

Alternatively the investigator may choose to count the number of individuals of each species in each sample. The cells of the matrix would then contain abundance data rather than presence-absence data. Our investigator may take ten replicate samples in old-growth forest and another ten in managed forest nearby. The species-by-sample matrix now has the columns, which represent different samples, organized in two groups. Species-by-sample matrices such as these are the fundamental data structure for ecological sampling (Pielou 1984). Replicate samples are represented by columns and species are represented by rows (or vice versa). The cell contents may be presence-absence data or they may be abundances. The replicate samples may have no particular order or grouping, or they may be stratified or grouped in various ways.

For very large data sets, a matrix remains the conceptual structure of the data, but not the best way to actually store them. A large matrix, most of the cells of which are empty, is a cumbersome way to store data. Data are better stored as a list, with each row representing a nonzero cell of the matrix. Thus each row contains species name or code number, sample number, any ecological grouping variables (e.g., old growth versus managed forest), and abundance.

Some species will be common in the data set, and others will be rare. Terminology for rare species will become important in some of the analyses discussed later. *Singletons* are species known from a single specimen, and *doubletons* are species known from two. *Uniques* are species that occur in only one sample (regardless of their abundance within the sample), and *duplicates* are species known from two samples (Colwell and Coddington 1994; Coddington et al. 1996; Silva and Coddington 1996; Chazdon et al. 1998).

Caveats to Ant Sampling

Caveat 1: Ants Are Spatially Clumped

Some analysis methods assume that within a spatially and temporally defined community all individuals in the sampling universe have an equal probability of being sampled. In other words, knowing the identity of an individual in a sample should not influence the probability of observing other members of the same species in that sample. This assumption is nearly always violated because spatial aggregation of organisms appears to be the rule rather than the exception (L. R. Taylor et al. 1978). Ants are social and are strongly aggregated when sampling methods capture colonies or portions of colonies. For this reason presence-absence data may be preferred over abundance data in analyses. An extreme example is a pitfall trap that catches 10,000 army ants. In terms of number of individuals, the army ant species might dwarf the abundance of all other species combined, whereas in terms of number of samples in which it occurred the army ant species might be among the rarest.

Caveat 2: Obtaining an Unbiased Sample of Arthropods (Including Ants) Is Nearly Impossible

Ideally the relative abundances of species in a sample should reflect the relative abundances in the community. This is possible if one takes a huge number of small, random volumetric samples from the environment and observes every arthropod in the sample. Arthropods are small and intricately embedded in other biotic and abiotic components of the environment (e.g., soil, wood, leaf litter, foliage, air). Direct searching for arthropods is extremely labor intensive and makes it possible to characterize only very small areas. Rarely is such an exhaustive search practical.