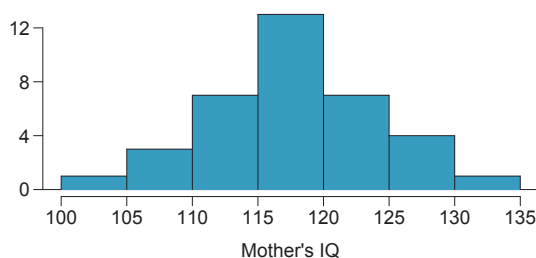
4.23 Waiting at an ER, Part III. The hospital administrator mentioned in Exercise 4.11 randomly selected 64 patients and measured the time (in minutes) between when they checked in to the ER and the time they were first seen by a doctor. The average time is 137.5 minutes and the standard deviation is 39 minutes. He is getting grief from his supervisor on the basis that the wait times in the ER increased greatly from last year's average of 127 minutes. However, the administrator claims that the increase is probably just due to chance.

- Are conditions for inference met? Note any assumptions you must make to proceed.
- Using a significance level of $\alpha = 0.05$, is the change in wait times statistically significant? Use a two-sided test since it seems the supervisor had to inspect the data before he suggested an increase occurred.
- Would the conclusion of the hypothesis test change if the significance level was changed to $\alpha = 0.01$?

4.24 Gifted children, Part II. Exercise 4.22 describes a study on gifted children. In this study, along with variables on the children, the researchers also collected data on the mother's and father's IQ of the 36 randomly sampled gifted children. The histogram below shows the distribution of mother's IQ. Also provided are some sample statistics.

- Perform a hypothesis test to evaluate if these data provide convincing evidence that the average IQ of mothers of gifted children is different than the average IQ for the population at large, which is 100. Use a significance level of 0.10.
- Calculate a 90% confidence interval for the average IQ of mothers of gifted children.
- Do your results from the hypothesis test and the confidence interval agree? Explain.

n	36
min	101
mean	118.2
sd	6.5
max	131



4.25 Nutrition labels. The nutrition label on a bag of potato chips says that a one ounce (28 gram) serving of potato chips has 130 calories and contains ten grams of fat, with three grams of saturated fat. A random sample of 35 bags yielded a sample mean of 134 calories with a standard deviation of 17 calories. Is there evidence that the nutrition label does not provide an accurate measure of calories in the bags of potato chips? We have verified the independence, sample size, and skew conditions are satisfied.

4.26 Find the sample mean. You are given the following hypotheses: $H_0: \mu = 34$, $H_A: \mu > 34$. We know that the sample standard deviation is 10 and the sample size is 65. For what sample mean would the p-value be equal to 0.05? Assume that all conditions necessary for inference are satisfied.

4.27 Testing for Fibromyalgia. A patient named Diana was diagnosed with Fibromyalgia, a long-term syndrome of body pain, and was prescribed anti-depressants. Being the skeptic that she is, Diana didn't initially believe that anti-depressants would help her symptoms. However after a couple months of being on the medication she decides that the anti-depressants are working, because she feels like her symptoms are in fact getting better.

- Write the hypotheses in words for Diana's skeptical position when she started taking the anti-depressants.
- What is a Type 1 error in this context?
- What is a Type 2 error in this context?
- How would these errors affect the patient?

4.28 Testing for food safety. A food safety inspector is called upon to investigate a restaurant with a few customer reports of poor sanitation practices. The food safety inspector uses a hypothesis testing framework to evaluate whether regulations are not being met. If he decides the restaurant is in gross violation, its license to serve food will be revoked.

- Write the hypotheses in words.
- What is a Type 1 error in this context?
- What is a Type 2 error in this context?
- Which error is more problematic for the restaurant owner? Why?
- Which error is more problematic for the diners? Why?
- As a diner, would you prefer that the food safety inspector requires strong evidence or very strong evidence of health concerns before revoking a restaurant's license? Explain your reasoning.

4.29 Errors in drug testing. Suppose regulators monitored 403 drugs last year, each for a particular adverse response. For each drug they conducted a single hypothesis test with a significance level of 5% to determine if the adverse effect was higher in those taking the drug than those who did not take the drug; the regulators ultimately rejected the null hypothesis for 42 drugs.

- Describe the error the regulators might have made for a drug where the null hypothesis was rejected.
- Describe the error regulators might have made for a drug where the null hypothesis was not rejected.
- Suppose the vast majority of the 403 drugs do not have adverse effects. Then, if you picked one of the 42 suspect drugs at random, about how sure would you be that the drug really has an adverse effect?
- Can you also say how sure you are that a particular drug from the 361 where the null hypothesis was not rejected does not have the corresponding adverse response?

4.30 Car insurance savings, Part I. A car insurance company advertises that customers switching to their insurance save, on average, \$432 on their yearly premiums. A market researcher at a competing insurance discounter is interested in showing that this value is an overestimate so he can provide evidence to government regulators that the company is falsely advertising their prices. He randomly samples 82 customers who recently switched to this insurance and finds an average savings of \$395, with a standard deviation of \$102.

- Are conditions for inference satisfied?
- Perform a hypothesis test and state your conclusion.
- Do you agree with the market researcher that the amount of savings advertised is an overestimate? Explain your reasoning.
- Calculate a 90% confidence interval for the average amount of savings of all customers who switch their insurance.
- Do your results from the hypothesis test and the confidence interval agree? Explain.

4.31 Happy hour. A restaurant owner is considering extending the happy hour at his restaurant since he would like to see if it increases revenue. If it does, he will permanently extend happy hour. He estimates that the current average revenue per customer is \$18 during happy hour. He runs the extended happy hour for a week and finds an average revenue of \$19.25 with a standard deviation \$3.02 based on a simple random sample of 70 customers.

