Direct ANOVA as an Alternative to REML in the Analysis of Nested Block Designs

Agnieszka Łacka

Department of Mathematical and Statistical Methods
Poznań University of Life Sciences

XIII Working Seminar on Statistical Methods in Variety Testing
08 – 11.09.2025

Introduction

- **Blocking in experimental designs** – groups units, controls variability, may cause confounding
- **Multi-stratum / nested designs** – analyzed with mixed models (REML for variances, GLS for treatments)
- **Estimation issue** – block variance often estimated as zero even if true value $\neq 0$
- **Alternative approaches** – Bayesian with priors; robust analysis with minimal assumptions
- **Orthogonal block structure (OBS)** – clear partition of variation, independence of treatment vs. block structure

Introduction

The aim of this presentation is to compare the results obtained using REML and the *direct* ANOVA method in the analysis of sugar beet yield data from a nested block design.

Introduction

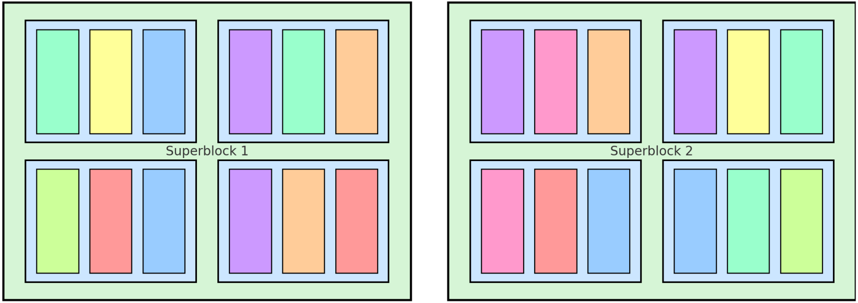
Definition 1. (from Section 2.2 in Houtman and Speed, 1983).

An experiment is said to have the **orthogonal block structure** (OBS) if the covariance (dispersion) matrix of the random variables observed on the experimental units (plots), $\boldsymbol{y} = [y_1, y_2, \dots, y_n]'$, has a representation of the form

$$\mathrm{D}(\boldsymbol{y}) = \sigma_1^2 \phi_1 + \sigma_2^2 \phi_2 + \cdots + \sigma_t^2 \phi_t,$$

where the $\{\phi_\alpha\}$, $\alpha = 1, 2, \dots, t$, are known symmetric, idempotent and pairwise orthogonal matrices, summing to the identity matrix, the last usually being of the form $\phi_t = n^{-1} \mathbf{1}_n \mathbf{1}_n'$.

Introduction – Nested Block Design



- Treatments:
  - v – number of treatments (varieties)
- Blocks:
  - $b = ab_0$ – total number of blocks
  - k – number of experimental units (plots) per block
- Superblocks:
  - a – number of superblocks
  - b_0 – number of blocks within each superblock

A randomization-derived model (NB design)

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\tau} + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{e}, \quad \text{with} \quad \mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}_1\boldsymbol{\tau}, \quad (1)$$

$\boldsymbol{y} = [\boldsymbol{y}'_1, \boldsymbol{y}'_2, \ldots, \boldsymbol{y}'_a]'$ – $n \times 1$ vector of observations, $n = ab_0k$

$\boldsymbol{y}_h = [y_{1h}, y_{2h}, \ldots, y_{n_0h}]'$, $n_0 = kb_0$, $h = 1, \ldots, a$ – responses in the h-th superblock

$\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_v]'$ – fixed treatment effects

$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_a]'$ – random superblock effects

$\boldsymbol{\beta} = [\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \ldots, \boldsymbol{\beta}'_a]'$, $\boldsymbol{\beta}_h = [\beta_{1(h)}, \ldots, \beta_{b_0(h)}]'$ – random block effects nested in superblocks

