Practice Exercises: Lesson 5.1

STAT 1201
Introduction to Probability and Statistics

ONLINE AND DISTANCE EDUCATION

Exercises

7.1 Identify the critical t. An independent random sample is selected from an approximately normal population with unknown standard deviation. Find the degrees of freedom and the critical t-value ($t^\star$) for the given sample size and confidence level.

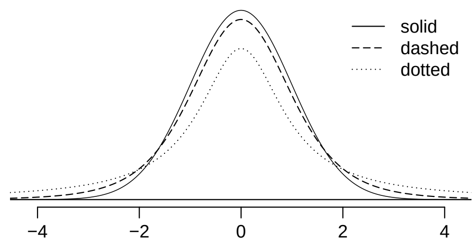
(a) $n = 6$, CL = 90%

(b) $n = 21$, CL = 98%

(c) $n = 29$, CL = 95%

(d) $n = 12$, CL = 99%

7.2 t-distribution. The figure on the right shows three unimodal and symmetric curves: the standard normal (z) distribution, the t-distribution with 5 degrees of freedom, and the t-distribution with 1 degree of freedom. Determine which is which, and explain your reasoning.



7.3 Find the p-value, Part I. An independent random sample is selected from an approximately normal population with an unknown standard deviation. Find the p-value for the given sample size and test statistic. Also determine if the null hypothesis would be rejected at $\alpha = 0.05$.

(a) $n = 11$, $T = 1.91$

(b) $n = 17$, $T = -3.45$

(c) $n = 7$, $T = 0.83$

(d) $n = 28$, $T = 2.13$

7.4 Find the p-value, Part II. An independent random sample is selected from an approximately normal population with an unknown standard deviation. Find the p-value for the given sample size and test statistic. Also determine if the null hypothesis would be rejected at $\alpha = 0.01$.

(a) $n = 26$, $T = 2.485$

(b) $n = 18$, $T = 0.5$

7.5 Working backwards, Part I. A 95% confidence interval for a population mean, μ, is given as (18.985, 21.015). This confidence interval is based on a simple random sample of 36 observations. Calculate the sample mean and standard deviation. Assume that all conditions necessary for inference are satisfied. Use the t-distribution in any calculations.

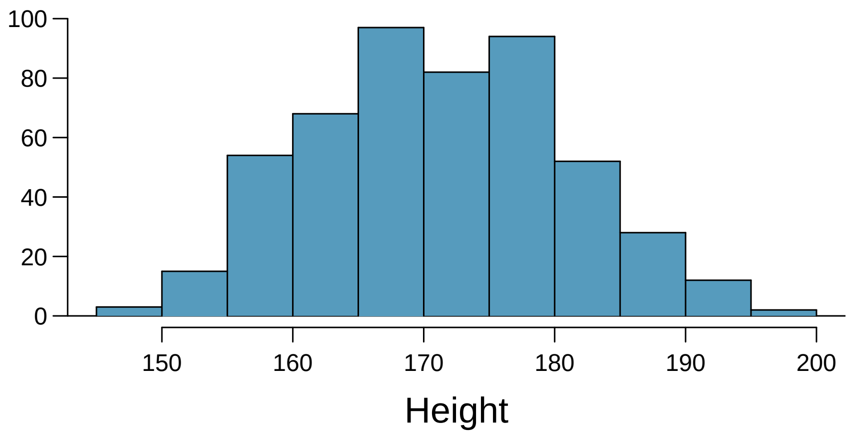
7.6 Working backwards, Part II. A 90% confidence interval for a population mean is (65, 77). The population distribution is approximately normal and the population standard deviation is unknown. This confidence interval is based on a simple random sample of 25 observations. Calculate the sample mean, the margin of error, and the sample standard deviation.