- Are conditions for inference satisfied?
- Perform a hypothesis test. Suppose the customers and their buying habits this week were no different than in any other week for this particular bar. (This may not always be a reasonable assumption.)
- Calculate a 90% confidence interval for the average revenue per customer.
- Do your results from the hypothesis test and the confidence interval agree? Explain.
- If your hypothesis test and confidence interval suggest a significant increase in revenue per customer, why might you still not recommend that the restaurant owner extend the happy hour based on this criterion? What may be a better measure to consider?

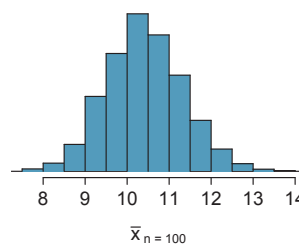
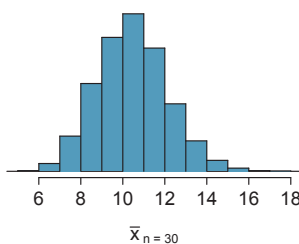
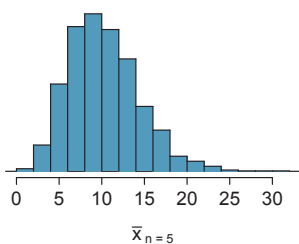
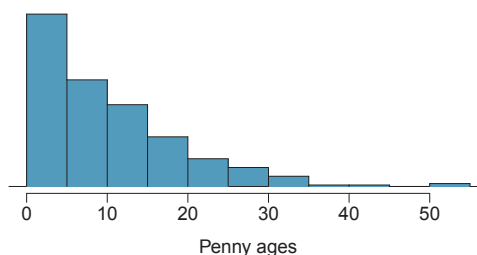
4.32 Speed reading, Part I. A company offering online speed reading courses claims that students who take their courses show a 5 times (500%) increase in the number of words they can read in a minute without losing comprehension. A random sample of 100 students yielded an average increase of 415% with a standard deviation of 220%. Is there evidence that the company's claim is false?

- Are conditions for inference satisfied?
- Perform a hypothesis test evaluating if the company's claim is reasonable or if the true average improvement is less than 500%. Make sure to interpret your response in context of the hypothesis test and the data. Use $\alpha = 0.025$.
- Calculate a 95% confidence interval for the average increase in the number of words students can read in a minute without losing comprehension.
- Do your results from the hypothesis test and the confidence interval agree? Explain.

4.7.4 Examining the Central Limit Theorem

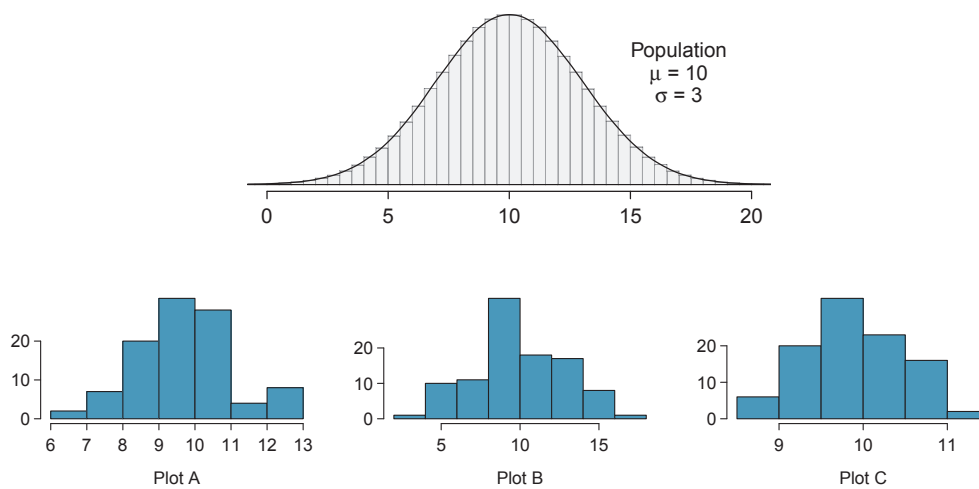
4.33 Ages of pennies, Part I. The histogram below shows the distribution of ages of pennies at a bank.

- Describe the distribution.
- Sampling distributions for means from simple random samples of 5, 30, and 100 pennies is shown in the histograms below. Describe the shapes of these distributions and comment on whether they look like what you would expect to see based on the Central Limit Theorem.

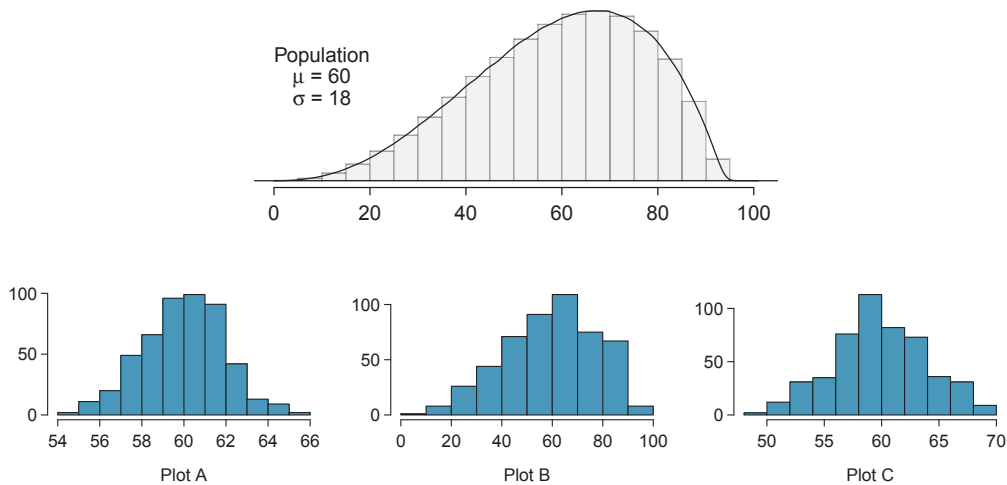


4.34 Ages of pennies, Part II. The mean age of the pennies from Exercise 4.33 is 10.44 years with a standard deviation of 9.2 years. Using the Central Limit Theorem, calculate the means and standard deviations of the distribution of the mean from random samples of size 5, 30, and 100. Comment on whether the sampling distributions shown in Exercise 4.33 agree with the values you compute.

