2.2 Characteristics of Data

"Data! Data! Data!" he cried impatiently.
"I can't make bricks without clay."

— Sherlock Holmes

data—a body or series of facts or information

It is evident from the definition that virtually anything can qualify as **data**. Although some data requires an immediate response and relatively little analysis (for example, the data your nervous system transmits to your brain after touching a hot stove), other data, specifically research data, takes years to collect and requires detailed analysis to understand it.

In this section, you will learn to refine your thesis question in order to collect specific kinds of information. You will then adapt your thesis question to gather data from a specific group.

POPULATION DATA VERSUS SAMPLE DATA

population—a group of individuals that is the focus of a study

sample—a selection of individuals taken from a population

A project that considers the attitudes and characteristics of students at your school is using the student body as the **population**. The collection of student records your school has is an example of population data. If you were to forecast the results of a federal election, the population would be the list of eligible voters. For most studies, it is impractical to collect data from the entire population because there are too many people. Instead, you would select a representative **sample** and study it instead.

When carrying out a study in a high school, the permission of the teachers involved and the principal is generally required.

Example 1 Identifying the Population

Suppose you had observed that the heights of Grade 9 students were quite varied. By the time these students reach Grade 12, of course, most are taller. State a thesis question that gives your study a focus and identifies the population.

Solution

Your first attempt at a thesis question might be *What is the pattern of growth (height) of students from the time they enter high school until they graduate?*

This is a good start, but it is not clear precisely who is the target of the study. Are any students to be considered, or only the students in their first and last year of high school? Are females and males treated alike for the purpose of this study? Can an **inference** be made about growth patterns of high school students beyond the borders of your school community?

inference—a conclusion about the population that is made from the sample

Project Connection
Be careful not to define a project that is too broad in scope.

The population needs to be defined more clearly in this question, and it would be wise to focus on only the students who are attending your high school. Also, you should define two populations of subjects: females and males. Now, restate your research question: *What is the pattern of growth (height) of females and males at my school from the time they enter high school until they graduate?*

CROSS-SECTIONAL AND LONGITUDINAL STUDIES

Before you can decide whether you will conduct a **census** or select a sample, you must determine what kind of study you wish to perform.

A **cross-sectional study** is carried out during a specific timeframe and focuses on a wide range of individuals. A **longitudinal study** typically focuses on a small group of individuals over an extended period. Often, research to determine the effects of a treatment for cancer is a longitudinal study that may extend over 20 years or more!

census—information gathered about an entire population

cross-sectional study—a study that considers individuals from different groups at the same time

longitudinal study—a study that considers individuals over a long period of time

Example 2 Identifying the Type of Study

The need for a better student centre has always been an issue in your school. Identify a thesis question for this topic that clearly defines the population, and then suggest a method of study.

Solution

The first draft of your thesis question is *How do the opinions about the student centre change among students from Grade 9 to Grade 12?*

You have clearly identified the population for your study as only the students who attend your school; however, it is unclear what type of study you wish to undertake. Do you want to ask students from each grade (a cross-sectional study), or do you intend to interview a selection of Grade 9 students and then return to ask them again next year when they are in Grade 10, and so on (a longitudinal study)?

A longitudinal study allows you to analyze the way a variable like a person's opinion changes over time. This form of data is called **time series data**. Longitudinal studies are usually expensive and can be very difficult to undertake. Despite these challenges, it is the method of choice for psychological and medical studies.

time series data—data that have accumulated over a long period of time

Since you are in Grade 12, it is impractical for you to return to your school for the next three years to collect data for a longitudinal study (also, your teacher is unlikely to approve such a study). A cross-sectional study is quicker, easier, and more suitable in this situation.

When outlining the type of study, you should also mention whether you plan to collect sample data or population data by taking a census. Since it is unlikely that you can interview all the students in your school, you must focus on a random sample of the students instead. You will learn more about sampling in Section 2.3.

A second draft of your thesis question is *How do the opinions about the student centre among a random sample of students in Grades 9 and 12 differ?*

QUALITATIVE AND QUANTITATIVE VARIABLES

quantitative variables—variables that can be measured numerically

qualitative variables—variables that cannot be measured numerically

discrete data—data that can be described using whole numbers. A count will always give discrete data.

continuous data—data that are only measurable with real numbers. A measure of quantity will always be continuous.

