Multitiered experiments and their analysis

Christopher Brien

School of Mathematics, University of South Australia,
North Terrace, Adelaide. South Australia 5000

Email: Chris.Brien@unisa.edu.au
WEB Home page: http://phoenix.levels.unisa.edu.au/staff/brien.c.j./home.html

Abstract

Multitiered experiments [Brien, *Biometrics* 39 (1983):51–9] include two-phase, superimposed and some plant and animal experiments. Their mixed model analysis is examined using an example; the method employed guarantees separate terms in the linear model and separate sources in the analysis of variance table for each of several confounded effects. As Harville [*J. Amer. Stat. Assoc.* 86 (1991):812–15] asserts, the custom of representing the sum of confounded effects by a single term or source is "overly restrictive" and "confusing". To ensure there is a term in the model and a source in the table for each confounded effect, it is essential that the randomization employed in the experiment is displayed in the table. We show that this can be achieved by building up the analysis in stages corresponding to the randomization. It will be demonstrated how the analysis can be achieved with software that performs conventional mixed model analyses, such as Minitab.

1. Multitiered experiments

Multitiered experiments (Brien, 1983) include:

- two-phase experiments
- superimposed experiments
- some plant and animal experiments

All these involve some form of multiple randomization.

1.1. A two-phase sensory experiment

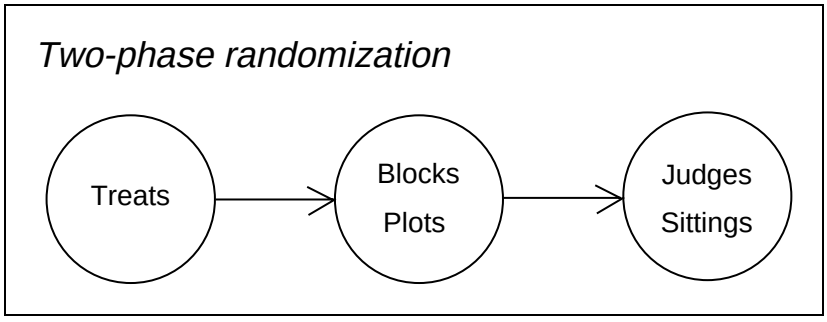
As an example consider a two-phase sensory experiment. In the first phase of the experiment, the field phase, three viticultural treatments were assigned to plots using a randomized complete block design with 5 blocks. The produce from each plot was made into a wine. In the second phase, the evaluation phase, the wines from the first phase were scored by four judges. Each judge participated in 15 sittings and at each sitting scored a wine, the presentation order of the 15 wines being randomized for each judge. A possible layout for the experiment is as follows:

A two-phase sensory experiment

		Judges			
		I	II	III	IV
	1	14	14	12	13
	2	3	2	3	8
	3	4	3	7	14
	4	13	1	9	7
	5	2	11	14	12
	6	10	7	8	4
Sittings	7	12	13	10	1
	8	8	10	6	9
	9	6	5	1	3
	10	11	9	4	11
	11	15	8	11	5
	12	5	6	13	10
	13	9	12	15	15
	14	1	4	5	6
	15	7	15	2	2

Numbers in cells specify block-plot combination

This experiment involves two randomizations, one in each phase of the experiment. These randomizations are represented in the following figure:



2. Mixed model analysis with multiple sources

Harville (1991) asserts that representing the sum of confounded effects by a single term in the model is "overly restrictive" and "confusing". This assertion encapsulates the very problem that our method eliminates — it produces a term in the model for each effect, including a term for each of several confounded effects. Because an ANOVA table is easier to interpret than a model we will concentrate on the ANOVA table here. There will be a source in the analysis for each term in the model and so there will be multiple sources for at least some contrasts in the ANOVA table. The crucial point is that randomization results in the confounding of effects, so that to guarantee multiple sources we must have the ANOVA table display randomization.