4.35 Identify distributions, Part I. Four plots are presented below. The plot at the top is a distribution for a population. The mean is 10 and the standard deviation is 3. Also shown below is a distribution of (1) a single random sample of 100 values from this population, (2) a distribution of 100 sample means from random samples with size 5, and (3) a distribution of 100 sample means from random samples with size 25. Determine which plot (A, B, or C) is which and explain your reasoning.



4.36 Identify distributions, Part II. Four plots are presented below. The plot at the top is a distribution for a population. The mean is 60 and the standard deviation is 18. Also shown below is a distribution of (1) a single random sample of 500 values from this population, (2) a distribution of 500 sample means from random samples of each size 18, and (3) a distribution of 500 sample means from random samples of each size 81. Determine which plot (A, B, or C) is which and explain your reasoning.



4.37 Housing prices, Part I. A housing survey was conducted to determine the price of a typical home in Topanga, CA. The mean price of a house was roughly \$1.3 million with a standard deviation of \$300,000. There were no houses listed below \$600,000 but a few houses above \$3 million.

- Is the distribution of housing prices in Topanga symmetric, right skewed, or left skewed? *Hint:* Sketch the distribution.
- Would you expect most houses in Topanga to cost more or less than \$1.3 million?
- Can we estimate the probability that a randomly chosen house in Topanga costs more than \$1.4 million using the normal distribution?
- What is the probability that the mean of 60 randomly chosen houses in Topanga is more than \$1.4 million?
- How would doubling the sample size affect the standard error of the mean?

4.38 Stats final scores. Each year about 1500 students take the introductory statistics course at a large university. This year scores on the final exam are distributed with a median of 74 points, a mean of 70 points, and a standard deviation of 10 points. There are no students who scored above 100 (the maximum score attainable on the final) but a few students scored below 20 points.

- Is the distribution of scores on this final exam symmetric, right skewed, or left skewed?
- Would you expect most students to have scored above or below 70 points?
- Can we calculate the probability that a randomly chosen student scored above 75 using the normal distribution?
- What is the probability that the average score for a random sample of 40 students is above 75?
- How would cutting the sample size in half affect the standard error of the mean?

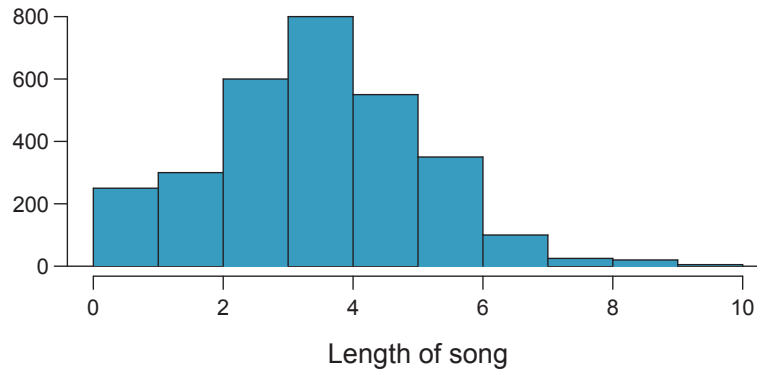
4.39 Weights of pennies. The distribution of weights of US pennies is approximately normal with a mean of 2.5 grams and a standard deviation of 0.03 grams.

- What is the probability that a randomly chosen penny weighs less than 2.4 grams?
- Describe the sampling distribution of the mean weight of 10 randomly chosen pennies.
- What is the probability that the mean weight of 10 pennies is less than 2.4 grams?
- Sketch the two distributions (population and sampling) on the same scale.
- Could you estimate the probabilities from (a) and (c) if the weights of pennies had a skewed distribution?

4.40 CFLs. A manufacturer of compact fluorescent light bulbs advertises that the distribution of the lifespans of these light bulbs is nearly normal with a mean of 9,000 hours and a standard deviation of 1,000 hours.

- What is the probability that a randomly chosen light bulb lasts more than 10,500 hours?
- Describe the distribution of the mean lifespan of 15 light bulbs.
- What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than 10,500 hours?
- Sketch the two distributions (population and sampling) on the same scale.
- Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had a skewed distribution?

4.41 Songs on an iPod. Suppose an iPod has 3,000 songs. The histogram below shows the distribution of the lengths of these songs. We also know that, for this iPod, the mean length is 3.45 minutes and the standard deviation is 1.63 minutes.



- Calculate the probability that a randomly selected song lasts more than 5 minutes.
- You are about to go for an hour run and you make a random playlist of 15 songs. What is the probability that your playlist lasts for the entire duration of your run? *Hint:* If you want the playlist to last 60 minutes, what should be the minimum average length of a song?
- You are about to take a trip to visit your parents and the drive is 6 hours. You make a random playlist of 100 songs. What is the probability that your playlist lasts the entire drive?

4.42 Spray paint. Suppose the area that can be painted using a single can of spray paint is slightly variable and follows a nearly normal distribution with a mean of 25 square feet and a standard deviation of 3 square feet.

- What is the probability that the area covered by a can of spray paint is more than 27 square feet?
- Suppose you want to spray paint an area of 540 square feet using 20 cans of spray paint. On average, how many square feet must each can be able to cover to spray paint all 540 square feet?
- What is the probability that you can cover a 540 square feet area using 20 cans of spray paint?
- If the area covered by a can of spray paint had a slightly skewed distribution, could you still calculate the probabilities in parts (a) and (c) using the normal distribution?