$\boldsymbol{\eta}$, $\boldsymbol{e}$ – unit and technical random errors $(n \times 1)$

$$\boldsymbol{X}_1 = [\boldsymbol{X}'_{11} : \boldsymbol{X}'_{12} : \cdots : \boldsymbol{X}'_{1a}]', \quad \boldsymbol{X}_A = \boldsymbol{I}_a \otimes \mathbf{1}_{n_0}, \quad \boldsymbol{X}_B = \mathrm{diag}[\boldsymbol{X}_{B1} : \boldsymbol{X}_{B2} : \cdots : \boldsymbol{X}_{Ba}]$$

$\boldsymbol{X}'_1\mathbf{1}_n = \boldsymbol{r} = [r_1, \ldots, r_v]'$ – vector of treatment replications,

$\boldsymbol{X}'_1\boldsymbol{X}_1 = \boldsymbol{r}^{\delta} = \mathrm{diag}[r_1, \ldots, r_v]$, $\quad \boldsymbol{r}^{-\delta}$ – inverse diagonal

A randomization-derived model (NB design)

- Random components denoted explicitly:

$\{\alpha_h\}$ – superblock effects, $h = 1, \ldots, a$
$\{\beta_{j(h)}\}$ – block effects nested in superblocks, $j = 1, \ldots, b_0$
$\{\eta_{\ell[j(h)]}\}$ – unit-level errors, $\ell = 1, \ldots, k$
$\{e_{\ell[j(h)]}\}$ – technical errors

- Zero expectations: $\mathrm{E}(\alpha_h) = \mathrm{E}(\beta_{j(h)}) = \mathrm{E}(\eta_{\ell[j(h)]}) = 0$

- Orthogonality among random terms:

$\mathrm{Cov}(\alpha_h, \beta_{j(h')}) = \mathrm{Cov}(\alpha_h, \eta_{\ell[j(h')]}) = 0$
$\mathrm{Cov}(\beta_{j(h)}, \eta_{\ell[j'(h')]}) = 0$ for all $h, h', j, j', \ell, \ell'$

- Let where N_A, B_H and K_H denote constants associated with the potential (pre-randomization) numbers of superblocks, blocks within superblocks, and units within blocks.

Covariance structure of random effects

- Superblock effects:

$$\mathrm{Cov}(\alpha_h, \alpha_{h'}) = \begin{cases} N_A^{-1}(N_A - 1)\sigma_A^2, & h = h' \\ -N_A^{-1}\sigma_A^2, & h \neq h' \end{cases}$$

- Block effects within superblocks:

$$\mathrm{Cov}(\beta_{j(h)}, \beta_{j'(h')}) = \begin{cases} B_H^{-1}(B_H - 1)\sigma_B^2, & h = h', j = j' \\ -B_H^{-1}\sigma_B^2, & h = h', j \neq j' \\ 0, & h \neq h' \end{cases}$$

- Unit-level errors:

$$\mathrm{Cov}(\eta_{\ell[j(h)]}, \eta_{\ell'[j'(h')]}) = \begin{cases} K_H^{-1}(K_H - 1)\sigma_U^2, & h = h', j = j', \ell = \ell' \\ -K_H^{-1}\sigma_U^2, & h = h', j = j', \ell \neq \ell' \\ 0, & j(h) \neq j'(h') \end{cases}$$

- Technical errors: $\{e_{\ell[j(h)]}\}$ are uncorrelated, zero mean, variance σ_e^2, independent of all other terms

A randomization-derived model (NB design)

Units (plots) → Blocks → Superblocks → Total exp. area

Thus, the observed vector $\boldsymbol{y}$ can be decomposed as

$$\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{y}_3 + \boldsymbol{y}_4, \text{ where } \boldsymbol{y}_1 = \boldsymbol{\phi}_1\boldsymbol{y},\ \boldsymbol{y}_2 = \boldsymbol{\phi}_2\boldsymbol{y},\ \boldsymbol{y}_3 = \boldsymbol{\phi}_3\boldsymbol{y},\ \boldsymbol{y}_4 = \boldsymbol{\phi}_4\boldsymbol{y}.$$

