Australian & New Zealand Journal of Statistics

The incremental progression from fixed to random factors in the analysis of variance: a new synthesis

Journal:	*Australian & New Zealand Journal of Statistics*
Manuscript ID	ANZJS-23-02-009.R2
Wiley - Manuscript type:	Original Article - Theory and Methods
Date Submitted by the Author:	08-Jun-2024
Complete List of Authors:	Anderson, Marti; Massey University, New Zealand Institute for Advanced Study Gorley, Ray Terlizzi, Antonio; University of Trieste, Department of Life Sciences
Primary Keywords:	Analysis of variance and covariance, Multivariate Analysis - general, Bootstrap and resampling, Ecological science, Mixed Models, Inference Hypothesis testing, Experimental design and analysis
Secondary Keywords:	Applied statistics, Biostatistics - general, Environmental statistics, Mixed models - Applied statistics
Abstract:	Classically, the distinction between a fixed vs a random factor in analysis of variance has been considered a binary choice. Here, we consider that any given factor can also occur along an incremental series of steps between these two extremes, depending on the sampling fraction of its levels from the wider population. Fixed factors occur where all possible levels are drawn and random factors occur in the limit as the population of possible levels approaches infinity. When some identifiable fraction of a finite population of possible levels are drawn, the factor can be thought of as something in between fixed and random, and can be analysed explicitly as finite directly within the ANOVA framework. Requiring explicit specification of the population size from which observed levels are drawn for each factor, we provide a unified approach to derive expectations of mean squares (EMS) in ANOVA for any types of factors along the entire graded progression from fixed to random, inclusive, that may be nested within or crossed with one another, from balanced, asymmetrical or unbalanced designs, including multi-level hierarchical sampling designs, mixed models and interactions. Implications for estimation of variance components, tailored bootstrap methods, and tests of hypotheses under minimal assumptions of exchangeability are described and further extended to multivariate dissimilarity-based settings.

Aust. N. Z. J. Stat. 2024 doi: 10.1111/j.1467-842X.XXX

The incremental progression from fixed to random factors in the analysis of variance: a new synthesis

Marti J. Anderson[1*], Ray N. Gorley[2] and Antonio Terlizzi[3]

Massey University, PRIMER-e and University of Trieste

Summary

Classically, the distinction between a fixed *vs* a random factor in analysis of variance has been considered a binary choice. Here, we consider that any given factor can also occur along an incremental series of steps between these two extremes, depending on the sampling fraction of its levels from the wider population. Fixed factors occur where all possible levels are drawn and random factors occur in the limit as the population of possible levels approaches infinity. When some identifiable fraction of a finite population of possible levels are drawn, the factor can be thought of as something in between fixed and random, and can be analysed explicitly as finite directly within the ANOVA framework. Requiring explicit specification of the population size from which observed levels are drawn for each factor, we provide a unified approach to derive expectations of mean squares (EMS) in ANOVA for any types of factors along the entire graded progression from fixed to random, inclusive, that may be nested within or crossed with one another, from balanced, asymmetrical or unbalanced designs, including multi-level hierarchical sampling designs, mixed models and interactions. Implications for estimation of variance components, tailored bootstrap methods, and tests of hypotheses under minimal assumptions of exchangeability are described and further extended to multivariate dissimilarity-based settings.

Key words: ANOVA, bootstrap, dissimilarities, expectations of mean squares, finite populations, mixed models, permutation tests, variance components

1. Introduction

Analysis of variance (ANOVA) is one of the most widely used statistical techniques (Speed 1987; Gelman 2005; Großmann 2014), providing a partitioning of the measured variation in a random variable in response to complex experimental designs and sampling programmes (Cochran 1977; Winer, Brown & Michels 1991; Bailey 2008). Derivation of expectations of mean squares (EMS), allowing unbiased estimation of variance components and hypothesis-testing of individual terms in ANOVA models (Searle, Casella & McCulloch

*Author to whom correspondence should be addressed.
[1]New Zealand Institute for Advanced Study, Massey University, Albany Campus, Private Bag 102 904, North Shore Mail Centre, Auckland 0745, New Zealand. Email: m.j.anderson@massey.ac.nz
[2] 14 Stone Park Avenue, Beckenham, Kent BR3 3LS, United Kingdom.
[3] Department of Life Sciences, University of Trieste, v. Giorgio, 10, 34127 Trieste, Italy.

Prepared using anzsauth.cls *[Version: 2018/01/30 Version 9]*

1992), has a long and distinguished history (e.g., Cornfield & Tukey 1956; Hartley 1967; Rao 1968; Searle 1971; Hartley, Rao & Lamotte 1978).

There remains, however, long-standing controversy concerning the 'correct' derivation of EMS, particularly for mixed models and unbalanced designs (Hartley & Searle 1969; Hocking 1973; McLean, Sanders & Stroup 1991; Nelder 2008). Inconsistencies appear largely to be driven by differences in the choice of constraints imposed on parameters of the linear ANOVA model, which is well-known to be overparameterised (e.g., Neter et al. 1996). Yet, different parameterisations must to some extent remain arbitrary, as they yield the same partitioning, sum of squares and mean squares for ANOVA. Nevertheless, Searle, Casella & McCulloch (1992) describe different EMS resulting from models where effect parameters associated with levels of a fixed factor are either assumed to sum to zero – referred to as a 'summation restriction' – or not (see Searle, Casella & McCulloch 1992, p. 121 *et seq.*). The EMS for mixed models that are presumed to hold in the absence of such a restriction are provided by default in certain widely-used statistical computer programs for mixed-effects balanced or unbalanced ANOVA, such as 'proc mixed' in SAS (Littell et al. 2006).

Voss (1999) offered a resolution to the so-called 'mixed models controversy' for the case of a two-way crossed design with interaction (where one factor is fixed and the other is random) by showing that if levels of the random factor are considered to be a finite subset of levels from a very large 'superpopulation' of possible levels, then the EMS obtained under a so-called 'unconstrained-parameter' (UP) model match those obtained under the 'constrained-parameter' (CP) model (i.e., a model with summation restrictions), hence justifying the use of the latter for these two-factor designs. Nevertheless, ongoing disagreements persist, even for this relatively simple case. Lencina, Singer & Stanek (2005) suggested that F test statistics constructed using EMS derived from UP *vs* CP models 'are really directed at different hypotheses'. In contrast, Nelder (1998, 2008) contended that imposing any summation restrictions on ANOVA parameters is a 'temptation' that '*must* be resisted'.

We consider that arbitrary aspects of the parameterisation should not affect the EMS for a given ANOVA design. Rather, we show that by articulating the size of the inference space to which each factor is intended to refer, treating each factor in its proper place along a gradual series of steps (i.e., from being fixed, where all possible levels are observed, to random, where only a small sample of possible levels from an effectively infinite population are observed), the entire debate is rather cleanly dissolved and the resulting EMS distill to a singular, unified solution. Our results hold for ANOVA models that are balanced or unbalanced, including multi-way interactions, mixtures of different types of factors, hierarchical (nested) factors, or any combination(s) of these things. In addition, they do not impose summation restrictions, nor any other arbitrary parameterisation of the ANOVA model in the sample space.

Our approach follows directly from Cornfield & Tukey (1956), who first articulated the concept of the experimenter sampling levels of factors from finite *vs* infinite populations and showed the resulting outcomes for EMS in two-way and three-way crossed balanced designs. We combine this with the landmark work of Hartley (1967), Rao (1968) and Hartley, Rao & Lamotte (1978) for balanced or unbalanced cases, thereby incorporating the sampling fraction from finite populations into the derivation of EMS by 'synthesis' for any general complex ANOVA design. Although Searle & Fawcett (1970) also considered EMS in variance-component models for finite populations, including unbalanced designs, their focus was on random-effect models and they did not articulate the ultimate logical consequence of a sampling fraction equal to 1, which naturally corresponds to a fixed effect.

Motivation for this work arises in (at least) two contexts. First, it resolves prior historical debates concerning EMS and how they may be derived. Although Bayesian estimation techniques are becoming more common (e.g., Gelman 2005), classical EMS in ANOVA models are still relied upon in many settings, including modern computer-intensive multivariate dissimilarity-based analyses, which do not assume normality, but only exchangeability of numerator and denominator in the construction of pseudo-F statistics for hypothesis-testing and estimation in non-Euclidean spaces (Anderson, Gorley & Clarke 2008). These approaches are widely used in ecology (e.g., Anderson et al. 2005; Terlizzi et al. 2007) and, increasingly, in genetics (e.g., Zapala & Schork 2006, 2012), where dissimilarity or distance matrices often form the fundamental starting point for analysis.

Second, we desire flexibility in the definitions of factors for situations where the sampling fraction is neither equal to 1, nor infinitely small. For example, in ecology, environmental impact study designs often contrast responses of organisms (usually counts of abundance or biomass) measured at a purportedly impacted location *vs* one or more 'control' (reference or unimpacted) locations (Underwood 1991, 1992). The control locations are correctly viewed as a random sample from a larger population of control locations that are similar to the impacted location in ways other than the purported impact (which may be, e.g., a sewage outfall, an oil-drilling platform, etc.) (Underwood 1994; Glasby 1997). To increase the power and the scope of inferences, it is desirable to have as many control locations as possible (Glasby 1997; Glasby & Underwood 1998); in practice, however, the possible number of control locations, particularly at large spatial scales, is likely to be both finite and limiting. We show here how such asymmetrical designs can be modeled directly so that a single impact location can be specified as fixed, while reference locations can be treated as either random or drawn from a finite population of a specified size.

In Section 2, we provide the derivation of EMS for ANOVA models that combines the notion of a progression of steps from fixed to random factors (Cornfield & Tukey 1956) with the use of $\mathbf{U}$ matrices (Hartley 1967; Rao 1968), beginning with the one-way design and

moving to two-way crossed and nested designs, including finite sampling of errors. Section 3 then provides a general formulation for any ANOVA model under this 'new synthesis', followed by an extension to dissimilarity-based approaches for multivariate analysis. The implications of the new synthesis for statistical inference, including construction of test-statistics, hypothesis-testing through exchangeability and estimation of variance components, are described in Section 4. An important practical consequence of allowing for finite populations in hierarchical sampling designs is to increase the power to detect effects of factors of interest. This will be demonstrated by example in Section 5 through a case-study of the potential environmental impact of a sewage outfall on assemblages of molluscs in the Mediterranean sea along the coast of Italy (Terlizzi et al. 2005). Section 6 concludes with a brief discussion of alternative methods, including hierarchical Bayesian modelling, and suggestions for future work.

2. Derivation of EMS

2.1. One-way design

Consider a one-way analysis of variance for a factor (say, 'factor A'). Suppose the inference space for factor A contains a total of A possible individual levels of the factor, each with its own effect, α_i^*. Let a sample of size a of values of effects, α_i^*, $i = 1, \ldots, a$, be drawn independently from the population of A possible effects in the full inference space. Note that, for what has classically been referred to as a *fixed* factor, one considers that one has drawn all possible levels in the entire inference space of interest, so $a = A$ and the sampling fraction is $a/A = 1$. In contrast, for what has classically been referred to as a *random* factor, the value of A is considered to be arbitrarily large, and the sampling fraction is treated in the limit as

$$\lim_{A\to\infty} a/A = 0.$$

It is easy to conceive, however, of a finite population of levels, where A is known and $A > a$. Thus, the sampling fraction a/A is neither trivially small, approaching zero (random), yet nor is it precisely equal to 1 (fixed). Thus, as in Cornfield & Tukey (1956), our approach conceptually replaces the binary contrast between a fixed *vs* a random factor with a gradual progression of incremental steps, which depends critically on specification of the number of levels drawn, a, by reference to the size of the inference space, A, and thus on the sampling fraction.

