Principles in the design of multiphase experiments with a later laboratory phase: orthogonal designs

Chris Brien¹, Bronwyn Harch², Ray Correll² & Rosemary Bailey³

¹University of South Australia, ²CSIRO Mathematics, Informatics & Statistics, ³Queen Mary University of London

http://chris.brien.name/multitier

Chris.brien@unisa.edu.au
Outline

1. Primary experimental design principles
2. Factor-allocation description for standard designs.
3. Principles for simple multiphase experiments.
4. Principles leading to complications, even with orthogonality.
5. Summary
1) Primary experimental design principles

- **Principle 1** (Evaluate designs with skeleton ANOVA tables)
  - Use whether or not data to be analyzed by ANOVA.

- **Principle 2** (Fundamentals): Use randomization, replication and blocking or local control.

- **Principle 3** (Minimize variance): Block entities to form new entities, within new entities being more homogeneous; assign treatments to least variable entity-type.

- **Principle 4** (Split units): confound some treatment sources with more variable sources if some treatment factors:
  - i. require larger units than others,
  - ii. are expected to have a larger effect, or
  - iii. are of less interest than others.
A standard athlete training example

- 9 training conditions — combinations of 3 surfaces and 3 intensities of training — to be investigated.

- Assume the prime interest is in surface differences
  - Intensities are only included to observe the surfaces over a range of intensities.

- Testing is to be conducted over 4 Months:
  - In each month, 3 endurance athletes are to be recruited.
  - Each athlete will undergo 3 tests, separated by 7 days, under 3 different training conditions.

- On completion of each test, the heart rate of the athlete will be measured.

- Randomize 3 intensities to 3 athletes in a month and 3 surfaces to 3 tests in an athlete.
  - A split-unit design, employing Principles 2, 3 and 4(iii).
2) Factor-allocation description for standard designs

- Standard designs involve a single allocation in which a set of treatments is assigned to a set of units:
  - treatments are whatever are allocated;
  - units are what treatments are allocated to;
  - treatments and units each referred to as a set of objects;
- Often do by randomization using a permutation of the units.
  - More generally treatments are allocated to units e.g. using a spatial design or systematically
- Each set of objects is indexed by a set of factors:
  - Unit or unallocated factors (indexing units);
  - Treatment or allocated factors (indexing treatments).
- Represent the allocation using factor-allocation diagrams that have a panel for each set of objects with:
  - a list of the factors; their numbers of levels; their nesting relationships.

(Nelder, 1965; Brien, 1983; Brien & Bailey, 2006)
Factor-allocation diagram for the standard athlete training experiment

- One allocation (randomization):
  - a set of training conditions to a set of tests.

![Diagram](image)

- The set of factors belonging to a set of objects forms a **tier**:
  - they have the same status in the allocation (randomization):
    - {Intensities, Surfaces} or {Months, Athletes, Tests}
    - Textbook experiments are two-tiered.

- A crucial feature is that diagram automatically shows EU and restrictions on randomization/allocation.
Some derived items

- Sets of generalized factors (terms in the mixed model):
  - Months, Months∧Athletes, Months∧Athletes∧Tests;
  - Intensities, Surfaces, Intensities∧Surfaces.

- Corresponding types of entities (groupings of objects):
  - month, athlete, test (last two are Eus);
  - intensity, surface, training condition (intensity-surface combination).

