Week 1 (Mon 1/14/2008—Thu 1/17/2008)

What is statistics?

Statistics is the science of collecting, organizing, summarizing, and analyzing data to draw conclusions with the help of the mathematics of probability.

(Quoted from our course syllabus) When studying a huge population, it is usually infeasible or prohibitively expensive to conduct an exhaustive "census". Statistics provides a collection of methods that enable us to make "educated guesses" about attributes of a population by analyzing only a sample of a limited size appropriately selected from the population. Guesses involve chance and likelihood. The mathematics of probability helps us understand and quantify such chance and likelihood. This allows us to correctly interpret the collected data to answer questions of interest about the population.

The Process of Statistics

1. Identify the research objective in terms of a question of interest about a population.
2. Collect the data needed by scrutinizing a sample drawn from the population in a proper way (randomness!!!)
3. Organize and summarize the data collected. ← Descriptive Statistics (Ch 1—4)
4. Draw conclusions about the population. ← Inferential Statistics. (Ch 9—12)
   This stop uses the mathematics of probability. ← Ch 5-8

A Typical Example

There are more than **200,000,000 adults in the U.S**. We would like to know the percentage of them that use Internet for shopping.

*Let **1025 randomly selected adults** be surveyed. Among them, 297 (i.e. 29%) said they used Internet for shopping.*

From such data, an "educated guess" is made by a statistician. It can take many forms:

- We are **99% confident** that the percentage being sought is **29% ± 3.6%.**
- We are **90% confident** that the percentage being sought is **29% ± 2.3%.**
- etc … (Note that the window narrows at the price of diminished confidence.)

Population, **Individuals**
Census
Population vs. **Sample**
Parameter vs. **Statistic**
(A parameter means an attribute of the population, whereas a statistic means a (similar) attribute of the sample.
In this example, we have

Population: the about 200,000,000 adults in the U.S.
Parameter: ???% in the entire population that use internet for shopping, which we are interested in knowing but which we will never know for sure.
Sample: the 1025 adults selected from the population.
Statistic: $\frac{297}{1025}$, *i.e. 29%*

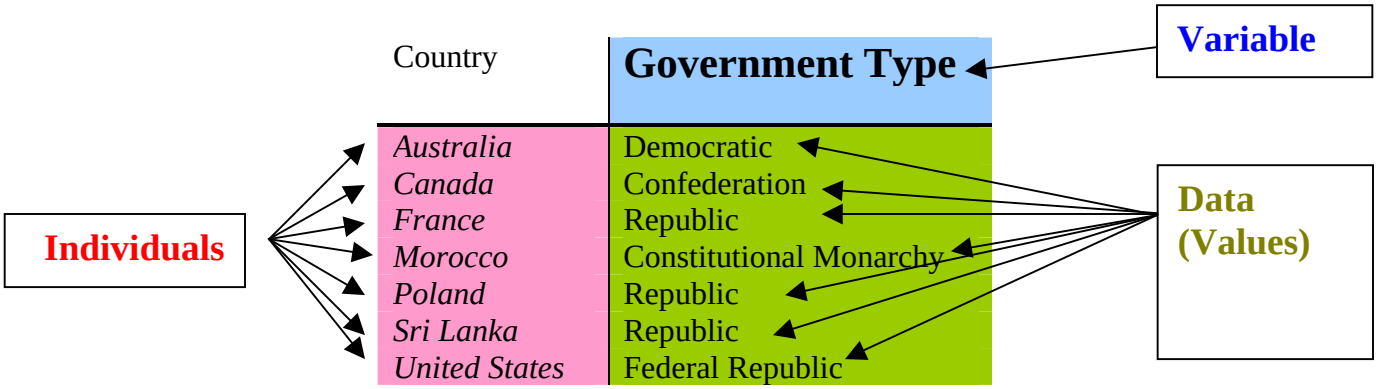
Variable vs. Data (Values)

Types of Variables

- **Qualitative Variables** (a.k.a. "**Categorical Variables):** values are "categories", or descriptions.

- **Quantitative Variables**: Values are numerical AND arithmetic operations make sense.
  - **Discrete** Quantitative Variable
  - **Continuous** Quantitative Variable

Import Remark: A qualitative variable may have its values coded by numbers for convenience, yet the variable is still qualitative because arithmetic operations don't make sense. Example: Social security numbers.



Country	Population (millions)
Australia	19.9
Canada	32.5
France	60.4
Morocco	32.2
Poland	38.6
Sri Lanka	19.9
United States	293.0

Country	Life Expectancy (years)
Australia	80.26
Canada	79.96
France	79.44
Morocco	70.35
Poland	74.16
Sri Lanka	72.89
United States	77.43

Country	Government Type	Life Expectancy (years)	Population (millions)
Australia	Democratic	80.26	19.9
Canada	Confederation	79.96	32.5
France	Republic	79.44	60.4
Morocco	Constitutional Monarchy	70.35	32.2
Poland	Republic	74.16	38.6
Sri Lanka	Republic	72.89	19.9
United States	Federal Republic	77.43	293.0

Chapter 2: Organizing and Summarizing Data

Section 2.1 Organizing Qualitative Data

- Frequency Table
- Relative Frequency Table

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Sum of all frequencies}}$$

- Bar Graphs
- Pareto Chart
- Side-by-Side Bar Graphs
- Pie Charts

Use a protractor if you would like to draw a pie chart by hand.

$$\text{Degree measure of a sector} = (\text{Relative Frequency}) \cdot 360°$$

Section 2.2 Organizing Quantitative Data

- **Organize Discrete Data in Tables**
- **Construct Histograms of Discrete Data**
- **Organize Continuous Data in Tables**
- **Construct Histograms of Continuous Data**
- **Identify the Shape of a Distribution (pp 80-81)**

(We will discuss stem-and-leaf plots, dot plots, and time-series graphs next time. We will also briefly mention 2.3.)