Experimental Design

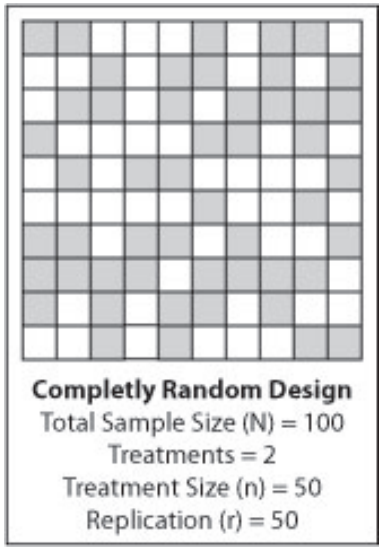
Unit 3

Werner G. Müller

Institute of Applied Statistics

March 19th 2025

3.1 Introduction to Completely randomized designs (CRD)



We are comparing t treatments, using n experimental units and will be considering experimental designs in the context of *unstructured* treatments, by which we simply mean a discrete collection of experimental conditions.

We determine a priori the number of experimental units to be assigned to each treatment, $n_1, n_2, n_3, \ldots n_t$, such that their sum is n, without any additional restrictions.

CRD does not contain blocks of units purposefully selected to be especially similar, any pair of units is viewed as being related in the same way as any other pair.

3.1.1 Example: radiation and rats

TABLE 3.1 Epinephrine Levels (Grams, g) in Rats Treated with Whole Body X-Irradiation, from Matsuu et al. (2005)

Treatment			
1	2	3	4
9.934	8.675	10.509	8.829
9.819	10.720	8.067	10.484
10.693	10.040	9.027	8.632
10.106	9.894	9.680	8.352
9.139	11.912	8.967	9.323

3.2 Models

CRD-models may be seen as models used in one-way analysis of variance. The *cell means model* can be written as:

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad i = 1, \ldots, t \quad j = 1, \ldots, n_i$$

$$\varepsilon_{ij} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ij}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ij}) = \sigma^2$$

ε often represents variation associated with multiple sources and we assume that the ε's are independent, so the experiment should

- assure that treatment-to-unit assignments are made randomly and independently for each unit,
- apply each treatment individually and independently to each of its allocated units,
- carry out any material handling or subsampling processes required for response evaluation independently for each unit, and
- apply the measurement process independently for each unit.

Models (cont.)

Experimental runs receiving the same treatment are actually artificially similar, the data collected in an experiment are realizations of random variables conditioned on all the specific circumstances.

Hence, μ_i actually represents the expectation of responses associated with treatment i, collected under very special, controlled circumstances.

Because the experimental treatments, and not the particular circumstances of experimental execution, are of primary interest, the *effects model* has interpretive advantages:

$$y_{ij} = \alpha + \tau_i + \varepsilon_{ij} \qquad i = 1, \ldots, t \quad j = 1, \ldots, n_i$$

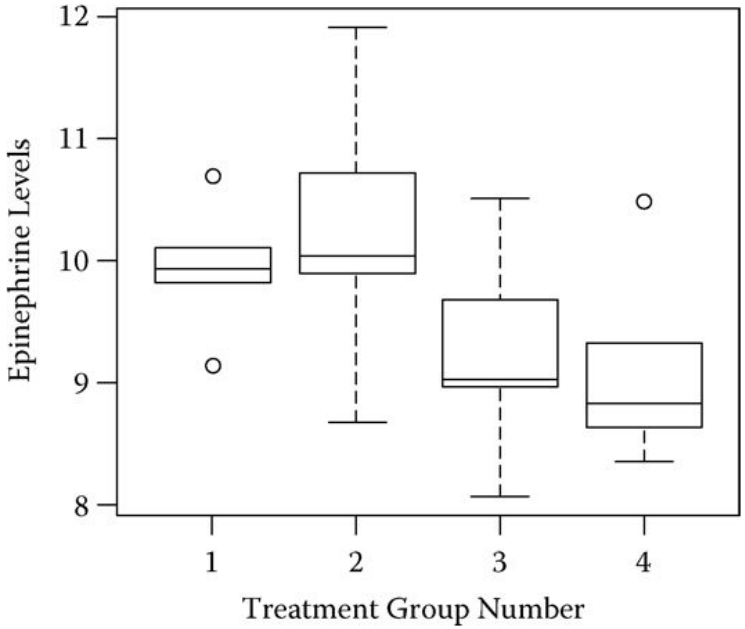
$$\varepsilon_{ij} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ij}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ij}) = \sigma^2$$

