Automatic analysis of variance for orthogonal plot structures

Heiko Grossmann
Queen Mary, University of London

Isaac Newton Institute Cambridge, 13 October 2011

1. Introduction
2. Theory
3. `AutomaticANOVA` package
4. Examples

1. INTRODUCTION

"Every statistician knows how to do the analysis of variance!"

"Every statistician knows how to do the analysis of variance!"

- ► True, but often only for a certain selection of models. →
  "Template approach"
- ► Less true, if model has more complicated blocking structure and
  includes random and fixed effects. → usually required in modern
  applications

"Every statistician knows how to do the analysis of variance!"

- ► True, but often only for a certain selection of models. →
  "Template approach"
- ► Less true, if model has more complicated blocking structure and
  includes random and fixed effects. → usually required in modern
  applications

Audience will be tested later!

"Customers" (e.g. biologists)

► are often familiar with one-way and two-way ANOVA for completely randomized designs, but not much more,

► usually do experiments for which no off-the-shelf analysis is available,

► often have good intuition about the design of a study, but think about it in terms different from those of the statistician,

► usually seek input from statisticians only after the experiment,

► love to use Excel.

Example 1 (Cider apples; Bailey (2008, p. 145))

- ▶ An orchard of cider-apple trees was divided into **6 blocks** of **5 trees** each.
- ▶ Each of **3 types of spray** was applied to **2 blocks**.
- ▶ Within each block the same **5 pruning methods** were used on different trees.
- ▶ The response was the number of apples fallen down, given as a percentage of the total for each tree.

Example 1 (Cider apples; Bailey (2008, p. 145))

- ► An orchard of cider-apple trees was divided into **6 blocks** of **5 trees** each.
- ► Each of **3 types of spray** was applied to **2 blocks**.
- ► Within each block the same **5 pruning methods** were used on different trees.
- ► The response was the number of apples fallen down, given as a percentage of the total for each tree.

> What is the correct analysis of variance?

Answers to how Example 1 should be analyzed appear to reveal a **divide** between the UK(+Australia) and the "rest of the world".

UK+AUS:
Emphasis on **plot (or block) structure** and **treatment structure**

RoW:
Emphasis on **model equation** and general linear model

- ▶ Nelder (1965a,b) developed ANOVA for **simple orthogonal block structures**.
- ▶ Theory further generalized by Bailey for **orthogonal plot (or block) structures** (see summary in Bailey (2008, Chap. 10)).
- ▶ **GenStat** appears to be the only stats software which implements this approach.

- ► Nelder (1965a,b) developed ANOVA for **simple orthogonal block structures**.
- ► Theory further generalized by Bailey for **orthogonal plot (or block) structures** (see summary in Bailey (2008, Chap. 10)).
- ► **GenStat** appears to be the only stats software which implements this approach.

All of the above is hardly known outside the UK.
↓
Slow take-up of the theory

Present and illustrate use of `AutomaticANOVA` package which implements the theory in Bailey (2008, Chap. 10)

`AutomaticANOVA` **package**

`AutomaticANOVA` package

- ▶ provides **single** function for analyzing any **orthognal design** as defined by Bailey (2008, p. 198);

`AutomaticANOVA` package

- ▶ provides **single** function for analyzing any **orthognal design** as defined by Bailey (2008, p. 198);
- ▶ requires **minimal** input
  - ▶ Excel file, list of column numbers for plot factors and similar list for treatment factors, response variable (if available);

`AutomaticANOVA` package

- ▶ provides **single** function for analyzing any **orthognal design** as defined by Bailey (2008, p. 198);
- ▶ requires **minimal** input
  - ▶ Excel file, list of column numbers for plot factors and similar list for treatment factors, response variable (if available);
- ▶ automatically **deduces appropriate model** from design, does **not** require a **model formula**!