4.7.5 Inference for other estimators

4.43 Spam mail, Part I. The 2004 National Technology Readiness Survey sponsored by the Smith School of Business at the University of Maryland surveyed 418 randomly sampled Americans, asking them how many spam emails they receive per day. The survey was repeated on a new random sample of 499 Americans in 2009.⁴⁷

- What are the hypotheses for evaluating if the average spam emails per day has changed from 2004 to 2009.
- In 2004 the mean was 18.5 spam emails per day, and in 2009 this value was 14.9 emails per day. What is the point estimate for the difference between the two population means?
- A report on the survey states that the observed difference between the sample means is not statistically significant. Explain what this means in context of the hypothesis test and the data.
- Would you expect a confidence interval for the difference between the two population means to contain 0? Explain your reasoning.

4.44 Nearsightedness. It is believed that nearsightedness affects about 8% of all children. In a random sample of 194 children, 21 are nearsighted.

- Construct hypotheses appropriate for the following question: do these data provide evidence that the 8% value is inaccurate?
- What proportion of children in this sample are nearsighted?
- Given that the standard error of the sample proportion is 0.0195 and the point estimate follows a nearly normal distribution, calculate the test statistic (the Z statistic).
- What is the p-value for this hypothesis test?
- What is the conclusion of the hypothesis test?

4.45 Spam mail, Part II. The National Technology Readiness Survey from Exercise 4.43 also asked Americans how often they delete spam emails. 23% of the respondents in 2004 said they delete their spam mail once a month or less, and in 2009 this value was 16%.

- What are the hypotheses for evaluating if the proportion of those who delete their email once a month or less (or never) has changed from 2004 to 2009?
- What is the point estimate for the difference between the two population proportions?
- A report on the survey states that the observed decrease from 2004 to 2009 is statistically significant. Explain what this means in context of the hypothesis test and the data.
- Would you expect a confidence interval for the difference between the two population proportions to contain 0? Explain your reasoning.

4.46 Unemployment and relationship problems. A USA Today/Gallup poll conducted between 2010 and 2011 asked a group of unemployed and underemployed Americans if they have had major problems in their relationships with their spouse or another close family member as a result of not having a job (if unemployed) or not having a full-time job (if underemployed). 27% of the 1,145 unemployed respondents and 25% of the 675 underemployed respondents said they had major problems in relationships as a result of their employment status.

- What are the hypotheses for evaluating if the proportions of unemployed and underemployed people who had relationship problems were different?
- The p-value for this hypothesis test is approximately 0.35. Explain what this means in context of the hypothesis test and the data.

⁴⁷Rockbridge, 2009 National Technology Readiness Survey SPAM Report.

4.7.6 Sample size and power

4.47 Which is higher? In each part below, there is a value of interest and two scenarios (I and II). For each part, report if the value of interest is larger under scenario I, scenario II, or whether the value is equal under the scenarios.

- The standard error of \bar{x} when $s = 120$ and (I) $n = 25$ or (II) $n = 125$.
- The margin of error of a confidence interval when the confidence level is (I) 90% or (II) 80%.
- The p-value for a Z statistic of 2.5 when (I) $n = 500$ or (II) $n = 1000$.
- The probability of making a Type 2 error when the alternative hypothesis is true and the significance level is (I) 0.05 or (II) 0.10.

4.48 True or false. Determine if the following statements are true or false, and explain your reasoning. If false, state how it could be corrected.

- If a given value (for example, the null hypothesized value of a parameter) is within a 95% confidence interval, it will also be within a 99% confidence interval.
- Decreasing the significance level (α) will increase the probability of making a Type 1 error.
- Suppose the null hypothesis is $\mu = 5$ and we fail to reject H_0 . Under this scenario, the true population mean is 5.
- If the alternative hypothesis is true, then the probability of making a Type 2 error and the power of a test add up to 1.
- With large sample sizes, even small differences between the null value and the true value of the parameter, a difference often called the effect size, will be identified as statistically significant.
- A cutoff of $\alpha = 0.05$ is the ideal value for all hypothesis tests.

4.49 Car insurance savings, Part II. The market researcher from Exercise 4.30 collected data about the savings of 82 customers at a competing car insurance company. The mean and standard deviation of this sample are \$395 and \$102, respectively. He would like to conduct another survey but have a margin of error of no more than \$10 at a 99% confidence level. How large of a sample should he collect?

4.50 Speed reading, Part II. A random sample of 100 students who took online speed reading courses from the company described in Exercise 4.32 yielded an average increase in reading speed of 415% and a standard deviation of 220%. We would like to calculate a 95% confidence interval for the average increase in reading speed with a margin of error of no more than 15%. How many students should we sample?

4.51 Waiting at the ER, Part IV. Exercise 4.23 introduced us to a hospital where ER wait times were being analyzed. The previous year's average was 128 minutes. Suppose that this year's average wait time is 135 minutes.

- Provide the hypotheses for this situation in plain language.
- If we plan to collect a sample size of $n = 64$, what values could \bar{x} take so that we reject H_0 ? Suppose the sample standard deviation from the earlier exercise (39 minutes) is the population standard deviation. You may assume that the conditions for the nearly normal model for \bar{x} are satisfied.
- Calculate the probability of a Type 2 error.

Chapter 5

Inference for numerical data

Chapter 4 introduced a framework for statistical inference based on confidence intervals and hypotheses. In this chapter, we encounter several new point estimates and scenarios. In each case, the inference ideas remain the same:

1. Determine which point estimate or test statistic is useful.
2. Identify an appropriate distribution for the point estimate or test statistic.
3. Apply the ideas from Chapter 4 using the distribution from step 2.

Each section in Chapter 5 explores a new situation: the difference of two means (5.1, 5.2); a single mean or difference of means where we relax the minimum sample size condition (5.3, 5.4); and the comparison of means across multiple groups (5.5). Chapter 6 will introduce scenarios that highlight categorical data.

