

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**.

↳ Do not have to be people

- **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.

From page 2 in book

Whenever you receive data, ask:

→ sample size?
population size?

- **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?

(W^SH)

Types of Variables

- **Categorical** – (Qualitative) Places individuals into one of several groups or categories
- **Quantitative** – (Numerical) Take on numerical values (ones that we can average)

example: Categorical: gender, grade level, zip code
Quantitative: Age, weight, salary

Example - Table on page 3.

Given below, you can read actual problem on page 3, it is more than just the table.

Province	Gender	Language spoken	Handed	Height (cm)	Wrist circum. (mm)	Preferred communication
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 - Categorical (Cat)
 - Gender - Cat
 - Dominant hand - Cat
 - Height Quantitative (Quant.)
 - Wrist circum. Quant.
 - Preferred communication - Cat

Describe the individual in the highlighted row?

Ontario male, speaks 4 languages, is left handed, 157.5 cm tall, has wrist circum. 147mm, and prefers text messaging

Pg 4 When examining data sets we are going to be concerned about the **distribution** of the variables in the data set.

We will spend a great deal of time determining distributions of data.

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

Pg 5 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.

→ understanding distributions will allow us to make inferences.

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