`AutomaticANOVA` package

- ▶ provides **single** function for analyzing any **orthognal design** as defined by Bailey (2008, p. 198);
- ▶ requires **minimal** input
  - ▶ Excel file, list of column numbers for plot factors and similar list for treatment factors, response variable (if available);
- ▶ automatically **deduces appropriate model** from design, does **not** require a **model formula**!
- ▶ automatically produces
  - ▶ **null ANOVA**, **skeleton ANOVA** or **full ANOVA** as appropriate,
  - ▶ **Hasse diagrams** for both plot and treatment factors,
  - ▶ estimates of **variance components** if data are available or **algebraic formulae** for ANOVA estimators of VCs otherwise;

Package features

`AutomaticANOVA` package

- provides **single** function for analyzing any **orthognal design** as defined by Bailey (2008, p. 198);
- requires **minimal** input
  - Excel file, list of column numbers for plot factors and similar list for treatment factors, response variable (if available);
- automatically **deduces appropriate model** from design, does **not** require a **model formula**!
- automatically produces
  - **null ANOVA**, **skeleton ANOVA** or **full ANOVA** as appropriate,
  - **Hasse diagrams** for both plot and treatment factors,
  - estimates of **variance components** if data are available or **algebraic formulae** for ANOVA estimators of VCs otherwise;
- eliminates **redundant** factors;

Package features

`AutomaticANOVA` package

- ▶ provides **single** function for analyzing any **orthognal design** as defined by Bailey (2008, p. 198);
- ▶ requires **minimal** input
    - ▶ Excel file, list of column numbers for plot factors and similar list for treatment factors, response variable (if available);
- ▶ automatically **deduces appropriate model** from design, does **not** require a **model formula**!
- ▶ automatically produces
    - ▶ **null ANOVA**, **skeleton ANOVA** or **full ANOVA** as appropriate,
    - ▶ **Hasse diagrams** for both plot and treatment factors,
    - ▶ estimates of **variance components** if data are available or **algebraic formulae** for ANOVA estimators of VCs otherwise;
- ▶ eliminates **redundant** factors;
- ▶ automatically adds **pseudo-factors** when needed;

Package features

`AutomaticANOVA` package

- ▶ provides **single** function for analyzing any **orthognal design** as defined by Bailey (2008, p. 198);
- ▶ requires **minimal** input
  - ▶ Excel file, list of column numbers for plot factors and similar list for treatment factors, response variable (if available);
- ▶ automatically **deduces appropriate model** from design, does **not** require a **model formula**!
- ▶ automatically produces
  - ▶ **null ANOVA**, **skeleton ANOVA** or **full ANOVA** as appropriate,
  - ▶ **Hasse diagrams** for both plot and treatment factors,
  - ▶ estimates of **variance components** if data are available or **algebraic formulae** for ANOVA estimators of VCs otherwise;
- ▶ eliminates **redundant** factors;
- ▶ automatically adds **pseudo-factors** when needed;
- ▶ deals with degrees of freedom equal to zero.

2. THEORY

Brief summary of the theory in Bailey (2008, Chap. 10).

No contribution from me.

Important concepts to be mentioned:

- ▶ **Finer relation** for factors
- ▶ **Infimum** and **supremum** of two factors
- ▶ **Hasse diagram**
- ▶ **Orthogonal plot structures**
- ▶ **Orthogonal treatment structures**
- ▶ **Orthogonal designs**
- ▶ **Null ANOVA** and **skeleton ANOVA**

Let $\mathbf{y} \in \mathbb{R}^N$ be a response vector in an ANOVA experiment.

1. An **orthogonal plot structure** $\mathcal{F}$ allows a decomposition of $\mathbb{R}^N$ into **mutually orthogonal** subspaces W_F, $F \in \mathcal{F}$. Further $\|\mathbf{y}\|^2$ is the sum of the sums of squares for the factors in $\mathcal{F}$.

2. An **orthogonal treatment structure** $\mathcal{G}$ gives rise to another set of **mutually orthogonal** subspaces $\{W_G : G \in \mathcal{G}\}$.

3. If the factors in $\mathcal{F}$ have **random effects**, then the W_F in 1. are the **eigenspaces** or **strata** of the covariance matrix $\mathrm{Cov}(\mathbf{Y})$ of the random vector which generates $\mathbf{y}$—assuming $\mathrm{Cov}(\mathbf{Y})$ has a specific form.

4. For an **orthogonal design** the effect of every treatment factor $G \in \mathcal{G}$ can be tested in the unique stratum W_F, $F \in \mathcal{F}$, for which $W_G \subseteq W_F$.

Let Ω be the set of N **observational units** and $\mathcal{T}$ be the set of **treatments** in an experiment.

A **plot (or block) factor** is a function $F : \Omega \rightarrow R_F$, where R_F is finite with $|R_F| = n_F$. Similarly, a **treatment factor** is a function $G : \mathcal{T} \rightarrow R_G$ with $|R_G| = n_G$.

