Chapters 1 & 2 Introduction and Data

What is Statistics?

Statistics is a science that involves the design of studies, data collection, summarizing and analyzing the data, interpreting the results and drawing conclusions.

Conclusions are made about specific phenomena on the basis of relatively limited information.

Data and Variables

Questions of interest:
How big and frequent are earthquakes in Canada?
How are earthquakes that occur in British Columbia compared to the rest of Canada?

Statistical process of investigation:

1. *Collect data:*
What data to collect?
How and where do we obtain data?

goal of this chapter

2. *Examine the data:*
How do we present and obtain useful information from the data?
We will learn the techniques in this chapter.

3. *Interpret the results and draw conclusions*

Solution: We can examine earthquake data during the past year that are available on the Natural Resources Canada website at:
`http://www.earthquakescanada.nrcan.gc.ca/recent/maps-cartes/index-eng.php`

A snapshot of the data was taken on Aug 14, 2012. The data available are during the period Aug 13, 2011 - Aug 13, 2012 in reverse chronological order.

Year	Month	Day	Depth (km)	Magnitude	Felt or not	Region
2012	8	13	5	1.8	No	ON
2012	8	13	18	1.9	No	QC
2012	8	13	5	2.3	No	NB
2012	8	12	18	1.8	No	QC
2012	8	12	18	2.4	No	QC
2012	8	10	20	2.8	No	NT
2012	8	10	10.1	0.9	No	QC
2012	8	10	8.8	2	No	QC
2012	8	10	20	4.1	No	NT
2012	8	9	2.6	0.5	No	BC
2012	8	9	7.7	1	No	QC
2012	8	9	6.7	1.5	No	BC
2012	8	9	26.8	1.4	No	BC
2012	8	9	10	2.3	No	BC
2012	8	9	19.9	0.3	No	BC
2012	8	9	6.5	1.2	No	QC

2012	8	8	10	2.7	No	BC
2012	8	8	18	2.7	No	QC
2012	8	8	10	2.6	No	BC
2012	8	8	10	2.8	No	BC
2012	8	8	18	3	No	NU
2012	8	8	18	2.1	No	QC
2012	8	8	18	2	No	QC
2012	8	8	10	1.9	No	BC
2012	8	8	20	0.8	No	BC
2012	8	8	30	1.1	No	YT
2012	8	8	28.5	1.9	No	BC
2012	8	8	20	1.8	No	BC
2012	8	8	10	2.7	No	BC
2012	8	8	20	0.6	No	BC
2012	8	8	20	1.7	No	BC
2012	8	8	20	3.7	Yes	BC
:	:	:	:	:	:	:
:	:	:	:	:	:	:

The full data set can be found on the lecture notes page on the course website.

In a typical data set, each row contains information corresponding to an individual or an object or an experimental unit.

A variable refers to a characteristic of interest, e.g. magnitude of an earthquake. A variable can be:

1. qualitative/categorical – categorical variables with categories that can be ordered are called ordinal variables. (gender - income as low, medium high)
2. quantitative (measured on a numerical scale) – units should be attached. (income in $)

Variable	Variable type	Unit
Depth	Quantitative	kilometers
Magnitude	Quantitative	dyne-cm (Richter scale)
Felt or not	Categorical	N/A
Region	Categorical	N/A

quantitative have Units

Categorical do not have Units

Understanding the Data (5 W + How)

- Who - subjects we wish to study about
- What - variables of interest
- Where - location in which the study is conducted / data are collected
- When - at what time point or over what time period are the data collected
- Why - purpose of doing a study / collecting the data
- How - method used to collect the data

* good to know as many as possible