

1 Statistics: The Art and Science of Learning from Data

Disclaimer: Unless otherwise stated, all statements in quotes are taken directly from our textbook Agresti/Franklin, 2007.

1.1 How Can You Investigate Using Data?

“The information we gather with *experiments* and with *surveys* is collectively called **data**.”

“**Statistics** is the art and science of designing studies and analyzing the data that those studies produce. Its ultimate goal is translating data into knowledge and understanding of the world around us. In short, **statistics** is the art and science of learning from data.”

Where is statistics used in the real

world?

Example: 2000 U.S. Senatorial election in New York, with 6,253,883 voters (out of 19 million New Yorkers, including children).

Based on an exit poll from sampling 2232 voters (0.036% of voters):

Democrat Hillary Rodham Clinton (56%), Republican Rick Lazio (44%)

Is it reasonable to conclude that Clinton won, based on that exit poll?

Example: Test marketing. Will revenues increase if a company starts marketing a new product?

Example: Suppose there is an 80% chance of rain.

Example: Global warming

Example: Technology stocks plummet. Airline stocks remain stable. Interest rates increase. How

should you invest your money?

Example: You work for an insurance company. What premium should be charged, and what benefits should be offered, for a person with a particular driving record?

Example: Different drugs and different dosages are tested on mice with cancer. Some mice die quickly; others do not.

Example: How should a farmer farm her/his land?

Example: DNA evidence.

Example:

In 1787–88 the *Federalist* papers were written.

These 85 essays were written between October 1787 and May 1788 to urge New Yorkers to ratify the proposed U.S. Constitution.

Example: How high should a dam be built to

avoid flooding?

Example: Sports and gambling? MOSTLY JUST ENTERTAINMENT.

□

Statistics consists of

- **“Design:** Planning how to obtain data to answer the questions of interest.”
- **“Description:** Summarizing the data that are obtained.”
- **“Inference:** Making decisions and predictions based on the data.”

1.2 We Learn about Populations Using Samples

“The **population** is the total set of subjects in which we are interested.”

“A **sample** is the subset of the population for whom we have (or plan to have) data.”

The field of *statistics* may be divided into the two categories:

- **descriptive statistics:** Summarizing data via graphs and numbers (such as averages and percentages).
- **inferential statistics:** Making decisions or predictions about a population based on a sample. Hence, the population is unknown, but a sample is known.

Example of inferential statistics:

In this course we study both *statistics* and *probability* (from section 1.1).

Regarding **probability** (described on section 1.1), the *population* is assumed known, and statements

are made regarding the likelihood of obtaining certain data values.

Example of probability:

“A **parameter** is a numerical summary of the *population*.”

“A **statistic** is a numerical summary of a *sample* taken from the population.”

Randomness and Variability

Random sampling implies that data are selected for the sample at random from the population.

1.3 What Role do Computers Play in Statistics

- Store large data files

- Analyze data
- Graph data
- Generate random samples
- Simulate complicated sampling scenarios