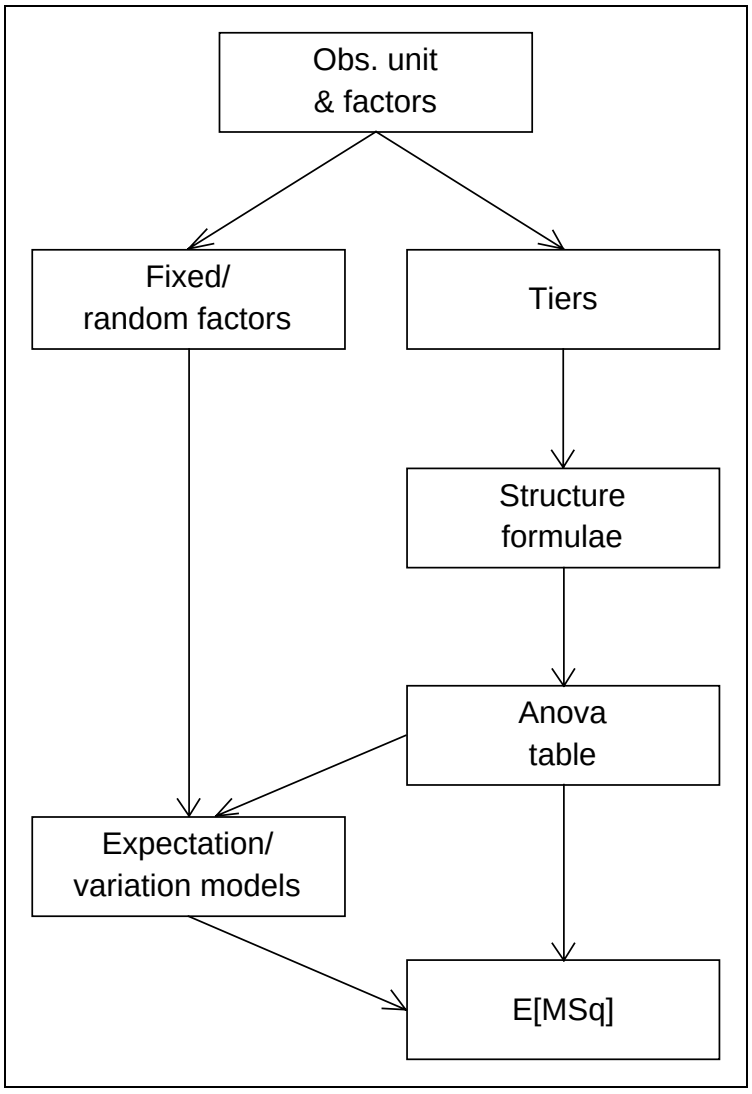
2.1. Procedure for ANOVA

The procedure to be used to derive the ANOVA table involves the following five steps:

Step 1 Divide factors into tiers and specify fixed and random factors
Step 2 Work out structure formulae to describe experiment
Step 3 Derive ANOVA table in stages
Step 4 Derive single structure to specify analysis to package
Step 5 Perform analysis

The process for the first three steps is illustrated in the following figure:

Framework for deriving ANOVA table and models



This procedure is a modified version of the procedure in Brien (1995). The crucial aspect of it is that one works out the formulae, plural, to describe the experiment. Each formula corresponds to different stages in the randomization and uses crossing

and nesting operators to describe the experiment. Then the ANOVA table is derived in stages, each stage incorporating another formula into the analysis.

This procedure is illustrated for the two-phase sensory experiment.

Step 1 Divide factors into tiers and specify fixed and random factors

It is useful to first identify the observational unit, the physical entity for which single observations are recorded. Then we divide the factors in tiers, a tier being a set of factors having the same randomization status. The factors are also classified as either fixed or random.

Observational unit: a judge at a sitting

Tiers:

unrandomized evaluation phase:	{Judges, Sittings}
unrandomized field phase, randomized evaluation phase:	{Blocks, Plots}
randomized field phase:	{Treatments}

Fixed and random factors:

fixed:	Treatments, Judges
random:	Sittings, Blocks, Plots

Step 2 Work out structure formulae to describe experiment

In this step, we first determine the structure ignoring all randomization and then add formulae one at a time to describe terms added at each stage in the randomization.

a) Structure ignoring all randomization

The inherent physical setup is that we have 4 Judges that scored a wine at each of 15 consecutive sittings. This setup is described using the two factors from the tier described as the unrandomized evaluation phase tier. These two factors would appear to be crossed since the sitting for the different judges are linked by the order in which occur. If we use the operator '|' to indicate that factors are crossed, the structure formula is:

