Unbalanced Two-way ANOVA

When we have an unequal number of replicates per cell, the estimation space $C(X)$ CANNOT be decomposed into a sum of orthogonal subspaces, in general. There are special cases under which $C(X)$ can be decomposed into a sum of orthogonal subspaces when we have an unequal number of replicates. We first discuss these special cases.

Consider the model

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \epsilon_{ijk} \tag{3}$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, n_{ij}$.

Defn

We say that a model has proportional numbers if for $i, i' = 1, \ldots, a$, and $j, j' = 1, \ldots, b$,

$$\frac{n_{ij}}{n_{ij'}} = \frac{n_{i'j}}{n_{i'j'}} \ .$$

It turns out that orthogonality of subspaces is preserved if we have proportional numbers. That is,

$$C(X) = M_\mu + M_\alpha + M_\eta \ ,$$

where M_μ, M_α and M_η are all mutually orthogonal.

The following theorem gives an equivalent condition for proportional numbers.

Theorem

Suppose the n_{rs}'s satisfy the proportional numbers definition. Then for any $r = 1, \ldots, a$, and $s = 1, \ldots, b$,

$$n_{rs} = \frac{n_{r.} n_{.s}}{n_{..}}$$

where

$$n_{r.} = \sum_{j=1}^{b} n_{rj}, \quad n_{.s} = \sum_{i=1}^{a} n_{is}, \quad \text{and} \quad n_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \ .$$

This theorem says that proportional numbers is equivalent to writing any n_{rs} as a product of the rth row sample size total times the sth column sample size total divided by total sample size.

Proof

Using the definition of proportional numbers, we have

$$\frac{n_{rs}}{n_{rj}} = \frac{n_{is}}{n_{ij}} \ ,$$

which implies

$$n_{ij} n_{rs} = n_{rj} n_{is} \ .$$

Now summing both sides of the equation over i and j, we have

$$\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}n_{rs} = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{rj}n_{is} \tag{4}$$

The left side of (4) reduces to

$$n_{rs}\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij} = n_{rs}n_{..}$$

The right side of (4) reduces to

$$\sum_{i=1}^{a}\sum_{j=1}^{b} n_{rj}n_{is} = \left(\sum_{j=1}^{b} n_{rj}\right)\left(\sum_{i=1}^{a} n_{is}\right) = n_{r.}n_{.s}\ .$$

Thus we have

$$n_{rs}n_{..} = n_{r.}n_{.s}$$

which implies

$$n_{rs} = \frac{n_{r.}n_{.s}}{n_{..}}\ .$$

We can write the design matrix of model (3) as

$$X = (J, X_1, \ldots, X_a, X_{a+1}, \ldots, X_{a+b})$$

where

$$X_r = (t_{ijk})$$

where $r = 1, \ldots, a$, $t_{ijk} = \delta_{ir}$, and $\delta_{ir} = 1$ if $i = r$, 0 otherwise. Also,

$$X_{a+s} = (u_{ijk})$$

$s = 1, \ldots, b$, $u_{ijk} = \delta_{js}$, and $\delta_{js} = 1$ if $j = s$ and 0 otherwise.

With proportional numbers, we can construct orthogonal subspaces as follows. As before, define

$$Z_r = X_r - \frac{n_{r.}}{n_{..}}J \quad r = 1, \ldots, a$$

$$Z_{a+s} = X_s - \frac{n_{.s}}{n_{..}}J \quad s = 1. \ldots, b\ .$$

We can see that for $r = 1, \ldots, a$, and $s = 1, \ldots, b$,

$$\begin{aligned} Z'_{a+s}Z_r &= (X_s - \frac{n_{.s}}{n_{..}}J)'(X_r - \frac{n_{r.}}{n_{..}}J) \\ &= X'_sX_r - \frac{n_{.s}}{n_{..}}(J'X_r) - \frac{n_{r.}}{n_{..}}(X'_sJ) + \frac{n_{r.}n_{.s}}{n_{..}^2}(J'J) \end{aligned}$$

$$
\begin{aligned}
&= n_{rs} - n_{.s}\frac{n_{r.}}{n_{..}} - n_{r.}\frac{n_{.s}}{n_{..}} + n_{..}\frac{n_{r.}n_{.s}}{n_{..}^2} \\
&= n_{rs} - \frac{n_{.s}n_{r.}}{n_{..}} - \frac{n_{r.}n_{.s}}{n_{..}} + \frac{n_{r.}n_{.s}}{n_{..}} \\
&= n_{rs} - \frac{n_{r.}n_{.s}}{n_{..}} \\
&= \frac{n_{r.}n_{.s}}{n_{..}} - \frac{n_{r.}n_{.s}}{n_{..}} = 0 \; .
\end{aligned}
$$

Similar results can be obtained if an interaction term is included in the model with proportional numbers.

The General Case

If we don't have proportional numbers, then $C(X)$ cannot be decomposed into a sum of orthogonal subspaces with each subspace corresponding to a term in the model. In this case, the sum of squares depend on which terms have been included in the model. That is, the sums of squares depend on the order of the inclusion of the effects. Let $R(\alpha \mid \mu, \eta)$ denote the sums of squares of the α treatment given that μ and η are in the model, and so forth. In this general case,

$$R(\alpha \mid \mu, \eta) \neq R(\alpha \mid \mu)$$

and

$$R(\eta \mid \mu, \alpha) \neq R(\eta \mid \mu)$$

In the general case, we analyze the unbalanced ANOVA model using the general theory of linear models. That is, we write the model as

$$Y = X\beta + \epsilon$$

and compute sums of squares of various models by finding the appropriate X_0 matrix, where $C(X_0) \subset C(X)$. Thus, we do tests and inference using the general nested versus full framework that we developed earlier.

For example, consider the two-way unbalanced ANOVA model with interaction, given by

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \epsilon_{ijk}$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots n_{ij}$. Also assume that the ϵ_{ijk}'s are $i.i.d.$ $N(0, \sigma^2)$ random variables. Suppose we wanted to test the hypothesis of no interaction. That is, we want to test

$$
\begin{aligned}
H_0 &: \gamma_{ij} = \text{constant for all } (i,j) \\
H_a &: \gamma_{ij} \neq \text{constant for at least one pair } (i,j)
\end{aligned}
$$

To conduct this test in the unbalanced case, we rewrite the hypotheses as

$$\begin{aligned} H_0 &: E(Y) \in C(X_0) \\ H_a &: E(Y) \in C(X) \cap C(X_0)^c \end{aligned}$$

where $C(X_0) \subset C(X)$, and X_0 corresponds to the matrix of the reduced model of no interaction. Thus the F test is

$$F = \frac{\|(M - M_0)Y\|^2 / r(M - M_0)}{\|(I - M)Y\|^2 / r(I - M)}$$

as before.

As a specific example, consider $a = 2$, $b = 3$, $n_{11} = 2$, $n_{12} = 1$, $n_{13} = 3$, $n_{21} = 1$, $n_{22} = 2$, and $n_{23} = 2$. It is clear that we don't have proportional numbers here. The total sample size is $n = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} = 11$. We can write the two-way ANOVA model with interaction as

$$Y = X\beta + \epsilon$$

where

$$Y = \begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{131} \\ Y_{132} \\ Y_{133} \\ Y_{211} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{pmatrix}_{11\times 1}, \beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \eta_1 \\ \eta_2 \\ \eta_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{pmatrix}_{12\times 1}$$

$$X = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{11\times 12}$$

The X_0 matrix corresponding to the no interaction model is

$$X_0 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}_{11\times 6} .$$