We shall assume that, in the inference space, the effects $\alpha_i^* \ \forall \ i$ are independent and identically distributed as random variables with mean μ_α and variance σ_α^2. Thus, $\mathrm{E}(\alpha_i^* \cdot \alpha_{i'}^*) = \mu_\alpha^2 \ \ \forall \ \ i \neq i'$ and $\mathrm{E}(\alpha_i^* \cdot \alpha_i^*) = \mu_\alpha^2 + \sigma_\alpha^2$. For simplicity in what follows, but

without loss of generality, the values of α_i^* can be centred in the inference space, as in Searle & Fawcett (1970), which gives

$$\alpha_i = \left(\alpha_i^* - \sum_{i=1}^{A} \alpha_i^* / A \right). \tag{1}$$

A key point here is that this does not impose any kind of summation restriction on the original effects α_i^*, which remain in all situations free to vary independently. An important consequence of this is, however, that the resulting α_i are not independent of one another, and this will be true regardless of the size of the sampling fraction inherent in the original design and inference space. Specifically, $\mathrm{E}(\alpha_i) = 0$ but

$$\mathrm{E}(\alpha_i \cdot \alpha_{i'}) = -\tfrac{1}{A} \cdot \sigma_\alpha^2 \ \ \forall \ \ i \neq i', \text{ and} \tag{2}$$

$$\mathrm{E}(\alpha_i \cdot \alpha_i) = \left(1 - \tfrac{1}{A}\right) \cdot \sigma_\alpha^2. \tag{3}$$

See Appendix I for proofs of (2) and (3). Note that $\mathrm{var}(\alpha_i) = \mathrm{var}(\alpha_i^*) = \sigma_\alpha^2$. The variance component, σ_α^2, is generally of greatest interest for estimation, hypothesis-testing and inference. Equations (2) and (3) are succinctly expressed by

$$\mathrm{E}(\alpha_i \cdot \alpha_{i'}) = \left(\delta_{ii'} - \tfrac{1}{A}\right) \cdot \sigma_\alpha^2 \tag{4}$$

where $\delta_{ii'} = 1$ for $i = i'$ and $\delta_{ii'} = 0$ for $i \neq i'$. This fundamental result allows derivation of expectations of mean squares (EMS) in ANOVA models that allows factors to occur anywhere along the fixed-to-random gradation. Equation (4) simultaneously embraces also the two natural end-points of the sequence; namely, for a fixed factor we have $A = a$, and $\sum_{i=1}^{a} \alpha_i = 0$ so $\sigma_\alpha^2 = \sum_{i=1}^{a} \alpha_i^2 / a$, while for a random factor, where $A \to \infty$, we have the standard results

$$\lim_{A \to \infty} \mathrm{E}(\alpha_i \cdot \alpha_{i'}) = 0 \ \ \forall \ \ i \neq i', \text{ and}$$

$$\lim_{A \to \infty} \mathrm{E}(\alpha_i \cdot \alpha_i) = \sigma_\alpha^2.$$

Consider a one-way ANOVA model for a sample of n_i independent observations on a response variable Y within each of the $i = 1, \ldots, a$ levels of factor A. Let $\mathbf{y}$ be a column vector of length $N = \sum_{i=1}^{a} n_i$ containing the observed response values and let $\mathbf{U}_\mathrm{A}$ be a $N \times a$ indicator matrix for the design, where the ith column $\mathbf{u}_i^{[\mathrm{A}]}$ contains a value of 1 in every row where the corresponding observation occurs in group i and zeros elsewhere. Each row of $\mathbf{U}_\mathrm{A}$ has only a single value of 1, indicating the group to which each observation

belongs. The model can be written as

$$\mathbf{y} = \mu\mathbf{1} + \sum_{i=1}^{a} \alpha_i \mathbf{u}_i^{[\mathrm{A}]} + \boldsymbol{\varepsilon} \tag{5}$$

where $\mathbf{1}$ is a $N \times 1$ vector of 1's, μ is an overall mean, and $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of errors, which are assumed to be independent and identically distributed random variables with expectation zero and variance σ_ε^2. We assume also that effects and errors are mutually independent of one another. Without loss of generality, let $\mathbf{y}$ be centred on the overall mean. The expectation of the total sum of squares (SS) under the model is then

$$\mathrm{E}\left(tr\left[\mathbf{y}\mathbf{y}^{\mathsf{T}}\right]\right) = \sigma_\alpha^2 \cdot tr\left[\sum_{i=1}^{a}\sum_{i'=1}^{a} \mathbf{u}_i^{[\mathrm{A}]}\mathbf{u}_{i'}^{[\mathrm{A}]\mathsf{T}}\left(\delta_{ii'} - \tfrac{1}{A}\right)\right] + \sigma_\varepsilon^2 \cdot tr[\mathbf{I}]$$

where "$tr[\cdot]$" indicates the trace of a matrix, superscript "T" indicates the transpose and $\mathbf{I}$ denotes a $N \times N$ identity matrix. We have chosen to consider the SS as the trace of an outer product matrix ($N \times N$) for simplicity in what will follow for the extension to dissimilarity matrices. This expectation holds for any balanced or unbalanced one-way model, does not assume normality, and clearly incorporates the population size (A) from which the observed levels (a) of factor A are drawn.

Next, consider the usual ANOVA partitioning of the total SS. Let $\mathbf{X}$ be a $N \times r$ matrix of full rank r containing orthogonal contrasts for a factor. For example, matrix $\mathbf{X}_\mathrm{A}$ of rank $r = (a-1)$ associated with factor A might be easily obtained by subtracting the last column of $\mathbf{U}_\mathrm{A}$ from each of its previous columns, so that the $(a-1)$ columns of $\mathbf{X}_\mathrm{A}$ contain contrast values ($+1$ *vs* -1) for the contrasts: group 1 *vs* group a, group 2 *vs* group a, ..., group $(a-1)$ *vs* a, and with zeros elsewhere. For the $\mathbf{X}$ matrix associated with any particular model (e.g., a single factor, or a single term, such as an interaction term, or a set of terms), a so-called "hat" matrix $\mathbf{H} = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ provides a classical linear projection of the response values onto the space spanned by the model contained in $\mathbf{X}$.

For the full one-way model in (5), including the overall mean, we have $\mathbf{X}_\text{full} = \left[\mathbf{1} \mid \mathbf{X}_\mathrm{A}\right]$ of rank a. The reduced model is that considered under some null hypothesis. In the one-way case, under the null hypothesis H_0: $\sigma_\alpha^2 = 0$ (or equivalently, H_0: $\alpha_i = 0\ \forall i$), the reduced model is $\mathbf{X}_\text{reduced} = \left[\mathbf{1}\right]$. The orthogonal projection matrix to use for factor A is then obtained by differencing: $\mathbf{H}_\mathrm{A} = \mathbf{H}_\text{full} - \mathbf{H}_\text{reduced}$. This general formulation allows for any type of partitioning (e.g., Type I, II or III SS, *sensu* Searle (2006)) to be obtained for individual terms in ANOVA models in balanced or unbalanced designs to be constructed according to relevant null hypotheses. For the one-way case, the expectation for the among-group SS for

factor A is

$$\mathrm{E}\left(tr\left[\mathbf{H}_{\mathrm{A}}\mathbf{y}\mathbf{y}^{\mathsf{T}}\right]\right) = \sigma_{\alpha}^{2} \cdot tr\left[\mathbf{H}_{\mathrm{A}} \sum_{i=1}^{a} \sum_{i'=1}^{a} \mathbf{u}_{i}^{[\mathrm{A}]} \mathbf{u}_{i'}^{[\mathrm{A}]\mathsf{T}} \left(\delta_{ii'} - \tfrac{1}{A}\right)\right] + \sigma_{\varepsilon}^{2} \cdot tr\left[\mathbf{H}_{\mathrm{A}}\right]$$

while the expectation for the within-group or residual SS is

$$\mathrm{E}\left(tr\left[\left(\mathbf{I} - \mathbf{H}_{\mathrm{full}}\right)\mathbf{y}\mathbf{y}^{\mathsf{T}}\right]\right) = \sigma_{\varepsilon}^{2} \cdot tr\left[\mathbf{I} - \mathbf{H}_{\mathrm{full}}\right].$$

Let $\mathbf{\Delta}_{\mathrm{A}}$ be the $a \times a$ matrix comprised of the elements $\{\Delta_{ii'}\} = \left\{\delta_{ii'} - \frac{1}{A}\right\}$, then the expectation of the mean square (EMS) for factor A is

$$\mathrm{E}\left(\mathrm{MS}_{\mathrm{A}}\right) = \tfrac{1}{(a-1)}\left\{tr\left[\mathbf{H}_{\mathrm{A}}\mathbf{U}_{\mathrm{A}}\mathbf{\Delta}_{\mathrm{A}}\mathbf{U}_{\mathrm{A}}^{\mathsf{T}}\right] \cdot \sigma_{\alpha}^{2} + tr\left[\mathbf{H}_{\mathrm{A}}\right] \cdot \sigma_{\varepsilon}^{2}\right\}$$

and the EMS for the residual is

$$\mathrm{E}\left(\mathrm{MS}_{\mathrm{R}}\right) = \tfrac{1}{(N-a)} tr\left[\mathbf{I} - \mathbf{H}_{\mathrm{full}}\right] \cdot \sigma_{\varepsilon}^{2}$$

These results mirror those obtained by Cornfield & Tukey (1956); however, our derivation on the basis of $\mathbf{U}$ matrices, as in Hartley (1967) and Hartley, Rao & Lamotte (1978) allows the key results of Cornfield & Tukey (1956) to be extended, so that EMS can be readily obtained for any term in any multi-way balanced or unbalanced ANOVA design, including explicit specification of the inference space being examined for every factor in the model.

2.2. Crossed design

Consider a two-way model with factor A having effects α_i^* for levels $i = 1, \ldots, a$ sampled from a population of size A and factor B having effects β_j^* for levels $j = 1, \ldots, b$ sampled from a population of size B. Interaction effects for each of the $a \times b$ combinations of levels of A and B are denoted γ_{ij}^*. The $N \times ab$ indicator matrix for the interaction term $\mathbf{U}_{\mathrm{AB}}$ will have column vectors $\mathbf{u}_{ij}^{[\mathrm{AB}]} = \mathbf{u}_{i}^{[\mathrm{A}]} \circ \mathbf{u}_{j}^{[\mathrm{B}]}$, the Hadamard products of every unique pair of column vectors $\mathbf{u}_{i}^{[\mathrm{A}]}$ and $\mathbf{u}_{j}^{[\mathrm{B}]}$ across the indicator matrices for the main factors $\mathbf{U}_{\mathrm{A}}$ and $\mathbf{U}_{\mathrm{B}}$, respectively. We further stipulate that the resulting columns of $\mathbf{U}_{\mathrm{AB}}$ shall be ordered as $[\mathbf{u}_{11}^{[\mathrm{AB}]}|\mathbf{u}_{12}^{[\mathrm{AB}]}|\ldots|\mathbf{u}_{1b}^{[\mathrm{AB}]}|\mathbf{u}_{21}^{[\mathrm{AB}]}|\mathbf{u}_{22}^{[\mathrm{AB}]}|\ldots|\mathbf{u}_{2b}^{[\mathrm{AB}]}|\ldots|\mathbf{u}_{ab}^{[\mathrm{AB}]}]$ and shall refer to this entire operation as the *concatenation of columnwise Hadamard products of vectors*, denoted hereafter by "$\boxdot$". The two-way crossed model may then be written as

$$\mathbf{y} = \mu\mathbf{1} + \sum_{i=1}^{a} \alpha_i \mathbf{u}_{i}^{[\mathrm{A}]} + \sum_{j=1}^{b} \beta_j \mathbf{u}_{j}^{[\mathrm{B}]} + \sum_{i=1}^{a} \sum_{j=1}^{b} \gamma_{ij} \mathbf{u}_{ij}^{[\mathrm{AB}]} + \boldsymbol{\varepsilon}$$

Table 1. Components to include in model matrices ($\mathbf{X}_{\text{full}}$ and $\mathbf{X}_{\text{reduced}}$) for each term (factor A, factor B and the interaction, A×B) in an ANOVA partitioning of a two-way crossed design using Type I, II or III SS. Note also that Type I SS produces a sequential fit, where the ordering of the terms matters. Below is shown a sequential fit of factor A, then B given A, and finally the interaction given the main effects. In the special case of a balanced design, $\mathbf{X}_{\text{A}}$, $\mathbf{X}_{\text{B}}$ and $\mathbf{X}_{\text{AB}}$ are mutually orthogonal, and the three Types of SS are equivalent.

		A	B	A × B
Type I SS	$\mathbf{X}_{\text{full}}$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}} \mid \mathbf{X}_{\text{AB}}]$
	$\mathbf{X}_{\text{reduced}}$	$[\mathbf{1}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}}]$
Type II SS	$\mathbf{X}_{\text{full}}$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}} \mid \mathbf{X}_{\text{AB}}]$
	$\mathbf{X}_{\text{reduced}}$	$[\mathbf{1} \mid \mathbf{X}_{\text{B}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}}]$
Type III SS	$\mathbf{X}_{\text{full}}$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}} \mid \mathbf{X}_{\text{AB}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}} \mid \mathbf{X}_{\text{AB}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}} \mid \mathbf{X}_{\text{AB}}]$
	$\mathbf{X}_{\text{reduced}}$	$[\mathbf{1} \mid \mathbf{X}_{\text{B}} \mid \mathbf{X}_{\text{AB}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{AB}}]$	$[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}}]$

where parameters β_j are derived from the effect parameters β_j^* for factor B in the same way as α_i were derived from the effect parameters α_i^* for factor A, such that

$$\mathrm{E}(\beta_j \cdot \beta_{j'}) = \left(\delta_{jj'} - \tfrac{1}{B}\right) \cdot \sigma_\beta^2$$

where $\delta_{jj'} = 1$ for $j = j'$, $\delta_{jj'} = 0$ for $j \neq j'$ and σ_β^2 is the variance component for factor B. Similarly, γ_{ij} are derived from the interaction effects γ_{ij}^*, such that $\sum_{i=1}^{a} \gamma_{ij} = 0 \ \forall \ j$ and $\sum_{j=1}^{b} \gamma_{ij} = 0 \ \forall \ i$, with σ_γ^2 denoting the corresponding variance component for the interaction, and we have

$$\mathrm{E}(\gamma_{ij} \cdot \gamma_{i'j'}) = \left(\delta_{ii'} - \tfrac{1}{A}\right)\left(\delta_{jj'} - \tfrac{1}{B}\right) \cdot \sigma_\gamma^2.$$

Next, for the partitioning, an orthogonal matrix of contrasts corresponding to the interaction term, $\mathbf{X}_{\text{AB}}$ can be obtained as the concatenation of columnwise Hadamard products of vectors of the $\mathbf{X}$ matrices for main effects, i.e., $\mathbf{X}_{\text{AB}} = \mathbf{X}_{\text{A}} \boxdot \mathbf{X}_{\text{B}}$. These are then used to produce appropriate projection matrices for each term in the model under a desired Type of SS (Table 1). For example, a projection matrix for the interaction term may be obtained by $\mathbf{H}_{\text{AB}} = \mathbf{H}_{\text{full}} - \mathbf{H}_{\text{reduced}}$, where $\mathbf{X}_{\text{full}} = \left[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}} \mid \mathbf{X}_{\text{AB}}\right]$ and $\mathbf{X}_{\text{reduced}} = \left[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B}}\right]$. This corresponds to either Type II or III SS (*sensu* Searle 2006), explicitly conditioning the interaction term on main effects[†].