This allows the expectation vector and the covariance (dispersion) matrix of $\boldsymbol{y}$ to be written as

$$\begin{aligned} \mathsf{E}(\boldsymbol{y}) &= \boldsymbol{\phi}_1\boldsymbol{X}_1\boldsymbol{\tau} + \boldsymbol{\phi}_2\boldsymbol{X}_1\boldsymbol{\tau} + \boldsymbol{\phi}_3\boldsymbol{X}_1\boldsymbol{\tau} + \boldsymbol{\phi}_4\boldsymbol{X}_1\boldsymbol{\tau} = \boldsymbol{X}_1\boldsymbol{\tau}, \\ \mathsf{D}(\boldsymbol{y}) \equiv \boldsymbol{V} &= \sigma_1^2\boldsymbol{\phi}_1 + \sigma_2^2\boldsymbol{\phi}_2 + \sigma_3^2\boldsymbol{\phi}_3 + \sigma_4^2\boldsymbol{\phi}_4 \end{aligned}$$

where the matrices

$$\boldsymbol{\phi}_1 = \boldsymbol{I}_n - k^{-1}\boldsymbol{X}_B\boldsymbol{X}_B',$$
$$\boldsymbol{\phi}_2 = k^{-1}\boldsymbol{X}_B\boldsymbol{X}_B' - n_0^{-1}\boldsymbol{X}_A\boldsymbol{X}_A',\ \boldsymbol{\phi}_3 = n_0^{-1}\boldsymbol{X}_A\boldsymbol{X}_A' - n^{-1}\mathbf{1}_n\mathbf{1}_n',$$
$$\text{and } \boldsymbol{\phi}_4 = n^{-1}\mathbf{1}_n\mathbf{1}_n'$$

are symmetric, idempotent and pairwise orthogonal, summing to the identity matrix, and the scalars $\sigma_1^2, \sigma_2^2, \sigma_3^2$ and σ_4^2 represent the relevant unknown stratum variances.

A randomization-derived model (NB design)

The associated variance components are defined as:

$$\sigma_1^2 = \sigma_U^2 + \sigma_e^2$$

$$\sigma_2^2 = k\sigma_B^2 + (1 - K_H^{-1}k)\sigma_U^2 + \sigma_e^2$$

$$\sigma_3^2 = n_0\sigma_A^2 + (k - B_H^{-1}n_0)\sigma_B^2 + (1 - K_H^{-1}k)\sigma_U^2 + \sigma_e^2$$

$$\sigma_4^2 = (n_0 - N_A^{-1}n)\sigma_A^2 + (k - B_H^{-1}n_0)\sigma_B^2 + (1 - K_H^{-1}k)\sigma_U^2 + \sigma_e^2$$

The stratum analysis

$$\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{y}_3 + \boldsymbol{y}_4$$

$$\boldsymbol{y}_1 = \phi_1 \boldsymbol{y}, \quad \boldsymbol{y}_2 = \phi_2 \boldsymbol{y}, \quad \boldsymbol{y}_3 = \phi_3 \boldsymbol{y}, \quad \boldsymbol{y}_4 = \phi_4 \boldsymbol{y}$$

$$\begin{aligned} \phi_1 &= \boldsymbol{I}_n - k^{-1}\boldsymbol{X}_B\boldsymbol{X}_B' & \phi_2 = k^{-1}\boldsymbol{X}_B\boldsymbol{X}_B' - n_0^{-1}\boldsymbol{X}_A\boldsymbol{X}_A' \\ \phi_3 &= n_0^{-1}\boldsymbol{X}_A\boldsymbol{X}_A' - n^{-1}\mathbf{1}_n\mathbf{1}_n' & \phi_4 = n^{-1}\mathbf{1}_n\mathbf{1}_n' \end{aligned}$$

