

# 1.28 Estimating components of variation

The EMS's also yield another important insight: they provide a direct method to get *unbiased estimates* of each of the components of variation in the model. PERMANOVA estimates these components using mean squares, in a directly analogous fashion to the unbiased univariate ANOVA estimators of variance components (e.g., [Searle, Casella & McCulloch \(1992\)](#)). In essence, this is achieved by setting the mean squares equal to their expectations and solving for the component of interest.

For example, by setting  $MS_{Ar}$  and  $MS_{Res}$  equal to their respective expectations (placing “hats” on the parameters to indicate that we are now talking about estimates of these things, rather than their true parameter values), we have:

$$MS_{Ar} = 1 \times \hat{V} \left( \text{Res} \right) + 5 \times \hat{V} \left( \text{Ar(Si(Lo))} \right)$$

$$MS_{Res} = 1 \times \hat{V} \left( \text{Res} \right)$$

Thus,

$$\hat{V} \left( \text{Res} \right) = MS_{Res} / 1$$

$$\hat{V} \left( \text{Ar(Si(Lo))} \right) = \left( MS_{Ar} - MS_{Res} \right) / 5$$

From the output, we therefore can calculate these estimates directly from the mean squares. The estimated component of variation for the residual is  $MS_{Res} = 2525.7$  and for areas this is  $\left( MS_{Ar} - MS_{Res} \right) / 5 = \left( 3111.8 - 2525.7 \right) / 5 = 117.2$ . Similar logic, when applied to the other terms in the analysis yields:

$$\hat{V} \left( \text{Si(Lo)} \right) = \left( MS_{Si} - MS_{Ar} \right) / 10 = 110.9$$

$$\hat{V} \left( \text{Lo} \right) = \left( MS_{Lo} - MS_{Si} \right) / 20 = 381.7$$

These estimates are all calculated automatically by the program and included in the output file in the column labeled ‘Estimate’ under the heading entitled ‘*Components of variation*’ (Fig. 1.29). For the species composition of molluscs in these kelp holdfast assemblages, the greatest component of variation occurred at the smallest spatial scale (the residual), followed by locations, and then areas and sites, with the latter two being comparable in size (Fig. 1.29).

An important point here is that these estimates are not actual “variance components” in the traditional sense *unless* one is analysing a single variable and the resemblance measure used is Euclidean distance. In addition, these are obviously not the same as variance-covariance matrices used in traditional multivariate statistics either (e.g. [Mardia, Kent & Bibby \(1979\)](#), [Seber \(1982\)](#)), because they do not include any estimation of covariance structure at all. Rather, they are

interpretable geometrically as measures of variability from a partitioning on the basis of the dissimilarity (or similarity) measure chosen.

These estimates (like their univariate counterparts of variance components) will be in terms of the *squared* units of the dissimilarity measure chosen. Thus, in order to put these back onto the original units, PERMANOVA also calculates their square root (provided in the column labeled 'Sq.root' in the results file). These values are akin to a standard deviation in a traditional univariate analysis. Thus, if the value of the dissimilarity measure used has a direct interpretation (such as the Jaccard or Bray-Curtis measures, which are both percentages), then these can be examined and interpreted as well. For example, the greatest variation in molluscan composition is at the level of individual replicate holdfasts, which (according to the square root of the estimated component of variation due to the residual of 50.3) may share only around 50% of their species, even though they may be separated by just a few metres. Over and above this, holdfasts in different areas may be an additional 10-11% dissimilar in their composition, on average, and so on.

Although the above design included all random factors, for which a discussion of estimating components of variation is a fairly natural one, we can also estimate the components of variation due to fixed effects. Recall that these are not measures of variance *per se*, but rather are sums of squared fixed effects divided by appropriate degrees of freedom (see the section on **Components of variation**). However, if we are interested in comparing the amount of variation that is attributable to different terms in the model, estimates of components for fixed and/or random factors are useful and are directly comparable. In fact, it is indeed these estimates of components of variation that should be used as a correct basis for comparing the *relative importance* of different terms in the model towards explaining overall variation ( [Underwood & Petraitis \(1993\)](#) ). In contrast, the raw sums of squares (whether alone or as a percentage of the total sum of squares) are *not* directly comparable, because different terms generally have different degrees of freedom (e.g., it would clearly be inappropriate to compare the percentage of the total sum of squares explained by a factor having only 1 degree of freedom *versus* some other factor that had 5 degrees of freedom).

The only potentially unsettling consequence of using analogues of the ANOVA estimators to estimate components of variation is the fact that these estimates (even in the univariate case on the basis of Euclidean distance) can sometimes turn out to be negative ( [Thompson \(1962\)](#) , [Searle, Casella & McCulloch \(1992\)](#) ). This is clearly illogical and is generally accompanied by there being little or no evidence against the null hypothesis for the term in question that its component is equal to zero (i.e., a large *P*-value). Although there are other methods available for estimating variance components (i.e., ML, REML, Bayesian, etc., see [Searle, Casella & McCulloch \(1992\)](#) ), the ANOVA estimators do have the attractive quality of being *unbiased*<sup>34</sup>. The best solution to this issue is often to re-analyse the data after removing that term from the model (e.g., [Thompson & Moore \(1963\)](#) , [Fletcher & Underwood \(2002\)](#) ). This leads naturally to a consideration of how to remove terms from a model, also referred to in some cases as *pooling* (see the following section).

---

<sup>34</sup> An *unbiased* estimator is one whose expectation is equal to the parameter it is trying to estimate.

---

Revision #25

Created 8 August 2022 09:30:46 by Arden

Updated 27 November 2024 01:41:01 by Abby Miller