SAMPLING HOMEWORK

TEXT: 1.2

LAST NAME	FIRST NAME	DATE

Determine the sampling method used in the following scenarios: *simple random*, *systematic*, *cluster*, *stratified*, or *convenience*. Identify and describe possible sources of bias in each sampling scenario.

1. The instructor interested in examining the population of STEM students in California takes her sample by gathering data on five randomly selected students from each Lake Tahoe Community College math class.

2. A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Jose. The first house in the neighborhood around the park was selected randomly and then every eighth house in the neighborhood around the park was interviewed.

Determine the sampling method used in the following scenarios: *simple random*, *systematic*, *cluster*, *stratified*, or *convenience*. Identify and describe possible sources of bias in each sampling scenario.

3. A woman in the airport is handing out questionnaires to travelers asking them to evaluate the airport's service. She does not ask travelers who are hurrying through the airport with their hands full of luggage, but instead asks all travelers who are sitting near gates and not taking naps while they wait.

4. The marketing manager for an electronics chain store wants information about the ages of its customers. Over the next two weeks, at each store location, 100 randomly selected customers are given questionnaires to fill out asking for information about age, as well as about other variables of interest.

Determine the sampling method used in the following scenarios: *simple random*, *systematic*, *cluster*, *stratified*, or *convenience*. Identify and describe possible sources of bias in each sampling scenario.

5. A teacher wants to know if her students are doing homework, so she randomly selects rows two and five and then calls on all students in row two and all students in row five to present the solutions to homework problems to the class.

6. A political party wants to know the reaction of voters to a debate between the candidates. The day after the debate, the party's polling staff calls 1,200 randomly selected phone numbers. If a registered voter answers the phone or is available to come to the phone, that registered voter is asked whom they intend to vote for and whether the debate changed their opinion of the candidates.

Determine the sampling method used in the following scenarios: *simple random*, *systematic*, *cluster*, *stratified*, or *convenience*. Identify and describe possible sources of bias in each sampling scenario.

7. The librarian at a public library wants to determine what proportion of the library users are children. The librarian has a tally sheet on which she marks whether books are checked out by an adult or a child. She records this data for every fourth patron who checks out books.

8. In order to maintain quality control, a pastry baker wants to randomly sample one mini-cupcake from each baking tray throughout his shift.

Describe the sampling techniques used at each stage of the following example of multistage sampling. Identify and describe possible sources of bias in each sampling scenario.

9. Household surveys conducted by the Australian Bureau of Statistics begin by dividing metropolitan regions into *collection districts* and selecting some of these districts. The selected collection districts are then divided into *blocks*, and a few blocks are chosen from within each selected collection district. Finally, *dwellings* are listed within each selected block, and some of these dwellings are selected. This method is preferable because it makes it unnecessary to create a list of every dwelling in the region.

10. A researcher working for the office of the city planning wants to collect some data about the public transit travel times over the course of the year. She selects 21 calendar dates spanning the whole year, with each pair of consecutive dates being 17 days apart. On each one of those dates, she selects a few bus routes at random, and has her assistants go out for a ride and log the running times for every 5th bus on each route.

For each of the following scenarios, describe probable sources of bias, and what kind of individuals are more likely than others to be chosen.

11. A study of stress levels among California Community College students is conducted on a sample of students randomly approached on a given day at CRC Library and Math Center.

12. A survey of Californians' online shopping habits is conducted on a sample of system**76** customers.

For each of the following scenarios, describe probable sources of bias, and what kind of individuals are more likely than others to be chosen.

13. A research project on American consumer preferences for a new car brand collects data from a predetermined number of South Sacramento residents.

14. A study on the experiences of remote workers in USA starts with a few call answering workers, and relies on them to refer their colleagues.

Use a paragraph or so to describe a sampling procedure you could use in real life to collect a specified sample. Try to maximize the randomness of your sample while keeping the procedure cheap and realistic. Discuss probable ways in which your sampling procedure is biased.

15. Take a sample of 100 or so cars out of the population of cars parked in your town on a given day.

16. Take a sample of 20 or so people out of the population of all the people you know and communicate with in-person or online.

Write an R command to generate a random sample with specified properties:

17. Three distinct days of the week out of the seven: `Mo, Tu, We, Th, Fr, Sa, Su`.

18. 20 pets, with possible repetitions, with possible choices being Dog, Cat, Fish, and Bunny.

19. 13 random rational numbers between 0 and 120.

20. One random integer between 0 and 120.

21. 45 directions, with possible repetitions, with possible choices being Left, Right, and Up.

22. Two distinct subjects of study, chosen among Math, English, History, and Music.

23. Ten distinct integers between 2000 and 2030.

24. A sample of random size between 5 and 10 consisting of random rational numbers between 3 and 6.

Answers

1. cluster

3. convenience

5. cluster

7. systematic

17. `sample(c("Mo", "Tu", "We", "Th", "Fr", "Sa", "Su"), 3)`

19. `runif(13, 0, 120)`

21. `sample(c("Left", "Right", "Up"), 45, replace=T)`

23. `sample(2000:2030, 10)`