For F on Ω the F-**class** corresponding to level $i \in R_F$ is

$$F[[i]] = \{\omega \in \Omega : F(\omega) = i\}.$$

For G on $\mathcal{T}$ the G-**class** corresponding to level $j \in R_G$ is

$$G[[j]] = \{t \in \mathcal{T} : G(t) = j\}.$$

Factors F and G on the same set are **equivalent**, written as $F \equiv G$, if the set of F-classes and set of G-classes are equal.

Let F and G be factors on the same set. F is **finer** than G, or G is **coarser** than F, if every F-class is a subset of a G-class, but $F \not\equiv G$.

If F is finer than G, write $F \prec G$ (or $G \succ F$). Also write $F \preceq G$ (or $G \succeq F$), if $F \prec G$ or $F \equiv G$.

Infima and Suprema

Definition 2.1 (see Bailey (2008, p. 172))

The **infimum** $F \wedge G$ of factors F and G on the same set is the coarsest factor which is finer than or equivalent to F and G, that is the factor which satisfies

(i) $F \wedge G \preceq F$ and $F \wedge G \preceq G$

(ii) if $H \preceq F$ and $H \preceq G$, then $H \preceq F \wedge G$

Definition 2.1 (see Bailey (2008, p. 172))

The **infimum** $F \wedge G$ of factors F and G on the same set is the coarsest factor which is finer than or equivalent to F and G, that is the factor which satisfies

(i) $F \wedge G \preceq F$ and $F \wedge G \preceq G$

(ii) if $H \preceq F$ and $H \preceq G$, then $H \preceq F \wedge G$

Definition 2.2 (see Bailey (2008, p. 172))

The **supremum** $F \vee G$ of factors F and G on the same set is the finest factor which is coarser than or equivalent to F and G, that is the factor which satisfies

(i) $F \preceq F \vee G$ and $G \preceq F \vee G$

(ii) if $F \preceq H$ and $G \preceq H$ then $F \vee G \preceq H$.

There are **two special factors** U and E on Ω (and similarly on $\mathcal{T}$), which are defined by

$U(\omega) = 1$ for every ω in Ω and

$E(\omega) = \omega$ for every ω in Ω.

For a set of inequivalent factors $\mathcal{F}$ on the same set the $\preceq$ relation on $\mathcal{F}$ is a **partial order**, which can be represented by a **Hasse diagram**:

► Draw a dot for each factor $F \in \mathcal{F}$.

► If $F \prec G$, then draw the dot for G above the dot for F and join F to G with a line (which may go through other dots).

► If $U \in \mathcal{F}$, then the dot for U is at the top. If $E \in \mathcal{F}$, the dot for E is at the bottom.

► If neither of F and G is finer than the other, and if $F \vee G \in \mathcal{F}$, respectively if $F \wedge G \in \mathcal{F}$, include dots as follows:
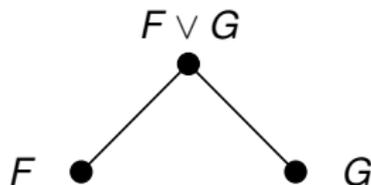
► Beside the dot for factor F, write its number of levels n_F

Orthogonal factors

For every factor F with n_F levels define

$V_F = \{$vectors which are constant on each F-class$\}$.

Definition 2.3 (see Bailey(2008, p. 179))

Factors F and G are **orthogonal** to each other if the subspace $V_F \cap V_{F \vee G}^{\perp}$ is orthogonal to the subspace $V_G \cap V_{F \vee G}^{\perp}$.

Theorem 2.1 (see Bailey(2008, p. 180))

Factors F and G are orthogonal to each other if and only if, within each class of $F \vee G$ separately

  (i) every F-class meets every G-class and
 (ii) all these intersections have size proportional to the product of the sizes of the relevant F-class and G-class.

Orthogonal decomposition

Theorem 2.2 (see Bailey(2008, p. 182))

Let $\mathcal{F}$ be a set of inequivalent factors on the same set. Suppose that

(a) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then $F \vee G \in \mathcal{F}$;

(b) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then F is orthogonal to G.

Theorem 2.2 (see Bailey(2008, p. 182))

Let $\mathcal{F}$ be a set of inequivalent factors on the same set. Suppose that

(a) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then $F \vee G \in \mathcal{F}$;

(b) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then F is orthogonal to G.

