

# Homework 2

## DATA 202 - Alexander - Fall 2023

Please submit **Homework 2** responses as a .pdf file on Canvas [here](#).

### Exercise 1.1

In the following questions, read each scenario and (a) describe the sampling method used and (b) determine whether the sampling method appears to be sound or flawed. Explain your reasoning in complete sentences.

- (Study hours): A group of students decides to collect data on the number of hours students in a university spend in the library per week. The researchers collect data by setting up a table outside of the library entrance.
- (Clinical trials): Researchers at a lab conduct a wide variety of clinical trials by using subjects who volunteer after reading advertisements hung on boards and light poles soliciting paid volunteers to participate in the study.
- (Covid-19): In an online survey with a sample of 957 subjects, the following question was posed on both Instagram and Twitter: “In your view, is the Covid-19 vaccine safe?” The survey respondents were internet users who chose to respond to the question posted on the social media accounts over the course of 24 hours.
- (Community data): A group of researchers decides to partner with a major company to examine environmental hazards at the neighborhood level. They code each region by zip code and randomly select zip codes in the metropolitan region. The researchers then plan to randomly select households within each of the selected zip codes.
- (Debt): In a survey of hospital workers, a total of 2,087 respondents were randomly selected and asked how much credit card debt they pay off each month. Survey results were used to generate population parameters.

### Exercise 1.2

In the following exercises, explain the issue with the study and sampling method. Use complete sentences.

- (Political party): In a research study conducted by a political party at their annual rally, a convenience sample of 1000 adults were asked to select their favorite political party, the favorite choice was the political party in question, which was selected by 92% of respondents.
- (Marijuana): Proponents of the legalization of marijuana in their state collected data using an electronic poll in various CBD and vape shops across the city, showing that 65% of those surveyed said that they “strongly agree” and 15% said that they “agree” with the legalization of marijuana.
- (Police Training Facility): A group study citizen beliefs about the “Cop City” facility in Atlanta collected data from 367 individuals at four city council meetings. In a report on their findings, in which they describe the use of advanced statistical methods, they state: “a majority of Atlanta citizens support the development of the training facility.”

### Exercise 1.3

(Discrete vs. continuous data): Identify which of the following is discrete vs continuous.

- The number of people surveyed in a national election poll
- The exact height of a random sample of students in a statistics course
- The exact times that drivers spend texting while driving over a 7 day period.
- The number of animals observed in a reserve on a given day.
- The temperature recorded by the National Weather Service.

### Exercise 1.4

(Levels of measurement): Identify the level of measurement for the variables below.

- College rankings in the U.S. News and World Report.
- Exit poll results for a presidential election where respondents were asked to identify political affiliation.
- The colors of shirts (e.g., red, green, blue, etc.) worn by a group of students listed in a data set.
- The amount of a virus in a sample of blood collected in a medical study.

### Exercise 1.5

In your own words, define the term “social justice” and describe how statistics can be used to support and advocate for social movement building around issues of injustice.

---

### For Exercises 1.6 through 1.10, you should complete all calculations in R.

Let  $P$  be a sample of payments (in thousands of dollars) for residents of a small rural community. These payments represent a random sample of 140 payments being made through the local city council. Payments are restitution for an environmental hazard. The hazard was the result of a major corporation’s new factory construction, which forced many of the town’s residents to relocate.

$$P = \{25.4, 27.6, 19.7, 18.1, 18.7, 65.6, 20.0, 21.7, 39.6, 17.2, 34.5, 32.7, 92.7, 12.3\}$$

### Exercise 1.6

Calculate the measures of center for  $P$ .

### Exercise 1.7

Calculate the measures of variation for  $P$ .

### Exercise 1.8

Calculate the IQR for  $P$  and the z-scores for the minimum and maximum values.

### Exercise 1.9

What do the descriptive statistics (measures of center, measures of variation, and measures of relative standing) on the sample data in  $P$  tell us about the payments?

### Exercise 1.10

Generate a descriptive and appropriate plot for the data in the set  $P$ , include labels.