Judges | Sittings

(b) Structure after randomizing Blocks-Plots combinations, ignoring field randomization

We now consider the factors in the tier designated unrandomized field phase, randomized evaluation phase: Blocks, Plots. For each Judge, the 14 Block-Plot combinations were randomized to their 15 Sittings. Now, because the randomization was done independently for each Judge and was not restricted to take into account the crossed relationship between the Judges and the Sittings, the relationship between Judges and Sittings becomes that Sittings is nested within Judges. Further, as their is no information to suggest that Blocks and Plots are crossed, Plots will be designated as nested within Blocks. Consequently, the structure formulae describing the experiment to this point are:

Judges / Sittings
Blocks / Plots

Note that the '/' indicates that factors to its right are nested within factors to its left.

(c) Structure incorporating all randomization (and Judge interactions)

The final tier consist of the factor Treatments which is the factor randomized in the field phase. There is only the one factor so there is no need to consider the relationship between factors in this tier. However, it often happens that judges differ in their scoring of wines and so we incorporate interactions between Judges and the filed-phase factors. The final structure formulae for the experiment are:

Judges / Sittings
(Blocks / Plots) | Judges
Treatments | Judges

Step 3 Derive ANOVA table in stages

The ANOVA table is derived in stages; in each stage the terms from one of the formulae are added to the analysis.

(a) ANOVA table for first formula

We must first expand Judges / Sittings to get the sources for table. Wilkinson & Rogers (1973) and Heiberger (1989) give rules for expanding such formulae. Using these rules we have:

Judges / Sittings = Judges + Sittings[Judges]

We now list these terms as sources in the ANOVA table, add their degrees of freedom and use standard rules for expected mean squares to determine theirs. The resulting ANOVA table is:

Source	df	E[MSq]	
		σ^2_{JS}	Q
Judges	3	1	Q(J)
Sittings[Judges]	56	1	
Total	59		

Note that the Q components in the E[MSq] are quadratic forms in the expectation parameters, the quadratic forms being the same as the sum of squares with the units' expected values substituted for their observed values.

(b) ANOVA table for first two formulae

Procedure to augment analysis:

1. Incorporate terms from second formula, placing them indented under terms from previous analysis with which they are confounded.
2. Add a Residual line for each source from the previous analysis that is being partitioned in augmenting the analysis, provided the Residual degrees of freedom is greater than zero. The Residual degrees of freedom is equal to the difference between the those of the original source and the sum of the degrees of freedom of the terms incorporated under it.
3. For each source from the previous analysis that is partitioned in augmenting the analysis, move their E[MSq] to sources into which it is partitioned.
4. Determine E[MSq] for terms from second formula (and a dummy error term if the factors in second formula do not uniquely index the units) using standard rules for expected mean squares and include these in the ANOVA table.

We first need to expand (Blocks / Plots) | Judges to obtain the terms that are to be added to the table.

(Blocks / Plots) | Judges
= Blocks + Plots[Blocks] + Judges + Blocks*Judges + Plots*Judges[Blocks]

The augmented ANOVA table is now formed as described above:

Source	df	E[MSq]					
		σ^2_{JS}	σ^2_{BPJ}	σ^2_{BP}	σ^2_{BJ}	σ^2_{B}	Q
Judges	3	1	1		3		Q(J)
Sittings[Judges]	56						
Blocks	4	1	1	4	3	12	
Plots[Blocks]	10	1	1	4			
Blocks*Judges	12	1	1		3		
Plots*Judges[Blocks]	30	1	1				
Total	59						

(c) ANOVA table for all formulae

Treatments | Judges = Treatments + Judges + Treatments*Judges

Source	df	E[MSq]					
		σ^2_{JS}	σ^2_{BPJ}	σ^2_{BP}	σ^2_{BJ}	σ^2_{B}	Q
Judges	3	1	1		3		Q(J)
Sittings[Judges]	56						
Blocks	4	1	1	4	3	12	
Plots[Blocks]	10						
Treatments	2	1	1	4			Q(T)
Residual	8	1	1	4			
Blocks*Judges	12	1	1		3		
Plots*Judges[Blocks]	30						
Treatments*Judges	6	1	1				Q(T*J)
Residual	24	1	1				
Total	59						

Notes:

- Each source in the analysis would have a term in either the expectation or variation models.
- ANOVA table has multiple sources for all contrasts except Judges and so shows confounding resulting from randomization.
- Block-treatment interactions: interaction between units and factors randomized to them
  - ⇒ Treatments*Blocks assumed zero, but not necessary
  - ⇒ Judges interactions with wines of block-treatment type not assumed zero.

