Informal data definitions

[**population**] The entire aggregate of items which is to be investigated.

[**sample**] A subset of the entire aggregate of items (the population), which is selected as a representative of the population.

[**variable**] A variable is a characteristic being measured. In a study or an experiment, it must be determined what variables (characteristics) are to be observed and measured.

[**measurement**] A measurement, observation, or value is a real number that represents a single response on an individual item or experimental unit.

Informal data definitions, cont'd

[**random variable**] A random variable is a variable which can be
assigned values or measurements that can be considered random.
That is it's value is observed at random with some distribution.

[**Process**] A process is a repeatable series of actions that results in an
observable characteristic or measurement.

[**Observation**] The collection of information in an experiment. or
actual values obtained on variables in an experiment.

[**Factors**] Controllable experimental variables that influence the
observed values of response variables.

Informal data definitions, cont'd

[**Simple Random Sample**] In an experimental setting, a simple random
sample of size n is obtained when items are selected from a fixed
population or a process in such a manner that every group of items of
size n has an equal chance of being selected as the sample.

[**Parameters and Statistics**] A parameter is a numerical characteristic
of a population or a process. A statistic is a numerical characteristic
that is computed from a sample of observations.

Data Description

The first (and often most important) thing to do with a set of data is to find a way to display it graphically. Graphs can summarize data sets, show typical and atypical values, highlight relationships between variables, and/or show how the data are spread out (what one would call the shape, or the distribution, of the data). Even for small sets of data, important features may be more obvious from a graph than from a list of numbers. We will discuss several ways to graph data, depending on how the data were collected.

(Silica Surface-Area Data)

Samples of a particular type of silica (a chemical product with many applications such as a filler in rubber products) were tested for their surface area (a key property). The resulting 32 measurements are listed below.

101.8	100.5	100.8	102.8	103.8	102.5	102.3	96.9
100.0	99.2	100.0	101.5	98.5	101.5	100.0	98.5
100.0	96.9	100.7	101.6	101.3	98.7	101.0	101.2
102.3	103.1	100.5	101.2	101.7	103.1	101.5	104.6

Class	Frequency	Relative Frequency
96.9–97.7	2	0.0625
97.7–98.5	0	0.0
98.5–99.3	4	0.125
99.3–100.1	4	0.125
100.1–100.9	4	0.125
100.9–101.7	8	0.25
101.7–102.5	4	0.125
102.5–103.3	4	0.125
103.3–104.1	1	0.03125
104.1–104.9	1	0.03125
Total	32	1.000

- All of the classes have the same width.

- The classes are adjoining, but not overlapping. This is so that every observation will be in one and only one class (i.e., it will get counted exactly once).

- The number of classes is somewhat arbitrary. Choosing too few classes will cause your graph to lose detail, while choosing too many may obscure the main features.

- Histograms for small data sets are more sensitive to the number of classes than histograms for large data sets.