$$\begin{aligned} \boldsymbol{C}_1 &= \boldsymbol{X}_1'\phi_1\boldsymbol{X}_1 = \boldsymbol{r}^{\delta} - k^{-1}\boldsymbol{N}\boldsymbol{N}' & \text{of rank } h_1 \\ \boldsymbol{C}_2 &= \boldsymbol{X}_1'\phi_2\boldsymbol{X}_1 = k^{-1}\boldsymbol{N}\boldsymbol{N}' - n_0^{-1}\boldsymbol{M}\boldsymbol{M}' & \text{of rank } h_2 \\ \boldsymbol{C}_3 &= \boldsymbol{X}_1'\phi_3\boldsymbol{X}_1 = n_0^{-1}\boldsymbol{M}\boldsymbol{M}' - n^{-1}\boldsymbol{r}\boldsymbol{r}' & \text{of rank } h_3 \\ \boldsymbol{C}_4 &= \boldsymbol{X}_1'\phi_4\boldsymbol{X}_1 = n^{-1}\boldsymbol{r}\boldsymbol{r}' & \text{of rank } 1 \end{aligned}$$

$$\boldsymbol{M} = \boldsymbol{X}_1'\boldsymbol{X}_A = [\boldsymbol{r}_1 : \boldsymbol{r}_2 : \cdots : \boldsymbol{r}_a] \qquad \boldsymbol{N} = \boldsymbol{X}_1'\boldsymbol{X}_B = [\boldsymbol{N}_1 : \boldsymbol{N}_2 : \cdots : \boldsymbol{N}_a]$$

The stratum analysis

$$\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{y}_3 + \boldsymbol{y}_4$$

$$\boldsymbol{y}_1 = \boldsymbol{\phi}_1\boldsymbol{y}, \quad \boldsymbol{y}_2 = \boldsymbol{\phi}_2\boldsymbol{y}, \quad \boldsymbol{y}_3 = \boldsymbol{\phi}_3\boldsymbol{y}, \quad \boldsymbol{y}_4 = \boldsymbol{\phi}_4\boldsymbol{y}$$

$$\begin{aligned} \phi_1 &= \boldsymbol{I}_n - k^{-1}\boldsymbol{X}_B\boldsymbol{X}_B' \qquad & \phi_2 = k^{-1}\boldsymbol{X}_B\boldsymbol{X}_B' - n_0^{-1}\boldsymbol{X}_A\boldsymbol{X}_A' \\ \phi_3 &= n_0^{-1}\boldsymbol{X}_A\boldsymbol{X}_A' - n^{-1}\mathbf{1}_n\mathbf{1}_n' \qquad & \phi_4 = n^{-1}\mathbf{1}_n\mathbf{1}_n' \end{aligned}$$

$$\begin{aligned} S_1^2 &= \boldsymbol{y}'\psi_1\boldsymbol{y}/(n-b-h_1) \qquad & \boldsymbol{Q}_\alpha = \boldsymbol{X}_1'\phi_\alpha\boldsymbol{y} \\ S_2^2 &= \boldsymbol{y}'\psi_2\boldsymbol{y}/(b-a-h_2) \qquad & \psi_\alpha = \phi_\alpha - \phi_\alpha\boldsymbol{X}_1\boldsymbol{C}_\alpha^-\boldsymbol{X}_1'\phi_\alpha \\ S_3^2 &= \boldsymbol{y}'\psi_3\boldsymbol{y}/(a-1-h_3) \qquad & \alpha = 1,2,3 \end{aligned}$$

