Why

We want language and notation for recording elements of a set.

Definition

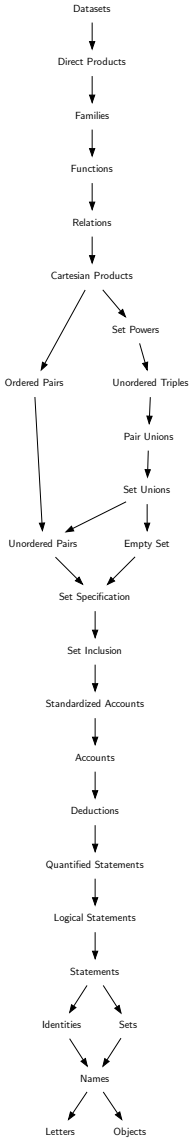
A *dataset* of size n *in* (or *from*) a non-empty set A is an element of the direct product A^n. In other words, it is an n-tuple. In the case that we speak of a dataset, we call its coordinates the records. So, for example, we call the first coordinate the first record and the second coordinate the second record. Of course the ith coordinate is the ith record, for $i = 1, \ldots, n$.

Although the terminology dataset is standard, datasets are not sets. Datasets, in theory and (usually) in practice, contain repeat elements. Thus we use the single word "dataset." Had we used "data set," the language would suggest that the object involved was a set. But it is not; it is a sequence. A minor other point is that "data" is the plural form of a latin noun, and so is analagous to saying something like "records set," instead of something like "record set." For these reasons we prefer the single word dataset.

In the case that the dataset is from a set A which is a product of two sets (say, $\mathcal{U}$ and $\mathcal{V}$, so that $A = \mathcal{U} \times \mathcal{V}$) we say that it is a dataset of *paired* records.

0.1 Notation

Let A be a non-empty set. Although we usually denote sequences by $(a_1, \ldots, a_n)$, we will denote datasets by $(a^1, \ldots, a^n)$; notably, using a superscript. Often A is itself a set of sequences, and so we will want to refer to the ith component of a^1 by a^1_i.


Datasets
Direct Products
Families
Functions
Relations
Cartesian Products
Set Powers
Ordered Pairs
Unordered Triples
Pair Unions
Set Unions
Unordered Pairs
Empty Set
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Identities
Sets
Names
Letters
Objects