## 1.2 Data, Sampling, and Variation in Data and Sampling

Definition	Examples
Qualitative data are data that can be characterized by certain non-numerical attributes. The data cannot be ordered or ranked.	• Eye color • Gender • Sport • Ethnicity

		Examples
Quantitative data are data that can be measured numerically and can be ordered or ranked.	*Discrete data* are data that can be counted.	• Number of children in a family • Number of students in a classroom
	Continuous data are data that can take on all real number values of a given interval.	• Height • Weight • Temperature

The diagram to the right is a visual aid to help you understand the hierarchy of the data types above.
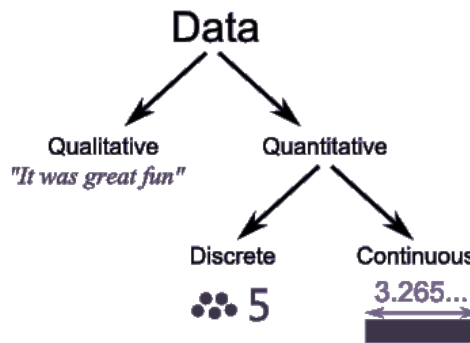
Note that continuous data must be rounded to a desired place value due to limitations in measuring devices such as scales, rulers, and thermometers. For example, a student grade average reported as 80% could actually be any score from 79.5% up to but not including 80.5%.

Demonstration Problems	*Practice Problems*
1. (a) Classify as qualitative or quantitative: (*i*) car model (*ii*) car cost **2. (a)** Classify as discrete or continuous (*i*) number of doors on a car (*ii*) number of gallons to fill the gas tank of a car	**1. (b)** Classify as qualitative or quantitative: (*i*) pine tree species (*ii*) pine tree age **2. (b)** Classify as discrete or continuous (*i*) number of pine needles in a cluster (*ii*) pine tree height
	Answers: **1. (b)** (*i*) qualitative; (*ii*) quantitative **2. (b)** (*i*) discrete; (*ii*) continuous.

Consider the following queries:

- An insurance company wants to know what percent of all adults over age 50 have high blood pressure.
- A libertarian candidate wants to know what percent of registered libertarians support the Affordable Health Care Act.
- A casino pit boss wants to know if a certain six-sided die is fair.

To answer these questions, a statistical study must be designed and performed. In designing a study, the method of collecting data and sampling technique must be determined.

Data Collection

	Advantages	Disadvantages
Telephone survey	• Low cost. • People may be more candid in their responses as there is no face-to-face contact.	• Some people do not have phones. • Some people may not answer when called. • Obtaining phone numbers may be difficult.
Mailed questionnaire	• Can cover a wide geographic area.	• People may not respond. • People may answer questions inappropriately. • Some people may have difficulty understanding the questions.
Online survey	• Low cost • Can cover a wide geographic area. • Can collect data from a large number of respondents. • Good statistical significance.	• Respondents may not feel comfortable providing honest, accurate answers. • People may feel too rushed to take time to complete a survey.
Personal interview	• Can obtain in-depth responses to questions.	• High cost due to necessary training and wages of interviewers. • Interviewer may be biased in selection of respondents.
Survey of records	• Ongoing records may be readily available. • Recorded data may contain multi-faceted information providing data for multiple studies.	• Records may be inaccessible. • Records may contain errors. • Can be costly and time consuming to examine large quantities of records.
Direct observation	• Data is accurate. • Data is current.	• Can be costly to observe a large quantity of data.

Sampling Techniques

	Method	Example
Simple Random Sampling	A sample of n elements is chosen from a population of N elements in which each of the possible samples has the same chance of selection.	• 10 computer generated random numbers between 1 and 100 are used to determine which passengers will be selected for an additional pre-boarding screening.
Systematic Sampling	The population is arranged according to a chosen order, then selects every k^{th} element starting at a randomly chosen element from $1 - k$.	• The 27th, 47th, 67th, 87th, and 107th passengers are selected for additional pre-boarding screening.
Stratified Sampling	A simple random sample is selected from each of a given number of subpopulations, or strata.	• From 30 women and 70 men, 20 women and 20 men are selected randomly to form a potential pool of jurors.
Cluster Sampling	A simple random sample of clusters is chosen from the available clusters in the population	• 10 elementary schools are chosen randomly in Shasta County and all of the teacher salaries of the selected schools are recorded.
Convenience Sampling	A selection is obtained from the population by any easy and simple method without random selection.	• A survey is given to students in this class to determine how many own a smartphone.

Demonstration Problems	*Practice Problems*
Define a population and the method of data collection that may have been used.	Define a population and the method of data collection that may have been used.
3. (a) According to the U. S. Census of 2010, 50.8% of all Americans are female.	**3. (b)** California demographers assert that 39% of all Californians are Latino.
4. (a) The average rainfall in Tehama County is 28.8 inches.	**4. (b)** The 2014 Sierra snowpack was 18% of average.
5. (a) The National Marines Fishery Service predicts that 634,650 fall-run king salmon from the Sacramento River will be out in the ocean this year.	**5. (b)** Each year the Coleman Fish Hatchery produces 12 million late fall salmon smolt.
Identify the sampling technique used.	Identify the sampling technique used.
6. (a) To determine how many bus riders are riding the bus to the Shasta College Tehama Campus, 10 riders were randomly selected and surveyed daily for the month of September.	**6. (b)** To inspect the cleanliness of patients beds in a hospital, an inspector numbers the beds, chooses a beginning number randomly, then inspect every 10th bed starting from the randomly chosen bed.
7. (a) Using employment records of state employees, 100 employees were chosen randomly from declared Hispanic Americans, 100 from declared African Americans, and 100 from declared Asian Americans to determine their average salaries.	**7. (b)** Ten cartons of eggs are chosen randomly from a truck load. All of the eggs from each carton are tested for traces of salmonella.

Answers: **3. (b)** Population: all Californians, data collection: mailed questionnaire;   **4. (b)** Population: all recorded history of annual snowpack of the Sierras, data collection: survey of records; **5. (b)** Population: all fish produced annually by Coleman Fish Hatchery, data collection: observation; **6. (b)** Systematic sampling; **7. (b)** Cluster sampling