



Practice Exercises: Lesson 1.2

Diez, D. M., Çetinkaya-Rundel, M., Barr, C. D. (2019). OpenIntro Statistics (4th ed.). OpenIntro.
<https://www.openintro.org/book/os/> CC BY-SA 3.0

STAT 1201
Introduction to Probability and Statistics

ONLINE AND DISTANCE EDUCATION

Exercises

1.13 Air pollution and birth outcomes, scope of inference. Exercise 1.3 introduces a study where researchers collected data to examine the relationship between air pollutants and preterm births in Southern California. During the study air pollution levels were measured by air quality monitoring stations. Length of gestation data were collected on 143,196 births between the years 1989 and 1993, and air pollution exposure during gestation was calculated for each birth.

- Identify the population of interest and the sample in this study.
- Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

1.14 Cheaters, scope of inference. Exercise 1.5 introduces a study where researchers studying the relationship between honesty, age, and self-control conducted an experiment on 160 children between the ages of 5 and 15. The researchers asked each child to toss a fair coin in private and to record the outcome (white or black) on a paper sheet, and said they would only reward children who report white. Half the students were explicitly told not to cheat and the others were not given any explicit instructions. Differences were observed in the cheating rates in the instruction and no instruction groups, as well as some differences across children's characteristics within each group.

- Identify the population of interest and the sample in this study.
- Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

1.15 Buteyko method, scope of inference. Exercise 1.4 introduces a study on using the Buteyko shallow breathing technique to reduce asthma symptoms and improve quality of life. As part of this study 600 asthma patients aged 18-69 who relied on medication for asthma treatment were recruited and randomly assigned to two groups: one practiced the Buteyko method and the other did not. Those in the Buteyko group experienced, on average, a significant reduction in asthma symptoms and an improvement in quality of life.

- Identify the population of interest and the sample in this study.
- Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

1.16 Stealers, scope of inference. Exercise 1.6 introduces a study on the relationship between socio-economic class and unethical behavior. As part of this study 129 University of California Berkeley undergraduates were asked to identify themselves as having low or high social-class by comparing themselves to others with the most (least) money, most (least) education, and most (least) respected jobs. They were also presented with a jar of individually wrapped candies and informed that the candies were for children in a nearby laboratory, but that they could take some if they wanted. After completing some unrelated tasks, participants reported the number of candies they had taken. It was found that those who were identified as upper-class took more candy than others.

- Identify the population of interest and the sample in this study.
- Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

1.17 Relaxing after work. The General Social Survey asked the question, "After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?" to a random sample of 1,155 Americans. The average relaxing time was found to be 1.65 hours. Determine which of the following is an observation, a variable, a sample statistic (value calculated based on the observed sample), or a population parameter.

- An American in the sample.
- Number of hours spent relaxing after an average work day.
- 1.65.
- Average number of hours all Americans spend relaxing after an average work day.

1.18 Cats on YouTube. Suppose you want to estimate the percentage of videos on YouTube that are cat videos. It is impossible for you to watch all videos on YouTube so you use a random video picker to select 1000 videos for you. You find that 2% of these videos are cat videos. Determine which of the following is an observation, a variable, a sample statistic (value calculated based on the observed sample), or a population parameter.

- Percentage of all videos on YouTube that are cat videos.
- 2%.
- A video in your sample.
- Whether or not a video is a cat video.

1.19 Course satisfaction across sections. A large college class has 160 students. All 160 students attend the lectures together, but the students are divided into 4 groups, each of 40 students, for lab sections administered by different teaching assistants. The professor wants to conduct a survey about how satisfied the students are with the course, and he believes that the lab section a student is in might affect the student's overall satisfaction with the course.