For every F in $\mathcal{F}$ define W_F by

$$W_F = V_F \cap \left(\sum\nolimits_{G \succ F} V_G \right)^{\perp}$$

and put $d_F = \dim(W_F)$.

Orthogonal decomposition

Theorem 2.2 (see Bailey(2008, p. 182))

Let $\mathcal{F}$ be a set of inequivalent factors on the same set. Suppose that

(a) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then $F \vee G \in \mathcal{F}$;

(b) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then F is orthogonal to G.

For every F in $\mathcal{F}$ define W_F by

$$W_F = V_F \cap \left(\sum_{G \succ F} V_G \right)^{\perp}$$

and put $d_F = \dim(W_F)$.

Then:

(i) If $F, G \in \mathcal{F}$ are different, then W_F is orthogonal to W_G.

(ii) If $F \in \mathcal{F}$ then $V_F = \bigoplus_{G \succeq F} W_G$; in particular, if $E \in \mathcal{F}$ then the whole space is equal to $\bigoplus_{F \in \mathcal{F}} W_F$.

Theorem 2.2 (see Bailey(2008, p. 182))

Let $\mathcal{F}$ be a set of inequivalent factors on the same set. Suppose that

(a) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then $F \vee G \in \mathcal{F}$;

(b) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then F is orthogonal to G.

For every F in $\mathcal{F}$ define W_F by

$$W_F = V_F \cap \left(\sum_{G \succ F} V_G\right)^{\perp}$$

and put $d_F = \dim(W_F)$.

Then:

(i) If $F, G \in \mathcal{F}$ are different, then W_F is orthogonal to W_G.

(ii) If $F \in \mathcal{F}$ then $V_F = \bigoplus_{G \succeq F} W_G$; in particular, if $E \in \mathcal{F}$ then the whole space is equal to $\bigoplus_{F \in \mathcal{F}} W_F$.

(iii) If $F \in \mathcal{F}$ then $d_F = n_F - \sum_{G \succ F} d_G$.

If **y** is the response vector and F a factor with n_F levels, then

- the **crude sum of squares** for F is defined as the squared length of the orthogonal projection of **y** onto V_F, that is $\text{CSS}(F) = \|P_{V_F}\mathbf{y}\|^2$,

- the **sum of squares** for F is defined as $\text{SS}(F) = \|P_{W_F}\mathbf{y}\|^2$.

If $\mathcal{F}$ is a set of inequivalent factors satisfying (a) and (b) in Thm. 2.2, then for every $F \in \mathcal{F}$

$$\text{SS}(F) = \text{CSS}(F) - \sum_{G \succ F} \text{SS}(G).$$

Thm. 2.2 implies degrees of freedom and sums of squares can be calculated by working down the Hasse diagram! Suppose that $U \in \mathcal{F}$.

Thm. 2.2 implies degrees of freedom and sums of squares can be calculated by working down the Hasse diagram! Suppose that $U \in \mathcal{F}$.

Calculation of degrees of freedom

Start at the top of the Hasse diagram. There is no factor coarser than U, so $d_U = n_U = 1$.

Work down the diagram. At each dot F calculate d_F by subtracting from n_F the degrees of freedom d_G for all dots G **above** F. Write d_F immediately to the right or immediately below n_F.

Thm. 2.2 implies degrees of freedom and sums of squares can be calculated by working down the Hasse diagram! Suppose that $U \in \mathcal{F}$.

Calculation of degrees of freedom

Start at the top of the Hasse diagram. There is no factor coarser than U, so $d_U = n_U = 1$.

Work down the diagram. At each dot F calculate d_F by subtracting from n_F the degrees of freedom d_G for all dots G **above** F. Write d_F immediately to the right or immediately below n_F.

Calculation of sums of squares

Start at the top of the Hasse diagram. There is no factor coarser than U, so $SS(U) = CSS(U) = SS(\text{mean})$.

At each dot F calculate $CSS(F)$. Find $SS(F)$ by subtracting from $CSS(F)$ the sums of squares $SS(G)$ for all dots G **above** F.

Definition 2.4

A factor is **uniform** if all its classes have the same size.

Definition 2.5 (see Bailey (2008, p. 194))

An **orthogonal plot structure** is a set $\mathcal{F}$ of inequivalent factors on Ω such that

(i) every factor in $\mathcal{F}$ is uniform

(ii) $U \in \mathcal{F}$

(iii) $E \in \mathcal{F}$

(iv) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then $F \vee G \in \mathcal{F}$

(v) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then $F \wedge G \in \mathcal{F}$

(vi) if $F \in \mathcal{F}$ and $G \in \mathcal{F}$ then F is orthogonal to G.