7.7 Sleep habits of New Yorkers. 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. Statistical summaries of these data are shown below. The point estimate suggests New Yorkers sleep less than 8 hours a night on average. Is the result statistically significant?

n	$\bar{x}$	s	min	max
25	7.73	0.77	6.17	9.78

(a) Write the hypotheses in symbols and in words.
(b) Check conditions, then calculate the test statistic, T, and the associated degrees of freedom.
(c) Find and interpret the p-value in this context. Drawing a picture may be helpful.
(d) What is the conclusion of the hypothesis test?
(e) If you were to construct a 90% confidence interval that corresponded to this hypothesis test, would you expect 8 hours to be in the interval?

7.8 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.[8]



Min	147.2
Q1	163.8
Median	170.3
Mean	171.1
SD	9.4
Q3	177.8
Max	198.1

(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.

7.9 Find the mean. You are given the following hypotheses:

$$H_0 : \mu = 60$$
$$H_A : \mu \neq 60$$

We know that the sample standard deviation is 8 and the sample size is 20. For what sample mean would the p-value be equal to 0.05? Assume that all conditions necessary for inference are satisfied.

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

7.10 $t^\star$ vs. $z^\star$. For a given confidence level, $t^\star_{df}$ is larger than $z^\star$. Explain how t^*_{df} being slightly larger than z^* affects the width of the confidence interval.

7.11 Play the piano. Georgianna claims that in a small city renowned for its music school, the average child takes less than 5 years of piano lessons. We have a random sample of 20 children from the city, with a mean of 4.6 years of piano lessons and a standard deviation of 2.2 years.

(a) Evaluate Georgianna's claim (or that the opposite might be true) using a hypothesis test.

(b) Construct a 95% confidence interval for the number of years students in this city take piano lessons, and interpret it in context of the data.

(c) Do your results from the hypothesis test and the confidence interval agree? Explain your reasoning.

7.12 Auto exhaust and lead exposure. Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 μg/l and a SD of 37.74 μg/l; a previous study of individuals from a nearby suburb, with no history of exposure, found an average blood level concentration of 35 μg/l.[9]

(a) Write down the hypotheses that would be appropriate for testing if the police officers appear to have been exposed to a different concentration of lead.

(b) Explicitly state and check all conditions necessary for inference on these data.

(c) Regardless of your answers in part (b), test the hypothesis that the downtown police officers have a higher lead exposure than the group in the previous study. Interpret your results in context.

7.13 Car insurance savings. A market researcher wants to evaluate car insurance savings at a competing company. Based on past studies he is assuming that the standard deviation of savings is \$100. He wants to collect data such that he can get a margin of error of no more than \$10 at a 95% confidence level. How large of a sample should he collect?

7.14 SAT scores. The standard deviation of SAT scores for students at a particular Ivy League college is 250 points. Two statistics students, Raina and Luke, want to estimate the average SAT score of students at this college as part of a class project. They want their margin of error to be no more than 25 points.

(a) Raina wants to use a 90% confidence interval. How large a sample should she collect?

(b) Luke wants to use a 99% confidence interval. Without calculating the actual sample size, determine whether his sample should be larger or smaller than Raina's, and explain your reasoning.

(c) Calculate the minimum required sample size for Luke.

[9] WI Mortada et al. "Study of lead exposure from automobile exhaust as a risk for nephrotoxicity among traffic policemen." In: *American journal of nephrology* 21.4 (2000), pp. 274–279.

Exercises

7.15 Air quality. Air quality measurements were collected in a random sample of 25 country capitals in 2013, and then again in the same cities in 2014. We would like to use these data to compare average air quality between the two years. Should we use a paired or non-paired test? Explain your reasoning.

7.16 True / False: paired. Determine if the following statements are true or false. If false, explain.

(a) In a paired analysis we first take the difference of each pair of observations, and then we do inference on these differences.

(b) Two data sets of different sizes cannot be analyzed as paired data.

(c) Consider two sets of data that are paired with each other. Each observation in one data set has a natural correspondence with exactly one observation from the other data set.

(d) Consider two sets of data that are paired with each other. Each observation in one data set is subtracted from the average of the other data set's observations.