α is a nuisance parameter reflecting the contributions of the relatively uninteresting circumstance-details of the experiment, and τ_i is the deviation from α associated with treatment i.
τ_i represents information only about treatment i relative to the other treatments.

3.2.1 Graphical logic

Because CRDs with unstructured treatments are relatively simple, the form of a graphical analysis to present experimental results can also be simple.

The main question is whether the distributions of data from the various treatment groups, especially their means or other measures of central tendency, are different. A set of parallel boxplots of the measured data, one boxplot generated from the data from each treatment group, is a useful presentation for this purpose.



3.3 Matrix formulation

A matrix expression of the cell means model representing all data in the experiment can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is the n-vector of responses, $\mathbf{X}$ is the $(n \times t)$-design matrix, $\boldsymbol{\mu}$ is the t-vector of treatment-means and $\boldsymbol{\varepsilon}$ is the n-vector of errors with $\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathsf{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$.

The effects model may be written in partitioned matrix form to represent data from the entire experiment as:

$$\mathbf{y} = \mathbf{X}_1\alpha + \mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}$$

where $\mathbf{X}_1$ is a n-vector of ones, α is the nuisance parameter, $\boldsymbol{\tau}$ is the t-vector of treatment effects and $\mathbf{X}_2$ is the $(n \times t)$-design matrix associated with the parameters of interest. $\mathbf{X}_2$ is exactly $\mathbf{X}$ from the cell means model - (overparametrization!), so there is no new „structure“ added to the data analysis with the effects model.

estimable contrasts

A linear combination $\mathbf{c}^T\boldsymbol{\tau}$ of treatment parameters of the effects model is estimable only if $\mathbf{c}$ can be written as a linear combination of the rows of $\mathbf{X}_{2|1} = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2$.

In the CRD we have $\mathbf{H}_1 = \frac{1}{n}\mathbf{J}$ where $\mathbf{J}$ is a $n \times n$-matrix of ones, hence

$$\mathbf{X}_{2|1} = \mathbf{X}_2 - \frac{1}{n}\begin{pmatrix} n_1\mathbf{1} & n_2\mathbf{1} & \cdots & n_t\mathbf{1} \end{pmatrix}$$

$\mathbf{X}_{2|1}$ is a „column-centered“ version of $\mathbf{X}_2$, i.e. the sum of the elements of each column of $\mathbf{X}_{2|1}$ is zero. Also each row of $\mathbf{X}_{2|1}$ has a zero sum, so this must also be true of $\mathbf{c}$ for any estimable $\mathbf{c}^T\boldsymbol{\tau}$.

Hence, the only linear combinations of $\boldsymbol{\tau}$'s that are estimable (in the effects model) are contrasts for which $\mathbf{c}^T\mathbf{1} = 0$.

estimable contrasts (cont.)

In the cell means model, $\mathbf{c}^T\boldsymbol{\mu}$ is estimable if $\mathbf{c}$ is a linear combination of the rows of $\mathbf{X}$. This is no restriction at all because any real-valued vector $\mathbf{c}$ can be formed as a linear combination of the rows of $\mathbf{X}$.

In order to eliminate any experiment-wide effects common to the means of all observations, *the only estimable functions of treatment-specific parameters are contrasts*, because these are the only linear functions that eliminate the common nuisance parameter α in the expectation through cancellation.

For information matrices from now on we drop the subscript „$2|1$“ used for partitioned models, all subsequent information matrices $\mathcal{I}$ presented will be for parameters associated with treatments (here $\boldsymbol{\tau}$), controlling for nuisance parameters (here α)

estimable contrasts (cont.)