Orthogonal treatment structures

Queen Mary
University of London

Definition 2.6 (see Bailey (2008, p. 190))

An **orthogonal treatment structure** is a set $\mathcal{G}$ of inequivalent factors on $\mathcal{T}$ such that

(i) $U \in \mathcal{G}$

(ii) if $G_1 \in \mathcal{G}$ and $G_2 \in \mathcal{G}$ then $G_1 \vee G_2 \in \mathcal{G}$

(iii) if $G_1 \in \mathcal{G}$ and $G_2 \in \mathcal{G}$ then G_1 is orthogonal to G_2

Orthogonal designs

Definition 2.7 (see Bailey (2008, p. 198))

Let $\mathcal{F}$ be a set of factors on Ω and $\mathcal{G}$ be a set of factors on $\mathcal{T}$. A design with plot structure $\mathcal{F}$ and treatment structure $\mathcal{G}$ that assigns treatment $T(\omega)$ to plot ω is an **orthogonal design** if

(a) $\mathcal{F}$ is an orthogonal plot structure

(b) $\mathcal{G}$ is an orthogonal treatment structure

(c) the function T is such that

    (i) if $G_1 \in \mathcal{G}$ and $G_2 \in \mathcal{G}$ then $G_1 \circ T$ and $G_2 \circ T$ are orthogonal on Ω.

    (ii) if $F \in \mathcal{F}$ and $G \in \mathcal{G}$ then F is orthogonal to $G \circ T$.

    (iii) if $F \in \mathcal{F}$ and $G \in \mathcal{G}$ then $F \vee (G \circ T)$ is equal to $H \circ T$ for some $H \in \mathcal{G}$.

Assume plot factors have **random** effects and only affect Cov(**Y**).

Model assumptions

Assume plot factors have **random** effects and only affect Cov(**Y**).

Theorem 2.3 (see Bailey (2008, p. 196))

Let $\mathcal{F}$ be an orthogonal plot structure. Suppose that

$$\text{Cov}(\mathbf{Y}) = \sigma^2 \sum_{F \in \mathcal{F}} \rho_F \mathbf{A}_F,$$

for unknown σ^2 and unknown correlations ρ_F for F in $\mathcal{F}$ (with $\rho_E = 1$). Then the W-subspaces defined in Thm. 2.2 are the eigenspaces of Cov(**Y**); that is, they are the strata.

Assume plot factors have **random** effects and only affect Cov(**Y**).

Theorem 2.3 (see Bailey (2008, p. 196))

Let $\mathcal{F}$ be an orthogonal plot structure. Suppose that

$$\text{Cov}(\mathbf{Y}) = \sigma^2 \sum_{F \in \mathcal{F}} \rho_F \mathbf{A}_F,$$

for unknown σ^2 and unknown correlations ρ_F for F in $\mathcal{F}$ (with $\rho_E = 1$). Then the W-subspaces defined in Thm. 2.2 are the eigenspaces of Cov(**Y**); that is, they are the strata.

Here $\mathbf{A}_F$ is an $N \times N$ matrix. Element $\mathbf{A}_F(\alpha, \beta) = 1$, if F is the **finest** plot factor with $F(\alpha) = F(\beta)$, and 0 otherwise.

Denote eigenvalue of Cov(**Y**) for $F \in \mathcal{F}$ by ξ_F.

Queen Mary
University of London

The **null ANOVA** has one line for each stratum, showing the name F and the degrees of freedom d_F for every plot factor.

This is expanded to **calculations ignoring treatments** by writing in the sums of squares SS(F).

The **null ANOVA** has one line for each stratum, showing the name F and the degrees of freedom d_F for every plot factor.

This is expanded to **calculations ignoring treatments** by writing in the sums of squares SS(F).

The **skeleton ANOVA**, shows which stratum each treatment subspace is put in. To this end use

Theorem 2.4 (see Bailey (2008, p. 198))

Let G be a treatment factor in an orthogonal design. Then there is a unique plot factor F finer than or equivalent to G such that $H \preceq F$ for every other plot factor H with $H \preceq G$. Moreover, $W_G \subseteq W_F$.

