Topics

Statistics

the mathematical science that deals with the **collection**, **analysis**, and **presentation** of data, which can then be used as a basis for **inference** and **induction**

Topics of our course:

- Data
- Displaying and Calculating Descriptive Statistics
- Introduction to Probabilities
- Discrete and Continuous Probability Distributions
- Sampling and Sampling Distributions
- Confidence Intervals
- Hypothesis Testing
- Regression Analysis

An Introduction to Statistical Analysis for Business and Economics

- Classifications of Data
- Descriptive and Inferential Statistics

- Reading: Chapter 1

Data

Data

values assigned to observations or measurements

All the data collected in a particular study are referred to as the *data set* for the study

Information

data that are transformed into useful facts that can be used for a specific purpose, such as making a decision

… it is more than just data

Data

Raw facts or measurements of interest

New and Used Passenger Cars Imported in the US, by country of origin

Each individual value is considered a **data point**

Country	1988	2007
Japan	2,123,051	2,300,913
Germany	264,249	466,458
Italy	6,053	5,650
France	15,990	1,746

Source: World Almanac and Book of Facts (2009)

Information

Analyzing the data can provide information for decision making

New and Used Passenger Cars Imported in the US, by country of origin

Country	1988	2007	% Change
Japan	2,123,051	2,300,913	8.38
Germany	264,249	466,458	76.52
Italy	6,053	5,650	-6.66
France	15,990	1,746	-89.08

Did the imports from country X increase (decrease)?

Elements, Variables, and Observations

- *Elements* are the entities on which data are collected

- A *variable* is a characteristic of interest for the elements

- The set of measurements obtained for a particular element is called an *observation*

  - A data set with n elements contains n observations

Data Tables - Example

Element

Variables

Sale	Customer	CC#	Item	Cost	Gender
1	Nathan	1234	Stats Book	130.00	Male
2	Susan	5555	Unscented Lotion	20.00	Female
3	Andrew	8989	Snicker's Bar	2.00	Male
4	Patrick	8734	Toothpaste	2.99	Male
	…	…	…	…	…
	…	…	…	…	…
10	Josh	2468	Zinc Supplement	11.00	Male

Observation

Data Set

The Sources of Data

Primary data

Definition:
   data that you have collected for your own use

Collection Methods:
- Direct Observations
- Experiments
- Surveys

Advantages:
- Collected by the person or organization who uses the data

Disadvantages:
- Can be expensive and time-consuming to gather

Secondary data

Definition:
   data collected by someone else

Advantages:
- Readily available
- Less expensive to collect

Disadvantages:
- No control over how the data was collected
- Less reliable unless collected and recorded accurately

Types of Data

Qualitative Data:

Classified by descriptive terms

(labels or names used to identify an attribute of elements)

Examples:

- Marital Status
- Political Party
- Eye Color
  (Defined categories)

Quantitative Data:

Described by numerical values
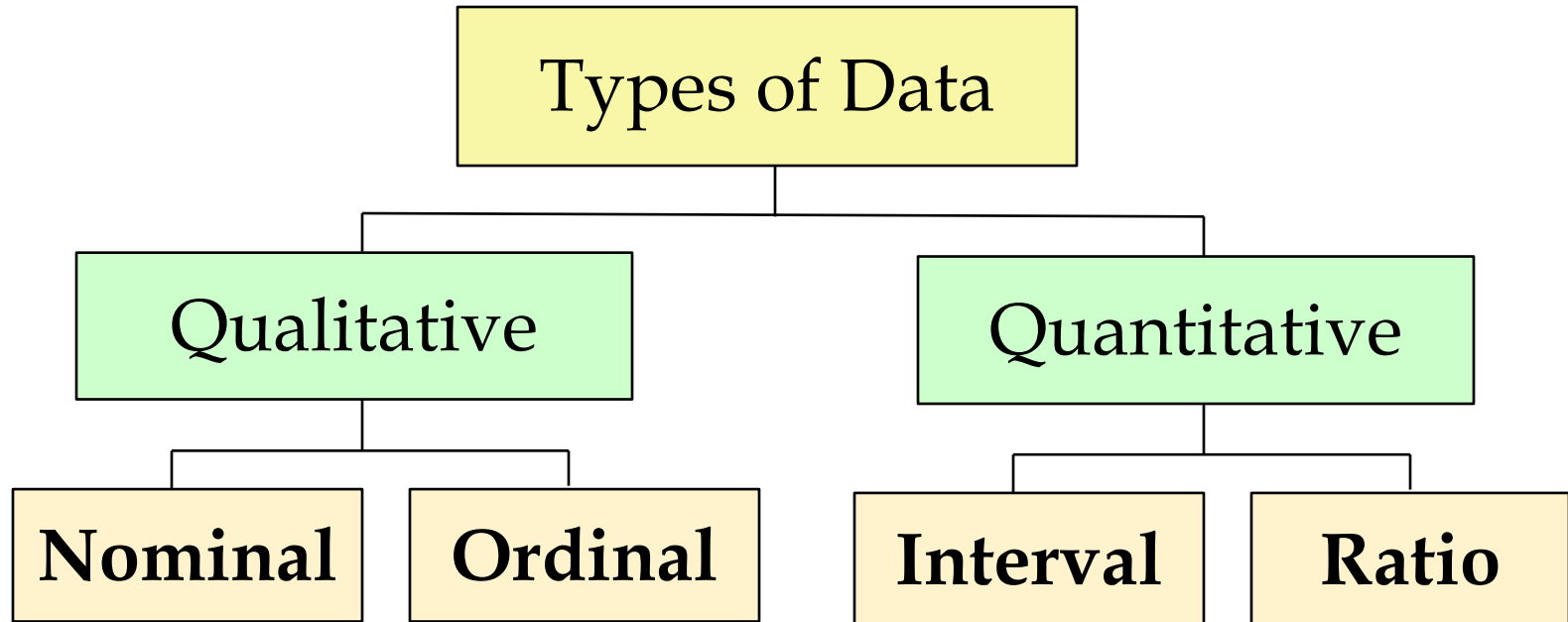
(how many or how much)

