Data

by Sophia

☰ WHAT'S COVERED

This lesson will introduce the collection and evaluation of data including:

1. Defining Data

2. Evaluating Types of Data

3. Gathering Data

1. Defining Data

Data is the pieces of information that we use in order to answer some statistical question. It could be a number or an attribute.

But ultimately, it's the pieces of information that we use to get a more accurate picture of a scenario. Every piece of data helps us to get a more accurate description, which begs the question, how do you obtain data? Where does it come from? Do you just make it up? Where is data?

🗋 TERM TO KNOW

Data
Information used in a study to answer a statistical question.

2. Evaluating Types of Data

There are two types of data to serve your purposes. It's possible that the easier route is to go with something someone else has already done. **Available data** is data that has already been collected by somebody.

Now, who collects data? Well, a lot of places collect data, such as:

- Government organizations
- Polling organizations

- News sources

- Government entities

- Private entities

The vast majority of sources are trustworthy. However, when using available data, it's important to think critically about what the information is trying to convey. It's essential to break apart the information and ask yourself these questions:

- Who collected it?

- Are they reputable?

- Are they trustworthy?

- When was it collected?

- How was it collected?

- Why did they collect it?

So, how do you know when you need to gather the information yourself? Gathering information yourself is called **raw data**. Obviously, if the population doesn't match your topic of interest, then it is of no value to you, so you need to gather it yourself.

But what about less obvious characteristics such as whether or not a source has an agenda? This is a key point here. Having an agenda, whether intentional or not, can introduce what's called **bias**.

Often, polling organizations and news organizations and government entities try to do the best job they can to get relevant information. It's not usually intentionally put out there. But sometimes it is when they're trying to push some kind of agenda.

> 📄 TERMS TO KNOW
>
> **Available Data**
> Data collected by some other entity - a government organization or private company.
>
> **Raw Data**
> Unorganized, unprocessed, and not summarized. Typically, this is data that is not already available.
>
> **Bias**
> The systematic favoring of certain outcomes in a study. There are many ways to introduce bias into a study.

3. Gathering Data

If you choose to collect your own data, you must think critically and ask yourself these questions:

- Who will receive this data?

- For whom is the data intended?

- How will you and others gain access to it?

Collecting data is important because it's the source of statistics. Think about data as the raw means of creating something useful. If you collect your data well, the statistics are going to be accurate. If you collect your data poorly, then your data is poor. There's no rescuing that.

☆ **BIG IDEA**

You can't make useful statistics out of poor data. Thinking critically will help you determine which type of data should be used for your purposes.

SUMMARY

This tutorial defined data as "information used in a study to answer a statistical question." We discussed how to evaluate types of data, available or raw, and questions focusing on the who, what, why, and how should be posed to help identify bias. When gathering your own data, it's important to understand your audience and consider how they will gain access to all your hard work.

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

TERMS TO KNOW

Available Data

Data collected by some other entity - a government organization or private company.

Bias

The systematic favoring of certain outcomes in a study. There are many ways to introduce bias into a study

Data

Information used in a study to answer a statistical question

Raw Data

Unorganized, unprocessed and not summarized.. Typically, this is data that is not already available