Write $E(\mathbf{Y}) = \boldsymbol{\tau}$ and $\boldsymbol{\tau}_G$ for $P_{W_G}\boldsymbol{\tau}$.

Write $E(\mathbf{Y}) = \boldsymbol{\tau}$ and $\boldsymbol{\tau}_G$ for $P_{W_G}\boldsymbol{\tau}$.

In the **full ANOVA**, the line for stratum F is split up to include

- each treatment factor G with $W_G \subseteq W_F$ and
- a residual calculated by subtraction.

Write $E(\mathbf{Y}) = \boldsymbol{\tau}$ and $\boldsymbol{\tau}_G$ for $P_{W_G}\boldsymbol{\tau}$.

In the **full ANOVA**, the line for stratum F is split up to include

- each treatment factor G with $W_G \subseteq W_F$ and
- a residual calculated by subtraction.

Stratum	source	df	SS	EMS	VR
F	G_1	d_{G_1}	$SS(G_1)$	$\frac{\|\boldsymbol{\tau}_{G_1}\|^2}{d_{G_1}} + \xi_F$	$\frac{MS(G_1)}{MS(\text{residual in } F)}$
	G_2	d_{G_2}	$SS(G_2)$	$\frac{\|\boldsymbol{\tau}_{G_2}\|^2}{d_{G_2}} + \xi_F$	$\frac{MS(G_2)}{MS(\text{residual in } F)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	residual	$\leftarrow$ by subtraction $\rightarrow$		ξ_F	–
	total	d_F	$SS(F)$		

Write $E(\mathbf{Y}) = \boldsymbol{\tau}$ and $\boldsymbol{\tau}_G$ for $P_{W_G}\boldsymbol{\tau}$.

In the **full ANOVA**, the line for stratum F is split up to include

- each treatment factor G with $W_G \subseteq W_F$ and
- a residual calculated by subtraction.

Stratum	source	df	SS	EMS	VR
F	G_1	d_{G_1}	$\mathrm{SS}(G_1)$	$\dfrac{\|\boldsymbol{\tau}_{G_1}\|^2}{d_{G_1}} + \xi_F$	$\dfrac{\mathrm{MS}(G_1)}{\mathrm{MS}(\text{residual in } F)}$
	G_2	d_{G_2}	$\mathrm{SS}(G_2)$	$\dfrac{\|\boldsymbol{\tau}_{G_2}\|^2}{d_{G_2}} + \xi_F$	$\dfrac{\mathrm{MS}(G_2)}{\mathrm{MS}(\text{residual in } F)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	residual	$\leftarrow$ by subtraction $\rightarrow$		ξ_F	–
	total	d_F	$\mathrm{SS}(F)$		

Note: Sum of squares $\mathrm{SS}(G)$ and degrees of freedom d_G for each tmt factor G are calculated using the Hasse diagram for the tmt factors.

3. `AutomaticANOVA` **PACKAGE**

For an Excel file containing a design/dataset in the form

PlotFac1	TmtFac1	TmtFac2	PlotFac2	PlotFac3	TmtFac3	Response
1	2	3	1	3	1	45.6
2	1	3	2	2	1	35.7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	2	2	3	1	49.3

automatically **deduce** a model and carry out the **appropriate** analysis of variance.

- ► Implement theory in Bailey (2008, Chap. 10) **literally**.
    - ► To this end, identify each factor F with its set of F-classes, that is work with **sets of sets**.
    - ► Heavily exploit finer relation.
- ► Use **Mathematica** since it
    - ► allows very flexible data structures,
    - ► supports set operations etc.,
    - ► provides a function for drawing the Hasse diagram from a partial order,
    - ► can do symbolic and numerical computations,
    - ► has many other desirable features.

Suppose plot factors $\mathcal{F}$ and $\mathcal{G}$ have been specified. Consider the factors in $\mathcal{G}$ as factors on Ω.