[†]For a detailed discussion of the Types of SS and what they measure under a variety of hypotheses, consult Searle (2006).

Having chosen the Type of SS and calculated an associated $\mathbf{H}_{\text{AB}}$ accordingly, the expectation of the SS for the interaction term is then

$$\mathrm{E}\left(tr\left[\mathbf{H}_{\text{AB}}\mathbf{y}\mathbf{y}^{\mathsf{T}}\right]\right) = \sigma^2_\gamma \cdot tr\left[\mathbf{H}_{\text{AB}} \sum_{i=1}^{a}\sum_{i'=1}^{a}\sum_{j=1}^{b}\sum_{j'=1}^{b} \mathbf{u}_{ij}^{[\text{AB}]}\mathbf{u}_{i'j'}^{[\text{AB}]\mathsf{T}} \left(\delta_{ii'} - \tfrac{1}{A}\right)\left(\delta_{jj'} - \tfrac{1}{B}\right)\right] + \sigma^2_\varepsilon \cdot tr[\mathbf{H}_{\text{AB}}].$$

Let $\mathbf{\Delta}_{\text{B}}$ be a $b \times b$ matrix of elements $\{\Delta_{jj'}\} = \left\{\delta_{jj'} - \frac{1}{B}\right\}$ and define $\mathbf{\Delta}_{\text{AB}} = \mathbf{\Delta}_{\text{A}} \otimes \mathbf{\Delta}_{\text{B}}$, a Kronecker product of dimension $ab \times ab$, then the EMS for the interaction term is

$$\mathrm{E}\left(\mathrm{MS}_{\text{AB}}\right) = \tfrac{1}{(a-1)(b-1)}\left\{tr\left[\mathbf{H}_{\text{AB}}\mathbf{U}_{\text{AB}}\mathbf{\Delta}_{\text{AB}}\mathbf{U}_{\text{AB}}^{\mathsf{T}}\right]\cdot\sigma^2_\gamma + tr\left[\mathbf{H}_{\text{AB}}\right]\cdot\sigma^2_\varepsilon\right\}.$$

The EMS for factor A is

$$\mathrm{E}\left(\mathrm{MS}_{\text{A}}\right) = \tfrac{1}{(a-1)}\left\{tr\left[\mathbf{H}_{\text{A}}\mathbf{U}_{\text{A}}\mathbf{\Delta}_{\text{A}}\mathbf{U}_{\text{A}}^{\mathsf{T}}\right]\cdot\sigma^2_\alpha \;+\; tr\left[\mathbf{H}_{\text{A}}\mathbf{U}_{\text{AB}}\mathbf{\Delta}_{\text{AB}}\mathbf{U}_{\text{AB}}^{\mathsf{T}}\right]\cdot\sigma^2_\gamma \;+\; tr\left[\mathbf{H}_{\text{A}}\right]\cdot\sigma^2_\varepsilon\right\}$$

and the EMS for factor B is

$$\mathrm{E}\left(\mathrm{MS}_{\text{B}}\right) = \tfrac{1}{(b-1)}\left\{tr\left[\mathbf{H}_{\text{B}}\mathbf{U}_{\text{B}}\mathbf{\Delta}_{\text{B}}\mathbf{U}_{\text{B}}^{\mathsf{T}}\right]\cdot\sigma^2_\beta \;+\; tr\left[\mathbf{H}_{\text{B}}\mathbf{U}_{\text{AB}}\mathbf{\Delta}_{\text{AB}}\mathbf{U}_{\text{AB}}^{\mathsf{T}}\right]\cdot\sigma^2_\gamma \;+\; tr\left[\mathbf{H}_{\text{B}}\right]\cdot\sigma^2_\varepsilon\right\}.$$

The EMS for each term in the model is a linear form in variance components, with coefficients on each component derived from appropriate $\mathbf{H}$, $\mathbf{U}$ and $\mathbf{\Delta}$ matrices. In each case, the $\mathbf{H}$ matrix identifies contrasts of sampled levels of factors, while $\mathbf{U}$ and $\mathbf{\Delta}$ matrices together reflect the underlying model and population parameter space.

2.3. Nested design

Consider factor B fully nested within factor A such that, for each level i of factor A, $i = 1, \ldots, a$, there are $j = 1, \ldots, b_i$ levels of factor B with associated effects $\beta^*_{j(i)}$ each drawn independently from a population of B_i possible levels. As before, we can assert without loss of generality a centering in the inference space $\beta_{j(i)} = \left(\beta^*_{j(i)} - \sum_{j=1}^{B_i}\beta^*_{j(i)}/B_i\right)\ \forall\ i$ and we have

$$\mathrm{E}(\beta_{j(i)}\cdot\beta_{j'(i)}) = \left(\delta_{jj'} - \tfrac{1}{B_i}\right)\cdot\sigma^2_{\beta(\alpha)} \quad \forall\ i$$

where $\sigma^2_{\beta(\alpha)}$ is the variance component for the nested factor†. Now, consider matrix $\mathbf{U}_{\text{B(A)}} = [\mathbf{u}^{\text{[B(A)]}}_{1(1)}|\mathbf{u}^{\text{[B(A)]}}_{2(1)}|\ldots|\mathbf{u}^{\text{[B(A)]}}_{b_1(1)}|\ldots|\mathbf{u}^{\text{[B(A)]}}_{1(a)}|\mathbf{u}^{\text{[B(A)]}}_{2(a)}|\ldots|\mathbf{u}^{\text{[B(A)]}}_{b_a(a)}]$, a set of ordered indicators for the effects of the nested factor. The two-way nested model is

$$\mathbf{y} = \mu\mathbf{1} + \sum_{i=1}^{a}\alpha_i\mathbf{u}^{\text{[A]}}_i + \sum_{i=1}^{a}\sum_{j=1}^{b_i}\beta_{j(i)}\mathbf{u}^{\text{[B(A)]}}_{j(i)} + \boldsymbol{\varepsilon}.$$

For the ANOVA partitioning, matrix $\mathbf{X}_{\text{B(A)}}$ is constructed by creating a set of contrasts for each set of b_i columns in $\mathbf{U}_{\text{B(A)}}$ separately within each of the $i = 1, \ldots, a$ levels of factor A, i.e., $\mathbf{x}^{\text{[B(A)]}}_{j(i)} = (\mathbf{u}^{\text{[B(A)]}}_{j(i)} - \mathbf{u}^{\text{[B(A)]}}_{b_i(i)}), j = 1, \ldots, (b-1), \forall\, i$ and the number of columns of $\mathbf{X}_{\text{B(A)}}$ is $\sum_{i=1}^{a}(b_i - 1)$. We generally would obtain $\mathbf{H}_{\text{B(A)}}$ as $\mathbf{H}_{\text{full}} - \mathbf{H}_{\text{reduced}}$ where $\mathbf{X}_{\text{full}} = \left[\mathbf{1} \mid \mathbf{X}_{\text{A}} \mid \mathbf{X}_{\text{B(A)}}\right]$ and $\mathbf{X}_{\text{reduced}} = \left[\mathbf{1} \mid \mathbf{X}_{\text{A}}\right]$, while for factor A, the projection matrix $\mathbf{H}_{\text{A}}$ is obtained most naturally using $\mathbf{X}_{\text{full}} = \left[\mathbf{1} \mid \mathbf{X}_{\text{A}}\right]$ and $\mathbf{X}_{\text{reduced}} = \left[\mathbf{1}\right]$, corresponding to a sequential Type I SS.

Let $\boldsymbol{\Delta}_{\text{B(A)}}$ be a block diagonal matrix obtained from the direct sum $[\boldsymbol{\Delta}_{\text{B(1)}} \oplus \boldsymbol{\Delta}_{\text{B(2)}} \oplus \ldots \oplus \boldsymbol{\Delta}_{\text{B}(a)}]$, where $\boldsymbol{\Delta}_{\text{B}(i)}$ is a $b_i \times b_i$ matrix of elements $\left\{\delta_{jj'} - \frac{1}{B_i}\right\}\ \forall\, i$. The expectation of the SS for the nested term is

$$\begin{aligned}\text{E}\left(tr\left[\mathbf{H}_{\text{B(A)}}\mathbf{y}\mathbf{y}^{\text{T}}\right]\right) = \sigma^2_{\beta(\alpha)} \cdot tr\left[\mathbf{H}_{\text{B(A)}}\sum_{i=1}^{a}\sum_{j=1}^{b_i}\sum_{j'=1}^{b_i}\mathbf{u}^{\text{[B(A)]}}_{j(i)}\mathbf{u}^{\text{[B(A)]T}}_{j'(i)}\left(\delta_{jj'} - \tfrac{1}{B_i}\right)\right] \\ + \sigma^2_{\varepsilon} \cdot tr[\mathbf{H}_{\text{B(A)}}].\end{aligned}$$

The EMS for factor A in this model is then

$$\begin{aligned}\text{E}\left(\text{MS}_{\text{A}}\right) = df_{\text{A}}^{-1}\Big\{&tr\left[\mathbf{H}_{\text{A}}\mathbf{U}_{\text{A}}\boldsymbol{\Delta}_{\text{A}}\mathbf{U}^{\text{T}}_{\text{A}}\right]\cdot\sigma^2_{\alpha} + \\ &tr\left[\mathbf{H}_{\text{A}}\mathbf{U}_{\text{B(A)}}\boldsymbol{\Delta}_{\text{B(A)}}\mathbf{U}^{\text{T}}_{\text{B(A)}}\right]\cdot\sigma^2_{\beta(\alpha)} + \\ &tr\left[\mathbf{H}_{\text{A}}\right]\cdot\sigma^2_{\varepsilon}\Big\}\end{aligned}$$

and the EMS for the nested term is

$$\begin{aligned}\text{E}\left(\text{MS}_{\text{B(A)}}\right) = df_{\text{B(A)}}^{-1}\Big\{&tr\left[\mathbf{H}_{\text{B(A)}}\mathbf{U}_{\text{B(A)}}\boldsymbol{\Delta}_{\text{B(A)}}\mathbf{U}^{\text{T}}_{\text{B(A)}}\right]\cdot\sigma^2_{\beta(\alpha)} + \\ &tr\left[\mathbf{H}_{\text{B(A)}}\right]\cdot\sigma^2_{\varepsilon}\Big\},\end{aligned}$$

where degrees of freedom for A and B(A) are denoted by $df_{\text{A}} = (a-1)$ and $df_{\text{B(A)}} = \sum_{i=1}^{a}(b_i - 1)$, respectively.

† We could also rather straighforwardly allow for different values of $\sigma^2_{\beta(\alpha)}$ within different levels of factor A, but this is not addressed in any further detail here.

2.4. Finite population sampling

The nested case immediately points to the obvious extension of sampling $l(k) = 1, \ldots, n_k$ units from a finite population of N_k possible sampling units within each of $k = 1, \ldots, c$ cells (combinations of all sampled levels of all factors) in a given ANOVA design, with $\sum_{k=1}^{c} n_k = N$. Each observation $y_{l(k)}$ has an associated unknown error $\varepsilon^*_{l(k)}$, and derivation of the EMS follows directly by considering this as a nested term in the ANOVA model, where the indicator matrix for the error term, $\mathbf{U}_{\mathrm{R}}$, is $\mathbf{I}$ and $\mathbf{\Delta}_{\mathrm{R}}$ is a $N \times N$ block diagonal matrix comprised of the direct sum $[\mathbf{\Delta}_{\mathrm{R}(1)} \oplus \mathbf{\Delta}_{\mathrm{R}(2)} \oplus \ldots \oplus \mathbf{\Delta}_{\mathrm{R}(c)}]$, where $\mathbf{\Delta}_{\mathrm{R}(k)}$ is a $n_k \times n_k$ matrix of elements $\left\{\delta_{ll'} - \frac{1}{N_k}\right\}\ \forall\ k$. For the one-way case under finite sampling, the EMS for factor A is

$$\mathrm{E}\left(\mathrm{MS}_{\mathrm{A}}\right) = df_{\mathrm{A}}^{-1}\left\{tr\left[\mathbf{H}_{\mathrm{A}}\mathbf{U}_{\mathrm{A}}\mathbf{\Delta}_{\mathrm{A}}\mathbf{U}_{\mathrm{A}}^{\mathsf{T}}\right]\cdot\sigma_{\alpha}^{2} + tr\left[\mathbf{H}_{\mathrm{A}}\mathbf{\Delta}_{\mathrm{R}}\right]\cdot\sigma_{\varepsilon}^{2}\right\}$$

and the EMS for the residual is

$$\mathrm{E}\left(\mathrm{MS}_{\mathrm{R}}\right) = df_{\mathrm{R}}^{-1}\left\{tr\left[\left(\mathbf{I} - \mathbf{H}_{\mathrm{full}}\right)\mathbf{\Delta}_{\mathrm{R}}\right]\cdot\sigma_{\varepsilon}^{2}\right\}$$

where, as above, $df_{\mathrm{A}} = (a-1)$ and $df_{\mathrm{R}} = \sum_{k=1}^{c}(n_k - 1)$ denote degrees of freedom for MS_{A} and MS_{R}, respectively.