7.17 Paired or not? Part I. In each of the following scenarios, determine if the data are paired.

(a) Compare pre- (beginning of semester) and post-test (end of semester) scores of students.

(b) Assess gender-related salary gap by comparing salaries of randomly sampled men and women.

(c) Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients.

(d) Assess effectiveness of a diet regimen by comparing the before and after weights of subjects.

7.18 Paired or not? Part II. In each of the following scenarios, determine if the data are paired.

(a) We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days, and record Intel's and Southwest's stock on those same days.

(b) We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items.

(c) A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school.

7.19 Global warming, Part I. Let's consider a limited set of climate data, examining temperature differences in 1948 vs 2018. We sampled 197 locations from the National Oceanic and Atmospheric Administration's (NOAA) historical data, where the data was available for both years of interest. We want to know: were there more days with temperatures exceeding 90°F in 2018 or in 1948?[12] The difference in number of days exceeding 90°F (number of days in 2018 - number of days in 1948) was calculated for each of the 197 locations. The average of these differences was 2.9 days with a standard deviation of 17.2 days. We are interested in determining whether these data provide strong evidence that there were more days in 2018 that exceeded 90°F from NOAA's weather stations.

(a) Is there a relationship between the observations collected in 1948 and 2018? Or are the observations in the two groups independent? Explain.

(b) Write hypotheses for this research in symbols and in words.

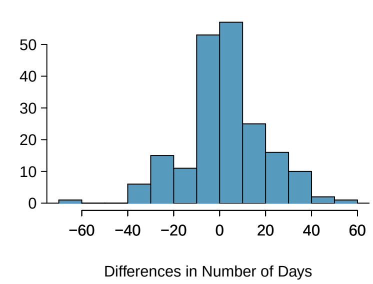
(c) Check the conditions required to complete this test. A histogram of the differences is given to the right.

(d) Calculate the test statistic and find the p-value.

(e) Use $\alpha = 0.05$ to evaluate the test, and interpret your conclusion in context.

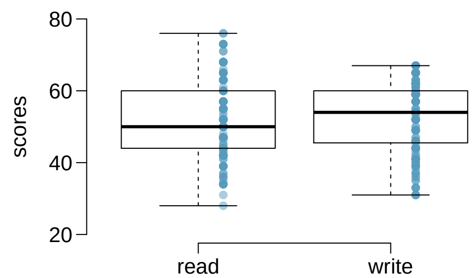
(f) What type of error might we have made? Explain in context what the error means.

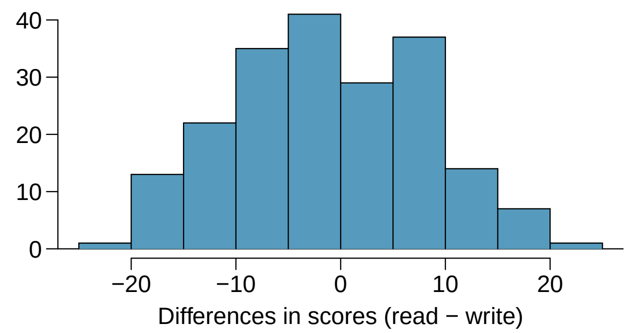
(g) Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the number of days exceeding 90°F from 1948 and 2018 to include 0? Explain your reasoning.



[12] NOAA, www.ncdc.noaa.gov/cdo-web/datasets, April 24, 2019.

7.20 High School and Beyond, Part I. The National Center of Education Statistics conducted a survey of high school seniors, collecting test data on reading, writing, and several other subjects. Here we examine a simple random sample of 200 students from this survey. Side-by-side box plots of reading and writing scores as well as a histogram of the differences in scores are shown below.





(a) Is there a clear difference in the average reading and writing scores?

(b) Are the reading and writing scores of each student independent of each other?

(c) Create hypotheses appropriate for the following research question: is there an evident difference in the average scores of students in the reading and writing exam?

(d) Check the conditions required to complete this test.