This experiment would traditionally be analysed using a split-plot analysis. The analysis of variance table would be something like the following table:

Source	df	E[MSq]			
		σ^2	σ^2_{BT}	σ^2_{B}	Q
Blocks	4	1	4	12	
Treatments	2	1	4		Q(T)
Main plot error	8	1	4		
Judges	3	1			Q(J)
Treatments*Judges	6	1			Q(T*J)
Subplot error	36	1			
Total	59				

The principal difference between this and the previous analyses is that Blocks*Judges is not included in the latter analysis. However, another important difference is that the confounding resulting from the randomization is not evident.

Step 4 Derive single structure to specify analysis to package

For orthogonal experiments, identify set of unconfounded terms that produce correct analysis. For the example, it is clear that the following formula specifies a set of unconfounded terms that will produce the correct partition of the total sum of squares:

Treatments | Blocks | Judges - Treatments*Blocks*Judges

Note this produces Treatments*Blocks as a computational convenience. In our analysis this interaction is assumed to be zero.

Step 5 Perform analysis

For example, Minitab

1. Set up columns for the factors Treatments, Blocks, Judges and the data.

2. Use ANOVA to get SSq for analysis

```
#  Analysis of data
anova y = Blocks | Treats | Judges   &
                        - Blocks*Treats*Judges;
random Blocks;
ems.
```

3. Perhaps use ANOVA to get E[MSq]

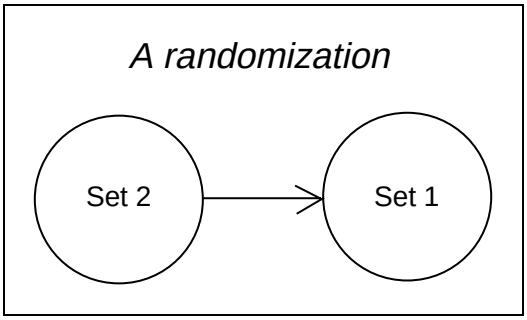
```
#Sequence of analyses to get E[MSq]
#Analysis to get E[MSq] for first formula
anova y = Judges;
ems.

#Analysis to get E[MSq] for second formula
anova y = Blocks Treats(Blocks) Blocks | Judges;
random Blocks;
ems.

#Analysis to get E[MSq] for third formulae
anova y = Treats | Judges;
ems.
```

3. Types of randomization

In this section, we formally define a randomization and identify a number of different kinds that can occur. One motivation for this is to better understand when multitiered experiments arise.

3.1. A randomization:

- All the combinations of the levels of the factors in set 1 have a combination of the levels of those in set 2 assigned to them.
- Each levels combination is assigned using a single random number.



3.2. Joint randomization:

Joint randomization occurs when set 2 in a randomization contains more than one factor. For example, factorial experiments in a CRD or RCBD or LS involve joint randomization: in a two-factor RCBD experiment the combinations of A and B would be randomized to the Plots within Blocks.

3.3. Multiple randomization:

Multiple randomization occurs when two or more randomizations are performed. Below we define some different types of multiple randomization. These are illustrated in the accompanying diagram.

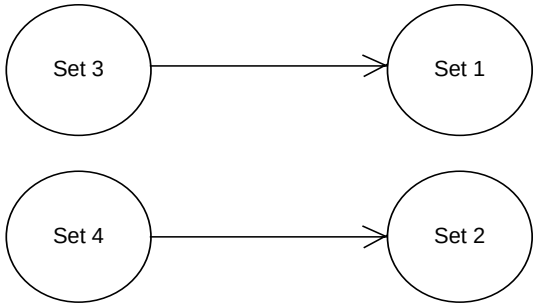
Independent multiple randomization:

The two sets of levels combinations of two sets of factors are randomized to those of two different sets of factors.

