Welcome to Eco 2470: Economic Statistics

Fall, 2019. Chaoyi Chen

(Chapter 1)

Statistical Practitioners and Theorists

- Practitioner: Uses existing methods to learn from data and improve decision making.
- Theorist: Attempts to better understand and improve on existing methods.
- Goal of Course: To become a practitioner.
- But, theory and practice can never be entirely separated.
  - Good practitioner understands the theory behind their methods.
  - Good theorist, has practical experience too.

What is Statistics?

- Art and Science of learning from data

Data

Information

Statistics

Strategy & other → Decisions
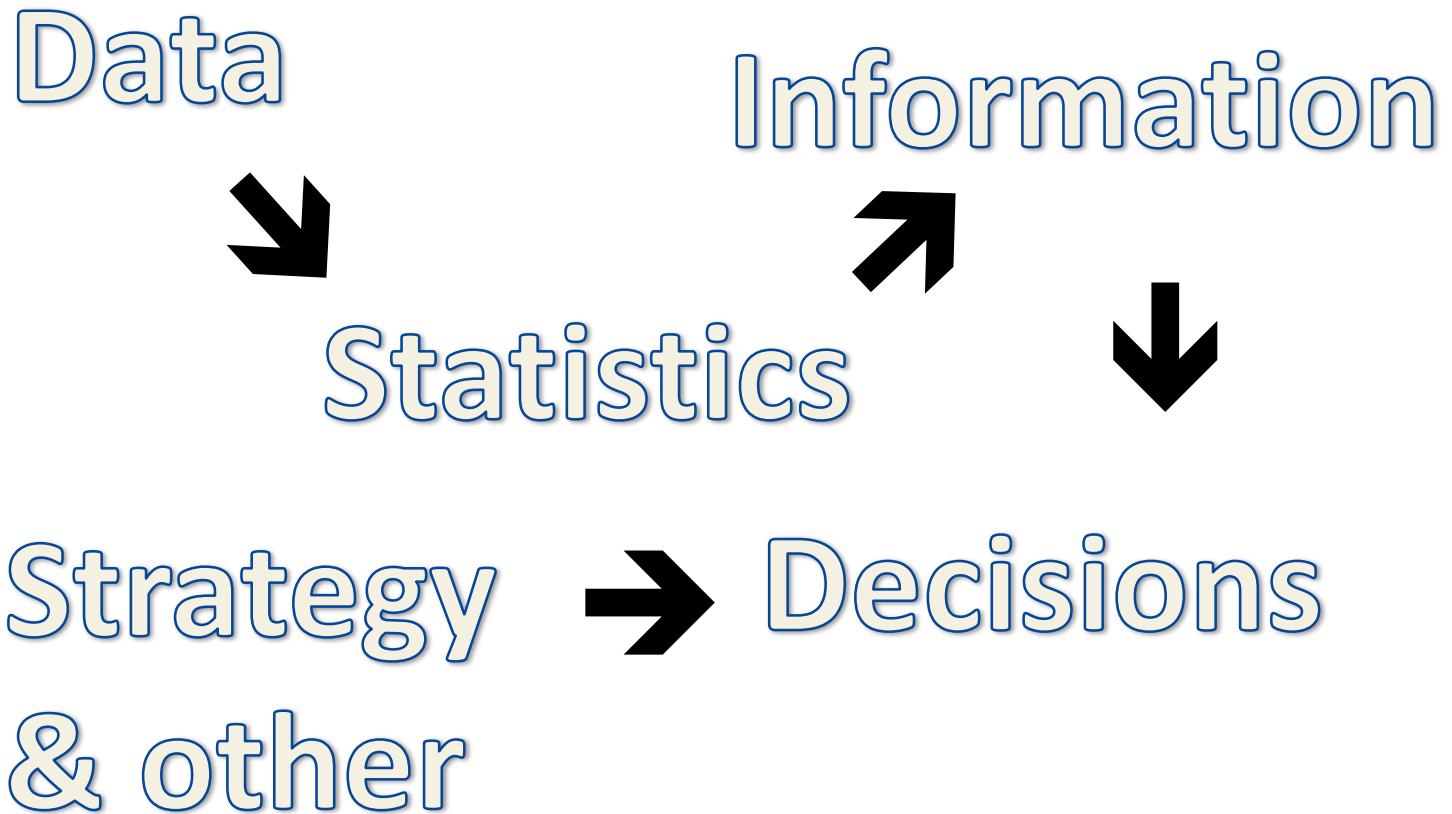
An Example

- Gallop Poll, August 15, 2012 (see link)

"Romney Sees No Immediate Bounce From Ryan V.P. Pick"

	Romney	Obama	N
Before Ryan announcement (Aug 7-10)	46%	45%	1,796
After Ryan announcement	47%	45%	1,863
Change	+1	0	

The Population

- Everyone we are interested in learning about.
- In our example: All U.S. voters
- Other examples:
  - All TV viewers
  - All Canadians
  - All Students
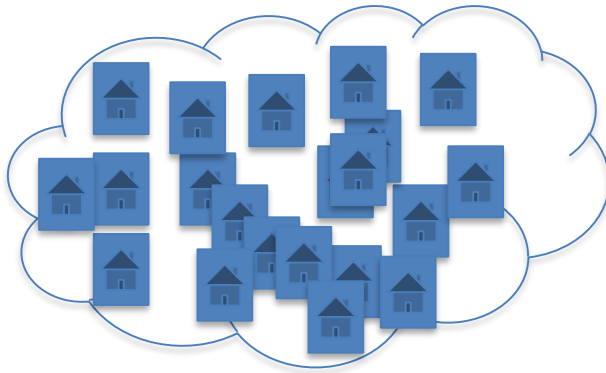 - All human beings past, present, & future