- Turn $\mathcal{F}$ into an orthogonal plot structure (Def 2.5), if possible:
  - Form its **closure** $\tilde{\mathcal{F}}$ under infima and suprema and add U and E.
  - Check **uniformity** for $\tilde{\mathcal{F}}$ (Def. 2.4).
  - Check **orthogonality** for $\tilde{\mathcal{F}}$ (Thm. 2.1).
- Turn $\mathcal{G}$ into an orthogonal treatment structure (Def 2.6), if possible:
  - If "Minimal", form its **closure** $\tilde{\mathcal{G}}$ under suprema and add U; if "Maximal" form closure under suprema and infima and add U (E on set of tmts will then be automatically included).
  - Check **orthogonality** for $\tilde{\mathcal{G}}$ (Thm. 2.1).
- Check **orthogonality** of $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{G}}$ (Thm. 2.1).
- Make sure the supremum of any factor in $\tilde{\mathcal{F}}$ and any factor in $\tilde{\mathcal{G}}$ is an element of $\tilde{\mathcal{G}}$. Add any necessary suprema to $\tilde{\mathcal{G}}$ (see Def. 2.7(c)(iii) and Bailey (2008, p. 198)).

If $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{G}}$ pass the above tests then they are the plot and treatment structures of an **orthogonal design** (Def. 2.7).

Analysis of variance:

- Use **finer relation** for $\tilde{\mathcal{F}}$ to get Hasse diagram and decomposition (Thm. 2.2) of SS and DF for plot factors.
- Use **finer relation** for $\tilde{\mathcal{G}}$ to get Hasse diagram and decomposition (Thm. 2.2) of SS and DF for tmt factors.
- Use **finer relation** for $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{G}}$ to allocate tmt factors to strata (Thm. 2.4).
- Use **finer relation** for $\tilde{\mathcal{F}}$ to estimate variance components (Thm. 2.3).

A single command can do everything!

Queen Mary
University of London

A single command can do everything!

AutomaticAnova[DataSet, PlotFactors, TreatmentFactors, Response,Type]

A single command can do everything!

AutomaticAnova[DataSet, PlotFactors, TreatmentFactors, Response,Type]

where **DataSet** := Excel file
       PlotFactors := list of plot factors, e.g. $\{1, 4, 5\}$
       TreatmentFactors := list of treatment factors, e.g. $\{2, 3, 6\}$
       Response := response variable in curly brackets, e.g. $\{7\}$
       Type := either "Minimal" or "Maximal"

4. EXAMPLES

Examples to be considered:

1. Cider apples (Bailey (2008, p. 145))
2. Bean weevils (Bailey (2008, p. 204))
3. Unwrinkled washing (Bailey (2008, pp. 208–209))
4. Molybdenum (Bailey (2008, p. 211))
5. Irrigated rice (Bailey (2008, p. 211))
6. Poset block structure (Bailey (2004, p. 257))

🔱 Queen Mary
University of London

Example 1 (Cider apples; Bailey (2008, p. 145))

- ▶ An orchard of cider-apple trees was divided into **6 blocks** of **5 trees** each.
- ▶ Each of **3 types of spray** was applied to **2 blocks**.
- ▶ Within each block the same **5 pruning methods** were used on different trees.
- ▶ The response was the number of apples fallen down, given as a percentage of the total for each tree.

Example 2 (Bean weevils; Bailey (2008, p. 204))

- ▶ Investigation of effect of pheromone and neem oil on beetles.
- ▶ A field is divided into **6 rows** and **6 columns**.
- ▶ **5 treatments** are allocated to the **36 plots** as follows

1	1	2	3	4	5
5	1	1	2	3	4
4	5	1	1	2	3
3	4	5	1	1	2
2	3	4	5	1	1
1	2	3	4	5	1

- ▶ Treatments are defined by **3 factors** as follows:

Treatment	1: Untreated	2: Pheromone 5μg per day	3: Pheromone 240μg per day	4: Neem oil in 10 litres	5: Neem oil in 25 litres
Type	1	2	2	3	3
Pheromone	1	2	3	4	4
Neem oil	1	2	2	3	4

Example 3 (Unwrinkled washing; Bailey (2008, pp. 208–209))

- ▶ Find best combination of wash and drying temperature.
- ▶ **4 wash temperatures** (WT) and **3 drying temperatures** (DT)
- ▶ **8 washing machines** (WM) and **6 dryers** (D)
- ▶ Procedure
    1. **48 sheets** randomized to WM (6 per machine)
    2. **4 WT** randomized to WM (2 machines run at each WT)
    3. After wash, **6 sheets** in each machine randomized to **6 D** (1 sheet per D)
    4. **3 DT** randomized to **6 D** (2 dryers run at each DT)

Further examples

Example 4 (Molybdenum; Bailey (2008, p. 211))