With the t-vector $\mathbf{n} = \begin{pmatrix} n_1 & n_2 & \cdots & n_t \end{pmatrix}^T$ the information matrix for the effects model is

$$\mathcal{I} = \mathbf{X}_{2|1}^T \mathbf{X}_{2|1} = \mathsf{diag}(\mathbf{n}) - \frac{1}{n}\mathbf{n}\,\mathbf{n}^T$$

In the special case of equal sample sizes for each treatment ($n_i = \frac{n}{t},\ i = 1, \ldots, t$), this reduces to $\mathcal{I} = \frac{n}{t}(\mathbf{I} - \frac{1}{t}\mathbf{J})$.

Solving the reduced normal equations for the CRD we get

$$\left(\mathsf{diag}(\mathbf{n}) - \frac{1}{n}\mathbf{n}\,\mathbf{n}^T\right)\ \hat{\boldsymbol{\tau}} = \mathbf{X}_2^T\mathbf{y} - \bar{y}\,\mathbf{n}$$

where $\bar{y}$ is the mean of all elements of $\mathbf{y}$.

estimable contrasts (cont.)

For the i-th scalar equation of the above set we get

$$\hat{\tau}_i - \bar{\hat{\tau}} = \bar{y}_i - \bar{y}$$

where $\bar{\hat{\tau}} = \frac{1}{n}\sum n_i \hat{\tau}_i$ and $\bar{y}_i$ is the mean of the n_i elements of $\mathbf{y}$ associated with treatment i.

For estimable functions $\mathbf{c}^T\boldsymbol{\tau}$ the sum of the elements of $\mathbf{c}$ has to be zero. Hence the constants $\bar{\hat{\tau}}$ and $\bar{y}$ are eliminated in

$$\widehat{\mathbf{c}^T\boldsymbol{\tau}} = \mathbf{c}^T\hat{\boldsymbol{\tau}} = \sum_{i=1}^{t} c_i\hat{\tau}_i = \sum_{i=1}^{t} c_i\bar{y}_i$$

so the least-squares estimate of any contrast of treatment parameters is the same contrast in the corresponding treatment data averages.

3.4 Influence of the design on estimation

The estimate $\mathbf{c}^T\hat{\boldsymbol{\tau}}$ is a linear combination of independent sample means with known variances, hence

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}}) = \sigma^2\,\mathbf{c}^T\mathcal{I}^-\mathbf{c}$$

Since $\mathsf{diag}(\mathbf{n})^{-1}$ is a generalized inverse of $\mathcal{I} = \mathbf{X}_2^T(\mathbf{I}-\mathbf{H}_1)\mathbf{X}_2$, for estimable $\mathbf{c}^T\boldsymbol{\tau}$ we get

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}}) = \sigma^2\,\mathbf{c}^T\,\mathsf{diag}(\mathbf{n})^{-1}\,\mathbf{c} = \sigma^2\sum_{i=1}^{t}\frac{c_i^2}{n_i}$$

the expected squared length of a $(1-\alpha)$-confidence interval for estimable $\mathbf{c}^T\boldsymbol{\tau}$ is

$$4\,\mathsf{t}^2_{1-\frac{\alpha}{2}}(n-\mathsf{rk}(\mathbf{X}))\,\sigma^2\,\mathbf{c}^T\mathcal{I}^-\mathbf{c} \;=\; 4\,\mathsf{t}^2_{1-\frac{\alpha}{2}}(n-t)\,\sigma^2\sum_{i=1}^{t}\frac{c_i^2}{n_i}$$

3.4.1 Allocation

We would prefer designs that lead to relatively small values of the above length

$$2\, \mathrm{t}_{1-\frac{\alpha}{2}}(n-t)\, \sigma \sqrt{\sum_{i=1}^{t} \frac{c_i^2}{n_i}}$$

either through large overall sample size (and so a relatively small t-quantile), or through *allocation* of relatively more units to groups for which the corresponding $|c_i|$ is large for contrasts of interest.

We often have the freedom to select the number of units to be used in each treatment group n_i. The CRD with equal group sizes has good overall properties, and is in fact optimal for many - but not all - experimental goals.

Allocation problems are formulated as constrained optimization problems in which the quantity to be optimized is, e.g. the variance of an estimator or some function of it. The constraint most often reflects the total number of units allowed.