- Corresponding sources (in an ANOVA):
  - Months, Athletes[M], Tests[M∧A];
  - Intensities, Surfaces, Intensities#Surfaces.
## Skeleton ANOVA

<table>
<thead>
<tr>
<th>Tests Tier</th>
<th>Source</th>
<th>df</th>
<th>Training Conditions Tier</th>
<th>Source</th>
<th>df</th>
<th>$E[MSq]$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Months</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>$\sigma^2_{MAT}$ $\sigma^2_{MA}$ $\sigma^2_M$</td>
</tr>
<tr>
<td></td>
<td>Athletes[M]</td>
<td>8</td>
<td>Intensities</td>
<td>2</td>
<td>1</td>
<td>$q_I(\mu)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Residual</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tests[M^A]</td>
<td>24</td>
<td>Surfaces</td>
<td>2</td>
<td>1</td>
<td>$q_S(\mu)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>l#S</td>
<td>4</td>
<td>1</td>
<td>$q_{lS}(\mu)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Residual</td>
<td>18</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- Intensities is confounded with the more-variable Athletes[M] & Surfaces with Tests[M^A].
3) Principles for simple multiphase experiments

- Suppose in the athlete training experiment:
  - in addition to heart rate taken immediately upon completion of a test,
  - the free haemoglobin is to be measured using blood specimens taken from the athletes after each test, and
  - the specimens are transported to the laboratory for analysis.

- The experiment is two phase: testing and laboratory phases.
  - The outcome of the testing phase is heart rate and a blood specimen.
  - The outcome of the laboratory phase is the free haemoglobin.

- How to process the specimens from the first phase in the laboratory phase?
Some principles

- **Principle 5** (Simplicity desirable): assign first-phase units to laboratory units so that each first-phase source is confounded with a single laboratory source.
  - Use composed randomizations with an orthogonal design.
- **Principle 6** (Preplan all): if possible.
- **Principle 7** (Allocate all and randomize in laboratory): always allocate all treatment and unit factors and randomize first-phase units and lab treatments.
- **Principle 8** (Big with big):
  - Confound big first-phase sources with big laboratory sources, provided no confounding of treatment with first-phase sources.
A simple two-phase athlete training experiment

- Simplest is to randomize specimens from a test to locations (in time or space) during the laboratory phase.

![Diagram](image)

<table>
<thead>
<tr>
<th>locations tier</th>
<th>tests tier</th>
<th>training conditions tier</th>
<th>E[MSq]</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>df</td>
<td>source</td>
<td>df</td>
</tr>
<tr>
<td>Locations</td>
<td>35</td>
<td>Months</td>
<td>3</td>
</tr>
<tr>
<td>Athletes[M]</td>
<td>8</td>
<td>Intensities</td>
<td>2</td>
</tr>
<tr>
<td>Tests[M^A]</td>
<td>24</td>
<td>Surfaces</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I#S</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Residual</td>
<td>18</td>
</tr>
</tbody>
</table>
A simple two-phase athlete training experiment (cont’d)

<table>
<thead>
<tr>
<th>locations tier</th>
<th>tests tier</th>
<th>training conditions tier</th>
<th>E[MSq]</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>source</td>
<td>source</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>df</td>
<td>df</td>
<td>σ_L^2</td>
</tr>
<tr>
<td>Locations 35</td>
<td>Months 3</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Athletes[M] 8</td>
<td>Intensities 2</td>
<td></td>
<td>1 1 3</td>
</tr>
<tr>
<td>Residual 6</td>
<td>Surfaces 2</td>
<td></td>
<td>1 1 3</td>
</tr>
<tr>
<td>Tests[M∧A] 24</td>
<td>I#S 4</td>
<td></td>
<td>1 1</td>
</tr>
<tr>
<td>Residual 18</td>
<td></td>
<td></td>
<td>1 1</td>
</tr>
</tbody>
</table>

- No. tests = no. locations = 36 and so tests sources exhaust the locations source.
- Cannot separately estimate locations and tests variability, but can estimate their sum.
- But do not want to hold blood specimens for 4 months.
A simple two-phase athlete training experiment (cont’d)

- Simplest is to align lab-phase and first-phase blocking.

![Diagram showing the experimental design with 3 Intensities, 3 Surfaces, 4 Months, 3 Athletes in M, 3 Tests in M, A, 4 Batches, and 9 Locations in B.]

