

7.1 Overview - Finite factors

ANOVA is one of the most widely used statistical techniques, providing a partitioning of the measured variation of a random variable in response to one or more factors in complex experimental designs and sampling programmes. A *factor* is a categorical variable that identifies several groups or *levels* that are of special interest to the researcher (e.g., treatment vs controls), or that are contributing a potentially important source of variation in the study design (e.g., sites). To make rigorous inferences in multi-factorial ANOVA settings, we need to ascertain, for each and every factor in a given experiment or sampling protocol, whether that factor is *fixed* or *random*. Classically, the levels of a fixed factor are viewed as being finite, while those of a random factor are viewed as being drawn randomly from an infinite (or, at least, an uncountably large) population of possible levels. The choice of whether any given factor is fixed or random is viewed as a dichotomy. The need to make appropriate decisions about this for every factor in the design before embarking on any statistical analysis is essential for dissimilarity-based PERMANOVA, just as it is for univariate ANOVA. There are important consequences of these choices on the results and the inferences that can be drawn from them.

What would happen if we have a factor that would typically be thought of and treated as random, but the population of possible levels is *finite* ? Well, if we can sample *all* of the levels, we might then just treat the random factor as fixed. However, what if we can't sample all of them, but we *can* sample a *substantial fraction* of them? [Anderson et al. \(2025\)](#) describe how the dichotomy of fixed vs random can, instead, be viewed as a *progression*, which depends on how much of the population of possible levels of a given factor has actually been sampled (i.e., the sampling fraction).

Finite factors tend to occur at large spatial scales. For example, suppose there is a cluster of 20 islands in a given region, and suppose 10 of these have undergone some intensive restoration of habitat. We may not be able to sample all of the islands, but perhaps we can sample 4 restored and 4 unrestored islands (in each case, out of a possible 10). By treating the factor of 'Islands' as 'finite', and specifying the size of the population and hence identifying the sampling fraction (the sampling fraction here is 4/10), we are able to get much stronger and more powerful inferences regarding the effectiveness of the restoration than we would otherwise obtain if we were to treat the factor of 'Islands' as random.

Revision #12

Created 22 April 2025 03:53:28 by Marti

Updated 30 November 2025 03:44:50 by Marti