5.1 Paired data

Are textbooks actually cheaper online? Here we compare the price of textbooks at UCLA's bookstore and prices at Amazon.com. Seventy-three UCLA courses were randomly sampled in Spring 2010, representing less than 10% of all UCLA courses.¹ A portion of this data set is shown in Table 5.1.

	dept	course	ucla	amazon	diff
1	Am Ind	C170	27.67	27.95	-0.28
2	Anthro	9	40.59	31.14	9.45
3	Anthro	135T	31.68	32.00	-0.32
4	Anthro	191HB	16.00	11.52	4.48
⋮	⋮	⋮	⋮	⋮	⋮
72	Wom Std	M144	23.76	18.72	5.04
73	Wom Std	285	27.70	18.22	9.48

Table 5.1: Six cases of the `textbooks` data set.

¹When a class had multiple books, only the most expensive text was considered.

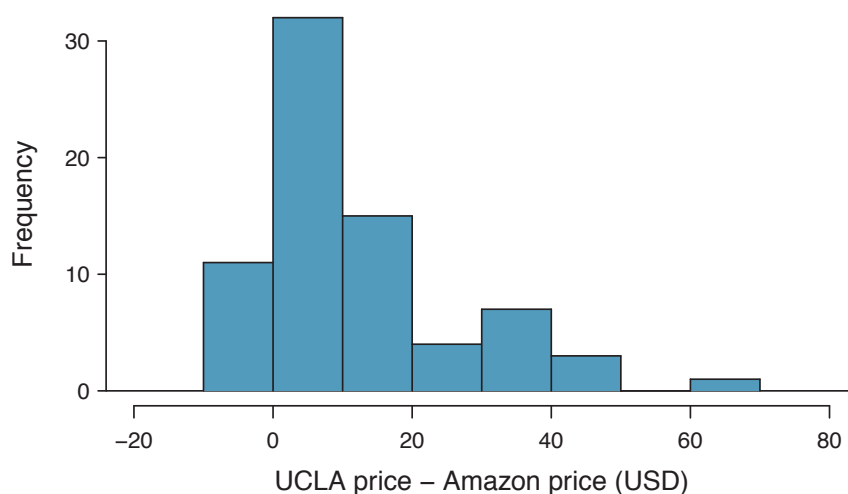


Figure 5.2: Histogram of the difference in price for each book sampled. These data are strongly skewed.

5.1.1 Paired observations and samples

Each textbook has two corresponding prices in the data set: one for the UCLA bookstore and one for Amazon. Therefore, each textbook price from the UCLA bookstore has a natural correspondence with a textbook price from Amazon. When two sets of observations have this special correspondence, they are said to be **paired**.

Paired data

Two sets of observations are *paired* if each observation in one set has a special correspondence or connection with exactly one observation in the other data set.

To analyze paired data, it is often useful to look at the difference in outcomes of each pair of observations. In the `textbook` data set, we look at the difference in prices, which is represented as the `diff` variable in the `textbooks` data. Here the differences are taken as

$$\text{UCLA price} - \text{Amazon price}$$

for each book. It is important that we always subtract using a consistent order; here Amazon prices are always subtracted from UCLA prices. A histogram of these differences is shown in Figure 5.2. Using differences between paired observations is a common and useful way to analyze paired data.

- ⊙ **Exercise 5.1** The first difference shown in Table 5.1 is computed as $27.67 - 27.95 = -0.28$. Verify the differences are calculated correctly for observations 2 and 3.²

5.1.2 Inference for paired data

To analyze a paired data set, we use the exact same tools that we developed in Chapter 4. Now we apply them to the differences in the paired observations.

²Observation 2: $40.59 - 31.14 = 9.45$. Observation 3: $31.68 - 32.00 = -0.32$.

n_{diff}	\bar{x}_{diff}	s_{diff}
73	12.76	14.26

Table 5.3: Summary statistics for the price differences. There were 73 books, so there are 73 differences.

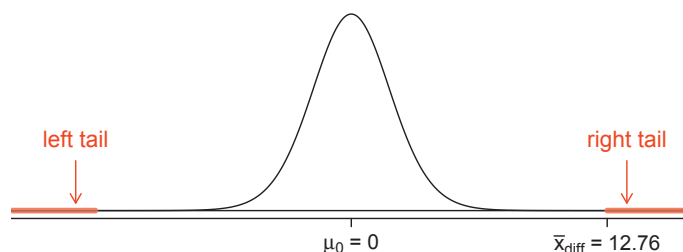


Figure 5.4: Sampling distribution for the mean difference in book prices, if the true average difference is zero.

- **Example 5.2** Set up and implement a hypothesis test to determine whether, on average, there is a difference between Amazon's price for a book and the UCLA bookstore's price.

There are two scenarios: there is no difference or there is some difference in average prices. The *no difference* scenario is always the null hypothesis:

H_0 : $\mu_{diff} = 0$. There is no difference in the average textbook price.

H_A : $\mu_{diff} \neq 0$. There is a difference in average prices.

Can the normal model be used to describe the sampling distribution of \bar{x}_{diff} ? We must check that the differences meet the conditions established in Chapter 4. The observations are based on a simple random sample from less than 10% of all books sold at the bookstore, so independence is reasonable; there are more than 30 differences; and the distribution of differences, shown in Figure 5.2, is strongly skewed, but this amount of skew is reasonable for this sized data set ($n = 73$). Because all three conditions are reasonably satisfied, we can conclude the sampling distribution of \bar{x}_{diff} is nearly normal and our estimate of the standard error will be reasonable.

We compute the standard error associated with \bar{x}_{diff} using the standard deviation of the differences ($s_{diff} = 14.26$) and the number of differences ($n_{diff} = 73$):

$$SE_{\bar{x}_{diff}} = \frac{s_{diff}}{\sqrt{n_{diff}}} = \frac{14.26}{\sqrt{73}} = 1.67$$

To visualize the p-value, the sampling distribution of \bar{x}_{diff} is drawn as though H_0 is true, which is shown in Figure 5.4. The p-value is represented by the two (very) small tails.

