

Chapters 1-5: Exercises

1.22 Stressed out, Part I. A study that surveyed a random sample of otherwise healthy high school students found that they are more likely to get muscle cramps when they are stressed. The study also noted that students drink more coffee and sleep less when they are stressed.

- What type of study is this?
- Can this study be used to conclude a causal relationship between increased stress and muscle cramps?
- State possible confounding variables that might explain the observed relationship between increased stress and muscle cramps.

1.26 Family size. Suppose we want to estimate household size, where a “household” is defined as people living together in the same dwelling, and sharing living accommodations. If we select students at random at an elementary school and ask them what their family size is, will this be a good measure of household size? Or will our average be biased? If so, will it overestimate or underestimate the true value?

1.42 Screens, teens, and psychological well-being. In a study of three nationally representative large-scale data sets from Ireland, the United States, and the United Kingdom ($n = 17,247$), teenagers between the ages of 12 to 15 were asked to keep a diary of their screen time and answer questions about how they felt or acted. The answers to these questions were then used to compute a psychological well-being score. Additional data were collected and included in the analysis, such as each child's sex and age, and on the mother's education, ethnicity, psychological distress, and employment. The study concluded that there is little clear-cut evidence that screen time decreases adolescent well-being.³⁵

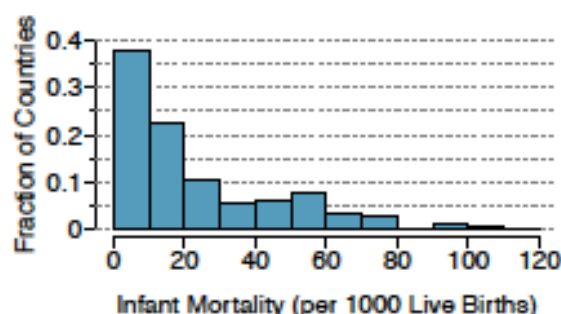
- What type of study is this?
- Identify the explanatory variables.
- Identify the response variable.
- Comment on whether the results of the study can be generalized to the population, and why.
- Comment on whether the results of the study can be used to establish causal relationships.

2.22 Views on immigration. 910 randomly sampled registered voters from Tampa, FL were asked if they thought workers who have illegally entered the US should be (i) allowed to keep their jobs and apply for US citizenship, (ii) allowed to keep their jobs as temporary guest workers but not allowed to apply for US citizenship, or (iii) lose their jobs and have to leave the country. The results of the survey by political ideology are shown below.²²

		<i>Political ideology</i>			<i>Total</i>
		<i>Conservative</i>	<i>Moderate</i>	<i>Liberal</i>	
<i>Response</i>	(i) Apply for citizenship	57	120	101	278
	(ii) Guest worker	121	113	28	262
	(iii) Leave the country	179	126	45	350
	(iv) Not sure	15	4	1	20
	<i>Total</i>	372	363	175	910

- What percent of these Tampa, FL voters identify themselves as conservatives?
- What percent of these Tampa, FL voters are in favor of the citizenship option?
- What percent of these Tampa, FL voters identify themselves as conservatives and are in favor of the citizenship option?
- What percent of these Tampa, FL voters who identify themselves as conservatives are also in favor of the citizenship option? What percent of moderates share this view? What percent of liberals share this view?
- Do political ideology and views on immigration appear to be independent? Explain your reasoning.

2.28 Infant mortality. The infant mortality rate is defined as the number of infant deaths per 1,000 live births. This rate is often used as an indicator of the level of health in a country. The relative frequency histogram below shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014.³¹



- Estimate Q1, the median, and Q3 from the histogram.
- Would you expect the mean of this data set to be smaller or larger than the median? Explain your reasoning.

3.10 Guessing on an exam. In a multiple choice exam, there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all and decides to randomly guess the answers. What is the probability that:

- the first question she gets right is the 5th question?
- she gets all of the questions right?
- she gets at least one question right?

3.12 School absences. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.²⁴

- What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?
- What is the probability that a student chosen at random misses no more than one day?
- What is the probability that a student chosen at random misses at least one day?
- If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.
- If a parent has two kids at a DeKalb County elementary school, what is the probability that both kids will miss some school, i.e. at least one day? Note any assumption you make.
- If you made an assumption in part (d) or (e), do you think it was reasonable? If you didn't make any assumptions, double check your earlier answers.

3.18 Assortative mating. Assortative mating is a nonrandom mating pattern where individuals with similar genotypes and/or phenotypes mate with one another more frequently than what would be expected under a random mating pattern. Researchers studying this topic collected data on eye colors of 204 Scandinavian men and their female partners. The table below summarizes the results. For simplicity, we only include heterosexual relationships in this exercise.⁴²

		Partner (female)			Total
		Blue	Brown	Green	
Self (male)	Blue	78	23	13	114
	Brown	19	23	12	54
	Green	11	9	16	36
	Total	108	55	41	204

- What is the probability that a randomly chosen male respondent or his partner has blue eyes?
- What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?
- What is the probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes? What about the probability of a randomly chosen male respondent with green eyes having a partner with blue eyes?
- Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

3.32 Is it worth it? Andy is always looking for ways to make money fast. Lately, he has been trying to make money by gambling. Here is the game he is considering playing: The game costs \$2 to play. He draws a card from a deck. If he gets a number card (2-10), he wins nothing. For any face card (jack, queen or king), he wins \$3. For any ace, he wins \$5, and he wins an *extra* \$20 if he draws the ace of clubs.