The stratum analysis – ANOVA in strata

Source	d.f.	Sum of Squares	Mean Square	F
Intra-block				
Treatments	h_1	$SS_V = \boldsymbol{Q}_1' \boldsymbol{C}_1^- \boldsymbol{Q}_1$	$MS_V = \dfrac{\boldsymbol{Q}_1' \boldsymbol{C}_1^- \boldsymbol{Q}_1}{h_1}$	$\dfrac{MS_V}{MS_R}$
Residuals	$n-b-h_1$	$SS_R = \boldsymbol{y}_1' \psi_1 \boldsymbol{y}_1$	$MS_R = \dfrac{\boldsymbol{y}_1' \psi_1 \boldsymbol{y}_1}{n-b-h_1}$	
Total	$n-b$	$SS_T = \boldsymbol{y}_1' \phi_1 \boldsymbol{y}_1$		
Inter-block-intra-superblock				
Treatments	h_2	$SS_V = \boldsymbol{Q}_2' \boldsymbol{C}_2^- \boldsymbol{Q}_2$	$MS_V = \dfrac{\boldsymbol{Q}_2' \boldsymbol{C}_2^- \boldsymbol{Q}_2}{h_2}$	$\dfrac{MS_V}{MS_R}$
Residuals	$b-a-h_2$	$SS_R = \boldsymbol{y}_2' \psi_2 \boldsymbol{y}_2$	$MS_R = \dfrac{\boldsymbol{y}_2' \psi_2 \boldsymbol{y}_2}{b-a-h_2}$	
Total	$b-a$	$SS_T = \boldsymbol{y}_2' \phi_2 \boldsymbol{y}_2$		
Inter-superblock				
Treatments	h_3	$SS_V = \boldsymbol{Q}_3' \boldsymbol{C}_3^- \boldsymbol{Q}_3$	$MS_V = \dfrac{\boldsymbol{Q}_3' \boldsymbol{C}_3^- \boldsymbol{Q}_3}{h_3}$	$\dfrac{MS_V}{MS_R}$
Residuals	$a-1-h_3$	$SS_R = \boldsymbol{y}_3' \psi_3 \boldsymbol{y}_3$	$MS_R = \dfrac{\boldsymbol{y}_3' \psi_3 \boldsymbol{y}_3}{a-1-h_3}$	
Total	$a-1$	$SS_T = \boldsymbol{y}_3' \phi_3 \boldsymbol{y}_3$		

Direct approach

$\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_v]'$,

$(\boldsymbol{I}_v - n^{-1}\mathbf{1}_v\boldsymbol{r}')\tau = [\tau_1 - \tau_., \tau_2 - \tau_., \ldots, \tau_v - \tau_.]'$, where $\tau_. = n^{-1}\sum_{i=1}^{v}(r_i\tau_i)$,

$(\boldsymbol{V}^{-1}$-orthogonal) projector

$$\boldsymbol{P}_{X_1(V^{-1})} = \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{V}^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{V}^{-1}.$$

We can decompose the analyzed data vector $\boldsymbol{y}$ into two uncorrelated parts, as

$$\boldsymbol{y} = \boldsymbol{P}_{X_1(V^{-1})}\boldsymbol{y} + (\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})\boldsymbol{y}.$$

Direct approach

$$\boldsymbol{y} = \boldsymbol{P}_{X_1(V^{-1})}\boldsymbol{y} + (\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})\boldsymbol{y}$$

Under the model (1), with presented properties, the first term of the partition provides the best linear unbiased estimator (BLUE) of $\boldsymbol{X}_1\boldsymbol{\tau}$, which can be expressed as

$$\widehat{\boldsymbol{X}_1\boldsymbol{\tau}} = \boldsymbol{P}_{X_1(V^{-1})}\boldsymbol{y}.$$

With regard to the second term, it can be seen as the residual vector, giving the residual sum of squares in the form

$$\begin{aligned} \|(\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})\boldsymbol{y}\|^2_{V^{-1}} &= \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})'\boldsymbol{V}^{-1}(\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})\boldsymbol{y} \\ &= \boldsymbol{y}'[\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{V}^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{V}^{-1}]\boldsymbol{y} \\ &= \boldsymbol{y}'\boldsymbol{V}^{-1}(\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})\boldsymbol{y}, \end{aligned}$$

with the residual degrees of freedom given by $\text{rank}(\boldsymbol{V} : \boldsymbol{X}_1) - \text{rank}(\boldsymbol{X}_1) = n - v$.

Direct approach

Also note that, as $\tau = \boldsymbol{r}^{-\delta}\boldsymbol{X}_1'\boldsymbol{X}_1\tau$, the BLUE of τ can be obtained as

$$\hat{\tau} = (\boldsymbol{X}_1'\boldsymbol{V}^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{V}^{-1}\boldsymbol{y}.$$

Its covariance (dispersion) matrix then takes the form

$$\begin{aligned} \mathsf{D}(\hat{\tau}) &= (\boldsymbol{X}_1'\boldsymbol{V}^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{V}^{-1}\mathsf{D}(\boldsymbol{y})\boldsymbol{V}^{-1}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{V}^{-1}\boldsymbol{X}_1)^{-1} \\ &= (\boldsymbol{X}_1'\boldsymbol{V}^{-1}\boldsymbol{X}_1)^{-1}. \end{aligned}$$