To find the tail areas, we compute the test statistic, which is the Z score of \bar{x}_{diff} under the null condition that the actual mean difference is 0:

$$Z = \frac{\bar{x}_{diff} - 0}{SE_{\bar{x}_{diff}}} = \frac{12.76 - 0}{1.67} = 7.59$$

This Z score is so large it isn't even in the table, which ensures the single tail area will be 0.0002 or smaller. Since the p-value corresponds to both tails in this case and the normal distribution is symmetric, the p-value can be estimated as twice the one-tail area:

$$\text{p-value} = 2 \times (\text{one tail area}) \approx 2 \times 0.0002 = 0.0004$$

Because the p-value is less than 0.05, we reject the null hypothesis. We have found convincing evidence that Amazon is, on average, cheaper than the UCLA bookstore for UCLA course textbooks.

- ◉ **Exercise 5.3** Create a 95% confidence interval for the average price difference between books at the UCLA bookstore and books on Amazon.³

5.2 Difference of two means

In this section we consider a difference in two population means, $\mu_1 - \mu_2$, under the condition that the data are not paired. The methods are similar in theory but different in the details. Just as with a single sample, we identify conditions to ensure a point estimate of the difference $\bar{x}_1 - \bar{x}_2$ is nearly normal. Next we introduce a formula for the standard error, which allows us to apply our general tools from Section 4.5.

We apply these methods to two examples: participants in the 2012 Cherry Blossom Run and newborn infants. This section is motivated by questions like “Is there convincing evidence that newborns from mothers who smoke have a different average birth weight than newborns from mothers who don't smoke?”

5.2.1 Point estimates and standard errors for differences of means

We would like to estimate the average difference in run times for men and women using the `run10Samp` data set, which was a simple random sample of 45 men and 55 women from all runners in the 2012 Cherry Blossom Run. Table 5.5 presents relevant summary statistics, and box plots of each sample are shown in Figure 5.6.

	men	women
\bar{x}	87.65	102.13
s	12.5	15.2
n	45	55

Table 5.5: Summary statistics for the run time of 100 participants in the 2009 Cherry Blossom Run.

The two samples are independent of one-another, so the data are not paired. Instead a point estimate of the difference in average 10 mile times for men and women, $\mu_w - \mu_m$, can be found using the two sample means:

$$\bar{x}_w - \bar{x}_m = 102.13 - 87.65 = 14.48$$

³Conditions have already verified and the standard error computed in Example 5.2. To find the interval, identify z^* (1.96 for 95% confidence) and plug it, the point estimate, and the standard error into the confidence interval formula:

$$\text{point estimate} \pm z^*SE \rightarrow 12.76 \pm 1.96 \times 1.67 \rightarrow (9.49, 16.03)$$

We are 95% confident that Amazon is, on average, between \$9.49 and \$16.03 cheaper than the UCLA bookstore for UCLA course books.

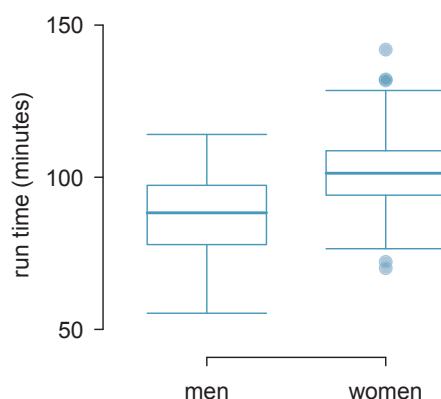


Figure 5.6: Side-by-side box plots for the sample of 2009 Cherry Blossom Run participants.

Because we are examining two simple random samples from less than 10% of the population, each sample contains at least 30 observations, and neither distribution is strongly skewed, we can safely conclude the sampling distribution of each sample mean is nearly normal. Finally, because each sample is independent of the other (e.g. the data are not paired), we can conclude that the difference in sample means can be modeled using a normal distribution.⁴

Conditions for normality of $\bar{x}_1 - \bar{x}_2$

If the sample means, \bar{x}_1 and \bar{x}_2 , each meet the criteria for having nearly normal sampling distributions and the observations in the two samples are independent, then the difference in sample means, $\bar{x}_1 - \bar{x}_2$, will have a sampling distribution that is nearly normal.

We can quantify the variability in the point estimate, $\bar{x}_w - \bar{x}_m$, using the following formula for its standard error:

$$SE_{\bar{x}_w - \bar{x}_m} = \sqrt{\frac{\sigma_w^2}{n_w} + \frac{\sigma_m^2}{n_m}}$$

We usually estimate this standard error using standard deviation estimates based on the samples:

$$\begin{aligned} SE_{\bar{x}_w - \bar{x}_m} &= \sqrt{\frac{\sigma_w^2}{n_w} + \frac{\sigma_m^2}{n_m}} \\ &\approx \sqrt{\frac{s_w^2}{n_w} + \frac{s_m^2}{n_m}} = \sqrt{\frac{15.2^2}{55} + \frac{12.5^2}{45}} = 2.77 \end{aligned}$$

Because each sample has at least 30 observations ($n_w = 55$ and $n_m = 45$), this substitution using the sample standard deviation tends to be very good.

⁴Probability theory guarantees that the difference of two independent normal random variables is also normal. Because each sample mean is nearly normal and observations in the samples are independent, we are assured the difference is also nearly normal.

Distribution of a difference of sample means

The sample difference of two means, $\bar{x}_1 - \bar{x}_2$, is nearly normal with mean $\mu_1 - \mu_2$ and estimated standard error

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (5.4)$$

when each sample mean is nearly normal and all observations are independent.

5.2.2 Confidence interval for the difference

When the data indicate that the point estimate $\bar{x}_1 - \bar{x}_2$ comes from a nearly normal distribution, we can construct a confidence interval for the difference in two means from the framework built in Chapter 4. Here a point estimate, $\bar{x}_w - \bar{x}_m = 14.48$, is associated with a normal model with standard error $SE = 2.77$. Using this information, the general confidence interval formula may be applied in an attempt to capture the true difference in means, in this case using a 95% confidence level:

$$\text{point estimate} \pm z^*SE \rightarrow 14.48 \pm 1.96 \times 2.77 \rightarrow (9.05, 19.91)$$

Based on the samples, we are 95% confident that men ran, on average, between 9.05 and 19.91 minutes faster than women in the 2012 Cherry Blossom Run.