Variables that can be measured numerically are called **quantitative** (e.g., height, distance); variables that cannot be measured numerically are called **qualitative** (e.g., eye colour, opinion). Quantitative data are said to be **discrete** if they can be described with whole numbers (e.g., number of students). Examples of quantitative data that are said to be **continuous** include the heights of students and the length of time a plant takes to germinate.

Example 3 Identifying Variables and Data Types

Imagine that, for years, you have been fascinated by the great performances of track-and-field athletes, and the high jump is your favourite event. Create a thesis question about this topic using your school's track team as the population. Identify the variables and data used.

Solution

Your first question might be *What attributes of an athlete's physique, training regimen, and competition history are predictors of success?*

In your question, you still need to state clearly what variables you will study. Perhaps part of your study will require you to compare quantitative data. A question might be *Is there a relationship between a high jumper's height and best jump this season?*

Another part of your project could study qualitative data, such as using a questionnaire to determine a description of the mental and physical preparation prior to jumping in a competition. A question might be *Are there common training exercises among the top five high jumpers on our track team?*

Statisticians believe that although it is helpful to collect qualitative data, quantitative data are much easier to study.

KEY IDEAS

population versus sample—the group being studied is called the *population;* a selection of individuals taken from the population is a *sample*. Data collected from the sample are called *sample data*. A *census* is a collection of population data.

inference—a conclusion about the population based on sample data

cross-sectional study—a study that considers individuals from different groups at the same time

longitudinal study—a study of a single group (or sample) over a long period of time

time series data—data that have accumulated over a long period of time

qualitative variables versus quantitative variables—all data can be characterized as either qualitative or quantitative. Quantitative data are numerical and qualitative data are non-numerical.

discrete data—data that result from a count (e.g., number of people, number of vehicles, etc.)

continuous data—data resulting from the measure of a quantity (e.g., mass, age, etc.)

2.2 Exercises

A

1. Identify each of the following variables as qualitative or quantitative.

(a) age
(b) favourite meal
(c) television viewing preferences
(d) volume of a radio
(e) colour of hair
(f) fabric texture
(g) pH of water samples
(h) seating capacity
(i) grades
(j) paint colours

2. For each quantitative variable mentioned in Question 1, identify whether it is continuous or discrete.

3. **Knowledge and Understanding** Identify the variables and their types, as well as the population for the following thesis questions.

(a) Is there a relationship between weather conditions and absenteeism in Grade 9 at your school?
(b) Is there a profile that describes people who generally buy used cars in Canada?
(c) Is there a relationship between the amount of television watched and the level of physical fitness among adult females?
(d) Do Grade 9 students who regularly eat breakfast perform better academically?
(e) Are teenage drivers who have been issued speeding tickets more likely to be males?
(f) What home conditions influence school-aged children in selecting a future career?
(g) When is the best time of day to find a parking space within 100 m of the mall?
(h) How much of their own money do students at your school spend on their clothes?

B

4. For each of the thesis questions in Question 3, would you collect a sample or conduct a census? Would each question require a cross-sectional study or a longitudinal study?

5. For each of the following scenarios,
- **(i)** determine the population;
- **(ii)** identify the key variables for the study;
- **(iii)** state whether the data will be quantitative or qualitative; and
- **(iv)** for the variables that are quantitative, state whether the data will be discrete or continuous.

- **(a)** You must get T-shirt sizes for the 42 members of your school's environment club (26 are female).
- **(b)** You are to canvass 200 households to determine the level of support that each of the candidates in a local by-election has.
- **(c)** You are studying biological succession in what was 45 hectares of a farmer's cornfield. You are trying to measure plant diversity by identifying the number of each species per hectare.
- **(d)** This summer, you have been hired to work with anthropologists from the Royal Ontario Museum. You are to gather data from the Aboriginal population on Manitoulin Island, Ojibways of Lake Huron, on how their family structures have changed in the last century.
- **(e)** You are collecting and analyzing suggestions for a new name for your school. Data must be gathered from present and former students, students in the feeder schools, past and present teachers and administrators, support staff, parents and guardians, as well as interested members of the community.
- **(f)** Your teacher has arranged your class in groups of three, and asked you to gather data, analyze them, and communicate whether teenagers today are economically worse off than teenagers were 20 years ago.