1. **Counted**

   Examples:

   - Number of Children
   - Defects per hour
     (Counted items)

2. **Measured**

   Examples:

   - Weight
   - Voltage
     (Measured characteristics)

Classifying Data by Level of Measurement

Types of Data

Qualitative

Quantitative

Nominal

Ordinal

Interval

Ratio

Classifying Data by Level of Measurement

The Four Levels of Data Measurement: A Summary

Level	Description	Example
Nominal	Arbitrary labels for data No ranking allowed	Eye Color, Zip Codes (19808, 76137)
Ordinal	Ranking allowed No measurable meaning to the number differences	Education level (Master's degree, doctorate degree)
Interval	Meaningful differences No true zero point (zero does not mean absence)	Calendar year (2009, 2010)
Ratio	Meaningful differences True zero point (zero means absence)	Income ($48,000, $0)

Time Series vs. Cross-Sectional Data
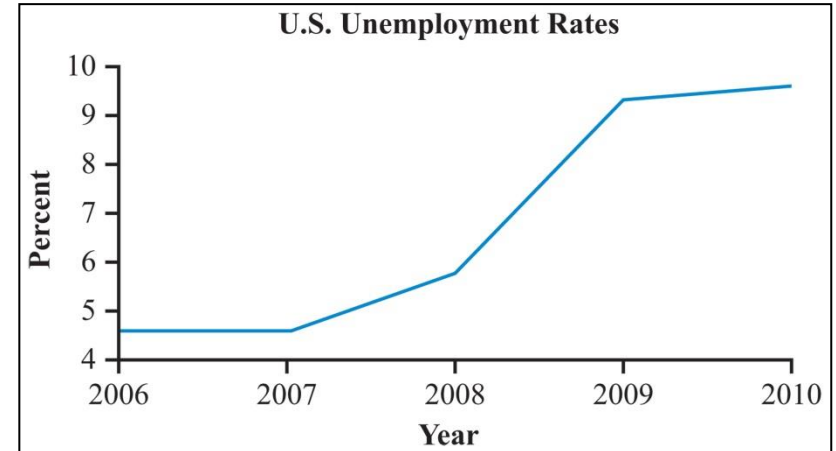
Time Series Data

- Values that correspond to specific measurements taken over a *range* of time periods

- Data can include hourly, daily, weekly, monthly, quarterly, or annual observations

Cross Section Data

- Values collected from a number of subjects during a *single* time period

- Subjects might include individuals, households, firms, industries, regions, countries, etc,

Time Series vs. Cross-Sectional Data

Time Series Graph of U.S. Unemployment Rates, 2006–2010

Cross-Sectional Graph of 2010 Unemployment Rates

Time Series vs. Cross-Sectional Data

New and Used Passenger Cars Imported in the US, by country of origin

Country	1988	2007	
Japan	2,123,051	2,300,913	**Time Series Data**
Germany	264,249	466,458	
Italy	6,053	5,650	
France	15,990	1,746	

Cross- Sectional Data

Source: World Almanac and Book of Facts (2009)

Descriptive vs. Inferential Statistics

Descriptive statistics

Collecting, summarizing, and displaying data (reported based on observations)

Inferential statistics

Making claims or conclusions about the population based on a sample of data
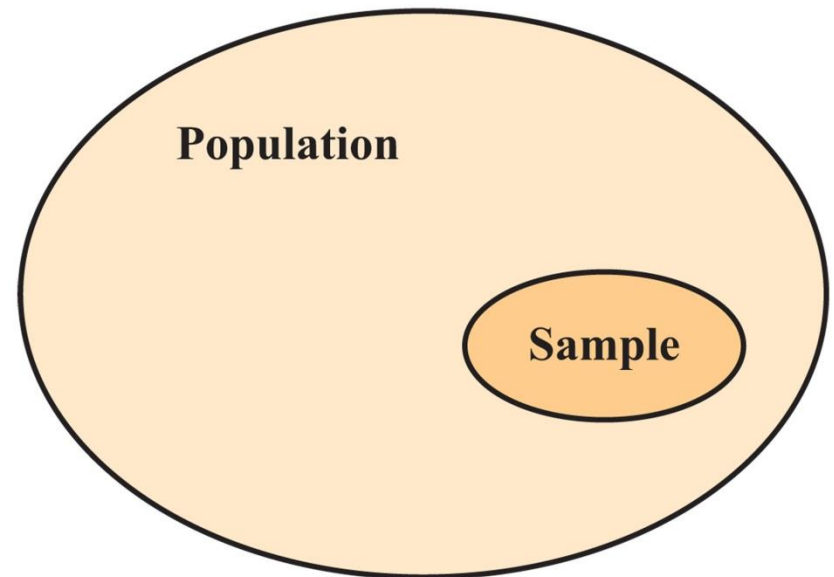
Recall the structure of our course…

