1. Introduction

1.1 The R Environment

R is an integrated suite of software which facilitates for data manipulation, calculation graphical display and analysis.

Among other things it has

- an effective data handling and storage facility
- a suite of operators for calculations on arrays, in particular matrices
- graphical facilities for data analysis and display either directly at the computer or on hardcopy
- a large, coherent, integrated collection of intermediate tools for data analysis

The most important feature of **R is that it is open to everybody. (Free and Open Source software).**

1.2 R and Statistics

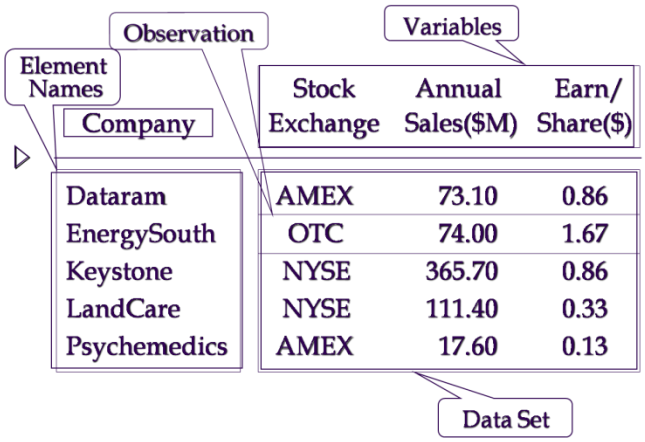
Many people use R as a statistics system. It is preferred to think of it as an environment within which many classical and modern statistical techniques have been implemented. A few of these are built into the base R environment, but many are supplied as packages. There are about 25 packages supplied with R (called "standard" and "recommended" packages) and many more are available through the CRAN family of Internet sites (via http://CRAN.R-project.org) and elsewhere. Most classical statistics and much of the latest methodology are available to use with R, but users may need to be prepared to do a little work to find it.

1.3 Elements, Variables and Observations

Data are the facts and figures collected, summarized and analyzed for presentation and interpretation. All the data collected in a particular study are referred to as the data set for the study.

The elements are the entities on which data are collected. A variable is a characteristic of interest for the elements. The set of measurements collected for a particular element is called an observation. The total number of data values in a data set is the number of elements multiplied by the number of variables.

Example: Data Set



1.4 Scales of Measurements

Data collection requires one of the four scales of measurements, *nominal, ordinal, interval or ratio*. The scale of measurement determines the amount of information contained in the data and indicates the most appropriate data summarization and statistical analyses.

- Nominal Scale: Data are labels or names used to identify an attribute of the element. A nonnumeric label or numeric code may be used.

  E.g.: Students of a university are classified by the school in which they are enrolled using a nonnumeric label such as Business, Humanities, Education, and so on. Alternatively, a numeric code could be used for the school variable (e.g. 1 denotes Business,2 denotes Humanities, 3 denotes Education, and so on).

- Ordinal Scale: The data have the properties of nominal data and the order or rank of the data is meaningful. A nonnumeric label or numeric code may be used.

  E.g.: Students of a university are classified by their class standing using a nonnumeric label such as first year, second year, third year , or fourth year. Alternatively, a numeric code could be used for the class standing variable (e.g. 1 denotes first year, 2 denotes second year, and so on).

- Interval Scale: The data have the properties of ordinal data, and the interval between observations is expressed in terms of a fixed unit of measure. Interval data are always numeric.

E.g.: Melissa has an SAT score of 1205, while Kevin has an SAT score of 1090. Melissa scored 115points more than Kevin.

- Ratio Scale: The data have all the properties of interval data and the ratio of two values is meaningful. Variables such as distance, height, weight, and time use the ratio scale. This scale must contain a zero value that indicates that nothing exists for the variable at the zero point.
  E.g.: Melissa's college record shows 36 credit hours earned, while Kevin's record shows 72 credit hours earned. Kevin has twice as many credit hours earned as Melissa.

1.5 Qualitative and Quantitative Data

Qualitative Data: Labels or names used to identify an attribute of each element are called qualitative data. These are often referred to as categorical data and use either the nominal or ordinal scale of measurements. Qualitative data can be either numeric or nonnumeric and appropriate statistical analyses for them are rather limited.

Quantitative data: Indicate how many or how much. If measuring how many, they are called discrete data and if measuring how much, they are called continuous data. Quantitative data are always numeric and ordinary arithmetic operations are meaningful for them.