6. For each of the scenarios in Question 5, complete the following.
- **(i)** Should a census or a sample be used? Explain.
- **(ii)** Would a cross-sectional or longitudinal study be most appropriate to draw conclusions? Explain.

7. Create a suitable thesis question for the following studies. Be sure to clearly identify the population in your question.
- **(a)** Customers leaving a local grocery store are asked how much they spent and how often they buy groceries.
- **(b)** A furniture store wishes to use existing data to determine trends in consumer buying habits over the last five years.
- **(c)** A company that sells books over the Internet will collect data for the next 12 months about those people who make online purchases.
- **(d)** A researcher from the Ontario Institute for Studies in Education at the University of Toronto wants to determine if calculator use in elementary school improves student confidence in doing math.
- **(e)** A medical officer of health needs to collect data about the frequency of senior citizen visits to doctors' offices over the last five years.
- **(f)** Domestic and foreign cars of various ages are tested to determine their minimum stopping distance when travelling at 90 km/h.

8. **Communication** Write a thesis question that a product manager might use to focus an analysis of consumer buying habits. Be sure to clearly identify the population in your question.

9. For each of the studies described in Question 7, complete the following.
(i) Is this study longitudinal or cross-sectional? Why?
(ii) Why would a sample be preferable to gathering census data for each?

10. A quality control officer at a manufacturing plant selects a number of integrated circuits to ensure that they meet company standards before they are shipped to customers.
(a) What is the population? **(b)** Describe the sample.

11. **Application** Using the Internet, find a recent study done about high school students.
(a) Identify the thesis question of the study.
(b) Was this study longitudinal or cross-sectional?
(c) Do you think the results of this study are reflective of the habits and attitudes of the students at your school? Explain.

12. Using the Internet or other media, find a recent longitudinal study.
(a) Identify the thesis question of the study.
(b) Why did the researcher choose to use a longitudinal study and not a cross-sectional study?

C

13. **Thinking, Inquiry, Problem Solving** Describe a topic for a statistical study for each scenario. Create a suitable thesis question for each.
(a) a longitudinal study that collects quantitative data from a sample of a population
(b) a cross-sectional study using a census that collects quantitative data
(c) a longitudinal census that collects qualitative data
(d) a cross-sectional study of a population sample that collects qualitative data

ADDITIONAL ACHIEVEMENT CHART QUESTIONS

14. **Knowledge and Understanding** Consider this thesis question: *In North America, do foreign cars depreciate in value faster than domestic cars?* Now answer the questions that follow.
(a) What is the population?
(b) What are the key variables that must be considered? Are these quantitative or qualitative?
(c) Should a census or a sample be used to collect the data?
(d) Are the data continuous or discrete?
(e) Is a cross-sectional or a longitudinal study more appropriate for drawing conclusions?

15. **Application** Find a recent study of consumer spending through online purchases on the Internet.

(a) Identify the thesis question or statement of the study.

(b) Identify whether the researcher used a cross-sectional or a longitudinal study.

(c) Are the results of this study reflective of the spending habits of your family and friends? Explain.

16. **Thinking, Inquiry, Problem Solving** Cross-sectional and longitudinal studies are often used in the field of medicine. Find an example of each from the field of medicine. Explain why both types of studies are necessary in medical research and identify the different types of information that each type of study can provide.

17. **Communication** Explain the differences between each pair of terms.

(a) population/sample

(b) cross-sectional study/longitudinal study

(c) quantitative variable/qualitative variable

(d) discrete data/continuous data

Chapter Problem

Mystery Most Mathematical—Part II

```
Subject:   Re: Puzzle - Part 2
Date:      Wednesday 17:27:11 -0700 (EDT)
From:      27182818@homework.com
To:        jto@coldmail.com

Use the research question that you designed after
receiving the first e-mail to
(a) identify the variables and indicate whether the
    variables are discrete or continuous
(b) identify the population
(c) suggest whether a cross-sectional or a
    longitudinal study would be more appropriate
    and why
```