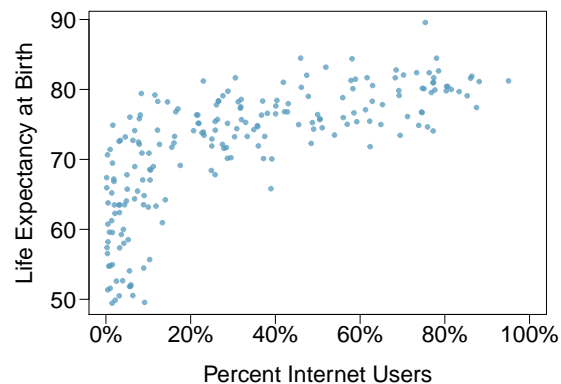
- What type of study is this?
- Suggest a sampling strategy for carrying out this study.

1.20 Housing proposal across dorms. On a large college campus first-year students and sophomores live in dorms located on the eastern part of the campus and juniors and seniors live in dorms located on the western part of the campus. Suppose you want to collect student opinions on a new housing structure the college administration is proposing and you want to make sure your survey equally represents opinions from students from all years.

- What type of study is this?
- Suggest a sampling strategy for carrying out this study.

1.21 Internet use and life expectancy. The following scatterplot was created as part of a study evaluating the relationship between estimated life expectancy at birth (as of 2014) and percentage of internet users (as of 2009) in 208 countries for which such data were available.²³

- Describe the relationship between life expectancy and percentage of internet users.
- What type of study is this?
- State a possible confounding variable that might explain this relationship and describe its potential effect.



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.

1.23 Evaluate sampling methods. A university wants to determine what fraction of its undergraduate student body support a new \$25 annual fee to improve the student union. For each proposed method below, indicate whether the method is reasonable or not.

- Survey a simple random sample of 500 students.
- Stratify students by their field of study, then sample 10% of students from each stratum.
- Cluster students by their ages (e.g. 18 years old in one cluster, 19 years old in one cluster, etc.), then randomly sample three clusters and survey all students in those clusters.

²³CIA Factbook, Country Comparisons, 2014.

1.24 Random digit dialing. The Gallup Poll uses a procedure called random digit dialing, which creates phone numbers based on a list of all area codes in America in conjunction with the associated number of residential households in each area code. Give a possible reason the Gallup Poll chooses to use random digit dialing instead of picking phone numbers from the phone book.

1.25 Haters are gonna hate, study confirms. A study published in the *Journal of Personality and Social Psychology* asked a group of 200 randomly sampled men and women to evaluate how they felt about various subjects, such as camping, health care, architecture, taxidermy, crossword puzzles, and Japan in order to measure their attitude towards mostly independent stimuli. Then, they presented the participants with information about a new product: a microwave oven. This microwave oven does not exist, but the participants didn't know this, and were given three positive and three negative fake reviews. People who reacted positively to the subjects on the dispositional attitude measurement also tended to react positively to the microwave oven, and those who reacted negatively tended to react negatively to it. Researchers concluded that "some people tend to like things, whereas others tend to dislike things, and a more thorough understanding of this tendency will lead to a more thorough understanding of the psychology of attitudes."²⁴

- What are the cases?
- What is (are) the response variable(s) in this study?
- What is (are) the explanatory variable(s) in this study?
- Does the study employ random sampling?
- Is this an observational study or an experiment? Explain your reasoning.
- Can we establish a causal link between the explanatory and response variables?
- Can the results of the study be generalized to the population at large?

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?

1.27 Sampling strategies. A statistics student who is curious about the relationship between the amount of time students spend on social networking sites and their performance at school decides to conduct a survey. Various research strategies for collecting data are described below. In each, name the sampling method proposed and any bias you might expect.

- He randomly samples 40 students from the study's population, gives them the survey, asks them to fill it out and bring it back the next day.
- He gives out the survey only to his friends, making sure each one of them fills out the survey.
- He posts a link to an online survey on Facebook and asks his friends to fill out the survey.
- He randomly samples 5 classes and asks a random sample of students from those classes to fill out the survey.