Direct approach

$$\tau_* = (\boldsymbol{I}_v - n^{-1}\mathbf{1}_v\boldsymbol{r}')\boldsymbol{\tau}$$

$$H_0 : \tau_* = \mathbf{0} \tag{2}$$

To verify H_0, the unknown stratum variances σ_i^2, for $i = 1,2,3,4$ must be estimated.
Since $\mathrm{E}\{\|\phi_i(\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})\boldsymbol{y}\|^2\} = \sigma_i^2 d_i'$, where $d_i' = \mathrm{tr}[\phi_i(\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})]$, the variances σ_i^2 can be estimated by solving the equations

$$\|\phi_i(\boldsymbol{I}_n - \boldsymbol{P}_{X_1(V^{-1})})\boldsymbol{y}\|^2 = \sigma_i^2 d_i', \tag{3}$$

for $i = 1,2,3$
Assumpion: $\boldsymbol{y} \sim N_n(\boldsymbol{X}_1\boldsymbol{\tau}, \boldsymbol{V})$ and, hence $\hat{\tau}_* \sim N_v[\tau_*, \mathrm{D}(\hat{\tau}_*)]$

Direct approach (NB design)

$$V^{-1} = \sigma_1^{-2}\phi_1 + \sigma_2^{-2}\phi_2 + \sigma_3^{-2}\phi_3 + \sigma_4^{-2}\phi_4$$

A desirable simplification can be obtained when the dispersion matrix $\boldsymbol{V}$ is replaced by the matrix

$$V_* = \sigma_1^2\phi_1 + \sigma_2^2\phi_2 + \sigma_3^2(I_n - \phi_1 - \phi_2).$$

The inverted matrix V_*^{-1} can be obtained as

$$V_*^{-1} = \sigma_1^{-2}\phi_1 + \sigma_2^{-2}\phi_2 + \sigma_3^{-2}(I_n - \phi_1 - \phi_2).$$

$$V = V_* + (\sigma_4^2 - \sigma_3^2) n^{-1} \mathbf{1}_n \mathbf{1}_n' \quad \text{and} \quad V^{-1} = V_*^{-1} + (\sigma_4^{-2} - \sigma_3^{-2}) n^{-1} \mathbf{1}_n \mathbf{1}_n'.$$

Now it can be shown that the BLUE of $\tau_* = (I_v - n^{-1}\mathbf{1}_v r')\tau$, i.e.,

$$\hat{\tau}_* = (I_v - n^{-1}\mathbf{1}_v r')\hat{\tau} = (I_v - n^{-1}\mathbf{1}_v r')(X_1' V^{-1} X_1)^{-1} X_1' V^{-1} y,$$

can equivalently be written as

$$\hat{\tau}_* = (I_v - n^{-1}\mathbf{1}_v r')(X_1' V_*^{-1} X_1)^{-1} X_1' V_*^{-1} y_*, \tag{4}$$

where $y_* = (I_n - n^{-1}\mathbf{1}_n \mathbf{1}_n') y$, for which
$\mathrm{E}(y_*) = (I_n - n^{-1}\mathbf{1}_n \mathbf{1}_n') X_1 \tau = X_1 (I_v - n^{-1}\mathbf{1}_v r')\tau = X_1 \tau_*$ and
$\mathrm{D}(y_*) = (I_n - n^{-1}\mathbf{1}_n \mathbf{1}_n') V_* (I_n - n^{-1}\mathbf{1}_n \mathbf{1}_n')$.
The dispersion matrix of $\hat{\tau}_*$, can be presented as

$$\mathrm{D}(\hat{\tau}_*) = (I_v - n^{-1}\mathbf{1}_v r')(X_1' V_*^{-1} X_1)^{-1}(I_v - n^{-1} r \mathbf{1}_v'). \tag{5}$$

