

A spiral-bound notebook with a light beige, textured cover. The metal spiral binding is visible on the left side. The text is centered on the page.

Section 1.1: How Can You Investigate Using Data?

What is Statistics?

Real Questions

- How can you evaluate evidence of global warming?
- What's the chance that your tax return will be audited?
- What “hot streaks” should you expect in basketball?
- How can you analyze whether a diet actually works?

What is Statistics?

- **Statistics** is the art and science of designing studies and analyzing the information that the studies produce.
- Simply, statistics helps us to take information and use it to better understand the world around us.

How is statistics used in the real world?

- Business
- Medical Research
- Sports
- Politics
- Sociology
- Sciences
- Journalism
-



To find answers to real questions we need to collect and analyze **DATA**.

- **Data is any information that we have gathered.**
- Fill out the Class Survey.

Components of Statistical Investigations

- What questions do you want to answer?
- What data would help you to address these questions?
- How can you collect this data in a responsible manner?
- How can you summarize this data?
- What will you do once you have the summary of the data?

How can we use the data?

- How would you describe the different kinds of data on the student survey form?
- What kinds of questions could we try and answer from data of this type?

Three Main Aspects of Statistics

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

A spiral-bound notebook with a light-colored, textured cover. The spiral binding is on the left side. The text is centered on the page.

Section 1.2 – We Learn about Populations Using Samples

Examples

- What is the percentage of American citizens living below the poverty line?
- What is the average class size at Shippensburg?
- What is the average time that college students spend on Facebook per day?

Population vs. Sample

- The **population** is the *total* set of subjects which we are interested in studying.
- A **sample** is a part of the population that we actually examine to gather information about the whole.

Example

- A sociologist is interested in determining the percentage of college students who have smoked marijuana so she randomly samples 50 college students and finds that 60% of the students in the sample respond “yes”.
 - **Population:** All college students.
 - **Sample:** 50 college students surveyed.

Example

- In 2003, a special election was held to determine whether CA Gov. Gray Davis should be recalled from office. The exit poll sampled 3160 voters of CA's 8 million voters and found that 54% said they voted to recall Davis.
 - What is the population?
 - What is the sample?

Descriptive vs. Inferential Statistics

- **Descriptive statistics** refers to the methods for summarizing the *sample* data.
 - May be graphs or numbers such as averages, percentages, etc.
- **Inferential statistics** refers to methods of making decisions or predictions about a *population* based on data obtained from a sample of that population.

Statistic vs. Parameter

- A **statistic** is a numerical measurement of a *sample* taken from the population.
- A **parameter** is a numerical summary of the *population*.

Example

- In the 1998 General Society Survey, a random sample of 2613 Americans were asked whether or not they had voted in the 1996 presidential election, and 1783 said yes they had. The Federal Election Commission reported that 49% of eligible voters actually voted in the 1996 election.
 - What is the **population** of interest? What is the **sample** of interest?
 - What **proportion** of the people surveyed said that they had voted?
 - Is the above number answer a **parameter** or a **statistic**? Why?
 - Is the data from the Federal Election Commission a **parameter** or a **statistic**?
 - What reasons could explain the differences between the survey data and the Federal Election Commission data?

Randomness

- **A sample tends to be a good reflection of the population if the sample is chosen randomly.**
 - To test whether or not the class understand a particular concept I can ask a question to 5 students randomly or I can ask 5 questions and see if 5 students will volunteer and correctly answer the questions.
 - Randomness in the sample helps to weed out other factors.

Variability

- Responses will vary quite a bit from person to person.
 - *Q: How many hours did you sleep last night?*
- Responses will still vary a bit from sample to sample.
 - *Q: What is the average number of hours of sleep that a college students got last night?*
- **Key idea:** With random sampling the amount of **variability** from sample to sample is actually quite predictable.