1.28 Reading the paper. Below are excerpts from two articles published in the *NY Times*:

- An article titled *Risks: Smokers Found More Prone to Dementia* states the following:²⁵

"Researchers analyzed data from 23,123 health plan members who participated in a voluntary exam and health behavior survey from 1978 to 1985, when they were 50-60 years old. 23 years later, about 25% of the group had dementia, including 1,136 with Alzheimer's disease and 416 with vascular dementia. After adjusting for other factors, the researchers concluded that pack-a-day smokers were 37% more likely than nonsmokers to develop dementia, and the risks went up with increased smoking; 44% for one to two packs a day; and twice the risk for more than two packs."

Based on this study, can we conclude that smoking causes dementia later in life? Explain your reasoning.
- Another article titled *The School Bully Is Sleepy* states the following:²⁶

"The University of Michigan study, collected survey data from parents on each child's sleep habits and asked both parents and teachers to assess behavioral concerns. About a third of the students studied were identified by parents or teachers as having problems with disruptive behavior or bullying. The researchers found that children who had behavioral issues and those who were identified as bullies were twice as likely to have shown symptoms of sleep disorders."

A friend of yours who read the article says, "The study shows that sleep disorders lead to bullying in school children." Is this statement justified? If not, how best can you describe the conclusion that can be drawn from this study?

²⁴Justin Hepler and Dolores Albarracín. "Attitudes without objects - Evidence for a dispositional attitude, its measurement, and its consequences". In: *Journal of personality and social psychology* 104.6 (2013), p. 1060.

²⁵R.C. Rabin. "Risks: Smokers Found More Prone to Dementia". In: *New York Times* (2010).

²⁶T. Parker-Pope. "The School Bully Is Sleepy". In: *New York Times* (2011).

Exercises

1.29 Light and exam performance. A study is designed to test the effect of light level on exam performance of students. The researcher believes that light levels might have different effects on males and females, so wants to make sure both are equally represented in each treatment. The treatments are fluorescent overhead lighting, yellow overhead lighting, no overhead lighting (only desk lamps).

- What is the response variable?
- What is the explanatory variable? What are its levels?
- What is the blocking variable? What are its levels?

1.30 Vitamin supplements. To assess the effectiveness of taking large doses of vitamin C in reducing the duration of the common cold, researchers recruited 400 healthy volunteers from staff and students at a university. A quarter of the patients were assigned a placebo, and the rest were evenly divided between 1g Vitamin C, 3g Vitamin C, or 3g Vitamin C plus additives to be taken at onset of a cold for the following two days. All tablets had identical appearance and packaging. The nurses who handed the prescribed pills to the patients knew which patient received which treatment, but the researchers assessing the patients when they were sick did not. No significant differences were observed in any measure of cold duration or severity between the four groups, and the placebo group had the shortest duration of symptoms.³²

- Was this an experiment or an observational study? Why?
- What are the explanatory and response variables in this study?
- Were the patients blinded to their treatment?
- Was this study double-blind?
- Participants are ultimately able to choose whether or not to use the pills prescribed to them. We might expect that not all of them will adhere and take their pills. Does this introduce a confounding variable to the study? Explain your reasoning.

1.31 Light, noise, and exam performance. A study is designed to test the effect of light level and noise level on exam performance of students. The researcher believes that light and noise levels might have different effects on males and females, so wants to make sure both are equally represented in each treatment. The light treatments considered are fluorescent overhead lighting, yellow overhead lighting, no overhead lighting (only desk lamps). The noise treatments considered are no noise, construction noise, and human chatter noise.

- What type of study is this?
- How many factors are considered in this study? Identify them, and describe their levels.
- What is the role of the sex variable in this study?

1.32 Music and learning. You would like to conduct an experiment in class to see if students learn better if they study without any music, with music that has no lyrics (instrumental), or with music that has lyrics. Briefly outline a design for this study.

1.33 Soda preference. You would like to conduct an experiment in class to see if your classmates prefer the taste of regular Coke or Diet Coke. Briefly outline a design for this study.