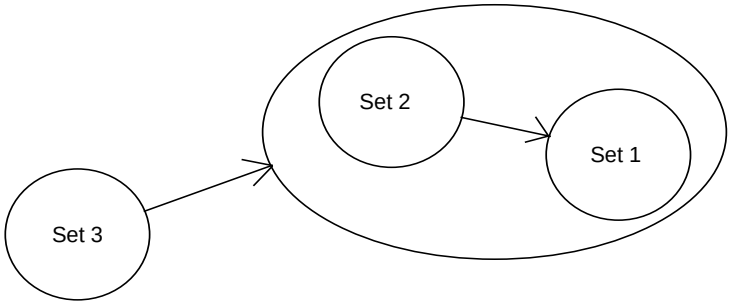
Example: standard split-plot

Types of multiple randomization

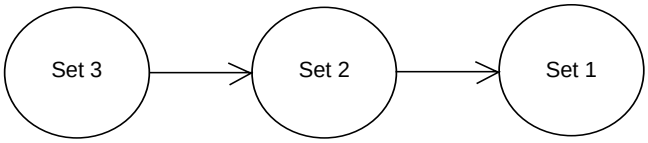
Independent multiple randomization



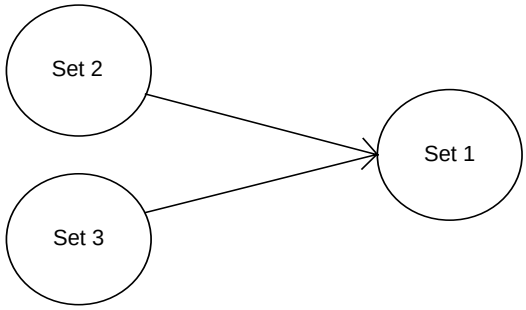
Conditional multiple randomization



Composite multiple randomization



Coincident multiple randomization



Conditional multiple randomization:

A set of factors that involves randomised factors and factors to which they are randomized has the levels combination of a set of factors randomised to their levels combinations.

Example: split-plots in a Latin square; superimposed experiments

Composite multiple randomization:

There is an initial randomization of one set of factors onto another set of factors. The latter set of factors are then randomized onto a third set of factors.

Example: two-phase experiments

Coincident multiple randomization:

The two sets of levels combinations of two sets of factors are randomized to those of the same set of factors.

Example: replicates and treatments are both randomized to the same units.

3.4. Representing randomization

- observational studies require 1 formula as no randomization;
- all randomized experiments with either single joint or only independent multiple randomization require 2 formulae;
- other randomized experiments require > 2 formulae.

All multiple randomizations, except independent multiple randomization, lead to multitiered experiments.

4. Summary

1. Have described the following ANOVA procedure:

   Step 1 Divide factors into tiers and specify fixed and random factors
   Step 2 Work out structure formulae to describe experiment
   Step 3 Derive ANOVA table in stages
   Step 4 Derive single structure to specify analysis to package
   Step 5 Perform analysis

2. The ANOVA table derived in this way portrays the randomization used in the experiment. For this to happen, two or more formulae must be used — one to describe the factors in the inherent physical setup and one or more to describe the randomized factors.

3. If ANOVA does portray randomization, the advantages include:

   - multiple sources for contrasts possible
   - different analyses for different randomizations (for example, the form of the analysis for the randomized complete block design and for the two-way factorial with no interaction differ)
   - inadequate replication becomes evident

Of the advantages listed we have demonstrated only the first in this paper. The second can be readily verified by deriving the ANOVA tables for the two experiments cited as examples. The third advantage is discussed in Brien (1995).

5. References

Brien, C. J. (1983) Analysis of variance tables based on experimental structure. *Biometrics*, 39, 51–59.

Brien, C. J. (1995) Multitiered experiments and mixed models. submitted for publication.

Harville, D.A. (1991) Contribution to the discussion of a paper by M.L. Samuels, G. Casella and G.P. McCabe. *Journal of the American Statistical Association*, 86, 812–815.

Heiberger, R.M. (1989) *Computation for the analysis of designed experiments*. Wiley, New York.

Samuels, M. L., Casella, G., and McCabe, G.P. (1991) Interpreting Blocks and Random Factors. *Journal of the American Statistical Association*, 86, 798–821.

Wilkinson, G. N. and Rogers, C. E. (1973) Symbolic description of factorial models for analysis of variance. *Applied Statistics*, 22, 392–399.