

Lecture 1

What Are Data?

§1.1

1. What Are Data?

Statistics begins with data and variation.

Data are recorded observations or measurements.

Variation means that the values in a data set differ from one another.

For example, the heights of five students might be

63, 65, 68, 70, 72 inches.

The different heights illustrate **variation**.

Data may consist of numbers, categories, dates, measurements, or other recorded information.

A useful summary of the statistical process is

Collect → Organize → Summarize → Interpret

2. Understanding Data in Context

A number by itself is often difficult to interpret.

For example,

72

could represent a test score, an age, a temperature, a height, or \$72. The meaning depends on the context.

When interpreting data, ask:

1. **Who or what was measured?**
2. **What was measured?**
3. **How was it measured?**
4. **What are the units?**

Context gives data meaning.

2.1 Individuals, Variables, and Data Values

Individual: a person, object, or event described by the data.

Variable: a characteristic recorded for each individual.

Data value: the particular value of a variable for one individual.

Consider the following data:

Student	Age	Study Hours	Exam Score
A	18	5	72
B	19	8	81
C	20	10	88
D	18	3	65
E	21	12	94

Here:

- the **individuals** are the five students;
- the **variables** are age, study hours, and exam score;
- Student C's study-hours value is 10;
- Student E's exam-score value is 94.

The differences among the students' values are examples of **variation**.

3. How Are Data Collected?

The method used to collect data affects what conclusions we can draw.

Method	What happens?	Example
Experiment	Researchers impose a treatment or condition.	Assign students to different study methods and compare exam scores.
Survey	Researchers ask people questions.	Ask students how many hours they study each week.
Observational study	Researchers observe individuals without imposing a treatment.	Record students' existing study habits and exam scores.
Existing data	Researchers analyze data that have already been collected.	Analyze university or government records.

Experiment: Researchers impose a treatment or condition and observe the results.

Survey: Researchers collect information by asking questions.

Observational study: Researchers collect information by observing individuals without imposing a treatment.

Key point: A large amount of data does not automatically mean that the data are reliable or useful. The quality of the collection method matters.

3.1 Why Does the Collection Method Matter?

Suppose a company claims:

“Students who use our program score 15 points higher on average.”

Before accepting the claim, ask:

- Who were the students?
- How many students were studied?
- How were they selected?
- Did students choose to use the program?
- Was there a comparison group?
- How were the scores measured?

The value 15 alone cannot answer these questions.

Important: Statistical calculations can summarize data, but they cannot automatically fix problems in how the data were collected.

4. Putting It All Together

Consider the statement:

A university reports that the average student spends 6.8 hours per week exercising.

Before interpreting 6.8, ask:

1. Who was measured?
2. What variable was measured?
3. What are the units?
4. How was exercise measured?
5. How were the students selected?

For example, students' estimates may differ from measurements obtained using a fitness tracker.

Before calculating anything, understand the data.

4.1 Class Activity

Suppose a report states:

“The average is 84.”

What information would you need before interpreting this number?

- a) Average of what?
- b) Who was measured?
- c) What are the units?
- d) How was the measurement obtained?
- e) How many observations were collected?
- f) How were the individuals selected?

5. Practice and Summary

5.1 Practice Questions

1. In your own words, define **data**.

Recorded observations or measurements about individuals, objects, or events.

2. In your own words, define **variation**.

The fact that observations differ from one another.

3. Why is variation important in statistics?

Statistics helps us describe and analyze differences in data.

4. Identify the individuals and variables in the following study:

A researcher records the age, major, GPA, and number of hours studied for 100 college students.

Individuals: the 100 college students. Variables: age, major, GPA, and hours studied.

5. A university reports that the “average income is \$42,000.” List four questions you would ask about the context.

Who was measured? What are the units? How was income measured? How were individuals selected?

6. Identify the data-collection method:

- a) Researchers assign patients to one of two treatments.

Experiment.

- b) Students are asked how many hours they sleep.

Survey.

- c) A researcher downloads ten years of weather records.

Existing data.

- d) A researcher records students' existing study habits without assigning a treatment.

Observational study.

7. Explain the difference between an experiment and a survey.

An experiment imposes a treatment or condition; a survey asks participants questions.

8. A company claims:

“Our customers lose an average of 20 pounds using our program.”

List four questions you would ask before accepting the claim.

Who were the customers? How many were studied? How were they selected? How was weight loss measured?

9. Consider

72, 81, 88, 65, 94.

Give one example of context you would need before interpreting these numbers.

For example, what the numbers measure and their units.