Statistics: Concepts and Controversies

Chapter 1: Where Do Data Come From?

Lecture Slides

Case Study: Statistics In the Media 1

You see statistics everywhere in the media. You hear:

Last month's unemployment rate was 5.4%.

88% of AAAS scientists versus 37% of U.S. residents agree it is generally safe to eat genetically modified foods.

A longer article says that low-income children who received high-quality day care did better on academic tests given years later and were more likely to go to college and hold good jobs than other similar children without high-quality day care.

Case Study: Statistics in the Media 2

Where do these data come from?

Can we trust these results? Why?

The first question to ask when someone throws a number at you is, "Where do the data come from?"
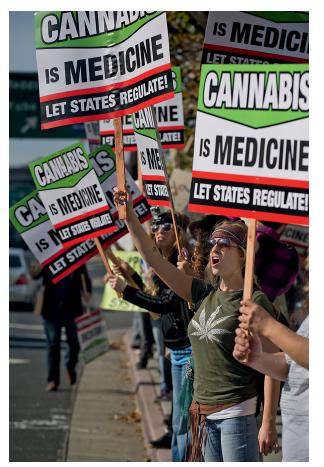
Case Study: Marijuana Legalization Poll 1

In January 2014 the Michigan news site MLive ran the story, "Take Our Online Poll: Should Michigan Legalize Marijuana?"

Of 9684 respondents,

7906 (81.84%) said, "Yes"

1190 (12.2%) said, "No"

588 (6.07%) said, "Decriminalize but don't legalize"

Sacramento Bee/Contributor/Getty Images

Case Study: Marijuana Legalization Poll 2

Do the results from the Mlive poll demonstrate overwhelming support for legalizing marijuana?

What can we say about data from this poll?

By the end of this chapter, you will learn basic questions to ask about the live poll.

Individuals and Data

Statistics is the science (or art) of data.

Individuals are the objects described by a set of data. Individuals may be people, but they may also be animals or things.

A *variable* is any characteristic of an individual. A variable can take different values for different individuals.

The actual measurements recorded for individuals are called *data.*

Example: Individuals and Variables

Name	Major	Points	Grade
Advani, Sura	Comm.	397	B
Barton, David	Hist.	323	C
Brown, Annette	Lit.	446	A
Chiu, Sun	Psyc.	405	B
Cortez, Maria	Psyc.	461	A

Individuals?

Data: 397, 321, etc ...

People in the study

Variables?

Major, points, grades

Categorical and Numerical Variables

A categorical variable simply places an individual into one of several groups or categories.

A numerical variable takes numerical values for which arithmetic operations such as adding and averaging make sense. A numerical variable is sometimes referred to as a quantitative variable.

Example: Who Recycles?

Researchers spent lots of time and money weighing the stuff put out for recycling in two neighborhoods in a California city; call them Upper Crust and Lower Mid.
Individuals?
Households (recycling pickup is done for residences, not for people, one at a time)
Variable?
Weight in pounds of recycling
Weight is a numerical (quantitative) variable.

Example: What's Your Race?

The U.S. census asks, "What is this person's race?" for every person in every household.

Variable?
Race is the variable.
Race is a categorical variable because it places people into a category.

Response Variable and Observational Study

A **response variable** measures an outcome or result of a study.

An **observational study** observes individuals and measures variables of interest but does not intervene in order to influence the responses.

The purpose of an observational study is to describe some group or situation.

Sample Survey

Sample surveys are an important kind of observational study.

They survey some group of individuals by studying only some of its members, selected not because they are of special interest but because they represent the larger group.

Populations and Samples

The **population** in a statistical study is the entire group of individuals about which we want information.

A **sample** is the part of the population from which we actually collect information and is used to draw conclusions about the whole.

Example: Public Opinion Polls

For a typical opinion poll:

Population?
U.S. residents 18 years of age and over. Noncitizens and even illegal immigrants are included.

Sample?
Between 1000 and 1500 people interviewed by telephone.

Example: Current Population Survey (CPS)

Most important government sample survey in the United States. Many variables recorded concern the employment status of everyone over 16 years old in a household. The government's monthly unemployment rate comes from the CPS.

Population?
More than 123 million U.S. households. (Individuals are households.)

Sample?
About 60,000 households interviewed each month.

Example: General Social Survey (GSS)

Every second year, by the National Opinion Research Center at the University of Chicago

The variables cover the subject's personal and family background, experiences and habits, and attitudes and opinions on different subjects.

Population?
Adults (aged 18 and over) living in households in the United States (does not include adults in institutions or persons who cannot be interviewed in English).

Sample? About 3000 adults, interviewed in person in their homes

Census

A **census** is a sample survey that attempts to include the entire population in the sample.

The U.S. Constitution requires a census of the American population every 10 years. Even a census can miss people.
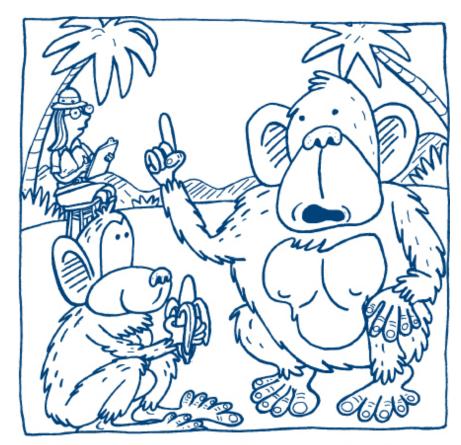
The Census Bureau estimated that the 2010 census overcounted the American population by 0.01% but undercounted the black population by 2.1%.

Experiments

An **experiment** deliberately imposes some treatment on individuals in order to observe their responses.

The purpose of an experiment is to study whether the treatment causes a change in the response.

"Now eat that banana. The nice statistician is watching us."

Moore/Notz, *Statistics: Concepts and Controversies*, 9e, © 2017 W. H. Freeman and Company

Example: Helping Welfare Mothers Find Jobs

Choose two similar groups of women as they apply for welfare.

Require one group to participate in a job training program but do not offer the program to the other group.

Compare the income and work record of the two groups after several years.

Statistics in Summary 1

- Any statistical study records data about some **individuals** (people, animals, or things) by giving the value of one or more **variables** for each individual.

- Some variables, such as age and income, take numerical values. Others, such as occupation and sex, do not. Be sure the variables in a study really do tell you what you want to know.

Statistics in Summary 2

- The most important fact about any statistical study is how the data were produced. **Observational studies** try to gather information without imposing a treatment on the individuals they are observing.

- **Sample surveys** are an important kind of observational study. A sample survey chooses a sample from a specific **population** and uses the **sample** to get information about the entire population.

Statistics in Summary 3

- A **census** attempts to measure every individual in a population.

- **Experiments** actually impose a treatment on the individuals in order to see how they respond. The goal of an experiment is usually to learn whether some treatment actually causes a certain response.