1.34 Exercise and mental health. A researcher is interested in the effects of exercise on mental health and he proposes the following study: Use stratified random sampling to ensure representative proportions of 18-30, 31-40 and 41- 55 year olds from the population. Next, randomly assign half the subjects from each age group to exercise twice a week, and instruct the rest not to exercise. Conduct a mental health exam at the beginning and at the end of the study, and compare the results.

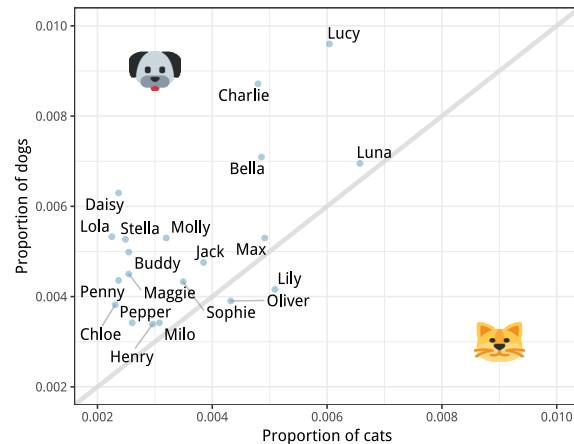
- What type of study is this?
- What are the treatment and control groups in this study?
- Does this study make use of blocking? If so, what is the blocking variable?
- Does this study make use of blinding?
- Comment on whether or not the results of the study can be used to establish a causal relationship between exercise and mental health, and indicate whether or not the conclusions can be generalized to the population at large.
- Suppose you are given the task of determining if this proposed study should get funding. Would you have any reservations about the study proposal?

³²C. Audera et al. "Mega-dose vitamin C in treatment of the common cold: a randomised controlled trial". In: *Medical Journal of Australia* 175.7 (2001), pp. 359–362.

Chapter exercises

1.35 Pet names. The city of Seattle, WA has an open data portal that includes pets registered in the city. For each registered pet, we have information on the pet's name and species. The following visualization plots the proportion of dogs with a given name versus the proportion of cats with the same name. The 20 most common cat and dog names are displayed. The diagonal line on the plot is the $x = y$ line; if a name appeared on this line, the name's popularity would be exactly the same for dogs and cats.

- Are these data collected as part of an experiment or an observational study?
- What is the most common dog name? What is the most common cat name?
- What names are more common for cats than dogs?
- Is the relationship between the two variables positive or negative? What does this mean in context of the data?



1.36 Stressed out, Part II. In a study evaluating the relationship between stress and muscle cramps, half the subjects are randomly assigned to be exposed to increased stress by being placed into an elevator that falls rapidly and stops abruptly and the other half are left at no or baseline stress.

- What type of study is this?
- Can this study be used to conclude a causal relationship between increased stress and muscle cramps?

1.37 Chia seeds and weight loss. Chia Pets – those terra-cotta figurines that sprout fuzzy green hair – made the chia plant a household name. But chia has gained an entirely new reputation as a diet supplement. In one 2009 study, a team of researchers recruited 38 men and divided them randomly into two groups: treatment or control. They also recruited 38 women, and they randomly placed half of these participants into the treatment group and the other half into the control group. One group was given 25 grams of chia seeds twice a day, and the other was given a placebo. The subjects volunteered to be a part of the study. After 12 weeks, the scientists found no significant difference between the groups in appetite or weight loss.³³

- What type of study is this?
- What are the experimental and control treatments in this study?
- Has blocking been used in this study? If so, what is the blocking variable?
- Has blinding been used in this study?
- Comment on whether or not we can make a causal statement, and indicate whether or not we can generalize the conclusion to the population at large.

1.38 City council survey. A city council has requested a household survey be conducted in a suburban area of their city. The area is broken into many distinct and unique neighborhoods, some including large homes, some with only apartments, and others a diverse mixture of housing structures. For each part below, identify the sampling methods described, and describe the statistical pros and cons of the method in the city's context.