3. A new synthesis

3.1. General result

Consider any ANOVA model comprising a partitioning among a total of T terms, including one or more main effects (e.g., A, B, C, etc.), possibly some interaction terms (e.g., A×B, A×B×C, etc.) and/or some nested terms (e.g., B(A), C(A×B), etc.), and a residual (R). Now, as outlined in detail in Section 2 and Appendix II, every term $t = 1, \ldots, T$, with degrees of freedom df_t, has an associated variance component σ_t^2 and an indicator matrix $\mathbf{U}_t$, according to the model, along with an associated matrix $\mathbf{\Delta}_t$ which will identify sizes of populations of levels for each factor according to the experimental design. In addition, every term $t = 1, \ldots, T$ has an identifiable contrast matrix $\mathbf{X}_t$ (of rank df_t) and associated projection matrix $\mathbf{H}_t$, which provides the desirable partitioning according to a chosen Type of SS. For any model where there exists replication of sampling units within one or more cells, we assert that the T^{th} term in the full ANOVA model will be the residual, that $\mathbf{U}_{t=T} = \mathbf{I}$, that $\mathbf{\Delta}_{t=T} = \mathbf{\Delta}_{\mathrm{R}}$ as in Section 2.4, and that $\mathbf{H}_{t=T} = (\mathbf{I} - \mathbf{H}_{\mathrm{full}})$ where $\mathbf{H}_{\mathrm{full}}$ is constructed from $\mathbf{X}_{\mathrm{full}} = \left[\mathbf{1} \mid \mathbf{X}_{t=1} \mid \mathbf{X}_{t=2} \mid \ldots \mid \mathbf{X}_{t=(T-1)}\right]$.

The EMS for any term of interest Ω is then simply given by

$$\mathrm{E}\left(\mathrm{MS}_{\Omega}\right) = df_{\Omega}^{-1}\left\{\sum_{t=1}^{T} tr\left[\mathbf{H}_{\Omega}\mathbf{U}_{t}\boldsymbol{\Delta}_{t}\mathbf{U}_{t}^{\mathsf{T}}\right]\cdot\sigma_{t}^{2}\right\}. \tag{6}$$

A proof of (6) is outlined in Appendix II. This general synthesis holds for any term in any balanced or unbalanced ANOVA design, for any types of factors (i.e., fixed, random, or anywhere along the spectrum of steps from fixed to random) in any partitioning that may be desired using Type I, II or III SS.

3.2. Multivariate extensions

We extend the above to multivariate analysis *via* redundancy analysis (RDA; Rao 1964; van den Wollenberg 1977) and distance-based redundancy analysis (dbRDA; Legendre & Anderson 1999; McArdle & Anderson 2001). Consider $N \times p$ matrix $\mathbf{Y}$ with column vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p$ being the observed values for each of $\ell = 1, \ldots, p$ response variables Y_ℓ obtained simultaneously from each of $i = 1, \ldots, N$ sampling units. Without loss of generality, suppose also that each variable is centred on its overall sample mean. Consider also a single ANOVA design and associated model, applicable to all response variables simultaneously, with separate individual vectors of parameters for each variable accordingly; e.g., in the one-way case, we may have, for each variable ℓ,

$$\mathbf{y}_{\ell} = \mu_{\ell}\mathbf{1} + \mathbf{U}_{\ell}\boldsymbol{\alpha}_{\ell} + \boldsymbol{\varepsilon}_{\ell}.$$

Correspondingly, for every variable ℓ there will be a variance component $\sigma_{\ell t}^2$ for each term $t = 1, \ldots, T$ in the ANOVA model. Indeed, for any term, there will be a full $p \times p$ variance-covariance matrix of parameters, $\boldsymbol{\Sigma}_t$. RDA is a type of multivariate multiple regression, where interest lies in explaining variation in matrix $\mathbf{Y}$ by reference to factors or explanatory variables held in $\mathbf{X}$. For further details regarding RDA, see chapter 11 in Legendre & Legendre (2012). RDA, however, effectively ignores covariance structures among the variables in $\mathbf{Y}$ and achieves a partitioning of the total SS, defined as $tr\left[\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\right] = \sum_{\ell=1}^{p} tr\left[\mathbf{y}_{\ell}\mathbf{y}_{\ell}^{\mathsf{T}}\right]$, given some projection matrix $\mathbf{H}$ (constructed from matrix $\mathbf{X}$), as

$$tr\left[\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\right] = tr\left[\mathbf{H}\,\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\right] + tr\left[\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\right].$$

Thus, the direct generalisation of 6 for RDA is clearly given by

$$\mathrm{E}\left(\mathrm{MS}_{\Omega}\right) = df_{\Omega}^{-1}\left\{\sum_{t=1}^{T} tr\left[\mathbf{H}_{\Omega}\mathbf{U}_{t}\boldsymbol{\Delta}_{t}\mathbf{U}_{t}^{\mathsf{T}}\right]\cdot tr\left[\boldsymbol{\Sigma}_{t}\right]\right\}. \tag{7}$$

To extend this to a dissimliarity-based setting, suppose distances or dissimilarities are calculated among every pair of sampling units based on all p variables, yielding $N \times N$ matrix $\mathbf{D} = \{d_{ii'}\}, (i = 1, \ldots, N, i' = 1, \ldots, N)$. Using the transformation of Gower (1966) to obtain $\mathbf{G} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathsf{T}}\right)\mathbf{A}\left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathsf{T}}\right)$, where $\mathbf{A} = \{a_{ii'}\} = \left\{-\frac{1}{2}d_{ii'}^2\right\}$, the dbRDA partitioning for any projection matrix $\mathbf{H}$ is then

$$tr\left[\mathbf{G}\right] = tr\left[\mathbf{H}\,\mathbf{G}\right] + tr\left[\left(\mathbf{I} - \mathbf{H}\right)\mathbf{G}\right]$$

and the generalisation of (6) for dbRDA is

$$\mathrm{E}\left(\mathrm{MS}_{\Omega}\right) = df_{\Omega}^{-1}\left\{\sum_{t=1}^{T} tr\left[\mathbf{H}_{\Omega}\mathbf{U}_t\mathbf{\Delta}_t\mathbf{U}_t^{\mathsf{T}}\right]\cdot\vartheta_t\right\} \tag{8}$$

where ϑ_t is a pseudo multivariate component of variation defined in the space of the chosen dissimilarity measure (see Section 4.2 below and Anderson et al. (2005) for an example and interpretation). Note that in the special case where $\mathbf{D}$ contains Euclidean distances calculated directly on $\mathbf{Y}$, then $\mathbf{Y}\mathbf{Y}^{\mathsf{T}} = \mathbf{G}$, dbRDA is equivalent to RDA and (8) reduces to (7) as $\vartheta_t = tr\left[\mathbf{\Sigma}_t\right]$. Furthermore, if $\mathbf{D}$ contains Euclidean distances and $p = 1$, then dbRDA is equivalent to a univariate ANOVA partitioning along a single axis, hence (8) reduces to (6) and $\vartheta_t = \sigma_t^2$. See Legendre & Anderson (1999), McArdle & Anderson (2001) and Anderson (2017) for more details regarding the relationship between dbRDA, RDA and ANOVA. Although other forms of partitioning multivariate systems based on dissimilarities are available (e.g., Anderson & Robinson 2003), we shall not pursue these further here.

4. Estimation and hypothesis-testing

4.1. The generalised F-test

We shall presume that interest lies in estimating ϑ_Ω and in testing hypotheses of the type H_0: $\vartheta_\Omega = 0$, where Ω signifies a term (or set of terms) of interest in an ANOVA model. We shall let Φ denote the set of all terms in the model. Recognizing that ϑ_Ω may represent a classical univariate variance component, or a multivariate measure of variation in the space of some chosen distance or dissimilarity measure, we shall refer to ϑ_Ω generally as a "component of variation" attributable to term Ω.

Let $\mathbf{W}_t = \mathbf{U}_t\mathbf{\Delta}_t\mathbf{U}_t^{\mathsf{T}}$ and suppose also that the coefficient on the component of variation for term t in the EMS for term Ω is given by $K_{\Omega t} = df_{\Omega}^{-1}\mathbf{H}_{\Omega}\mathbf{W}_t$, then our synthesis is

$$\mathrm{E}\left(\mathrm{MS}_{\Omega}\right) = \sum_{t=1}^{T} K_{\Omega t}\cdot\vartheta_t$$

which is the sum of two parts, namely

$$K_{\Omega\Omega} \cdot \vartheta_\Omega \quad \text{and} \quad \sum_{t=1}^{T_{\Omega-}} K_{\Omega t} \cdot \vartheta_t, \quad t \in \Phi_{\Omega-}$$

where $\Phi_{\Omega-}$ is the set of all terms, but omitting the term Ω itself, and $T_{\Omega-}$ is the number of terms in the set $\Phi_{\Omega-}$.

Now, let $\boldsymbol{M}$ denote a vector of length T containing the mean squares of all terms in Φ. There exist two vectors, $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$, each of length T, that contain strictly non-negative values such that

$$\mathrm{E}\left(\boldsymbol{l}_1^\mathsf{T}\boldsymbol{M}\right) = \mathrm{E}\left(\boldsymbol{l}_2^\mathsf{T}\boldsymbol{M}\right) + K_{\Omega\Omega} \cdot \vartheta_\Omega. \tag{9}$$

We then readily build a pseudo F statistic

$$F_\Omega = \frac{\boldsymbol{l}_1^\mathsf{T}\boldsymbol{M}}{\boldsymbol{l}_2^\mathsf{T}\boldsymbol{M}}$$

as a suitable criterion to test H_0: $\vartheta_\Omega = 0$. Even for univariate cases, a ratio of linear combinations of mean squares is not necessarily distributed as a classical F statistic (e.g., Searle, Casella & McCulloch 1992), so we shall generally seek to calculate approximate p values empirically using resampling methods. However, for many balanced designs, there may be a single mean square $\mathrm{MS}_{\text{denom}}$ available from the ANOVA model partitioning such that

$$\mathrm{E}\left(\mathrm{MS}_\Omega\right) = \mathrm{E}\left(\mathrm{MS}_{\text{denom}}\right) + K_{\Omega\Omega} \cdot \vartheta_\Omega$$

and hence

$$F_\Omega = \frac{\mathrm{MS}_\Omega}{\mathrm{MS}_{\text{denom}}}, \tag{10}$$

all other values in $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ corresponding to other mean squares being equal to zero, respectively. Note further that, in the special case where $p = 1$ and $\mathbf{D}$ contains Euclidean distances calculated directly on $\mathbf{Y}$, F_Ω in (10) is the usual classical univariate F statistic, hence with known distribution under an additional assumption of normality of the errors in this case.

Values of F_Ω under a true null hypothesis may be generated under a simple assertion of exchangeability of appropriate exchangeable units, following and extending the work of Anderson & ter Braak (2003). Specifically, let Ψ denote the set of all terms *other* than Ω having non-zero values in either $\boldsymbol{l}_1$ or $\boldsymbol{l}_2$ (i.e., the collection of all terms other than Ω appearing in either the numerator or denominator of F_Ω), and let Ψ_F denote the set of factors (main effects) that appear *within* any of the terms listed in Ψ (e.g., A and B both appear within the term A×B; they also both appear within the term B(A), *et cetera*). The exchangeable units

for the test are then identifiable as the combination of all sampled levels of factors listed in Ψ_F. For example, suppose $\Omega = B$ and we discover that $\Psi = \{A\times B, C(B), A\times C(B), R\}$, then $\Psi_F = \{A, B, C\}$ and the exchangeable units are the abc cells corresponding to all combinations of sampled levels of factors A, B and C. All of the individual observations within each cell "travel together" as a single unit under permutation. If $\Psi = \{R\}$, then the exchangeable units are simply the original N sampling units.

Having identified exchangeable units, let F_Ω^π be a value of F_Ω obtained after a random equiprobable permutation of the exchangeable units across the full study design[§]; a p-value associated with the test of H_0: $\vartheta_\Omega = 0$ is obtained directly and empirically as $P(F_\Omega^\pi \geq F_\Omega)$.

4.2. Components of variation

The result in (9) above also provides directly an unbiased ANOVA-type estimator (*sensu* Searle, Casella & McCulloch 1992) for ϑ_Ω as

$$\hat{\vartheta}_\Omega = \frac{(\boldsymbol{l}_1 - \boldsymbol{l}_2)^\mathsf{T}\, \boldsymbol{M}}{K_{\Omega\Omega}}. \tag{11}$$

Note that $\hat{\vartheta}_\Omega$ in the multivariate case is distinguishable from a classical estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ because the former is a scalar that ignores covariance structures. For a univariate case, $\sqrt{\hat{\vartheta}_\Omega}$ is a standard deviation, while, for dissimilarity spaces, $\sqrt{\hat{\vartheta}_\Omega}$ is interpretable as a standard "distance-to-centroid" in the same units as the chosen dissimilarity measure.

Although (11) can produce a negative estimate for ϑ_Ω, such cases usually go hand-in-hand with large p-values in the associated test of H_0: $\vartheta_\Omega = 0$. In such cases, a logical course of action is generally to remove the term Ω from the model[¶] and re-estimate the EMS and components of variation for the remaining terms (Fletcher & Underwood 2002).

With respect to estimation, we note in passing that there is considerable literature regarding situations when best linear unbiased estimators (BLUEs) of estimable functions are unchanged by a change in covariance structure. For example, BLUEs are unchanged when a linear model with uncorrelated errors is instead fitted using an error covariance matrix with equi-correlated structure. For more details, see Haslett & Puntanen (2013), Haslett, Puntanen

[§]We note that permutation of residuals under either a full or reduced model, as described by Anderson & ter Braak (2003) would also be possible here; however, the identification of exchangeable units is always of paramount importance, and must be observed whether one permutes the original observations or some form of residuals.

[¶]If there is more than one such term in a given model, this is done sequentially in a step-wise fashion, beginning with removal of the term having the smallest MS, and with re-estimation of all estimates at each step.

& Arendacká (2015), Haslett et al. (2023) and references to earlier work by C. R. Rao and others therein.