Table 2. Analysis of variance for an experiment in an NB design

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments	$v-1$	$\widehat{\mathsf{SS}}_V = \hat{\boldsymbol{\tau}}_*' \boldsymbol{X}_1' \hat{\boldsymbol{V}}_*^{-1} \boldsymbol{X}_1 \hat{\boldsymbol{\tau}}_*$	$\widehat{\mathsf{MS}}_V = \frac{\widehat{\mathsf{SS}}_V}{(v-1)}$
Residuals	$n-v$	$\widehat{\mathsf{SS}}_R = \boldsymbol{y}_*'[\hat{\boldsymbol{V}}_*^{-1} - \hat{\boldsymbol{V}}_*^{-1}\boldsymbol{X}_1(\boldsymbol{X}_1'\hat{\boldsymbol{V}}_*^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\hat{\boldsymbol{V}}_*^{-1}]\boldsymbol{y}_* = n-v$	1
Total	$n-1$	$\widehat{\mathsf{SS}}_T = \boldsymbol{y}_*' \hat{\boldsymbol{V}}_*^{-1} \boldsymbol{y}_*$	—

Verification of the hypothesis $H_0 : \tau_* = \mathbf{0}$ will be based on the formulae presented in Table 2, which correspond to the statistics

$$\widehat{F} = \frac{n-v}{v-1} \frac{\widehat{\mathsf{SS}}_V}{n-v} = \frac{\widehat{\mathsf{SS}}_V}{v-1}, \tag{6}$$

where the estimated mean square has, under H_0, an approximated $\chi^2(v-1,0)/(v-1)$ distribution.

Dataset Description

The dataset contains results from sugar beet field trials conducted under the supervision of COBORU. The data include anonymized experimental designs from multi-location trials, covering different years and series of experiments.
Total number of experiments: 38
Overview of experimental designs:

Trial (yield)	Treatments	Superblocks	Blocks	Units	Units/block	Series
PDO trials (2022–2023)	18	4	8	72	9	4 + 5
2018 R trials	27	3	9	81	9	8
2020 R trials	32–37	3	12	96	8	8
2020 R2 trials	65	3	15	195	13	7
2024 R2 trials	66	3	18	198	11	6

Dataset Description

Direct approach:

- custom R code using the `Matrix` and `matlib` packages
- `max_iter <- 100` (maximum number of iterations)
- `tolerance <- 1e-5`

Standard approach (R, lme4 / lmerTest):

- Linear mixed model (LMM) fitted with `lmer()`
- Estimation method: REML
- Significance testing of fixed effects: Kenward–Roger; F-tests (`drop1`, `test = "F"`, `ddf = "Kenward-Roger"`)
- Model diagnostics: singular fit warning indicated near-zero variance for one random effect

Computer specifications:

- Processor: Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz (6 cores, 12 threads, 2592 MHz)
- Installed RAM: 32.0 GB
- System model: DELL G5 5590

Problems

Table: Problems encountered with the direct ANOVA approach and REML in the analysis of nested block design (sugar beet yield data).

	Direct approach	**REML**
Cases with problems	1 out of 38	18 out of 38
Type of problem	More iterations necessary	Singular fit
Details	–	SUPERBLOCK:BLOCK – 8 SUPERBLOCK – 10

Table: Results of hypothesis testing at significance level $\alpha = 0.05$

	Direct approach	**REML**
Rejected H_0	38	36
Not rejected H_0	0	2

Table: Problematic cases in the analysis of sugar beet yield data (nested block design).

Trial (site 4)	**Trt**	**SB**	**B**	**B/SB**	**Units**	**U/B**	p-**DA**	**Sing.**	p-**REML**
2023 PDO	18	4	8	2	72	9	0.048	FALSE	0.11
2020 R	32	3	12	4	96	8	0.030	TRUE	0.23