Population vs. Sample

Population

Represents all possible subjects that are of interest in a particular study
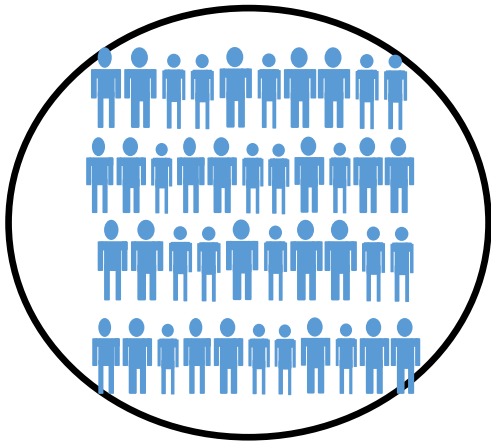
Sample

Refers to a portion of the population that is representative of the population from which it was selected
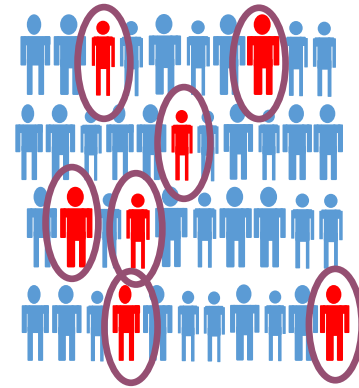
Parameter vs. Statistic

- **Parameter** – a described characteristic about a population
- **Statistic** – a described characteristic about a sample

Population

Values calculated using population data are called parameters

Sample

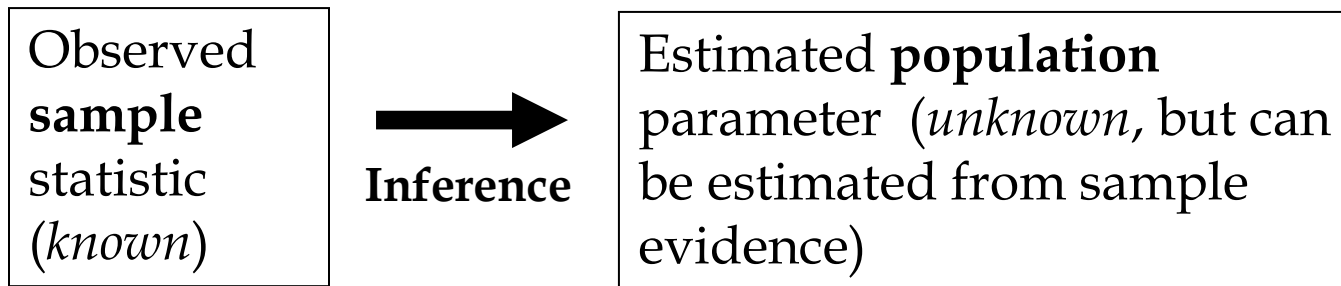Values computed from sample data are called statistics

The Need for Sampling

Reasons for sampling from the population:

- Too expensive to gather information on the entire population

- Too time-consuming to gather information on the entire population

- Often impossible to gather information on the entire population

Inferential Statistics

Sample statistic is calculated from the sample data and is used to make inferences about the unknown population parameter

| Observed **sample** statistic (*known*) | → **Inference** | Estimated **population** parameter (*unknown*, but can be estimated from sample evidence) |

Example:

A statistics professor asked students **in the class** about their ages. On the basis of this information, the professor states that the average age of **ALL the students in the university** is 24 years.

Practice Time

Identify the type of data (quantitative/qualitative):

1. The average monthly rainfall in Bloomington
2. The education level of survey respondents (High School, Bachelor's Degree, Master's Degree)
3. The marital status (Single, Married, Divorced)
4. The ages of the respondents in the survey
5. iPhone price
6. The number on the mailbox in the post office

Can you identify the level of measurement for the above data?

Practice Time

The Department of Transportation of a city has noted that on the average there are 17 accidents per day. The average number of accidents is an example of descriptive or inferential statistic?

Based on a survey, it was concluded that households with children under the age of 18 are more likely to have access to the Internet than family households with no children. Is it an example of descriptive or inferential statistic?

Practice Time

The table represents the results from a survey that collected annual household income in 2012. What type of data was used to construct this table - time series or cross-sectional data?

Household Income ($000)	# of Households
Under $30	67
$30 to under $40	111
$40 to under $50	125
$50 to under $60	21
$60 to under $70	38
Over $70	40