4.3. A tailor-made bootstrap

Confidence intervals for individual components of variation can be estimated using the bootstrap (Efron 1982), with due consideration of the structure inherent in the sampling design. Bootstrap estimates of a variance are known to be biased (e.g., see Efron (1982); see also Appendix B in Anderson et al. (2017)), so some form of correction for the bias is desirable (e.g., Hall 1992; Efron & Tibshirani 1993). We suggest that bootstrap realisations be constructed *via* independent resampling (with replacement) of bias-adjusted residuals and effects (e.g., Davison & Hinkley 1997; Carpenter, Goldstein & Rasbash 2003; Field & Welsh 2007).

First, consider a one-way univariate ANOVA model, as in (5). Let $\hat{\alpha}_i$ denote estimated effects and $\hat{\varepsilon}_{ij}$ denote residuals (estimated errors), for $i = 1, \ldots, a$ and $j = 1, \ldots, n_i$. The average of the squared estimated effects is given by

$$S_{\mathrm{A}} = \sum_{i=1}^{a} \hat{\alpha}_i^2 / a$$

To adjust for bias under bootstrapping, we can find a constant κ_{A} that transforms the individual estimated effects such that their average square is equal to the unbiased estimated variance component $\hat{\sigma}_\alpha^2 = \hat{\vartheta}_{\mathrm{A}}$ from (11). Specifically, provided $\hat{\vartheta}_{\mathrm{A}} > 0$,

$$\kappa_{\mathrm{A}} = (\hat{\vartheta}_{\mathrm{A}} / S_{\mathrm{A}})^{\frac{1}{2}}$$

yielding transformed effects

$$\tilde{\alpha}_i = \kappa_{\mathrm{A}} \hat{\alpha}_i.$$

(If $\hat{\vartheta}_{\mathrm{A}} \leq 0$, then we may assume $\sigma_\alpha^2 = 0$ is true and set $\tilde{\alpha}_i = 0 \; \forall \; i$). Generally, if $\sigma_\alpha^2 > 0$, then we expect $\hat{\vartheta}_{\mathrm{A}} > S_{\mathrm{A}}$, so transformed effects are "inflated" to counter the bias under bootstrapping. Smilarly, we have transformed residuals

$$\tilde{\varepsilon}_{ij} = \kappa_{\mathrm{R}} \hat{\varepsilon}_{ij}$$

where $\kappa_{\mathrm{R}} = (\hat{\vartheta}_{\mathrm{R}} / S_{\mathrm{R}})^{\frac{1}{2}} = (N/(N-a))^{\frac{1}{2}}$.

Now, we draw separate independent bootstrap resamples with replacement of the transformed effects $\tilde{\alpha}_i$ (a times) and $\tilde{\varepsilon}_{ij}$ (N times) to obtain α_i^{ℓ} and ε_{ij}^{ℓ}. A new bootstrap

realisation is then constructed as

$$y_{ij}^{\ell} = \hat{\mu} + \alpha_i^{\ell} + \varepsilon_{ij}^{\ell} \tag{12}$$

where $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}$. Bootstrap estimates of variance components σ_α^2 and σ_ε^2 are given directly through (11) as $\hat{\vartheta}_{\mathrm{A}}^{\ell}$ and $\hat{\vartheta}_{\mathrm{R}}^{\ell}$, respectively. Repeating the bootstrapping procedure a large number of times yields distributions of each of $\hat{\vartheta}_{\mathrm{A}}^{\ell}$ and $\hat{\vartheta}_{\mathrm{R}}^{\ell}$. The $(\varrho/2)$ and $(1-\varrho/2)$ quantiles of each of these empirical distributions can be used directly to provide a $(1-\varrho)\%$ confidence interval for each of σ_α^2 and σ_ε^2, respectively. If desired, the full bootstrap distribution (as a histogram or density) can also optionally be plotted (Schweder & Hjort 2016; Fletcher, Dillingham & Zeng 2019).

If the sampling fraction $a/A = 1$, then we neither transform nor bootstrap resample the estimated effects $\hat{\alpha}_i$, and the new bootstrap sample is constructed as

$$y_{ij}^{\ell} = \hat{\mu} + \hat{\alpha}_i + \varepsilon_{ij}^{\ell}$$

In this case, the bootstrap distribution of $\hat{\vartheta}_{\mathrm{A}}^{\ell}$ may be adjusted *a posteriori* for any bias by, for example, centering its median on $\hat{\vartheta}_{\mathrm{A}}$.

Next, suppose A is known and finite such that the sampling fraction a/A is less than 1 but is not trivially small. In this case, we can begin by expanding the a levels out to A levels by sampling transformed effects $\tilde{\alpha}_i$ a total of $(A-a)$ times *with* replacement. These, alongside the a original values of $\tilde{\alpha}_i$, provide a finite population of bootstrapped effects that is now of length A (i.e., $\alpha_{i'}^{\ell}$, $i' = 1, \ldots, A$). We then sample a times from these A levels *without* replacement and apply (12). The limiting case of $a = A$ yields the same result as the fixed factor case, while the limiting case as $A \to \infty$ yields the random factor case. We note that our proposed resampling approach here conceptually resembles the 'mirror-match' bootstrap method of Sitter (1992a,b) for stratified random samples without replacement. Results regarding formal theoretical properties and the empirical behaviour of our proposed approach across a variety of contexts, including comparisons with other potential methods (e.g., Sitter 1992b; Booth, Butler & Hall 1994; Saigo 2010) remains a topic for future research.

4.4. Bootstrapping in the multivariate dissimilarity-based context

Consider the partitioning of a data cloud defined by an $N \times N$ dissimilarity matrix $\mathbf{D}$ according to any multi-factor ANOVA model with T terms. Unbiased estimators $\hat{\vartheta}_t$ are obtained for components of variation associated with each term in the model $t = 1, \ldots, T$, using (11). Let the $N \times N$ matrix $\mathbf{Q}$ denote the principal coordinate (PCO) axes

(Gower 1966), obtained by finding latent roots (eigenvalues) $\lambda_1, \lambda_2, \ldots, \lambda_N$ and associated eigenvectors of matrix $\mathbf{G}$, each eigenvector $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N$ being scaled such that its sum of squares is equal to the absolute value of its corresponding latent root.

Restricting ourselves here to the use of a Type I (sequential) SS in the partitioning and construction of appropriate projection matrices $\mathbf{H}_t$, the estimated effects $\hat{\mathbf{T}}_t$ associated with each term t, having individual columns $\boldsymbol{\tau}_{t\ell}$, $\ell = 1, \ldots, N$, one for each PCO axis, are given by

$$\hat{\mathbf{T}}_t = \mathbf{H}_t \mathbf{Q}_t$$

So, the $1 \times N$ vector $\boldsymbol{S}_t = \{S_{t\ell}\}$ of average squared estimated effects for term t on each of the $\ell = 1, \ldots, N$ PCO axes is given by

$$\boldsymbol{S}_t = \left(\tfrac{1}{N}\right) \mathbf{1}^{\mathsf{T}} \hat{\mathbf{T}}_t \circ \hat{\mathbf{T}}_t$$

Next we can estimate components of variation $\hat{\vartheta}_{t\ell}$ for each term t along each PCO axis $\mathbf{q}_\ell$ by first calculating the mean squares of the ANOVA model

$$\mathrm{MS}_{t\ell} = df_t^{-1} tr\left[\mathbf{H}_t \mathbf{q}_\ell \mathbf{q}_\ell^{\mathsf{T}} \mathbf{H}_t\right]$$

and repeatedly applying (11) for each PCO axis in turn.

The N-length vector of required transformation constants $\boldsymbol{\kappa}_t = \{\kappa_{t\ell}\}$ has elements

$$\kappa_{t\ell} = \left(\hat{\vartheta}_{t\ell} / S_{t\ell}\right)^{\frac{1}{2}} \quad \text{if } \hat{\vartheta}_{t\ell} \geq 0,$$

else $\kappa_{t\ell} = 1$. Transformed effects are then given by

$$\tilde{\mathbf{T}}_t = \hat{\mathbf{T}}_t \cdot \mathrm{diag}\left(\boldsymbol{\kappa}_t\right).$$

We perform the bootstrap operation paying close attention to the original sampling design. Each effect is sampled independently of other effects and the number of bootstrap samples taken for a given effect is equal to the number of sampled levels for the associated term in the original study design. Furthermore, the inherent multivariate structure is maintained empirically under bootstrapping by sampling the entire row of effects simultaneously across all PCO axes in a given bootstrap draw (e.g., for factor A in a one-way design, we will sample a rows of the $N \times N$ matrix $\tilde{\mathbf{T}}_{\mathrm{A}}$ then replicate each row an appropriate number of times (n_i) to obtain bootstrap matrix $\mathbf{T}_{\mathrm{A}}^{\ell b}$.

Let Φ denote the set of terms having a sampling fraction of 1 (i.e., fixed effects) and Υ denote the set of terms having a sampling fraction < 1 and we generate bootstrap samples for

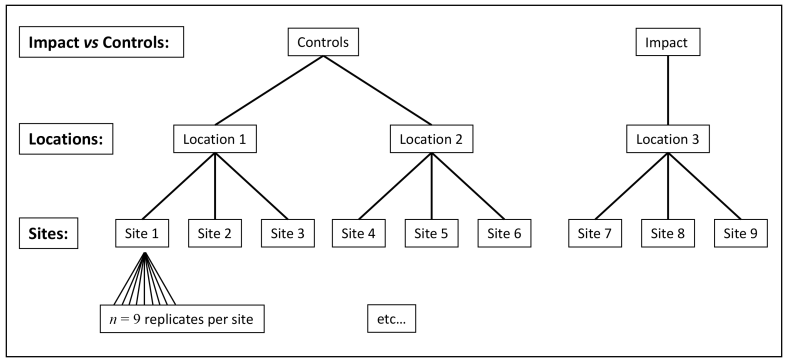

Figure 1. Schematic diagram of an hierarchical sampling design used by Terlizzi et al. (2005) to examine effects of a sewage outfall on subtidal molluscan assemblages.

the full design in the space of the full set of N PCO axes as

$$\mathbf{Q}^{\oint} = \sum_{t \in \Phi} \hat{\mathbf{T}}_t + \sum_{t \in \Upsilon} \tilde{\mathbf{T}}_t^{\oint}$$

To generate the bootstrap dissimilarity matrix, we must potentially carefully consider the PCO axes in two separate sets $\mathbf{Q}^{\oint} = \left[\mathbf{Q}^{\oint +} \mid \mathbf{Q}^{\oint -}\right]$; namely, those corresponding to positive eigenvalues ($\lambda_\ell \geq 0$, denoted $\mathbf{Q}^{\oint +}$) and those corresponding to strictly negative eigenvalues ($\lambda_\ell < 0$, denoted $\mathbf{Q}^{\oint -}$) (see McArdle & Anderson 2001). Let the separate $N \times N$ Euclidean distance matrices based on $\mathbf{Q}^{\oint +}$ and $\mathbf{Q}^{\oint -}$ be denoted by $\mathbf{D}^{\oint +}$ and $\mathbf{D}^{\oint -}$, respectively. The bootstrap dissimilarity matrix is then given by $\mathbf{D}^{\oint}$ with elements

$$d_{ii'}^{\oint} = \left| \left(d_{ii'}^{\oint +} \right)^2 - \left(d_{ii'}^{\oint -} \right)^2 \right|^{\frac{1}{2}}$$

We calculate $\mathbf{G}^{\oint}$ directly from $\mathbf{D}^{\oint}$ in the usual way and readily obtain $\hat{\vartheta}_t^{\oint}$ for each term t by applying (11) accordingly, and confidence intervals based on percentiles are directly accessible, just as for the univariate case.

5. Example: Mediterranean molluscs

We demonstate the effect of finite inference spaces in mixed models and the increased power afforded by our proposed approach through the multivariate analysis of 151 molluscan species in Mediterranean rocky subtidal habitats (depth = 3–4 m) located along the south-western coast of Apulia, Ionian Sea, Italy in response to potential impacts of a sewage outfall

Table 2. PERMANOVA using the new synthesis and Type I (sequential) SS, based on Bray-Curtis dissimilarities for 151 species of mollusc according to the experimental design described by Terlizzi et al. (2005), with p values obtained using 9999 random permutations of exchangeable units; "Unique F^{π}" indicates the number of unique values of F obtained under permutation for each term in the model.

Source (Ω)	*df*	SS	F_{Ω}	P	Unique F^{π}	$\hat{\vartheta}_{\Omega}$
IvC	1	30352	3.794	0.0221	839	620.85
Locations(IvC)	1	10507	1.466	0.1985	830	123.77
Sites(Locations(IvC))	6	42992	4.105	0.0001	9804	602.17
Residual	72	125690				1745.7
Total	80	209540				

(Terlizzi et al. 2005). Counts of abundance of each species were obtained from each of $n = 9$ replicate 20 cm $\times$ 20 cm quadrats in each of three random sites (separated by 80–100 m) at the outfall location ("I", putatively impacted) and at each of two control locations ("C1" and "C2"). Control locations were chosen randomly from a set of eight possible locations separated by at least 2.5 km that provided comparable environmental conditions (in terms of slope, wave exposure, type of substrate) to those occurring at the outfall (Glasby & Underwood 1998). This yielded an asymmetrical sampling design (Fig. 1) having a total of $N = 81$ sampling units, where greatest interest lies in examining the (fixed) contrast of I *vs* Cs (Glasby 1997).