Estimated variance components

Trial	σ_1^2	σ_2^2	σ_3^2	σ_4^2	Residual	SB:Blk	SB	DIFF
2022 PDO site 1	11.63	0.94	16.57	1	11.04	**0.0000**	0.3076	0.60
2022 PDO site 3	29.51	5.27	307.82	1	28.13	**0.0000**	15.54	1.38
2023 PDO site 1	7.21	48.57	24.39	1	7.23	3.3356	**0.0000**	-0.02
2023 PDO site 2	22.03	7.68	158.54	1	21.21	**0.0000**	7.63	0.82
2023 PDO site 3	56.19	623.60	190.93	1	56.25	41.57	**0.0000**	-0.06
2018 R site 1	1801.46	3592.64	3384.22	1	1802.25	190.91	**0.0000**	-0.79
2018 R site 3	1808.03	18.30	6751.01	1	1671.01	**0.0000**	188.15	137.02
2018 R site 4	718.35	3414.75	1441.25	1	720.31	234.05	**0.0000**	-1.96
2018 R site 8	524.19	1751.23	1306.85	1	524.87	120.95	**0.0000**	-0.68
2020 R site 3	4278.05	5709.44	3763.97	1	4295.94	100.58	**0.0000**	-17.89
2020 R site 4	18714.55	14754.07	54310.17	1	18331.28	**0.0000**	1124.34	383.27
2020 R site 5	2480.33	4381.16	3505.96	1	2484.64	206.99	**0.0000**	-4.31
2020 R site 7	1872.44	4474.64	3041.18	1	1876.50	279.83	0.0002	-4.06
2020 R2 site 4	5176.24	3339.62	21807.39	1	5060.88	**0.0000**	257.64	115.36
2020 R2 site 7	2698.54	3540.45	6467.90	1	2617.82	0.0006	**0.0000**	80.72
2024 R2 site 4	2382.29	6728.74	4879.65	1	2384.49	367.64	**0.0000**	-2.20
2024 R2 site 5	1064.74	1913.30	1044.84	1	1067.00	62.28	**0.0000**	-2.26
2024 R2 site 6	4806.48	5661.93	24314.84	1	4805.01	84.66	282.46	1.47

Time (s)

tr.	Direct approach				REML			
	min	max	median	mean	min	max	median	mean
18	0.010	0.120	0.030	0.044	0.030	0.060	0.050	0.049
27	0.040	0.450	0.060	0.107	0.040	0.070	0.050	0.051
32	0.080	0.160	0.120	0.121	0.040	0.070	0.060	0.057
37	0.110	0.110	0.110	0.110	0.070	0.070	0.070	0.070
65	0.480	1.370	0.800	0.813	0.090	0.110	0.090	0.099
66	0.660	0.880	0.755	0.770	0.100	0.110	0.110	0.107

Iterations

treatments	Direct approach				REML			
	min	max	median	mean	min	max	median	mean
18	7	40	9.00	15.11	20	31	24.00	24.44
27	9	14	13.00	12.43	19	33	27.00	26.71
32	13	24	16.00	16.57	18	32	27.00	25.29
37	14	14	14.00	14.00	34	34	34.00	34.00
65	9	26	15.00	14.71	22	47	34.00	34.43
66	12	16	14.00	14.17	17	26	23.50	22.50

Concluding remarks

The discovered results concerning the proposed approach to ANOVA for experiments with orthogonal block structure seem to be useful.

References

Caliński T., Siatkowski I. (2017): On a new approach to the analysis of variance for experiments with orthogonal block structure. I. Experiments in proper block designs. Biometrical Letters 54: 91-122.

Caliński T., Siatkowski I. (2018): On a new approach to the analysis of variance for experiments with orthogonal block structure. II. Experiments in nested block designs. Biometrical Letters 55: 147-178.

Caliński, T., Łacka, A., Siatkowski, I. (2019): On a new approach to the analysis of variance for experiments with orthogonal block structure. III. Experiments in row-column designs. Biometrical Letters 56: 183–213.

Caliński, T., Łacka, A., Siatkowski, I. (2020): On a new approach to the analysis of variance for experiments with orthogonal block structure. IV. Experiments in split-plot designs. Biometrical Letters 57: 183–213.

Łacka, A. NRC Designs – New Tools for Successful Agricultural Experiments. Agronomy **2021**, *11(12)*, 2406.

Thank you for your attention

Thank you!