Factors in designed experiments

R. A. Bailey
University of St Andrews

School of Mathematics and Statistics,
Research Day, 20 January 2015

Example 1: a real recent experiment

In 2012 the UK's Food and Environment Research Agency conducted an experiment to find out "the effects of neonicotinoid seed treatments on bumble bee colonies under field conditions" (from a DEFRA report available on the web, Crown copyright 2013).

`fera.co.uk/ccss/documents/defraBumbleBeereportPS2371V4A.pdf`

Example 1: a real recent experiment

In 2012 the UK's Food and Environment Research Agency conducted an experiment to find out "the effects of neonicotinoid seed treatments on bumble bee colonies under field conditions" (from a DEFRA report available on the web, Crown copyright 2013).

`fera.co.uk/ccss/documents/defraBumbleBeereportPS2371V4A.pdf`

Site	Treatment of oilseed rape seeds
Site A, near Lincoln	no treatment
Site B, near York	Modesto™
Site C, near Scunthorpe	Chinook™

Example 1: a real recent experiment

In 2012 the UK's Food and Environment Research Agency conducted an experiment to find out "the effects of neonicotinoid seed treatments on bumble bee colonies under field conditions" (from a DEFRA report available on the web, Crown copyright 2013).

`fera.co.uk/ccss/documents/defraBumbleBeereportPS2371V4A.pdf`

Site	Treatment of oilseed rape seeds
Site A, near Lincoln	no treatment
Site B, near York	Modesto™
Site C, near Scunthorpe	Chinook™

Twenty colonies of bumble bees were placed at each site. Various outcomes were measured on each colony.

Factors and partitions

There is factor Site with three levels.
This gives a partition of the set of 60 colonies into three parts.

Factors and partitions

There is factor Site with three levels.
This gives a partition of the set of 60 colonies into three parts.

There is a factor Colony with 60 parts.

$$\text{Colony} \preccurlyeq \text{Site}.$$

This means that each colony is in a single site.

Factors and partitions

There is factor Site with three levels.
This gives a partition of the set of 60 colonies into three parts.

There is a factor Colony with 60 parts.

$$\text{Colony} \preccurlyeq \text{Site}.$$

This means that each colony is in a single site.

There is a factor Treatment with three levels.

$$\text{Site} \not\preccurlyeq \text{Treatment} \quad \text{and} \quad \text{Treatment} \not\preccurlyeq \text{Site}.$$

Factors and partitions

There is factor Site with three levels.
This gives a partition of the set of 60 colonies into three parts.

There is a factor Colony with 60 parts.

$$\text{Colony} \preccurlyeq \text{Site}.$$

This means that each colony is in a single site.

There is a factor Treatment with three levels.

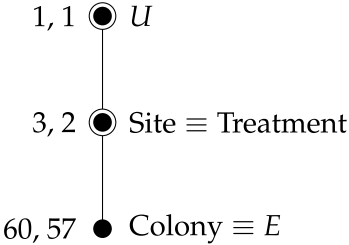
$$\text{Site} \preccurlyeq \text{Treatment} \quad \text{and} \quad \text{Treatment} \preccurlyeq \text{Site}.$$

This means that

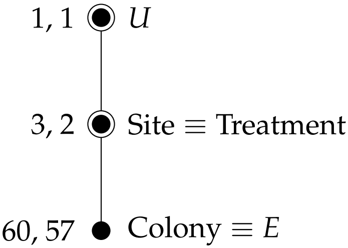
$$\text{Site} \equiv \text{Treatment}$$

(Site is aliased with Treatment),
because they give the same partition.

Example 1: Hasse diagram and skeleton anova



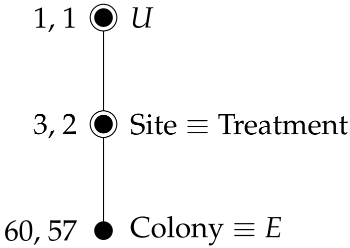
Example 1: Hasse diagram and skeleton anova



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

Example 1: Hasse diagram and skeleton anova

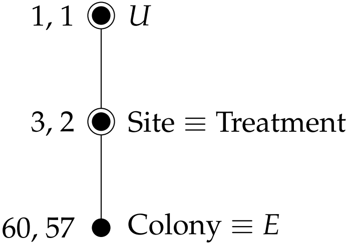


Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites.

Example 1: Hasse diagram and skeleton anova

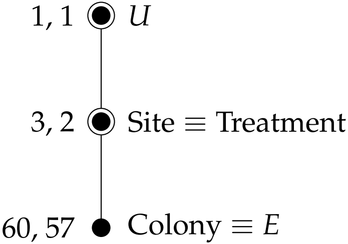


Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites. Therefore, there is no way of giving confidence intervals for the estimates of treatment differences, or of giving P values for testing the hypothesis of no treatment difference.

Example 1: Hasse diagram and skeleton anova

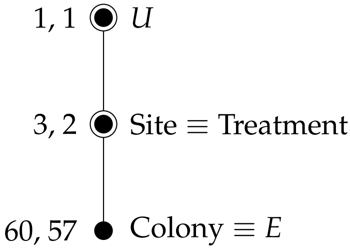


Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites. Therefore, there is no way of giving confidence intervals for the estimates of treatment differences, or of giving P values for testing the hypothesis of no treatment difference. The official report does claim to give confidence intervals and P values.