The ANOVA model arising from this design has three factors: Impact *vs* Controls ("I*v*C", fixed with $a = 2$ levels), Locations ("L", nested in I*v*C, with $b_1 = 2$ controls and $b_2 = 1$ impact) and Sites ("S", random and nested in L with $c_{j(i)} = 3 \ \forall \ j = 1, \ldots, b_i$; $i = 1, \ldots, a$). The $b_1 = 2$ control locations were chosen randomly from a finite population of size $B_1 = 8$, yielding a sampling fraction of $b_1/B_1 = \frac{1}{4}$, while there was only one (fixed) impact location, hence $b_2/B_2 = 1$. This example thus serves also to show that our new synthesis caters easily to asymmetry not just in sampled numbers of levels, but also in sizes of inference spaces for different sub-sets of an hierarchical design.

A partitioning of the multivariate variation according to the full ANOVA model on the basis of a Bray-Curtis dissimilarity matrix was done using the new synthesis and Type I (sequential) SS, with associated tests obtained using 9999 random permutations of appropriate exchangeable units for each term (i.e., PERMANOVA; Anderson 2001; Anderson, Gorley & Clarke 2008). The analysis provides evidence against H_0: $\vartheta_{\mathrm{IvC}} = 0$, suggesting a significant impact of sewage on mollusc assemblages (Table 2, p = 0.0221). A clear effect is also evidenced in a metric multi-dimensional scaling (MDS) ordination plot of the centroids for each site (Fig. 2).

The relative sizes of components of variation for each term in the model, along with their associated empirically derived confidence intervals, based on the tailor-made bootstrap, are

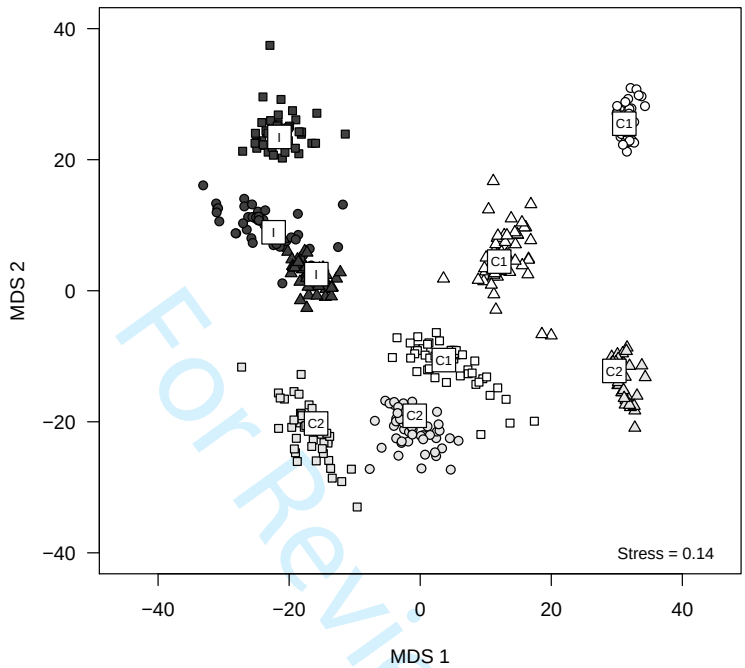

Figure 2. Metric MDS plot of the centroids in Bray-Curtis space for molluscan assemblages at each of the nine sites, with three sites at each location ("I", "C1" and "C2"). Also shown (as smaller-scale points) are the centroids obtained from 50 bootstrap samples taken within each site (see Clarke & Gorley 2015, for details), providing a visual assessment of within-site variation.

captured visually by the diagram shown in Fig. 3. This graphic is highly similar in its purpose to that suggested by Gelman (2005) in a Bayesian context for univariate ANOVA. Here, however, we propose its use more generally for visualising and comparing the contributing sources of variation for models of multivariate dissimilarity-based data clouds, with minimal assumptions. In the present example, small-scale residual variation in assemblage structure (from one sampling unit to the next) is the largest estimated component, followed by the effect of the impact, then site-level variation (Fig. 3). Consistent with the results of the hypothesis test shown in Table 2 ($P = 0.1985$), the bootstrap confidence interval associated with the component of variation at the spatial scale of Locations includes the value of 0.

Historical wisdom for such a design would treat Locations as a random factor (Underwood 1992; Glasby 1997), hence $B_1 = \infty$. The consequence of this choice (Table 3), as in many environmental impact study designs (Hewitt, Thrush & Cummings 2001), is to

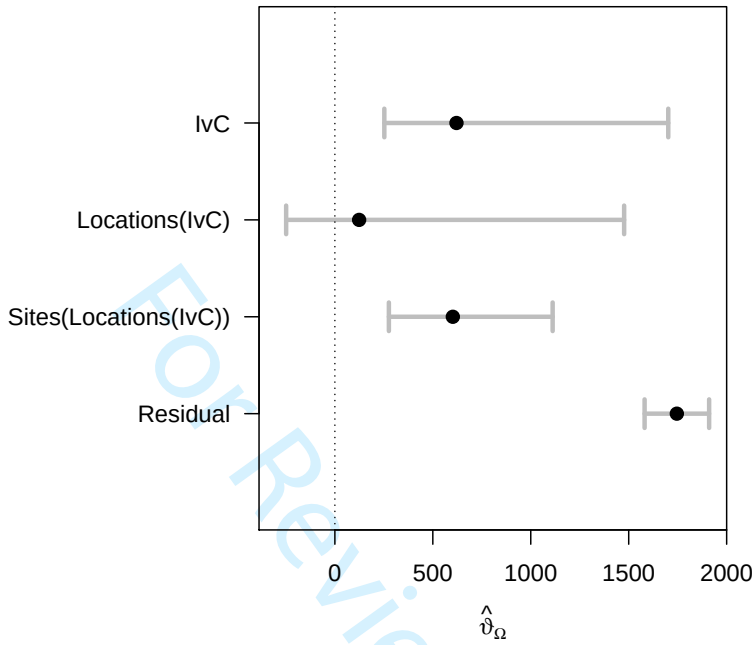

Figure 3. Estimated component of variation for each term in the three-factor hierarchical model for molluscan assemblages, based on Bray-Curtis dissimilarities, with 95% confidence intervals obtained using 10,000 bootstrap resamples generated from tailor-made bias-adjusted estimated effects and residuals.

yield a test for potential impact (H_0: $\vartheta_{IvC} = 0$) that lacks power – quite unwise in light of the precautionary principle (Fairweather 1991). However, under the new synthesis, by gradually decreasing the size of the population of locations to which inferences are to be drawn (B_1), we observe: (i) a gradual change in the size of the coefficient ($K_{\Omega L}$) attending the component of variation for Locations (ϑ_L) in the EMS for I*v*C, (ii) a gradual increase in the relative importance of MS_S compared to MS_L in the denominator for F_Ω and (iii) concomitant increases in F_{IvC}, $\hat{\vartheta}_{IvC}$ and hence, power (Table 3). To reduce the inference space so far as to treat Locations as fixed ($B_1 = 2$, hence using MS_S alone as denominator) clearly trivialises the analysis – inferences in this case would be limited to a comparison of location I with just these two other locations (C1 and C2) and no others. A correct analysis relies on the new synthesis and the logic attending the choice of control locations, where $B_1 = 8$ was

Table 3. Effects of changes to the size of the inference space (B_1, the number of control locations from which $b_1 = 2$ were drawn) on the test of H_0: $\vartheta_{IvC} = 0$. For $\Omega = IvC$, we have $E(MS_\Omega) = \vartheta_R + 9\vartheta_S + K_{\Omega L}\vartheta_L + 36\vartheta_{IvC}$ and $K_{\Omega L}$ depends on B_1. All p values were obtained using 9999 random permutations of exchangeable units.

B_1	$K_{\Omega L}$	Denominator	F_{IvC}	P	$\hat{\vartheta}_{IvC}$
∞ (random)	27	MS_L	2.89	0.335[a]	551.2
30	8.4	$0.69 \cdot MS_S + 0.31 \cdot MS_L$	3.70	0.022	615.2
20	8.1	$0.70 \cdot MS_S + 0.30 \cdot MS_L$	3.72	0.021	616.2
10	7.2	$0.73 \cdot MS_S + 0.27 \cdot MS_L$	3.77	0.026	619.3
8	6.75	$0.75 \cdot MS_S + 0.25 \cdot MS_L$	3.79	0.022	620.9
6	6.0	$0.78 \cdot MS_S + 0.22 \cdot MS_L$	3.84	0.019	623.4
4	4.5	$0.83 \cdot MS_S + 0.17 \cdot MS_L$	3.93	0.019	628.6
3	3	$0.89 \cdot MS_S + 0.11 \cdot MS_L$	4.03	0.013	633.8
2 (fixed)	0	MS_S	4.24	0.016	644.1

[a] For $B_1 = \infty$, there are only three unique value of F_Ω under permutation of the exchangeable units, which are whole locations in this case. For all other cases (i.e., where B_1 is finite and the denominator includes a mixture of MS_S and MS_L), the number of unique values of F_Ω is 839.

known *a priori* (Tables 2, 3). This affords greater power, yet retains meaningful broad-scale conclusions for the ecological system (the Italian coast) under study.

6. Discussion

We provide a new synthesis for deriving expectations of mean squares for any ANOVA design. Unlike previous methods, which require complex sets of rules, governed by what appear to be arbitrary parameterizations, our method is direct and has no explicit assumptions in this regard, only the fundamental notion of ANOVA being a linear model with a partitioning of variation attributable to factors of interest and encapsulated by the calculation of sums of squares.

The approach we offer gives the experimenter flexibility to define any given factor as being fixed, random or somewhere in between, when the levels drawn are a known fraction of a finite population of possible levels. It is clear that the role of finite population corrections (and, more implicitly, sample survey methodology in general) is key to our approach. When a factor is intermediate between fixed and random, the link between sampling theory and experimental design is explicit. We considered here only simple random sampling of levels without replacement and with equal probability. There is clearly scope similarly to explore the implications for complex ANOVA designs of other sampling strategies for factor levels, such as stratified sampling (Hartley & Rao 1969), systematic sampling (Hartley 1966) or sampling with unequal probabilities (Hartley & Rao 1962; Rao, Hartley & Cochran 1962).

Our approach gives results equivalent to those provided by Cornfield & Tukey (1956) for the two- and three-way balanced crossed designs considered by them; these demonstrate that the EMS that attend identification of factors purely as either fixed or random (the extremes at either end of the gradual progression) are easily obtained merely as special cases of this more general formulation (see Table 1, p. 926 in Cornfield & Tukey 1956). In addition, the treatment of factors as either fixed or random under the new synthesis produces EMS equal to those obtained by Searle, Casella & McCulloch (1992) when summation restrictions are used, yet we neither applied any such restriction in our derivation, nor have we asserted any "rules" to govern which coefficients on variance components should (or should not) be set to zero[‖].

Furthermore, under finite population sampling of levels of random factors in either balanced or unbalanced, crossed or nested designs, our approach gives results equivalent to those provided by Searle & Fawcett (1970), who similarly began with a centering of effect parameters in the inference space (*cf.* our eq. (1) with eq. (1) on p. 243 in Searle & Fawcett 1970). Shifting parameters by a constant *in the inference space* has no effect on the variance component being estimated. The new synthesis thus cleanly resolves previous debates regarding the 'correct' EMS for ANOVA designs.

We would agree with Nelder (1998, 2008) that no constraints should be made on ANOVA parameters, whether they be fixed or random. We hasten to add, however, that a virtually ubiquitous (and non-contentious) assumption in statistical linear models that include an intercept is that the errors $\varepsilon_{ij..}$ are assumed to be independent and identically distributed random variables with $\mathrm{E}(\varepsilon_{ij..}) = 0 \,\forall\, i, j, ..$ and variance σ^2_ε. Similarly, we assume $\mathrm{E}(\alpha_i) = 0$ $\forall\ i$ for the unknown effect parameters α_i associated with individual levels of any factor. In other words, both the effects and the errors are already defined by the ANOVA model itself as *deviations* (e.g., as per (5) above), which forms the fundamental basis for constructing sums of squares. It follows that if all possible levels α_i from the entire population were conceptually identifiable (i.e., if the population of levels is finite), then $\sum_i \alpha_i = 0$. This is not some un-natural constraint, but rather follows directly from the ANOVA model that corresponds mathematically to what is known regarding the population being sampled (e.g., Wilk & Kempthorne 1955).

The renovation of fixed *vs* random factors from a dichotomy to a gradual incremental progression adds a new degree of refinement, sophistication and flexibility to ANOVA models not previously afforded. To a large extent, it obviates the confusion surrounding what is

[‖]This begs the question of what role, if any, the purported EMS derived in the absence of summation restrictions (as implemented, e.g., in commonly used software such as 'proc mixed' in SAS) could possibly serve.

"meant" by a fixed *vs* a random factor. This comes, however, with a dose of responsibility. The onus is on the researcher to provide clarity on the number of sampled (or observed or chosen) levels relative to the size of the inference space – a logical thing to require for the rigorous statistical assessment of any factor. In practice, there may be cases where the size of the population of levels for a given factor (A), hence the sampling fraction (a/A), is unknown. Typically, this would occur when the number of levels is large/not easily counted, in which case treating the factor as random will generally be the most appropriate, albeit the most conservative, approach. One might also examine results across a range of plausible values for A to witness directly the implications of this choice on inferences. In addition, for a given set of data, the population size from which levels are drawn for a given factor and to which inferences are ultimately intended might indeed change with changes in the underlying conceptual logic and context of the study. A new analysis altering population sizes can easily be constructed from the same dataset, accordingly.

We agree with Gelman (2005) that analysis of variance is more important than ever. He suggested that, for any source of variation Ω there are two "natural" variance parameters to estimate: the (super)population variance (σ^2_Ω) and the finite-population variance based on the observed (estimated) effects (s^2_Ω)[††]. He suggested that there is no difference between a fixed and a random factor, but instead loosely aligns a random factor with interest in σ^2_Ω and a fixed factor with interest in s^2_Ω.

