1. 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.

   (a) 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.

   (b) 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.

   (c) 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.

2. A company considers candidates from two recruiting agencies: 60% of candidates come from Agency 1, 40% from Agency 2. 20% of candidates from Agency 1 are hired and only 10% from Agency 2 are hired.

   (a) What is the probability that a randomly selected candidate is hired? You should explicitly define events for me (e.g. $(A = ..., B = ...)$, then use probability notation to solve this problem.

   (b) What is the probability that someone is not hired?

   (c) If I know someone was hired, what's the probability they came from Agency 1?

3. 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."

   (a) What are the cases?

   (b) What is (are) the response variable(s) in this study?

   (c) What is (are) the explanatory variable(s) in this study?

   (d) Does the study employ random sampling?

   (e) Is this an observational study or an experiment? Explain your reasoning.

   (f) Can we establish a causal link between the explanatory and response variables?

   (g) Can the results of the study be generalized to the target population at large?

4. Suppose we want to construct a probability distribution table for household income in the United States. Only one of the following (a) - (d) could be correct. Which one must it be? What is wrong with the others?

   (a)

Income Range	$0-25k	$25k-50k	$50k-100k	$100k+
Probability	0.38	-0.27	0.52	0.37

   (b)

Income Range	$0-25k	$25k-50k	$50k-100k	$100k+
Probability	0.18	0.39	0.33	0.16

   (c)

Income Range	$0-25k	$25k-50k	$50k-100k	$100k+
Probability	0.28	0.27	0.29	0.16

   (d)

| Income Range	$0-25k	$24k-50k	$50k-100k	$100k+
Probability	0.28	0.27	0.29	0.16

5. The following summary table shows the number of space launches in the US by the type of launching agency and the outcome of the launch (success or failure).

	1957-1999		2000-2018	
	Failure	Success	Failure	Success
Private	13	295	10	562
State	281	3751	33	711
Startup	-	-	5	65

   (a) What variables were collected on each launch in order to create to the summary table above?

(b) State whether each variable is numerical or categorical. If numerical, state whether it is continuous or discrete. If categorical, state whether it is ordinal or not.

(c) Suppose we wanted to study how the success rate of launches vary between launching agencies and over time. In this analysis, which variable would be the response variable and which variable would be the explanatory variable?

6. In Spring 2024, Prof. Tang asked her STAT 311 students to fill out a mid-semester survey. She was particularly interested in the the amount of hours her STAT 311 students were spending per week on the course. The average time spent on the course was found to be 10.2 hours per week.

(a) Based on this information, identify each of the following: observation, variable, parameter, and statistic.

(b) Was this survey a census or (just) a sample? Why?

7. 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?

8. Chia seeds have gained reputation as a diet supplement. In one 2009 study, 38 men and 38 women were recruited (i.e. specifically chosen) and then divided randomly into two groups: treatment or control. One group was given 25 grams of chia seeds twice a day to consume, and the other was given a placebo. The seeds and placebo were designed to look the same. The subjects volunteered to be a part of the study. After 12 weeks, the scientists found no statistically discernible difference between the groups in appetite or weight loss.

(a) What type of study is this?

(b) What are the experimental and control treatments in this study?

(c) Has blocking been used in this study? If so, what is the blocking variable?

(d) Is there any blinding in the study?

(e) Comment on whether we can make a causal statement, and indicate whether we can generalize the conclusions from this study to the population at large.

9. After an introductory statistics course, 80% of students can successfully construct box plots. Of those who can construct box plots, 86% passed, while only 65% of those students who could not construct box plots passed. Calculate the probability that a student is able to construct a box plot if it is known that they passed. You can draw a tree diagram if that is helpful for you!

10. 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 histogram below shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014. In particular, this is a *relative frequency* histogram, which shows proportions instead of raw counts on the y-axis:
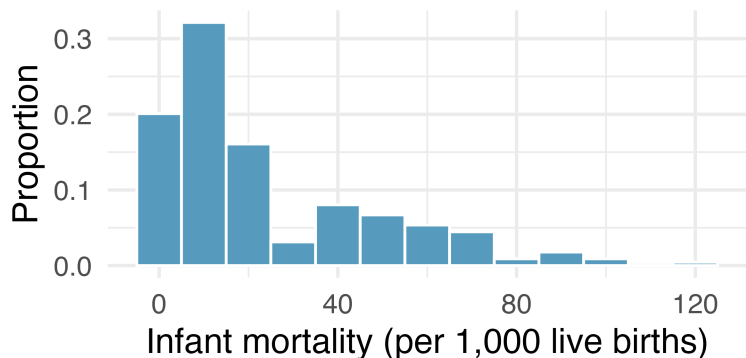
Estimate Q_1, the median m, and Q_3 from the histogram.

11. 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 the following data on eye colors of 204 Scandinavian men and their female partners. For simplicity, we only include heterosexual relationships in this exercise.

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

(a) Find the probability that a randomly chosen male respondent or his partner has blue eyes.

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?

(c) 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?

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.