- Randomly sample 200 households from the city.
- Divide the city into 20 neighborhoods, and sample 10 households from each neighborhood.
- Divide the city into 20 neighborhoods, randomly sample 3 neighborhoods, and then sample all households from those 3 neighborhoods.
- Divide the city into 20 neighborhoods, randomly sample 8 neighborhoods, and then randomly sample 50 households from those neighborhoods.
- Sample the 200 households closest to the city council offices.

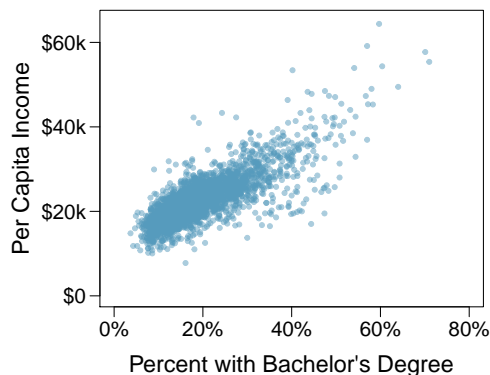
³³D.C. Nieman et al. "Chia seed does not promote weight loss or alter disease risk factors in overweight adults". In: *Nutrition Research* 29.6 (2009), pp. 414–418.

1.39 Flawed reasoning. Identify the flaw(s) in reasoning in the following scenarios. Explain what the individuals in the study should have done differently if they wanted to make such strong conclusions.

- Students at an elementary school are given a questionnaire that they are asked to return after their parents have completed it. One of the questions asked is, “Do you find that your work schedule makes it difficult for you to spend time with your kids after school?” Of the parents who replied, 85% said “no”. Based on these results, the school officials conclude that a great majority of the parents have no difficulty spending time with their kids after school.
- A survey is conducted on a simple random sample of 1,000 women who recently gave birth, asking them about whether or not they smoked during pregnancy. A follow-up survey asking if the children have respiratory problems is conducted 3 years later. However, only 567 of these women are reached at the same address. The researcher reports that these 567 women are representative of all mothers.
- An orthopedist administers a questionnaire to 30 of his patients who do not have any joint problems and finds that 20 of them regularly go running. He concludes that running decreases the risk of joint problems.

1.40 Income and education in US counties. The scatterplot below shows the relationship between per capita income (in thousands of dollars) and percent of population with a bachelor’s degree in 3,143 counties in the US in 2010.

- What are the explanatory and response variables?
- Describe the relationship between the two variables. Make sure to discuss unusual observations, if any.
- Can we conclude that having a bachelor’s degree increases one’s income?



1.41 Eat better, feel better? In a public health study on the effects of consumption of fruits and vegetables on psychological well-being in young adults, participants were randomly assigned to three groups: (1) diet-as-usual, (2) an ecological momentary intervention involving text message reminders to increase their fruits and vegetable consumption plus a voucher to purchase them, or (3) a fruit and vegetable intervention in which participants were given two additional daily servings of fresh fruits and vegetables to consume on top of their normal diet. Participants were asked to take a nightly survey on their smartphones. Participants were student volunteers at the University of Otago, New Zealand. At the end of the 14-day study, only participants in the third group showed improvements to their psychological well-being across the 14-days relative to the other groups.³⁴

- What type of study is this?
- Identify the explanatory and response variables.
- Comment on whether the results of the study can be generalized to the population.
- Comment on whether the results of the study can be used to establish causal relationships.
- A newspaper article reporting on the study states, “The results of this study provide proof that giving young adults fresh fruits and vegetables to eat can have psychological benefits, even over a brief period of time.” How would you suggest revising this statement so that it can be supported by the study?

³⁴Tamlin S Conner et al. “Let them eat fruit! The effect of fruit and vegetable consumption on psychological well-being in young adults: A randomized controlled trial”. In: *PLoS one* 12.2 (2017), e0171206.