Method of Lagrangian multipliers

Suppose we want to estimate p linear contrasts of $\boldsymbol{\tau}$, $\mathbf{C}\,\boldsymbol{\tau}$ where $\mathbf{C}$ is a $(p \times t)$-matrix of coefficients. The variance matrix of the least-squares estimate is:

$$\mathsf{Var}(\widehat{\mathbf{C}^T\boldsymbol{\tau}}) \;=\; \sigma^2\,\mathbf{C}^T\,\mathcal{I}^-\,\mathbf{C} \;=\; \sigma^2\,\mathbf{C}^T\,\mathsf{diag}(\mathbf{n})^{-1}\,\mathbf{C}$$

If we want to minimize the average variance of these estimates, this is equivalent to minimization of:

$$\mathsf{tr}(\mathbf{C}^T\,\mathsf{diag}(\mathbf{n})^{-1}\,\mathbf{C}) = \mathsf{tr}(\mathbf{C}^T\mathbf{C}\,\mathsf{diag}(\mathbf{n})^{-1})$$

With unconstrained optimization we would simply choose the largest possible value for each n_i. The more realistic constrained problem can be solved using the *Method of Lagrangian Multipliers*.

Optimal design problems are often solved as if $n_1, \ldots, n_t$ were actually continuous variables, and the resulting solution is „rounded“ to integer values if necessary.

Method of Lagrangian multipliers (review)

Suppose we wish to maximize or minimize a differentiable function $f(\mathbf{n})$ with respect to $\mathbf{n}$, a real-valued vector of t arguments, subject to the constraint $g(\mathbf{n}) = G$ for a specified differentiable function g and scalar value G. Now introduce a new scalar variable L, and define a function of $t+1$ arguments:

$$h(\mathbf{n}, L) = f(\mathbf{n}) + L\,(g(\mathbf{n}) - G)$$

The technique calls for solving the gradient of h:

$$\nabla h = \left(\begin{array}{cccc} \frac{\partial h}{\partial n_1} & \cdots & \frac{\partial h}{\partial n_t} & \frac{\partial h}{\partial L} \end{array} \right)^T = \mathbf{0}$$

The „Lagrangian Multiplier“ L introduces the desired constraint through

$$\frac{\partial h}{\partial L} = g(\mathbf{n}) - G = 0$$

3.4.2 Overall experiment size

The allocation problem is to find optimal values of n_i, $i = 1, \ldots, t$, under the constraint of a specified value for n.

Now suppose the proportion of units to be used with each treatment has been determined, and call these p_i, $i = 1, \ldots, t$, $\sum p_i = 1$. How large would n have to be, to assure that some experimental objectives are met?

With a known information matrix $\mathcal{I}$ the *per-observation information matrix* is defined as $\mathcal{I}^1 = \frac{1}{n}\mathcal{I}$.

For CRDs with fixed proportions of units assigned to each treatment, $\mathcal{I}^1$ would be the same matrix regardless of the value of n used to compute $\mathcal{I}$. The per observation information matrix for the effects model is

$$\mathcal{I}^1 = \frac{1}{n}\mathbf{X}_{2|1}^T\mathbf{X}_{2|1} = \frac{1}{n}(\mathsf{diag}(\mathbf{n}) - \frac{1}{n}\mathbf{n}\,\mathbf{n}^T) = \mathsf{diag}(\mathbf{p}) - \mathbf{p}\,\mathbf{p}^T$$

per-observation information matrix

A generalized inverse of the per-observation information matrix is:
$(\mathcal{I}^1)^- = \mathsf{diag}(\mathbf{p})^{-1}$

Then, for any value of n we have $\mathcal{I}^- = \frac{1}{n}\,(\mathcal{I}^1)^- = \frac{1}{n}\,\mathsf{diag}(\mathbf{p})^{-1}$

The variance of the estimate of a particular linear contrast $\mathbf{c}^T\boldsymbol{\tau}$ is a function of n:

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}}) \;=\; \sigma^2\,\mathbf{c}^T\mathcal{I}^-\mathbf{c} \;=\; \frac{\sigma^2}{n}\,\mathbf{c}^T\mathsf{diag}(\mathbf{p})^{-1}\mathbf{c}$$

If we want a design for which the square root of this variance will be small relative to $\mathbf{c}^T\boldsymbol{\tau}$, n should be large enough to make

$$\frac{\mathbf{c}^T\boldsymbol{\tau}}{\sqrt{\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}})}} \;=\; \sqrt{n}\,\frac{\mathbf{c}^T\boldsymbol{\tau}/\sigma}{\sqrt{\mathbf{c}^T\mathsf{diag}(\mathbf{p})^{-1}\mathbf{c}}}$$

acceptably large.

3.5 Influence of design on hypothesis testing

Tests of hypotheses on the treatment parameters τ are based on the variance decomposition

$$\mathsf{TSS} = \mathsf{SST} + \mathsf{SSE}$$

where TSS is the total sum of squares, SST is the sum of squares associated with the treatment and SSE is the sum of squares associated with the error, i.e. the sum of squares not explained by the treatment. For this model we have

$$\mathsf{TSS} = \sum_{i,j}(y_{ij} - \bar{y})^2 \qquad \mathsf{SST} = \sum_{i=1}^{t} n_i(\bar{y}_i - \bar{y})^2 \qquad \mathsf{SSE} = \sum_{i,j}(y_{ij} - \bar{y}_i)^2$$

The decomposition also holds for the associated degrees of freedom

$$\mathsf{df}_{\mathsf{TSS}} = n - 1 \qquad \mathsf{df}_{\mathsf{SST}} = t - 1 \qquad \mathsf{df}_{\mathsf{SSE}} = n - t$$

The mean squares are just the sum of squares divided by their degrees of freedom

$$\mathsf{TMS} = \frac{TSS}{n-1} \qquad \mathsf{MST} = \frac{SST}{t-1} \qquad \mathsf{MSE} = \frac{SSE}{n-t}$$

Influence of design on hypothesis testing (cont.)

We want to test the variation associated with $\boldsymbol{\tau}$, i.e. the null hypothesis

$$H_0 : \ \tau_1 = \tau_2 = \cdots = \tau_t$$

Under H_0 the F-test-statistic

$$\mathsf{F} \ = \ \frac{\mathsf{MST}}{\mathsf{MSE}} \ = \ \frac{\frac{1}{t-1}\sum_{i=1}^{t} n_i(\bar{y}_i - \bar{y})^2}{\frac{1}{n-t}\sum_{i,j}(y_{ij} - \bar{y}_i)^2}$$

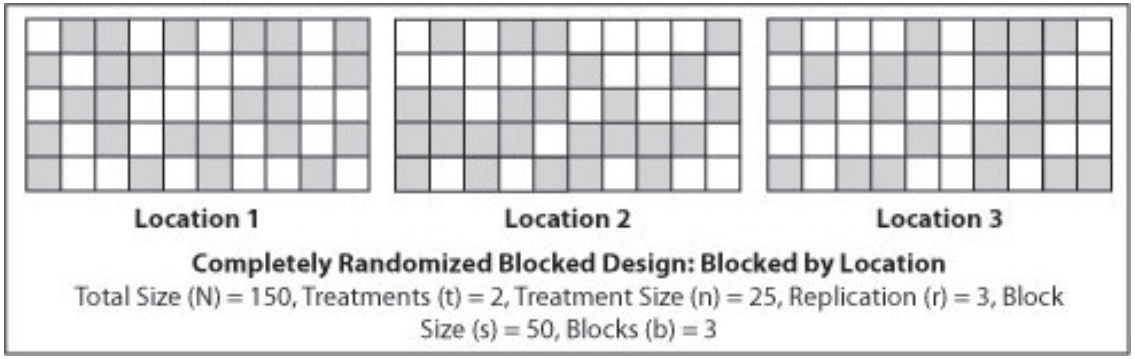
has a central F-distribution with $t-1$ and $n-t$ degrees of freedom.
Under H_1 F has a noncentral F-distribution depending on the value of $\boldsymbol{\tau}$, the noncentrality parameter for this test is

$$\lambda \ = \ \frac{1}{\sigma^2}\boldsymbol{\tau}^T\mathcal{I}\boldsymbol{\tau} \ = \ \frac{1}{\sigma^2}\boldsymbol{\tau}^T\left(\mathsf{diag}(\mathbf{n}) - \frac{1}{n}\mathbf{n}\,\mathbf{n}^T\right)\boldsymbol{\tau}$$