- Note Months confounded with Batches (i.e. Big with Big).

<table>
<thead>
<tr>
<th>locations tier</th>
<th>source</th>
<th>df</th>
<th>tests tier</th>
<th>source</th>
<th>df</th>
<th>training conditions tier</th>
<th>source</th>
<th>df</th>
<th>E[MSq]</th>
<th>( \sigma_{BL}^2 )</th>
<th>( \sigma_{B}^2 )</th>
<th>( \sigma_{MAT}^2 )</th>
<th>( \sigma_{MA}^2 )</th>
<th>( \sigma_{M}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>locations tier</td>
<td></td>
<td></td>
<td>tests tier</td>
<td></td>
<td></td>
<td>training conditions tier</td>
<td></td>
<td></td>
<td>E[MSq]</td>
<td>( \sigma_{BL}^2 )</td>
<td>( \sigma_{B}^2 )</td>
<td>( \sigma_{MAT}^2 )</td>
<td>( \sigma_{MA}^2 )</td>
<td>( \sigma_{M}^2 )</td>
</tr>
<tr>
<td>source</td>
<td>df</td>
<td></td>
<td>source</td>
<td>df</td>
<td></td>
<td>source</td>
<td>df</td>
<td></td>
<td></td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Batches</td>
<td>3</td>
<td></td>
<td>Months</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Locations[B]</td>
<td>32</td>
<td></td>
<td>Athletes[M]</td>
<td>8</td>
<td></td>
<td>Intensities</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>( q_i(\mu) )</td>
<td>( q_i(\mu) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Residual</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tests[M&amp;A]</td>
<td>24</td>
<td></td>
<td>Surfaces</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>( q_S(\mu) )</td>
<td>( q_S(\mu) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I#S</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>( q_{IS}(\mu) )</td>
<td>( q_{IS}(\mu) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Residual</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The multiphase law

- DF for sources from a previous phase can never be increased as a result of the laboratory-phase design.
- However, it is possible that first-phase sources are split into two or more sources, each with fewer degrees of freedom than the original source.

<table>
<thead>
<tr>
<th>locations tier</th>
<th>tests tier</th>
<th>training conditions tier</th>
<th>E[MSq]</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>df</td>
<td>source</td>
<td>df</td>
</tr>
<tr>
<td>Batches</td>
<td>3</td>
<td>Months</td>
<td>3</td>
</tr>
<tr>
<td>Locations[B]</td>
<td>32</td>
<td>Athletes[M]</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Residual</td>
<td>6</td>
</tr>
<tr>
<td>Tests[M&amp;A]</td>
<td>24</td>
<td>Surfaces</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I#S</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Residual</td>
<td>18</td>
</tr>
</tbody>
</table>

- DF for first phase sources unaffected.
4) Principles leading to complications, even with orthogonality

- **Principle 9** (Use pseudofactors):
  - An elegant way to split sources (as opposed to introducing grouping factors unconnected to real sources of variability).

- **Principle 10** (Compensating across phases):
  - Sometimes, if something is confounded with more variable first-phase source, can confound with less variable lab source.

- **Principle 11** (Laboratory replication):
  - Replicate laboratory analysis of first-phase units if lab variability much greater than 1st-phase variation;
  - Often involves splitting **product** from the first phase into **portions** (e.g. batches of harvested crop, wines, blood specimens into aliquots, drops, lots, samples and fractions).

- **Principle 12** (Laboratory treatments):
  - Sometimes treatments are introduced in the laboratory phase and this involves extra randomization.
5) Summary

- Have provided 4 standard principles and 8 principles specific to orthogonal, multiphase designs.

- In practice, will be important to have some idea of likely sources of laboratory variation.

- Are laboratory treatments to be incorporated?

- Will laboratory replicates be necessary?
References


Web address for link to Multitiered experiments site:  [http://chris.brien.name/multitier](http://chris.brien.name/multitier)