- ⊙ **Exercise 5.5** What does 95% confidence mean?⁵
- ⊙ **Exercise 5.6** We may be interested in a different confidence level. Construct the 99% confidence interval for the population difference in average run times based on the sample data.⁶

5.2.3 Hypothesis tests based on a difference in means

A data set called `baby_smoke` represents a random sample of 150 cases of mothers and their newborns in North Carolina over a year. Four cases from this data set are represented in Table 5.7. We are particularly interested in two variables: `weight` and `smoke`. The `weight` variable represents the weights of the newborns and the `smoke` variable describes which mothers smoked during pregnancy. We would like to know, is convincing evidence that newborns from mothers who smoke have a different average birth weight than newborns from mothers who don't smoke? We will use the North Carolina sample to try to answer this question. The smoking group includes 50 cases and the nonsmoking group contains 100 cases, represented in Figure 5.8.

⁵If we were to collect many such samples and create 95% confidence intervals for each, then about 95% of these intervals would contain the population difference, $\mu_w - \mu_m$.

⁶The only thing that changes is z^* : we use $z^* = 2.58$ for a 99% confidence level. (If the selection of z^* is confusing, see Section 4.2.4 for an explanation.) The 99% confidence interval: $14.48 \pm 2.58 \times 2.77 \rightarrow (7.33, 21.63)$. We are 99% confident that the true difference in the average run times between men and women is between 7.33 and 21.63 minutes.

	fAge	mAge	weeks	weight	sexBaby	smoke
1	NA	13	37	5.00	female	nonsmoker
2	NA	14	36	5.88	female	nonsmoker
3	19	15	41	8.13	male	smoker
⋮	⋮	⋮	⋮	⋮	⋮	
150	45	50	36	9.25	female	nonsmoker

Table 5.7: Four cases from the `baby_smoke` data set. The value “NA”, shown for the first two entries of the first variable, indicates that piece of data is missing.

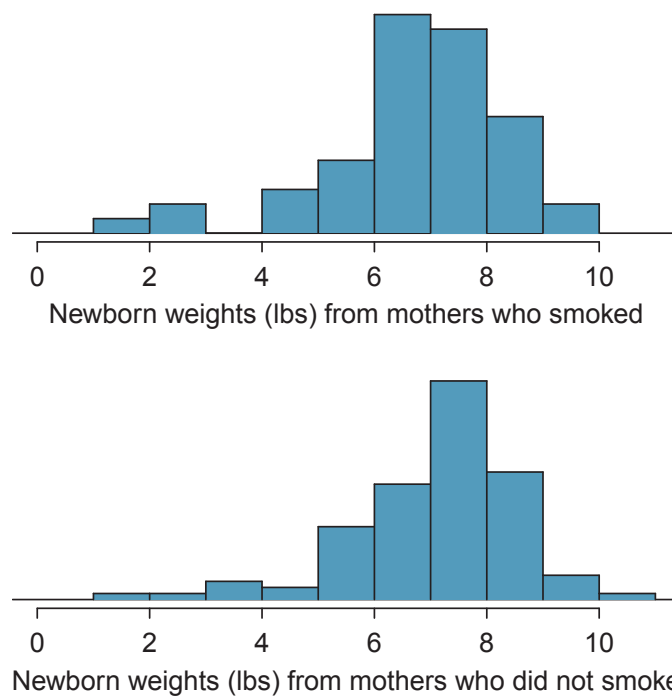


Figure 5.8: The top panel represents birth weights for infants whose mothers smoked. The bottom panel represents the birth weights for infants whose mothers who did not smoke. Both distributions exhibit strong skew.

- **Example 5.7** Set up appropriate hypotheses to evaluate whether there is a relationship between a mother smoking and average birth weight.

The null hypothesis represents the case of no difference between the groups.

H_0 : There is no difference in average birth weight for newborns from mothers who did and did not smoke. In statistical notation: $\mu_n - \mu_s = 0$, where μ_n represents non-smoking mothers and μ_s represents mothers who smoked.

H_A : There is some difference in average newborn weights from mothers who did and did not smoke ($\mu_n - \mu_s \neq 0$).

Summary statistics are shown for each sample in Table 5.9. Because the data come from a simple random sample and consist of less than 10% of all such cases, the observations are independent. Additionally, each group's sample size is at least 30 and the skew in each sample distribution is strong (see Figure 5.8). The skew is reasonable for these sample sizes of 50 and 100. Therefore, each sample mean is associated with a nearly normal distribution.

	smoker	nonsmoker
mean	6.78	7.18
st. dev.	1.43	1.60
samp. size	50	100

Table 5.9: Summary statistics for the `baby_smoke` data set.

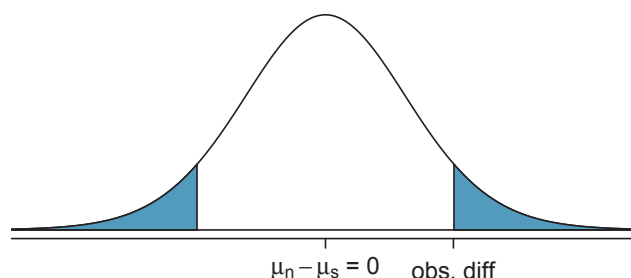
- **Exercise 5.8** (a) What is the point estimate of the population difference, $\mu_n - \mu_s$? (b) Can we use a normal distribution to model this difference? (c) Compute the standard error of the point estimate from part (a).⁷
- **Example 5.9** If the null hypothesis from Example 5.7 was true, what would be the expected value of the point estimate? And the standard deviation associated with this estimate? Draw a picture to represent the p-value.