The power of this test of equal treatment effects, for given values of $\boldsymbol{\tau}$ and σ^2, is

$$Pr\left(W > \mathsf{F}_{1-\alpha}(t-1; n-t)\right) \qquad \text{where} \qquad W \sim \mathsf{F}(t-1; n-t; \lambda)$$

4.1 Introduction to randomized complete blocks and related designs (CBD)



The fundamental assumption of CRD is that the available collection of experimental units is homogeneous - no predictable or systematic differences are expected in the collected data other than those that are attributable to the treatments.

The CRD is simple, popular, and frequently used, but its application is unrealistic or impractical in many settings.

Experimental units made from special subpopulations often regarded as being more similar than a collection of units made from the whole population.

Introduction to CBD (cont.)

With *Complete Block Designs* the experiment is executed in blocks, or „sub-experiments“, each using only the experimental material from one subpopulation. Each treatment is applied once using the material in each block, making a complete unreplicated sub-experiment, and this pattern is repeated using b such blocks.

In a *Randomized* CBD, treatments are randomly applied to units within each block, but these assignments are not completely random as in a CRD because they are restricted to balance across the units within each block.

4.1.1 Example: structural reinforcement bars

TABLE 4.1 Tensile Strength (Kilograms Per Square Inch, ksi) of Steel Reinforcement Bars, from Kocaoz et al. (2005)

	Coating			
Block	1	2	3	4
1	136	147	138	149
2	136	143	122	153
3	150	142	131	136
4	155	148	130	129
5	145	149	136	139
6	150	149	147	144
7	147	150	125	140
8	148	149	118	145

4.2 Model

CBD suggests two potential systematic patterns in the data. The pattern of primary interest is associated with the applied treatments, the other is associated with blocks of units. An effects model for t treatments and b blocks is

$$y_{ij} = \alpha + \beta_i + \tau_j + \varepsilon_{ij} \qquad i = 1, \ldots, b \quad j = 1, \ldots, t$$

$$\varepsilon_{ij} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ij}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ij}) = \sigma^2$$

In CBD blocks are treated as fixed effects (random block effects are discussed later).

The effect of blocks is additive, i.e. CBD include no block-by-treatment interaction terms.

4.2.1 Graphical logic

Parallel boxplots of response values from a CBD should be constructed with care because each data value contains a contribution from a specific block as well as a specific treatment.

Instead we could use boxplots of „block-corrected“ observations for each treatment group. For treatment j, summarize data by a boxplot of b values:

$$y_{ij}^* = y_{ij} - \bar{y}_i \qquad i = 1, \ldots, b$$

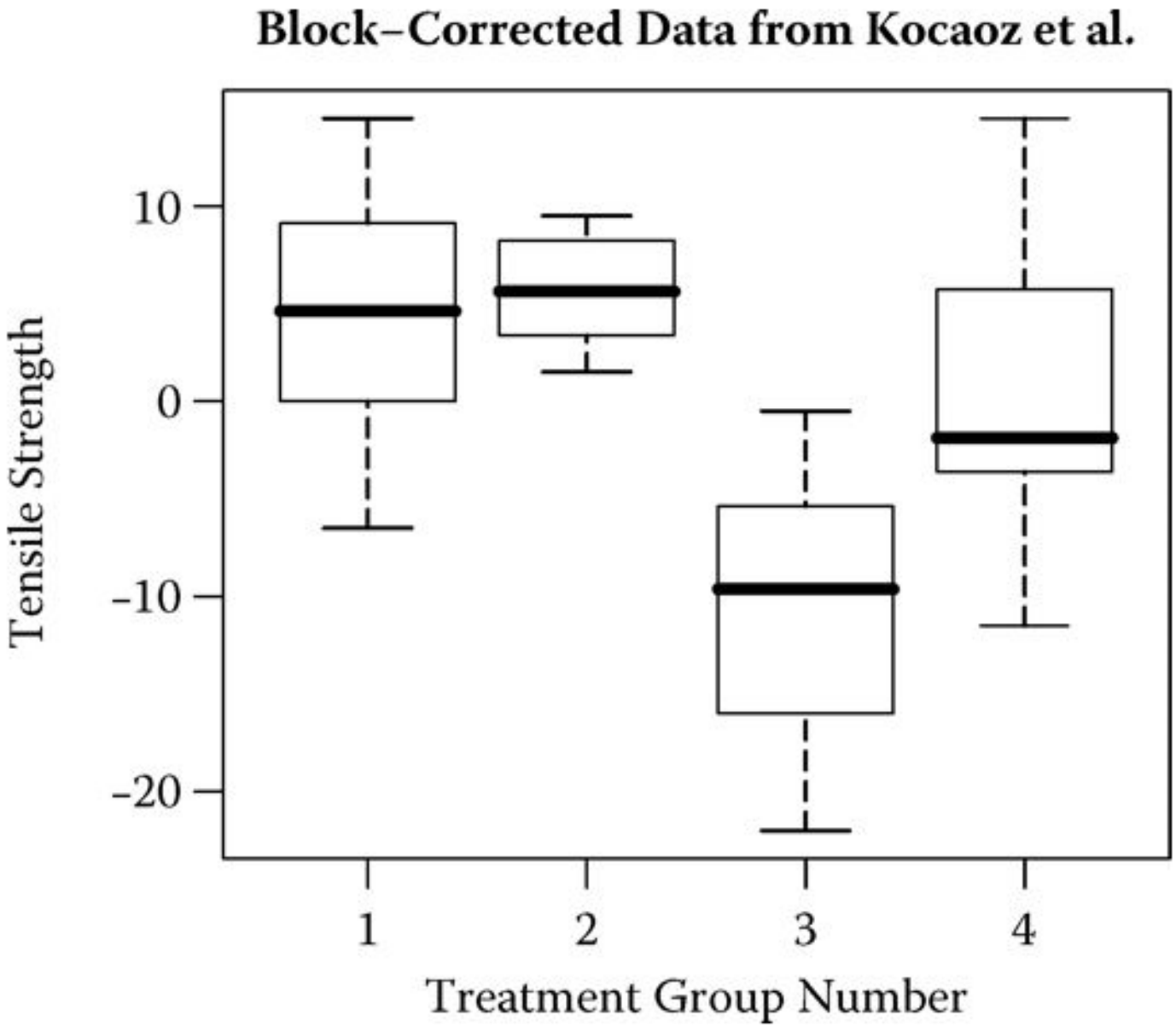
Using the model equation above we get $\bar{y}_i = \alpha + \beta_i + \bar{\tau} + \bar{\varepsilon}_i$ and

$$y_{ij}^* = (\tau_j - \bar{\tau}) + (\varepsilon_{ij} - \bar{\varepsilon}_i)$$

y_{ij}^* reflects only the contribution of treatment j, relative to the average effect of all treatments. y_{ij}^* has smaller variance than y_{ij}:

$$\mathsf{Var}(y_{ij}^*) = \sigma^2 \left(1 - \frac{1}{t}\right)$$

Graphical logic (cont.)



Graphical logic (cont.)

We may check the assumed additive effects of treatments and blocks using block-and-treatment-corrected data (residuals)

$$r_{ij} = y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y} = y_{ij} - \hat{y}_{ij}$$

with

$$\mathsf{E}(r_{ij}) = 0 \qquad \text{and} \qquad \mathsf{Var}(r_{ij}) = \sigma^2 \frac{(t-1)(b-1)}{t\,b}$$

The residuals are correlated, however, any outliers appearing in a boxplot of these values may be indicators of model inadequacy, such as possible treatment-block interaction.

