Australian Bureau of Statistics

Statistical Language - What are Data?

Statistical Language

What are Data?

Content on this page requires Adobe Flash Player to be viewed.

Download Adobe Flash player

Alternatively, read the transcripts, attached below, containing a text version of the information displayed in the Flash Animation.

This animation explains the concept of data. If you are unable to access the video a Transcript (.doc 29kb) *has been provided. The animation requires* Adobe Flash Player *to run. The animation contains no audio.*

What are data?

Data are measurements or observations that are collected as a source of information. There are a variety of different types of data, and different ways to represent data.

The number of people in Australia, the countries where people were born, number of calls received by the emergency services each day, the value of sales of a particular product, or the number of times Australia has won a cricket match, are all examples of data.

A data unit is one entity (such as a person or business) in the population being studied, about which data are collected. A data unit is also referred to as a unit record or record.

A data item is a characteristic (or attribute) of a data unit which is measured or counted, such as height, country of birth, or income. A data item is also referred to as a **variable** because the characteristic may vary between data units, and may vary over time.

An observation is an occurrence of a specific data item that is recorded about a data unit. It may also be referred to as **datum**, which is the singular form of data. An observation may be numeric or non-numeric (categorical). For example, 173 is a numeric observation of the data item 'height (cm)', whereas 'Australia' is a non-numeric (categorical) observation of the data item 'country of birth'.

A dataset is a complete collection of all observations.

The following table is an example of a dataset:

	age (years)	sex	income ($)	→ Data Items
Person 1 (John Smith)	18	m	50000	
Person 2 (Joe Bloggs)	18	m	48000	→ Data Unit - Person 2.
Person 3 (Sally Jones)	20	f	55000	
Person 4 (Linda Lee)	22	f	50000	→ Numeric observation of the data item 'income'
Person 5 (Harry James)	19	m	35000	→ Non-numeric (categorical) observation of the data item 'sex'

Return to Statistical Language Homepage

This page last updated 3 July 2013