If the null hypothesis was true, then we expect to see a difference near 0. The standard error corresponds to the standard deviation of the point estimate: 0.26. To depict the p-value, we draw the distribution of the point estimate as though H_0 was true and shade areas representing at least as much evidence against H_0 as what was observed. Both tails are shaded because it is a two-sided test.

⁷(a) The difference in sample means is an appropriate point estimate: $\bar{x}_n - \bar{x}_s = 0.40$. (b) Because the samples are independent and each sample mean is nearly normal, their difference is also nearly normal. (c) The standard error of the estimate can be estimated using Equation (5.4):

$$SE = \sqrt{\frac{\sigma_n^2}{n_n} + \frac{\sigma_s^2}{n_s}} \approx \sqrt{\frac{s_n^2}{n_n} + \frac{s_s^2}{n_s}} = \sqrt{\frac{1.60^2}{100} + \frac{1.43^2}{50}} = 0.26$$

The standard error estimate should be sufficiently accurate since the conditions were reasonably satisfied.



- **Example 5.10** Compute the p-value of the hypothesis test using the figure in Example 5.9, and evaluate the hypotheses using a significance level of $\alpha = 0.05$.

Since the point estimate is nearly normal, we can find the upper tail using the Z score and normal probability table:

$$Z = \frac{0.40 - 0}{0.26} = 1.54 \quad \rightarrow \quad \text{upper tail} = 1 - 0.938 = 0.062$$

Because this is a two-sided test and we want the area of both tails, we double this single tail to get the p-value: 0.124. This p-value is larger than the significance value, 0.05, so we fail to reject the null hypothesis. There is insufficient evidence to say there is a difference in average birth weight of newborns from North Carolina mothers who did smoke during pregnancy and newborns from North Carolina mothers who did not smoke during pregnancy.

- **Exercise 5.11** Does the conclusion to Example 5.10 mean that smoking and average birth weight are unrelated?⁸
- **Exercise 5.12** If we made a Type 2 Error and there is a difference, what could we have done differently in data collection to be more likely to detect such a difference?⁹

5.2.4 Summary for inference of the difference of two means

When considering the difference of two means, there are two common cases: the two samples are paired or they are independent. (There are instances where the data are neither paired nor independent.) The paired case was treated in Section 5.1, where the one-sample methods were applied to the differences from the paired observations. We examined the second and more complex scenario in this section.

When applying the normal model to the point estimate $\bar{x}_1 - \bar{x}_2$ (corresponding to unpaired data), it is important to verify conditions before applying the inference framework using the normal model. First, each sample mean must meet the conditions for normality; these conditions are described in Chapter 4 on page 168. Secondly, the samples must be collected independently (e.g. not paired data). When these conditions are satisfied, the general inference tools of Chapter 4 may be applied.

For example, a confidence interval may take the following form:

$$\text{point estimate} \pm z^* SE$$

⁸Absolutely not. It is possible that there is some difference but we did not detect it. If this is the case, we made a Type 2 Error.

⁹We could have collected more data. If the sample sizes are larger, we tend to have a better shot at finding a difference if one exists.

When we compute the confidence interval for $\mu_1 - \mu_2$, the point estimate is the difference in sample means, the value z^* corresponds to the confidence level, and the standard error is computed from Equation (5.4) on page 217. While the point estimate and standard error formulas change a little, the framework for a confidence interval stays the same. This is also true in hypothesis tests for differences of means.

In a hypothesis test, we apply the standard framework and use the specific formulas for the point estimate and standard error of a difference in two means. The test statistic represented by the Z score may be computed as

$$Z = \frac{\text{point estimate} - \text{null value}}{SE}$$

When assessing the difference in two means, the point estimate takes the form $\bar{x}_1 - \bar{x}_2$, and the standard error again takes the form of Equation (5.4) on page 217. Finally, the null value is the difference in sample means under the null hypothesis. Just as in Chapter 4, the test statistic Z is used to identify the p -value.

5.2.5 Examining the standard error formula

The formula for the standard error of the difference in two means is similar to the formula for other standard errors. Recall that the standard error of a single mean, \bar{x}_1 , can be approximated by

$$SE_{\bar{x}_1} = \frac{s_1}{\sqrt{n_1}}$$

where s_1 and n_1 represent the sample standard deviation and sample size.

The standard error of the difference of two sample means can be constructed from the standard errors of the separate sample means:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{SE_{\bar{x}_1}^2 + SE_{\bar{x}_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (5.13)$$

This special relationship follows from probability theory.

- ⊙ **Exercise 5.14** Prerequisite: Section 2.4. We can rewrite Equation (5.13) in a different way:

$$SE_{\bar{x}_1 - \bar{x}_2}^2 = SE_{\bar{x}_1}^2 + SE_{\bar{x}_2}^2$$

Explain where this formula comes from using the ideas of probability theory.¹⁰

5.3 One-sample means with the t distribution

The motivation in Chapter 4 for requiring a large sample was two-fold. First, a large sample ensures that the sampling distribution of \bar{x} is nearly normal. We will see in Section 5.3.1 that if the population data are nearly normal, then \bar{x} is also nearly normal regardless of the

¹⁰The standard error squared represents the variance of the estimate. If X and Y are two random variables with variances σ_x^2 and σ_y^2 , then the variance of $X - Y$ is $\sigma_x^2 + \sigma_y^2$. Likewise, the variance corresponding to $\bar{x}_1 - \bar{x}_2$ is $\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2$. Because $\sigma_{\bar{x}_1}^2$ and $\sigma_{\bar{x}_2}^2$ are just another way of writing $SE_{\bar{x}_1}^2$ and $SE_{\bar{x}_2}^2$, the variance associated with $\bar{x}_1 - \bar{x}_2$ may be written as $SE_{\bar{x}_1}^2 + SE_{\bar{x}_2}^2$.