More About Populations

- It depends on the question you ask.
- Typically large
- Sometimes infinitely large
- Typically we do NOT observe the population
- Why? Too difficult & expense to collect data on the entire population.
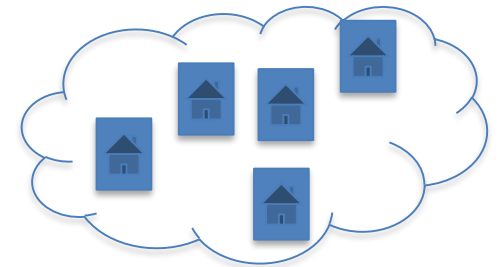- So, what do we do instead?

The Sample

- A smaller number of individuals drawn from our population.

population   sample

Gallop Poll Sample

"Results are based on telephone interviews conducted as part of Gallup Daily tracking Aug. 11-14, 2012, with a random sample of 1,863 registered voters, aged 18 and older, living in all 50 U.S. states and the District of Columbia."

The Sample: Why?

- Cheaper and easier to collect data on the sample than on the population

- Gallop could not possibly interview all U.S. voters within 3 days

- And, if they tried, they would go bankrupt.

Inferential Statistics

- What is it? Using the sample (we do see) to learn about the population (we don't see).

- In our Example: Use interviewed voters (sample) to learn about all voters (population)
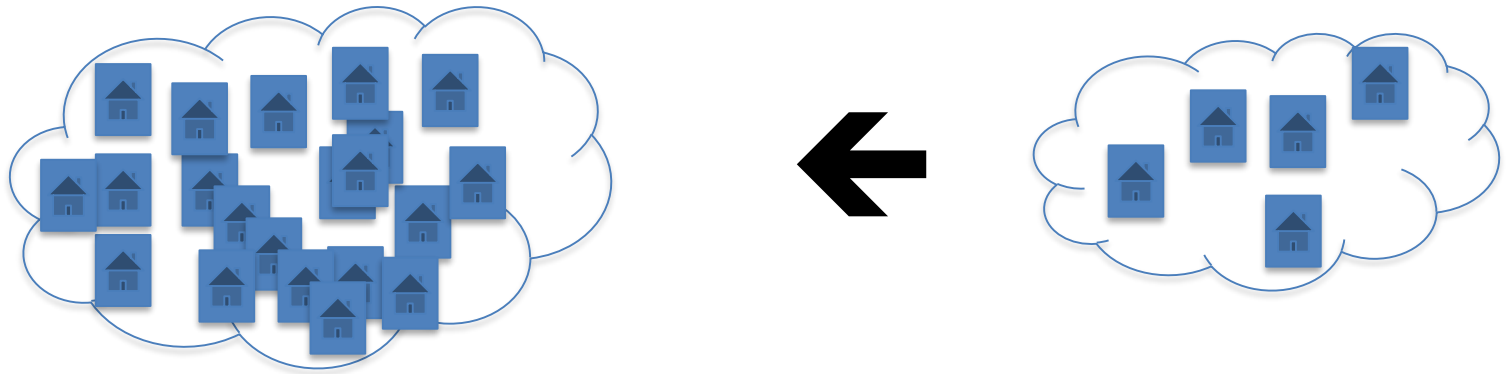
Gallop Poll's Statistical Inference

"Mitt Romney's standing in the presidential election has not changed materially in the immediate days after his announcement of Wisconsin Rep. Paul Ryan as his vice-presidential running mate"

- Surveyed 1,863 voters
- Make claims about all voters!

Creating a sample that is representative of the population?

1. Random Sampling: Draw sample at random from population.

    – Example of Non-Random Sample: Suppose Gallop conducted all interviews at fancy golf clubs.

2. Adequate Sample Size:

    –   Gallop surveyed 1,863 voters

    –   Suppose they just surveyed 1 or 2 voters?

Measures of Reliability

- Still, our sample may NOT be representative of the population?

- Can NEVER draw conclusions with FULL 100% confidence.

- Instead, we provide measures how confident we are in the conclusions we make.

Measures of Reliability (Continued)

- <u>Confidence Level</u>: How often our statistical procedure would lead to correct conclusions.

- <u>Significance Level</u>: How often our statistical procedures would lead to false conclusions.

Gallop Poll Confidence Level:

"For results based on the total sample of registered voters, one can say with 95% confidence that the maximum margin of sampling error is + or − 3 percentage points."

What Do They Mean?

1. Providing a range, not an exact number:

	Romney	Obama	N
Before Ryan announcement (Aug 7-10)	(43%-49%)	(42%-48%)	1,796
After Ryan announcement	(44%-50%)	(42%-48%)	1,863

1. And even this range is not guaranteed:
   - Correct 19 times out of 20.
   - Wrong 1 time out 20.

Parameter and Statistic

1. <u>Parameter</u>: A descriptive measure of the <u>population</u>

   e.g. Percent of U.S. voters who plan to vote for Obama.

2. <u>Statistic</u>: A descriptive measure of the <u>sample</u>

   e.g. Percent of the 1,863 voters polled who plan to vote for Obama

Descriptive Statistics

- Just describe the sample

- Without worrying about the population

- Includes:

  - Graphical methods

  - Numerical methods

  Chapters 2-4 Cover Descriptive Statistics.