Between-curve distances

Another useful application of multivariate methods was touched on at the end of Section 4, namely the analysis of structured sets of curves or (pseudo-)frequency distributions, generically referred to as sample *profiles*. These include particle- or body-size analyses, or growth curves, with several replicate profiles from each of a number of sites, times, treatments etc. Simple univariate statistical treatment of the size variable is often impossible because of the inherent serial correlation problems (*repeated measures*) of, for example, tracking the body size of a single organism through time, or the lack of a proper frequency distribution structure in histograms of particle sizes (in no sense are we counting independent particles entering the sampling device, to give multinomial frequencies). A viable multivariate alternative is to treat the independent units as the whole profiles and define distances among them, taking these pairwise resemblances into, say, the ANOSIM tests discussed in Section 9. Suitable distance measures between pairs of curves include Euclidean distance D_1 (or its square), the Manhattan distance D_7 and, specifically for comparing cumulative curves:

$D^{\max} = \max_i | y_{i1} - y_{i2} | \text{ \hspace{35mm} Maximum distance,} $

which is also a Distance/dissimilarity option on the •Others list. The maximum departure of two cumulative frequency curves from each other, taken over all the size categories, is the basis of the Kolmogorov-Smirnov test, but the testing structure there relies on real (multinomial) frequencies. Where this is not the case, as often, maximum departure may still be a sensible distance measure of two curves to feed into multivariate analysis, though Manhattan (or Euclidean) distance is likely to be at least as good, since it sums positive contributions across the entire size range.

Revision #3
Created 10 June 2024 23:00:28 by Arden
Updated 15 January 2025 02:44:07 by Abby Miller