

Lecture Notes & Examples Chapter 1 Introduction

Chapter 1 – Exploring Data

Introduction (pp. 2-7)

Statistics is the science of data. It is the practice or science of collecting and analyzing numerical data in large quantities, esp. for the purpose of inferring proportions in a whole from those in a representative sample.

- Collect information (data - sample)
- Analyze the information (compute statistics, make plots, etc)
- Make conclusions (infer characteristics of a population based upon a sample)

Statistics is “**customer driven**” – always a question to be answered.

Any set of data contains info on **individuals**. The characteristics of individuals are referred to as **variables**.

- **Individuals** are the objects described by a set of data. People, animals, things.
- **Variables** are characteristics of an individual. Can take on different values for different individuals.

Example – A high school’s student data base includes data about every currently enrolled student.

The individuals are the students described by the data set

The variables are age, gender, grade point average, homeroom, and grade level.

Whenever you receive data, ask:

- **Who** are the individuals described by the data? How many are there?
 - **What** are the variables? What units are involved?
 - We will eventually extend the questioning to **Why, when, where,** and **how** were the data produced?
-

Types of Variables

- **Categorical** –
- **Quantitative** –

Example – Table on page 3.

Province	Gender	Language		Height (cm)	Wrist circum. (mm)	Preferred communication
		spoken	Handed			
Saskatchewan	Male	1	Right	175	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178	174	Facebook
Ontario	Male	2	Right	169	160	Cell phone
Ontario	Female	2	Right	166	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text Messaging
Ontario	Female	2	Right	150.5	187	Text Messaging
Ontario	Female	1	Right	171	180	Text Messaging

- **Who?** 10 Canadian students who took the survey.
- **What** variables?
 - **Province**
 - **Gender**
 - **Dominant hand**
 - **Height**
 - **Wrist circum.**
 - **Preferred communication**

Describe the individual in the highlighted row?

When examining data sets we are going to be concerned about the **distribution** of the variables in the data set.

Distribution – tells us what values the variable takes on and how often it does so.

In Statistics we are going to be interested in drawing conclusions that go beyond the data at hand. This is called **inference** – the 3rd step in Statistics.

Homework: pp. 6-7 1, 3, 5, 7, 8 (Problems are attached)

1. **Protecting wood** How can we help wood surfaces resist weathering, especially when restoring historic wooden buildings? In a study of this question, researchers prepared wooden panels and then exposed them to the weather. Here are some of the variables recorded: type of wood (yellow poplar, pine, cedar); type of water repellent (solvent-based, water-based); paint thickness (millimeters); paint color (white, gray, light blue); weathering time (months). Identify each variable as categorical or quantitative.

3. **A class survey** Here is a small part of the data set that describes the students in an AP Statistics class. The data come from anonymous responses to a questionnaire filled out on the first day of class.

Gender	Hand	Height (in)	Homework time (min)	Favorite music	Pocket change (cents)
F	L	65	200	Hip-hop	50
M	L	72	30	Country	35
M	R	62	95	Rock	35
F	L	64	120	Alternative	0
M	R	63	220	Hip-hop	0
F	R	58	60	Alternative	76
F	R	67	150	Rock	215

- (a) What individuals does this data set describe?
- (b) Clearly identify each of the variables. Which are quantitative? In what units are they measured?
- (c) Describe the individual in the highlighted row.

5. **Ranking colleges** Popular magazines rank colleges and universities on their “academic quality” in serving undergraduate students. Describe two categorical variables and two quantitative variables that you might record for each student. Give the units of measurement for the quantitative variables.

Multiple choice: Select the best answer.

Exercises 7 and 8 refer to the following setting. At the Census Bureau Web site, you can view detailed data collected by the American Community Survey. The table below includes data for 10 people chosen at random from the more than one million people in households contacted by the survey. “School” gives the highest level of education completed.

Weight (lb)	Age (yr)	Travel to work (min)	School	Gender	Income last year (\$)
187	66	0	Ninth grade	1	24,000
158	66	n/a	High school grad	2	0
176	54	10	Assoc. degree	2	11,900
339	37	10	Assoc. degree	1	6,000
91	27	10	Some college	2	30,000
155	18	n/a	High school grad	2	0
213	38	15	Master's degree	2	125,000
194	40	0	High school grad	1	800
221	18	20	High school grad	1	2,500
193	11	n/a	Fifth grade	1	0

- 7.** The individuals in this data set are
- (a) households.
 - (b) people.
 - (c) adults.
 - (d) 120 variables.
 - (e) columns.
- 8.** This data set contains
- (a) 7 variables, 2 of which are categorical.
 - (b) 7 variables, 1 of which is categorical.
 - (c) 6 variables, 2 of which are categorical.
 - (d) 6 variables, 1 of which is categorical.
 - (e) None of these.