- Create a probability model and find Andy's expected profit per game.
- Would you recommend this game to Andy as a good way to make money? Explain.

3.36 European roulette. The game of European roulette involves spinning a wheel with 37 slots: 18 red, 18 black, and 1 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money.

- Suppose you play roulette and bet \$3 on a single round. What is the expected value and standard deviation of your total winnings?
- Suppose you bet \$1 in three different rounds. What is the expected value and standard deviation of your total winnings?
- How do your answers to parts (a) and (b) compare? What does this say about the riskiness of the two games?

3.38 Income and gender. The relative frequency table below displays the distribution of annual total personal income (in 2009 inflation-adjusted dollars) for a representative sample of 96,420,486 Americans. These data come from the American Community Survey for 2005-2009. This sample is comprised of 59% males and 41% females.⁶³

	<i>Income</i>	<i>Total</i>
(a) Describe the distribution of total personal income.	\$1 to \$9,999 or less	2.2%
(b) What is the probability that a randomly chosen US resident makes less than \$50,000 per year?	\$10,000 to \$14,999	4.7%
(c) What is the probability that a randomly chosen US resident makes less than \$50,000 per year and is female? Note any assumptions you make.	\$15,000 to \$24,999	15.8%
	\$25,000 to \$34,999	18.3%
	\$35,000 to \$49,999	21.2%
	\$50,000 to \$64,999	13.9%
(d) The same data source indicates that 71.8% of females make less than \$50,000 per year. Use this value to determine whether or not the assumption you made in part (c) is valid.	\$65,000 to \$74,999	5.8%
	\$75,000 to \$99,999	8.4%
	\$100,000 or more	9.7%

3.42 Twins. About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex – half are males and the other half are females. One-quarter of fraternal twins are both male, one-quarter both female, and one-half are mixes: one male, one female. You have just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

4.18 Chicken pox, Part I. The National Vaccine Information Center estimates that 90% of Americans have had chickenpox by the time they reach adulthood.³²

- Suppose we take a random sample of 100 American adults. Is the use of the binomial distribution appropriate for calculating the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood? Explain.
- Calculate the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood.
- What is the probability that exactly 3 out of a new sample of 100 American adults have *not* had chickenpox in their childhood?
- What is the probability that at least 1 out of 10 randomly sampled American adults have had chickenpox?
- What is the probability that at most 3 out of 10 randomly sampled American adults have *not* had chickenpox?

4.20 Chickenpox, Part II. We learned in Exercise 4.18 that about 90% of American adults had chickenpox before adulthood. We now consider a random sample of 120 American adults.

- How many people in this sample would you expect to have had chickenpox in their childhood? And with what standard deviation?
- Would you be surprised if there were 105 people who have had chickenpox in their childhood?
- What is the probability that 105 or fewer people in this sample have had chickenpox in their childhood? How does this probability relate to your answer to part (b)?

4.24 Sickle cell anemia. Sickle cell anemia is a genetic blood disorder where red blood cells lose their flexibility and assume an abnormal, rigid, “sickle” shape, which results in a risk of various complications. If both parents are carriers of the disease, then a child has a 25% chance of having the disease, 50% chance of being a carrier, and 25% chance of neither having the disease nor being a carrier. If two parents who are carriers of the disease have 3 children, what is the probability that

- (a) two will have the disease?
- (b) none will have the disease?
- (c) at least one will neither have the disease nor be a carrier?
- (d) the first child with the disease will be the 3rd child?

4.26 Male children. While it is often assumed that the probabilities of having a boy or a girl are the same, the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 kids.

- (a) Use the binomial model to calculate the probability that two of them will be boys.
- (b) Write out all possible orderings of 3 children, 2 of whom are boys. Use these scenarios to calculate the same probability from part (a) but using the addition rule for disjoint outcomes. Confirm that your answers from parts (a) and (b) match.
- (c) If we wanted to calculate the probability that a couple who plans to have 8 kids will have 3 boys, briefly describe why the approach from part (b) would be more tedious than the approach from part (a).

5.4 Unexpected expense. In a random sample 765 adults in the United States, 322 say they could not cover a \$400 unexpected expense without borrowing money or going into debt.

- (a) What population is under consideration in the data set?
- (b) What parameter is being estimated?
- (c) What is the point estimate for the parameter?
- (d) What is the name of the statistic can we use to measure the uncertainty of the point estimate?
- (e) Compute the value from part (d) for this context.
- (f) A cable news pundit thinks the value is actually 50%. Should she be surprised by the data?
- (g) Suppose the true population value was found to be 40%. If we use this proportion to recompute the value in part (e) using $p = 0.4$ instead of \hat{p} , does the resulting value change much?

5.8 Twitter users and news, Part I. A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter.¹² The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion. Construct a 99% confidence interval for the fraction of U.S. adult Twitter users who get some news on Twitter, and interpret the confidence interval in context.

5.13 Website registration. A website is trying to increase registration for first-time visitors, exposing 1% of these visitors to a new site design. Of 752 randomly sampled visitors over a month who saw the new design, 64 registered.

- (a) Check any conditions required for constructing a confidence interval.
- (b) Compute the standard error.
- (c) Construct and interpret a 90% confidence interval for the fraction of first-time visitors of the site who would register under the new design (assuming stable behaviors by new visitors over time).