In practice, however, there is little to distinguish them under Gelman's scheme, as these two "variances" are given the same point estimate, and although σ^2_Ω has more uncertainty, Gelman's paper focused on ANOVA partitioning to achieve estimates of s^2_Ω (which he terms "finite-population variances"). He computed EMS and estimated s^2_Ω for each term, indiscriminantly treating all factors as random and providing measures of uncertainty *via* simulation under highly specific (i.e., normal) assumptions.

The distinction between fixed and random factors is more tangible in Bayesian modeling by differences in their priors on the effect parameters. Generally, $\alpha^*_i \sim \mathrm{N}(0, \sigma^2_\Omega)\ \forall i$ and for fixed effects $\sigma^2_\Omega = \infty^*$, while for random effects σ^2_Ω is given a scaled inverse-χ^2 conjugate hyperprior distribution (Gelman et al. 2013)[**]. Gelman (2005) advocated setting up hierarchical models "automatically" with all terms being treated as random effects for each row of the ANOVA table as an effective strategy. Resulting graphics, showing relative effect sizes and uncertainty associated with each ANOVA term, are very appealing.

[††]Although we would prefer to use Greek letters for unknown parameters, we are using Gelman's own notation here.

[*]In practice, the prior will not be specifed as having an infinite variance, simply a very large one.

[**]i.e., $\sigma^2_\Omega \sim$ Inv-$\chi^2(\nu_\Omega, \sigma^2_{0\Omega})$. The non-informative prior density is generally taken as uniform on σ_Ω, which corresponds to $\nu_\Omega = -1$ and $\sigma_{0\Omega} = 0$. For further details, consult Gelman et al. (2013).

There is nothing offered by such a scheme, however, to acknowledge potential real limitations in the sizes of inference spaces associated with individual factors. It is also not clear how even the usual Bayesian models for fixed *vs* random factors might accommodate such genuine finite-ness; this is clearly a fruitful topic for future research.

Gelman's (2005) exposition is at least partially motivated by perceived difficulties in performing classical ANOVA for complicated data structures with nesting, crossing and lack of balance. Such difficulties are here fully resolved by the new synthesis, which, when coupled with simple exchangeability arguments and appropriate re-sampling methods, provides for hypothesis-testing and estimation alike. Unlike classical, maximum likelihood or Bayesian approaches, specific assumptions regarding distributions of effects (e.g., normality) are not necessary here.

Extension to multivariate analysis (RDA, dbRDA) further generalises the synthesis for rigorous assessment of components of variation based on distance or dissimilarity matrices *via* PERMANOVA. Future research may work towards more directed multivariate modeling of correlation structures and handling variation in the shapes of dispersion clouds across different levels of factors or cells in ANOVA designs. Nevertheless, overall, we expect the new synthesis given here will result in an increase in the utility, importance and power of analysis of variance in applied research.

Appendix I

Our purpose here is to prove the validity of (2) and (3). Suppose factor A has a total population of possible levels A, and each level has an associated effect α_i^*, $i = 1, \ldots, A$. In a given study, the researcher samples a levels from the population of possible levels. If $a = A$, (i.e., if all of the levels of interest for statistical inference are included in the study), then factor A is *fixed* and $a/A = 1$ and $1/A = 1/a$. If A is arbitrarily large relative to a, so far as to be considered effectively infinite, then factor A is *random* and we consider the sampling fraction in the limit as

$$\lim_{A\to\infty} a/A = 0.$$

In addition, for this case of a random factor, we have

$$\lim_{A\to\infty} 1/A = 0.$$

Assume that the effects $\alpha_i^* \ \forall \ i$ are independent and identically distributed as random variables with mean μ_α and variance σ_α^2. Thus, $\mathrm{E}(\alpha_i^* \cdot \alpha_{i'}^*) = \mu_\alpha^2 \ \ \forall \ \ i \neq i'$ and $\mathrm{E}(\alpha_i^* \cdot \alpha_i^*) = \mu_\alpha^2 + \sigma_\alpha^2$. Let $\alpha_i = \left(\alpha_i^* - \sum_{i=1}^{A} \alpha_i^*/A\right)$ be the centred effects in the inference space. So

636 $\mathrm{E}(\alpha_i) = 0$. We therefore have:

$$\mathrm{E}\left(\alpha_i \cdot \alpha_{i'}\right) = \mathrm{E}\left\{\left[\alpha_i^* - \tfrac{1}{A}\left(\alpha_1^* + \alpha_2^* + \ldots + \alpha_A^*\right)\right] \times \left[\alpha_{i'}^* - \tfrac{1}{A}\left(\alpha_1^* + \alpha_2^* + \ldots + \alpha_A^*\right)\right]\right\}$$

637 which can be re-written as

$$\mathrm{E}\left(\alpha_i \cdot \alpha_{i'}\right) = \mathrm{E}\left\{\alpha_i^* \alpha_{i'}^* - \tfrac{2}{A}\left[\left(\alpha_i^*\right)^2 + (A-1)\,\alpha_i^* \alpha_{i'}^*\right] + \tfrac{1}{A^2}\left[A\left(\alpha_i^*\right)^2 + A\,(A-1)\,\alpha_i^* \alpha_{i'}^*\right]\right\}$$

638 and then simplified to

$$\mathrm{E}\left(\alpha_i \cdot \alpha_{i'}\right) = \mathrm{E}\left\{\alpha_i^* \alpha_{i'}^* - \tfrac{1}{A}\left[\left(\alpha_i^*\right)^2 + (A-1)\,\alpha_i^* \alpha_{i'}^*\right]\right\},$$

639 yielding the following

$$\mathrm{E}\left(\alpha_i \cdot \alpha_{i'}\right) = \mathrm{E}\left\{\tfrac{1}{A}\alpha_i^* \alpha_{i'}^* - \tfrac{1}{A}\left(\alpha_i^*\right)^2\right\}.$$

640 Now, by substitution, we have

$$\mathrm{E}\left(\alpha_i \cdot \alpha_{i'}\right) = \tfrac{1}{A}\left(\mu_\alpha^2\right) - \tfrac{1}{A}\left(\mu_\alpha^2 + \sigma_\alpha^2\right),$$

641 which proves (2), i.e.,

$$\mathrm{E}(\alpha_i \cdot \alpha_{i'}) = -\tfrac{1}{A} \cdot \sigma_\alpha^2 \;\; \forall \;\; i \neq i',$$

642 Next, we can similarly consider $\mathrm{E}\left(\alpha_i^2\right)$ as

$$\mathrm{E}\left(\alpha_i \cdot \alpha_i\right) = \mathrm{E}\left\{\left[\alpha_i^* - \tfrac{1}{A}\left(\alpha_1^* + \alpha_2^* + \ldots + \alpha_A^*\right)\right] \times \left[\alpha_i^* - \tfrac{1}{A}\left(\alpha_1^* + \alpha_2^* + \ldots + \alpha_A^*\right)\right]\right\}$$

643 Expanding the square, as we have done before, we can write

$$\mathrm{E}\left(\alpha_i \cdot \alpha_i\right) = \mathrm{E}\left\{\left(\alpha_i^*\right)^2 - \tfrac{2}{A}\left[\left(\alpha_i^*\right)^2 + (A-1)\,\alpha_i^* \alpha_{i'}^*\right] + \tfrac{1}{A}\left[\left(\alpha_i^*\right)^2 + (A-1)\,\alpha_i^* \alpha_{i'}^*\right]\right\},$$

644 and simplifying this yields

$$\mathrm{E}\left(\alpha_i \cdot \alpha_i\right) = \mathrm{E}\left\{\left(1 - \tfrac{1}{A}\right)\left(\alpha_i^*\right)^2 - \left(1 - \tfrac{1}{A}\right)\alpha_i^* \alpha_{i'}^*\right\}.$$

645 By substitution, we can then write

$$\mathrm{E}\left(\alpha_i \cdot \alpha_i\right) = \left(1 - \tfrac{1}{A}\right)\left(\mu_\alpha^2 + \sigma_\alpha^2\right) - \left(1 - \tfrac{1}{A}\right)\mu_\alpha^2,$$

which proves (3), i.e.,

$$\mathrm{E}\left(\alpha_i \cdot \alpha_i\right) = \left(1 - \tfrac{1}{A}\right) \cdot \sigma_\alpha^2 \;\; \forall \;\; i$$

Q.E.D.

Appendix II

Our purpose here is to outline a proof of (6). Define the ANOVA linear model for centred response data, $\mathbf{y}$, as:

$$\mathbf{y} = \sum_t \sum_i \alpha_{ti} \mathbf{u}_{ti}, \;\; t = 1, \ldots, T; \;\; i = 1, \ldots, a_t \tag{13}$$

where α_{ti} is the centred effect of level (or cell) i for term t and $\mathbf{u}_{ti}$ indicates by 1 (or 0) the presence (or absence) of an individual sampling unit within level (or cell) i for term t, respectively. More specifically, each term has $i = 1, \ldots, a_t$ levels (or cells) and consists of one of the following: (i) a single *main effect*; (ii) an *interaction* among two or more crossed terms; or (iii) a factor that is *nested*, either within another single factor or within a combination of levels of other factors. Note that (iii) accommodates the inclusion of an error term as a factor nested in all combinations of levels of the other factors in the model. We shall assume effects for a given term have a common variance, denoted as $\mathrm{Var}(\alpha_{ti}) = \sigma_t^2$.

Main effects. For any given term t comprised of any single factor standing as a main effect in the model, we consider the individual effects α^*_{ti}, associated with individual levels of the factor $(i = 1, \ldots a_t)$ to have been drawn randomly from a population of size A_t, so $a_t \leq A_t$. Considering the two ends of the step-wise spectrum, at one end we may have $a_t = A_t$, corresponding to a fixed factor. At the other end, A_t is arbitrarily large, and may be treated in the limit to approach infinity, corresponding to a random factor. For each main effect, we assert a centering in the inference space (as in Section 2.1), so

$$\alpha_{ti} = \alpha^*_{ti} - \sum_{i=1}^{A_t} \alpha^*_{ti}/A_t$$

and thus, for any main effect term t,

$$\mathrm{E}(\alpha_{ti} \cdot \alpha_{ti'}) = \left(\delta_{tii'} - \tfrac{1}{A_t}\right) \cdot \sigma_t^2 \tag{14}$$

where $\delta_{tii'} = 1$ for $i = i'$ and $\delta_{tii'} = 0$ for $i \neq i'$. Furthermore, we shall let $\mathbf{\Delta}_t$ be an $a_t \times a_t$ matrix comprised of the elements $\{\Delta_{tii'}\} = \left\{\delta_{tii'} - \tfrac{1}{A_t}\right\}$.

Interactions. For any given term t comprised of an interaction among constituent terms t', t'', t''', ... etc., the total number of cells, denoted a_t, will be equal to the number of

combinations of levels of all constituent terms involved in the interaction. Each cell will have a corresponding centred interaction effect α_{ti}, $(i = 1, \ldots, a_t)$. Specifically, we assert a centering of the interaction effects in the inference space (as in Section 2.2) across each margin of the constituent terms involved in the interaction t', t'', t''', ... etc., and we define $\mathbf{\Delta}_t$ for an interaction term t to be constructed as a Kronecker product of the $\mathbf{\Delta}$ matrices of its constituent terms, i.e.,

$$\mathbf{\Delta}_t = \mathbf{\Delta}_{t'} \otimes \mathbf{\Delta}_{t''} \otimes \mathbf{\Delta}_{t'''} \otimes \ldots \quad , \text{etc.} \tag{15}$$

Nested factors. For any given term t comprised of a factor that is fully nested in one or more other factors or their interaction (the upper-level term, t'), there will be $j = 1, \ldots, b_{ti'}$ levels of the nested term, drawn randomly from a population of $B_{ti'}$ possible levels, separately and independently within each of $i' = 1, \ldots, a_{t'}$ levels (or cells) of the upper-level factor, yielding a total of $a_t = \sum_{i'=1}^{a_{t'}} b_{ti'}$ cells for the nested term. Each cell of the nested term t shall have a centred effect $\beta_{tj(i')} = \alpha_{ti}$, $i = 1, \ldots, a_t$. Specifically, for the individual effects $\beta^*_{tj(i')}$ associated wtih each cell of the nested term, we assert a separate centering in the inference space within each level (or cell) of the upper-level term (as in Section 2.3); that is,

$$\beta_{tj(i')} = \beta^*_{tj(i')} - \sum_{j=1}^{B_{ti'}} \beta^*_{tj(i')} / B_{ti'}$$

We shall assume that centred effects $\beta_{tj(i')}$ associated with individual levels of a nested term t within a given level (or cell) i' of some upper-level term t' are independent of those within some other level (or cell) i'' of the upper-level term, hence $\mathrm{E}(\beta_{tj(i')} \cdot \beta_{tj(i'')}) = 0$ for all $i \neq i''$. Also, for centred effects within a given cell i' of the upper-level term, we have, for all t and for all i',

$$\mathrm{E}(\beta_{tj(i')} \cdot \beta_{tj'(i')}) = \left(\delta_{tjj'(i')} - \tfrac{1}{B_{ti'}}\right) \cdot \sigma^2_t \tag{16}$$

where $\delta_{tjj'(i')} = 1$ for $j = j'$ and $\delta_{tjj'(i')} = 0$ for $j \neq j'$. We then define $\mathbf{\Delta}_t$ for a nested term t to be a block diagonal matrix obtained from the direct sum $\mathbf{\Delta}_t = \sum_{i'=1}^{a_{t'}} \mathbf{\Delta}_{i'}$, where $\mathbf{\Delta}_{i'}$ is defined as a $b_{ti'} \times b_{ti'}$ matrix having elements $\left\{\Delta_{tjj'(i')}\right\} = \left\{\delta_{tjj'(i')} - \frac{1}{B_{ti'}}\right\}$.