Example 1: Hasse diagram and skeleton anova



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites. Therefore, there is no way of giving confidence intervals for the estimates of treatment differences, or of giving P values for testing the hypothesis of no treatment difference. The official report does claim to give confidence intervals and P values.

The Hasse diagram can clearly show such false replication before the experiment is carried out.

Example 2: a made-up experiment based on a real one

This example is based on a real experiment carried out, at some expense of time and work, by some biologists. The details are confidential (because I was referee for their submitted paper), so I have changed the setting but preserved the mathematical structure.

Example 2: a made-up experiment based on a real one

This example is based on a real experiment carried out, at some expense of time and work, by some biologists. The details are confidential (because I was referee for their submitted paper), so I have changed the setting but preserved the mathematical structure.

Every so often, there is a chocolate-cake-baking contest. Different cooks use different recipes, and bring their cakes to the common room, where Valerie tastes each one and gives it a mark out of 100.

Example 2: a made-up experiment based on a real one

This example is based on a real experiment carried out, at some expense of time and work, by some biologists. The details are confidential (because I was referee for their submitted paper), so I have changed the setting but preserved the mathematical structure.

Every so often, there is a chocolate-cake-baking contest. Different cooks use different recipes, and bring their cakes to the common room, where Valerie tastes each one and gives it a mark out of 100.

Are differences in the marks caused by differences between cooks or by differences between recipes?

Example 2: ✓ shows the combinations which occur

cooks 1–12

	ck 1	ck 2	ck 3	ck 4	ck 5	ck 6	ck 7	ck 8	ck 9	ck 10	ck 11	ck 12
recipe 1		✓	✓	✓								
recipe 2	✓		✓	✓								
recipe 3	✓	✓		✓								
recipe 4	✓	✓	✓									
recipe 5						✓	✓	✓				
recipe 6					✓		✓	✓				
recipe 7					✓	✓		✓				
recipe 8					✓	✓	✓					
recipe 9										✓	✓	✓
recipe 10									✓		✓	✓
recipe 11									✓	✓		✓
recipe 12									✓	✓	✓	

The supremum of two factors

The supremum $A \vee B$ of factors A and B is defined to satisfy:

- $A \preccurlyeq A \vee B$, and $B \preccurlyeq A \vee B$;
- if there is any other factor C
  with $A \preccurlyeq C$ and $B \preccurlyeq C$,
  then $A \vee B \preccurlyeq C$.

Each part of factor $A \vee B$ is a union of parts of A and is also a union of parts of B, and is as small as possible subject to this.

Example 2: silly data (rows are recipes, columns are cooks)

	63	65	62								
64		61	68								
67	63		62								
64	68	63									
					15	19	12				
				13		13	16				
				17	11		18				
				10	14	17					
									81	88	85
								87		82	81
								83	85		82
								86	83	86	

Example 2: silly data (rows are recipes, columns are cooks)

	63	65	62								
64		61	68								
67	63		62								
64	68	63									
					15	19	12				
				13		13	16				
				17	11		18				
				10	14	17					
									81	88	85
								87		82	81
								83	85		82
								86	83	86	

Statistician 1: There are differences between cooks. Fit Cook and subtract, then there are essentially no differences between recipes.

Example 2: silly data (rows are recipes, columns are cooks)

	63	65	62								
64		61	68								
67	63		62								
64	68	63									
					15	19	12				
				13		13	16				
				17	11		18				
				10	14	17					
									81	88	85
								87		82	81
								83	85		82
								86	83	86	

Statistician 1: There are differences between cooks. Fit Cook and subtract, then there are essentially no differences between recipes.

Statistician 2: There are differences between recipes. Fit Recipe and subtract, then there are essentially no differences between cooks.

Example 2: silly data (rows are recipes, columns are cooks)

	63	65	62								
64		61	68								
67	63		62								
64	68	63									
					15	19	12				
				13		13	16				
				17	11		18				
				10	14	17					
									81	88	85
								87		82	81
								83	85		82
								86	83	86	

Statistician 1: There are differences between cooks. Fit Cook and subtract, then there are essentially no differences between recipes.

Statistician 2: There are differences between recipes. Fit Recipe and subtract, then there are essentially no differences between cooks.

Statistician 3: Recipe $\vee$ Cook $=$ Square.

Example 2: silly data (rows are recipes, columns are cooks)

	63	65	62								
64		61	68								
67	63		62								
64	68	63									
					15	19	12				
				13		13	16				
				17	11		18				
				10	14	17					
									81	88	85
								87		82	81
								83	85		82
								86	83	86	

Statistician 1: There are differences between cooks. Fit Cook and subtract, then there are essentially no differences between recipes.

Statistician 2: There are differences between recipes. Fit Recipe and subtract, then there are essentially no differences between cooks.

Statistician 3: Recipe $\vee$ Cook $=$ Square.
There are differences between squares: after fitting Square there are essentially no further differences between cooks or between recipes. We cannot tell whether the differences between squares are caused by cooks or recipes, because they are confounded.

And the moral is . . .

Most statistical software does not explicitly facilitate the calculation of suprema.

And the moral is . . .

Most statistical software does not explicitly facilitate the calculation of suprema.

Authors' response to referee: "The statistical software R does not allow you to fit nested models."

And the moral is . . .

Most statistical software does not explicitly facilitate the calculation of suprema.

Authors' response to referee: "The statistical software R does not allow you to fit nested models."

Problem: how to promote understanding of factors and their relationships so that they are properly taken into account in both the design and analysis of experiments?