(e) The average observed difference in scores is $\bar{x}_{read-write} = -0.545$, and the standard deviation of the differences is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams?

(f) What type of error might we have made? Explain what the error means in the context of the application.

(g) Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the reading and writing scores to include 0? Explain your reasoning.

7.21 Global warming, Part II. We considered the change in the number of days exceeding 90°F from 1948 and 2018 at 197 randomly sampled locations from the NOAA database in Exercise 7.19. The mean and standard deviation of the reported differences are 2.9 days and 17.2 days.

(a) Calculate a 90% confidence interval for the average difference between number of days exceeding 90°F between 1948 and 2018. We've already checked the conditions for you.

(b) Interpret the interval in context.

(c) Does the confidence interval provide convincing evidence that there were more days exceeding 90°F in 2018 than in 1948 at NOAA stations? Explain.

7.22 High school and beyond, Part II. We considered the differences between the reading and writing scores of a random sample of 200 students who took the High School and Beyond Survey in Exercise 7.20. The mean and standard deviation of the differences are $\bar{x}_{read-write} = -0.545$ and 8.887 points.

(a) Calculate a 95% confidence interval for the average difference between the reading and writing scores of all students.

(b) Interpret this interval in context.

(c) Does the confidence interval provide convincing evidence that there is a real difference in the average scores? Explain.

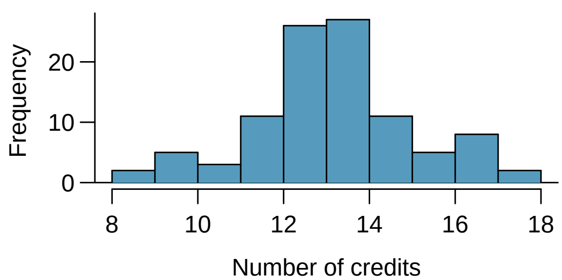
Chapter exercises

7.47 Gaming and distracted eating, Part I. A group of researchers are interested in the possible effects of distracting stimuli during eating, such as an increase or decrease in the amount of food consumption. To test this hypothesis, they monitored food intake for a group of 44 patients who were randomized into two equal groups. The treatment group ate lunch while playing solitaire, and the control group ate lunch without any added distractions. Patients in the treatment group ate 52.1 grams of biscuits, with a standard deviation of 45.1 grams, and patients in the control group ate 27.1 grams of biscuits, with a standard deviation of 26.4 grams. Do these data provide convincing evidence that the average food intake (measured in amount of biscuits consumed) is different for the patients in the treatment group? Assume that conditions for inference are satisfied.[39]

7.48 Gaming and distracted eating, Part II. The researchers from Exercise 7.47 also investigated the effects of being distracted by a game on how much people eat. The 22 patients in the treatment group who ate their lunch while playing solitaire were asked to do a serial-order recall of the food lunch items they ate. The average number of items recalled by the patients in this group was 4. 9, with a standard deviation of 1.8. The average number of items recalled by the patients in the control group (no distraction) was 6.1, with a standard deviation of 1.8. Do these data provide strong evidence that the average number of food items recalled by the patients in the treatment and control groups are different?

7.49 Sample size and pairing. Determine if the following statement is true or false, and if false, explain your reasoning: If comparing means of two groups with equal sample sizes, always use a paired test.

7.50 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.



Min	8
Q1	13
Median	14
Mean	13.65
SD	1.91
Q3	15
Max	18

(a) 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.

(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.

[39] R.E. Oldham-Cooper et al. "Playing a computer game during lunch affects fullness, memory for lunch, and later snack intake". In: *The American Journal of Clinical Nutrition* 93.2 (2011), p. 308.

7.51 Hen eggs. The distribution of the number of eggs laid by a certain species of hen during their breeding period has a mean of 35 eggs with a standard deviation of 18.2. Suppose a group of researchers randomly samples 45 hens of this species, counts the number of eggs laid during their breeding period, and records the sample mean. They repeat this 1,000 times, and build a distribution of sample means.

(a) What is this distribution called?