- ► Factorial experiment on beans.
- ► **3 treatment factors**: Cultivar (**3 levels**), Seeding (**2 levels**), Mo (**2 levels**).
- ► **4 blocks** each of which was a **3 × 4 rectangle**.
- ► Rows of rectangles were called **Strips**, columns of rectangles were called **Quarters** and grouped into **2 halves** per block.
- ► In a typical block, treatments were applied as follows:

	sown conventionally		sown with no tillage	
cultivar 2 →				
cultivar 1 →				
cultivar 3 →				
	↑ Mo applied	↑ No Mo	↑ No Mo	↑ Mo applied
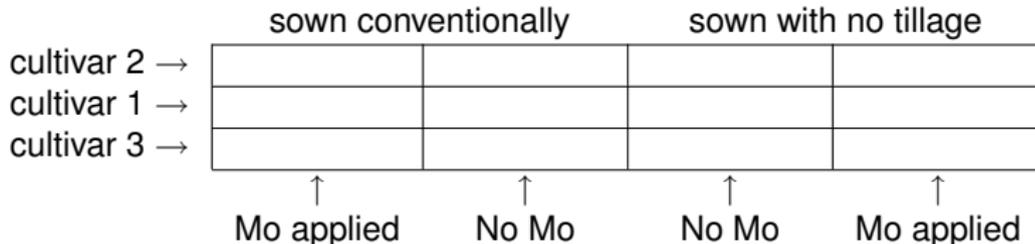
Example 5 (Irrigated rice; Bailey (2008, p. 211))

- A rice field is divided into **8 rows** and **4 columns**.
- Pairs of consecutive rows are grouped into **4 blocks**.
- Intersections of rows and blocks give **16 cells**.
- **8 treatments** $A, B, \ldots, H$ are applied to the cells as follows:

A	B	C	D
E	F	G	H
D	A	B	C
H	E	F	G
C	D	A	B
G	H	E	F
B	C	D	A
F	G	H	E

Example 6 (Poset block structure; Bailey (2004, p. 257))

- ▶ Over **2 weeks** samples of milk are tested for bacteria. Testing involves **2 labs**, each of which employs **2 technicians**.
- ▶ Procedure:
    - ▶ Each week **2 samples** of milk are sent to each lab.
    - ▶ At the lab each sample is split into **2 subsamples**, one of which is tested by each technician.
- ▶ Subsamples can be labelled by elements (x_1, x_2, x_3, x_4) of $\Omega = \{1, 2\}^4$, with components corresponding to **weeks**, **labs**, **technicians** and **samples**.
- ▶ For $L \subseteq \{1, 2, 3, 4\}$ let $F(L)$ be the factor on Ω whose classes are the quadruples which have the same x_i values for $i \in L$.
- ▶ The closure of $\mathcal{F} = \{F(\{1\}), F(\{2\}), F(\{2, 3\}), F(\{1, 2, 4\})\}$ under infima and suprema is an orthogonal plot structure, which **cannot** be obtained by iterated crossing and nesting.

- ► Provision of web application for online access to `AutomaticANOVA` package

- ► Automatic generation of syntax for **GenStat** block and treatment structures

- ► Confidence intervals for variance components

- ► Imputation methods for dealing with missing values

Discussion

- ▶ Theory in Bailey (2008, Chap 10) provides a **unifying** perspective on ANOVA which covers a large number of designs/models.
- ▶ Applying the theory "by hand" is often **laborious**.
- ▶ **User-friendly** `AutomaticANOVA` package may help in popularizing the approach among statisticians and practitioners.
- ▶ Although presentation here has been in terms of plot and treatment factors, the package can also be used for **mixed models** in a model-based framework under standard assumptions about the random factors (zero mean, uncorrelated, normal). To this end, all random factors need to be specified as plot factors and all fixed factors as treatment factors.
- ▶ A **restrictive** assumption made in this talk is the uniformity of the plot factors. This is more crucial than it may seem at first sight and precludes generalization of the theory to plot structures where Thm. 2.2 can still be applied, but which are not uniform.

Bailey, R. A. (2004). Association schemes. Designed experiments, algebra and combinatorics. Cambridge University Press, Cambridge, UK.

Bailey, R. A. (2008). Design of comparative experiments. Cambridge University Press, Cambridge, UK.

Nelder, J. A. (1965a). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proceedings of the Royal Society of London, Series A*, 283, 147–162.

Nelder, J. A. (1965b). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. *Proceedings of the Royal Society of London, Series A*, 283, 163–178.

If you have any examples you would like to share, please get in touch.

Thank you