Partitioning. For the ANOVA partitioning, we construct for every term an orthogonal contrast matrix of full rank $\mathbf{X}_t$, having df_t columns, and its associated projection matrix $\mathbf{H}_t$ for a chosen Type of sum of squares (Searle 2006). The expectation for the sum of squares for any chosen term, Ω, is therefore

$$\mathrm{E}\left(\mathrm{SS}_\Omega\right) = \mathrm{E}\left(tr\left[\mathbf{H}_\Omega \mathbf{y}\mathbf{y}^\mathsf{T}\right]\right)$$

Replacing $\mathbf{y}$ with the sum in (13), we have

$$\mathrm{E}\left(\mathrm{SS}_{\Omega}\right) = \mathrm{E}\left\{ tr\left[\mathbf{H}_{\Omega}\sum_{t}\sum_{i}\alpha_{ti}\mathbf{u}_{ti}\cdot\left[\sum_{t'}\sum_{i'}\alpha_{t'i'}\mathbf{u}_{t'i'}\right]^{\mathsf{T}}\right]\right\} \tag{17}$$

We shall assume that all effects α_{ti} for any given term t are independent of all effects $\alpha_{t'i'}$ associated with any other term t', so $\mathrm{E}(\alpha_{ti}\cdot\alpha_{t'i'}) = 0 \;\; \forall \;\; t \neq t'$ and $\forall \;\; i, i'$. Thus, the right-hand side of (17) reduces to sums of products of effects for individual terms, i.e., where $t = t'$. Furthermore, centering in the inference space as described above for main effects, interactions and nested terms ensures that all individual effects are counted as *deviations* from expectations under a true null hypothesis associated with that term in the model; that is, for term t, we have the explicit null hypothesis

$$\mathrm{H}_0^{[t]}: \;\; \alpha_{ti} = 0 \;\; \forall \;\; i \;\; . \tag{18}$$

Given the fundamental results provided by (14) for main effects, (15) for interaction terms and (16) for nested terms provided above, the contribution of any individual term t to the sum on the right-hand side of (17) is

$$tr\left[\mathbf{H}_{\Omega}\mathbf{U}_t\mathbf{\Delta}_t\mathbf{U}_t^{\mathsf{T}}\right]\cdot\sigma_t^2 \;\; .$$

As the expectation of the sum is equal to the sum of expectations, and as every mean square is constructed as $\mathrm{MS}_{\Omega} = \mathrm{SS}_{\Omega}/df_{\Omega}$, we have proven (6), i.e.

$$\mathrm{E}\left(\mathrm{MS}_{\Omega}\right) = df_{\Omega}^{-1}\left\{\sum_{t=1}^{T} tr\left[\mathbf{H}_{\Omega}\mathbf{U}_t\mathbf{\Delta}_t\mathbf{U}_t^{\mathsf{T}}\right]\cdot\sigma_t^2\right\}.$$

Q.E.D.

References

ANDERSON, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**, 32–46.

ANDERSON, M.J. (2017). Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online* **stat07841**, 1–15.

ANDERSON, M.J., CONNELL, S.D., GILLANDERS, B.M., DIEBEL, C.E., BLOM, W.M., SAUNDERS, J.E. & LANDERS, T.J. (2005). Relationships between taxonomic resolution and spatial scales of multivariate variation. *Journal of Animal Ecology* **74**, 636–646.

ANDERSON, M.J., GORLEY, R.N. & CLARKE, K.R. (2008). *PERMANOVA+ for PRIMER: Guide to software and statistical methods.* Plymouth, UK: PRIMER-E Ltd.

ANDERSON, M.J. & ROBINSON, J. (2003). Generalized discriminant analysis based on distances. *Australian & New Zealand Journal of Statistics* **45**, 301–318.

ANDERSON, M.J. & TER BRAAK, C.J.F. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation* **73**, 85–113.

ANDERSON, M.J., WALSH, D.C.I., CLARKE, K.R., GORLEY, R.N. & GUERRA-CASTRO, E. (2017). Some solutions to the multivariate Behrens-Fisher problem for dissimilarity-based analyses. *Australian & New Zealand Journal of Statistics* **59**, 57–79.

BAILEY, R.A. (2008). *Design of comparative experiments.* Cambridge, UK: Cambridge University Press.

BOOTH, J.G., BUTLER, R.W. & HALL, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association* **89**, 1282–1289.

CARPENTER, J.R., GOLDSTEIN, H. & RASBASH, J. (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *Applied Statistics* **52**, 431–443.

CLARKE, K.R. & GORLEY, R.N. (2015). *PRIMER v7: User manual/Tutorial.* Plymouth, UK: PRIMER-E.

COCHRAN, W.G. (1977). *Sampling techniques.* New York: John Wiley & Sons.

CORNFIELD, J. & TUKEY, J. (1956). Average values of mean squares in factorials. *The Annals of Mathematical Statistics* **27**, 907–949.

DAVISON, A.C. & HINKLEY, D.V. (1997). *Bootstrap methods and their application.* Cambridge, UK: Cambridge University Press.

EFRON, B. (1982). *The jackknife, the bootstrap and other resampling plans.* Philadelphia: Society for Industrial and Applied Mathematics.

EFRON, B. & TIBSHIRANI, R.J. (1993). *An introduction to the bootstrap.* Boca Raton, FL: Chapman & Hall.

FAIRWEATHER, P.G. (1991). Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research* **42**, 555–567.

FIELD, C.A. & WELSH, A.H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society, Series B* **69**, 369–390.

FLETCHER, D.J., DILLINGHAM, P.W. & ZENG, J. (2019). Model-averaged confidence distributions. *Environmental and Ecological Statistics* **26**, 367–384.

FLETCHER, D.J. & UNDERWOOD, A.J. (2002). How to cope with negative estimates of components of variance in ecological field studies. *Journal of Experimental Marine Biology and Ecology* **273**, 89–95.

GELMAN, A. (2005). Analysis of variance - why it is more important than ever. *The Annals of Statistics* **33**, 1–53.

GELMAN, A., CARLIN, J.B., STERN, H.S., DUNSON, D.B., VEHTARI, A. & RUBIN, D.B. (2013). *Bayesian data analysis, 3rd edition.* Boca Raton, FL: (Chapman & Hall/CRC Texts in Statistical Science.

GLASBY, T.M. (1997). Analysing data from post-impact studies using asymmetrical analyses of variance: a case study of epibiota on marinas. *Australian Journal of Ecology* **22**, 448–459.

GLASBY, T.M. & UNDERWOOD, A.J. (1998). Determining positions for control locations in environmental studies of estuarine marinas. *Marine Ecology Progress Series* **171**, 1–14.

GOWER, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–328.

GROSSMANN, H. (2014). Automating the analysis of variance of orthogonal designs. *Computational Statistics and Data Analysis* **70**, 1–18.

HALL, P. (1992). *The bootstrap and Edgeworth expansion.* New York, NY: Springer-Verlag.

HARTLEY, H.O. (1966). Systematic sampling with unequal probability and without replacement. *Journal of the American Statistical Association* **61**, 739–748.

HARTLEY, H.O. (1967). Expectations, variances and covariances of ANOVA mean squares by 'synthesis'. *Biometrics* **23**, 105–114.

HARTLEY, H.O. & RAO, J.N.K. (1962). Sampling with unequal probabilities and without replacement. *The Annals of Mathematical Statistics* **33**, 350–374.

HARTLEY, H.O. & RAO, J.N.K. (1969). A new estimation theory for sample surveys. II. In *New developments in survey sampling*. New York, NY: Wiley-Interscience, pp. 147–169.

HARTLEY, H.O., RAO, J.N.K. & LAMOTTE, L.R. (1978). A simple 'synthesis'-based method of variance component estimation. *Biometrics* **34**, 233–242.

HARTLEY, H.O. & SEARLE, S.R. (1969). A discontinuity in mixed model analyses. *Biometrics* **25**, 573–576.

HASLETT, S.J., ISOTALO, J., MARKIEWICZ, A. & PUNTANEN, S. (2023). Further remarks on permissible covariance structures for simultaneous retention of BLUEs in linear models. *Acta Et Commentationes Universitatis Tartuensis De Mathematica* **27**, 101–112.

HASLETT, S.J. & PUNTANEN, S. (2013). A review of conditions under which BLUEs and/or BLUPs in one linear mixed model are also BLUEs and/or BLUPs in another. *Calcutta Statistical Association Bulletin* **65**, 257–260.

HASLETT, S.J., PUNTANEN, S. & ARENDACKÁ, B. (2015). The link between the mixed and fixed linear models revisited. *Statistical Papers* **56**, 849–861.

HEWITT, J.E., THRUSH, S.E. & CUMMINGS, V.J. (2001). Assessing environmental impacts: effects of spatial scale and temporal variability at likely impact scales. *Ecological Applications* **11**, 1502–1516.

HOCKING, R.R. (1973). A discussion of the two-way mixed model. *The American Statistician* **27**, 148–152.

LEGENDRE, P. & ANDERSON, M.J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69**, 1–24.

LEGENDRE, P. & LEGENDRE, L. (2012). *Numerical Ecology, 3rd English edition.* Amsterdam, the Netherlands: Elsevier.

LENCINA, V.B., SINGER, J.M. & STANEK, E.J. (2005). Much ado about nothing: the mixed models controversy revisited. *International Statistical Review* **73**, 9–20.

LITTELL, R.C., MILLIKEN, G.A., STROUP, W.W., WOLFINGER, R.D. & SCHABENBERGER, O. (2006). *SAS for mixed models, 2nd edition.* Cary, North Carolina: SAS Institute Inc.

MCARDLE, B.H. & ANDERSON, M.J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**, 290–297.

MCLEAN, R.A., SANDERS, W.L. & STROUP, W.W. (1991). A unified approach to mixed linear models. *The American Statistician* **45**, 54–64.

NELDER, J.A. (1998). The great mixed-model muddle is alive and flourishing, alas! *Food Quality and Preference* **9**, 157–159.

NELDER, J.A. (2008). What is the mixed-models controversy? *International Statistical Review* **76**, 134–139.

NETER, J., KUTNER, M.H., NACHTSHEIM, C.J. & WASSERMAN, W. (1996). *Applied linear statistical models, 4th edition.* Chicago: Irwin.

RAO, J.N.K. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya A* **26**, 329–358.

RAO, J.N.K. (1968). On expectations, variances, and covariances of ANOVA mean squares by 'synthesis'. *Biometrics* **24**, 963–978.

RAO, J.N.K., HARTLEY, H.O. & COCHRAN, W.G. (1962). On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society, Series B* **24**, 482–491.

SAIGO, H. (2010). Comparing four bootstrap methods for stratified three-stage sampling. *Journal of Official Statistics* **26**, 193–207.

SCHWEDER, T. & HJORT, N.L. (2016). *Confidence, likelihood, probability: statistical inference with confidence distributions*. Cambridge, UK: Cambridge University Press.

SEARLE, S.R. (1971). Topics in variance component estimation. *Biometrics* **27**, 1–76.

SEARLE, S.R. (2006). *Linear models for unbalanced data*. New York: Wiley - Interscience.

SEARLE, S.R., CASELLA, G. & MCCULLOCH, C.E. (1992). *Variance components*. New York: John Wiley & Sons.

SEARLE, S.R. & FAWCETT, R.F. (1970). Expected mean squares in variance components models having finite populations. *Biometrics* **26**, 243–254.

SITTER, R.R. (1992a). A resampling procedure for complex survey data. *Journal of the American Statistical Association* **87**, 755–765.

SITTER, R.R. (1992b). Comparing three bootstrap methods for survey data. *The Canadian Journal of Statistics* **20**, 135–154.

SPEED, T.P. (1987). What is an analysis of variance? *The Annals of Statistics* **15**, 885–910.

TERLIZZI, A., ANDERSON, M.J., FRASCHETTI, S. & BENEDETTI-CECCHI, L. (2007). Scales of spatial variation in mediterranean subtidal sessile assemblages at different depths. *Marine Ecology Progress Series* **332**, 25–39.

TERLIZZI, A., SCUDERI, D., FRASCHETTI, S. & ANDERSON, M.J. (2005). Quantifying effects of pollution on biodiversity: a case study of highly diverse molluscan assemblages in the mediterranean. *Marine Biology* **148**, 293–305.

UNDERWOOD, A.J. (1991). Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Australian Journal of Marine and Freshwater Research* **42**, 569–587.

UNDERWOOD, A.J. (1992). Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *Journal of Experimental Marine Biology and Ecology* **161**, 145–178.

UNDERWOOD, A.J. (1994). On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecological Applications* **4**, 3–15.

VAN DEN WOLLENBERG, A.L. (1977). Redundancy analysis: an alternative for canonical correlation analysis. *Psychometrika* **42**, 207–219.

VOSS, D.T. (1999). Resolving the mixed models controversy. *The American Statistician* **53**, 352–356.

WILK, M.B. & KEMPTHORNE, O. (1955). Fixed, mixed, and random models. *Journal of the American Statistical Association* **50**, 1144–1167.

WINER, B.J., BROWN, D.R. & MICHELS, K.M. (1991). *Statistical principles in experimental design, 3rd edition*. New York: McGraw-Hill.

ZAPALA, M.A. & SCHORK, N.J. (2006). Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 19430–19435.

ZAPALA, M.A. & SCHORK, N.J. (2012). Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Frontiers in Genetics* **3**, 190.