(b) Would you expect the shape of this distribution to be symmetric, right skewed, or left skewed? Explain your reasoning.

(c) Calculate the variability of this distribution and state the appropriate term used to refer to this value.

(d) Suppose the researchers' budget is reduced and they are only able to collect random samples of 10 hens. The sample mean of the number of eggs is recorded, and we repeat this 1,000 times, and build a new distribution of sample means. How will the variability of this new distribution compare to the variability of the original distribution?

7.52 Forest management. Forest rangers wanted to better understand the rate of growth for younger trees in the park. They took measurements of a random sample of 50 young trees in 2009 and again measured those same trees in 2019. The data below summarize their measurements, where the heights are in feet:

	2009	2019	Differences
$\bar{x}$	12.0	24.5	12.5
s	3.5	9.5	7.2
n	50	50	50

Construct a 99% confidence interval for the average growth of (what had been) younger trees in the park over 2009-2019.

7.53 Experiment resizing. At a startup company running a new weather app, an engineering team generally runs experiments where a random sample of 1% of the app's visitors in the control group and another 1% were in the treatment group to test each new feature. The team's core goal is to increase a metric called *daily visitors*, which is essentially the number of visitors to the app each day. They track this metric in each experiment arm and as their core experiment metric. In their most recent experiment, the team tested including a new animation when the app started, and the number of daily visitors in this experiment stabilized at +1.2% with a 95% confidence interval of (-0.2%, +2.6%). This means if this new app start animation was launched, the team thinks they might lose as many as 0.2% of daily visitors or gain as many as 2.6% more daily visitors. Suppose you are consulting as the team's data scientist, and after discussing with the team, you and they agree that they should run another experiment that is bigger. You also agree that this new experiment should be able to detect a gain in the daily visitors metric of 1.0% or more with 80% power. Now they turn to you and ask, "How big of an experiment do we need to run to ensure we can detect this effect?"

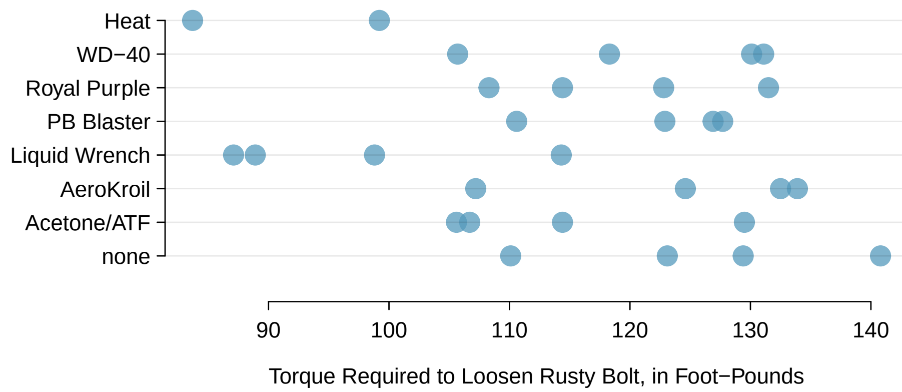
(a) How small must the standard error be if the team is to be able to detect an effect of 1.0% with 80% power and a significance level of $\alpha = 0.05$? You may safely assume the percent change in daily visitors metric follows a normal distribution.

(b) Consider the first experiment, where the point estimate was +1.2% and the 95% confidence interval was (-0.2%, +2.6%). If that point estimate followed a normal distribution, what was the standard error of the estimate?

(c) The ratio of the standard error from part (a) vs the standard error from part (b) should be 2.03. How much bigger of an experiment is needed to shrink a standard error by a factor of 2.03?

(d) Using your answer from part (c) and that the original experiment was a 1% vs 1% experiment to recommend an experiment size to the team.

7.54 Torque on a rusty bolt. Project Farm is a YouTube channel that routinely compares different products. In one episode, the channel evaluated different options for loosening rusty bolts.[40] Eight options were evaluated, including a control group where no treatment was given ("none" in the graph), to determine which was most effective. For all treatments, there were four bolts tested, except for a treatment of heat with a blow torch, where only two data points were collected. The results are shown in the figure below:



(a) Do you think it is reasonable to apply ANOVA in this case?

(b) Regardless of your answer in part (a), describe hypotheses for ANOVA in this context, and use the table below to carry out the test. Give your conclusion in the context of the data.

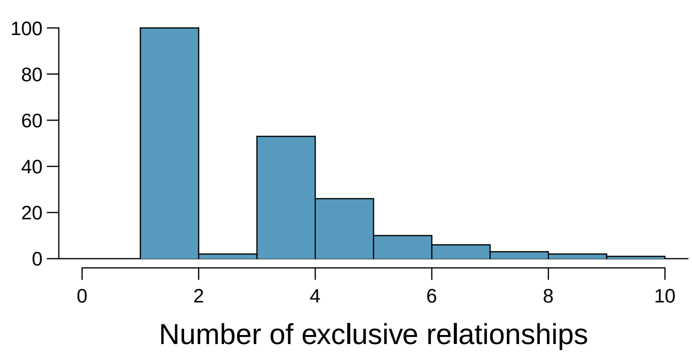
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	7	3603.43	514.78	4.03	0.0056
Residuals	22	2812.80	127.85		

(c) The table below are p-values for pairwise t-tests comparing each of the different groups. These p-values have not been corrected for multiple comparisons. Which pair of groups appears most likely to represent a difference?

	AeroKroil	Heat	Liquid Wrench	none	PB Blaster	Royal Purple	WD-40
Acetone/ATF	0.2026	0.0308	0.0476	0.1542	0.3294	0.5222	0.3744
AeroKroil		0.0027	0.0025	0.8723	0.7551	0.5143	0.6883
Heat			0.5580	0.0020	0.0050	0.0096	0.0059
Liquid Wrench				0.0017	0.0053	0.0117	0.0065
none					0.6371	0.4180	0.5751
PB Blaster						0.7318	0.9286
Royal Purple							0.8000

(d) There are 28 p-values shown in the table in part (c). Determine if any of them are statistically significant after correcting for multiple comparisons. If so, which one(s)? Explain your answer.

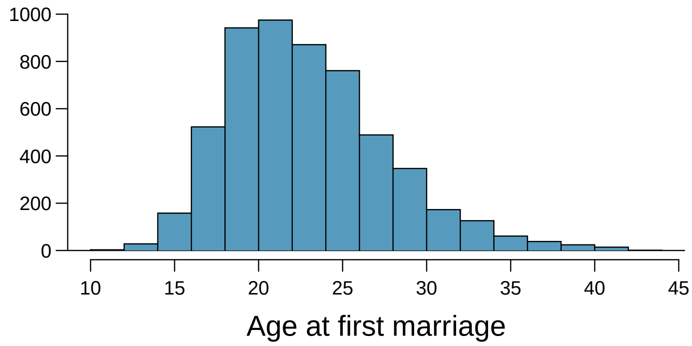
7.55 Exclusive relationships. A survey conducted on a reasonably random sample of 203 undergraduates asked, among many other questions, about the number of exclusive relationships these students have been in. The histogram below shows the distribution of the data from this sample. The sample average 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.

[40]Project Farm on YouTube, youtu.be/xUEob2oAKVs, April 16, 2018.

7.56 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.[41]



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

7.57 Online communication. A study suggests that the average college student spends 10 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 13.5 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.

$$H_0 : \bar{x} < 10 \ hours$$
$$H_A : \bar{x} > 13.5 \ hours$$

7.58 Age at first marriage, Part II. Exercise 7.56 presents the results of a 2006 - 2010 survey showing that the average age of women at first marriage is 23.44. Suppose a social scientist thinks this value has changed since the survey was taken. Below is how she set up her hypotheses. Indicate any errors you see.

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

[41] Centers for Disease Control and Prevention, National Survey of Family Growth, 2010.