4.3 Matrix formulation

We use a partitioned model for all data from the experiment:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}; \sigma^2\mathbf{I})$$

- $\boldsymbol{\beta}$ is the $(b+1)$-vector of nuisance parameters $\alpha, \beta_1, \ldots, \beta_b$
- $\boldsymbol{\tau}$ is the t-vector of treatment parameters
- $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are n-vectors of responses and random errors where $n = b \cdot t$
-

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_t & \mathbf{1}_t & \mathbf{0}_t & \cdots & \mathbf{0}_t \\ \mathbf{1}_t & \mathbf{0}_t & \mathbf{1}_t & \cdots & \mathbf{0}_t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_t & \mathbf{0}_t & \mathbf{0}_t & \cdots & \mathbf{1}_t \end{pmatrix} \qquad \mathbf{X}_2 = \begin{pmatrix} \mathbf{I}_t \\ \mathbf{I}_t \\ \vdots \\ \mathbf{I}_t \end{pmatrix}$$

where $\mathbf{1}_t$, $\mathbf{0}_t$, and $\mathbf{I}_t$ refer to t-vectors of ones and zeros, and the $(t \times t)$-identity matrix, respectively.

Matrix formulation (cont.)

The reduced normal equations for the treatment parameters $\boldsymbol{\tau}$ are

$$\mathbf{X}_2^T(\mathbf{I}-\mathbf{H}_1)\mathbf{X}_2\,\hat{\boldsymbol{\tau}} \;=\; \mathbf{X}_2^T(\mathbf{I}-\mathbf{H}_1)\mathbf{y}$$

Some of these matrices have different structure than their counterparts in the CRD.

With some matrix algebra we get

$$\mathbf{X}_{2|1} \;=\; (\mathbf{I}-\mathbf{H}_1)\mathbf{X}_2 \;=\; \mathbf{X}_2 - \frac{1}{t}\mathbf{1}_{(n\times t)}$$

where $\mathbf{1}_{(n\times t)}$ is an $(n\times t)$-matrix of ones.

$\mathsf{rk}(\mathbf{X}_1) = b$, so the nuisance part of the model is overparametrized. For the purpose of inferences about t, we might have omitted α or one of the β's from the model to get exactly the same $\mathbf{X}_{2|1}$.

Matrix formulation (cont.)

For this model, $\mathbf{X}_2$ and $\mathbf{X}_{2|1}$ are the same as with a CRD with b units in each treatment group. So the reduced normal equations for the CBD take the same form as those for the CRD:

$$\hat{\tau}_j - \bar{\hat{\tau}} = \bar{y}_j - \bar{y} \qquad j = 1, \ldots, t$$

The least-squares point estimators of estimable functions of treatment effects can be constructed for CBDs by simply ignoring the blocks. This is due to symmetry properties of the CBD, specifically, that each treatment is applied to exactly one unit in each block (balanced design).

So the CBD is a specific case of a CRD with $n_i = b$, $i = 1, \ldots, t$.

The generalized inverse of the information matrix is $\mathcal{I}^- = \frac{1}{b}\mathbf{I}$ and for estimable linear combinations $\mathbf{c}^T\boldsymbol{\tau}$ the vector $\mathbf{c}$ must satisfy $\mathbf{c}^T\mathbf{1} = 0$

4.4 Influence of design on estimation

The least-squares estimator of any set of estimable functions $\mathbf{C}^T\boldsymbol{\tau}$ is

$$\widehat{\mathbf{C}^T\boldsymbol{\tau}} = \mathbf{C}\begin{pmatrix}\bar{y}_1\\ \vdots \\ \bar{y}_t\end{pmatrix} \quad \text{with} \quad \text{Var}(\widehat{\mathbf{C}^T\boldsymbol{\tau}}) = \frac{\sigma^2}{b}\,\mathbf{C}\,\mathbf{C}^T$$

This variance function has the same form as those for the CRD with $n_i = b$.

BUT: σ^2 represents uncontrolled variation among all units in a CRD, while it represents only uncontrolled variation among units from a common block in a CBD.

Hence the sampling variance of treatment contrasts may be substantially smaller for a CBD than for a CRD of the same size if blocking is „effective“, i.e. if it results in greater homogeneity among units-within-blocks than can be expected within a larger collection of unblocked units.

Influence of design on estimation (cont.)

An experiment can be executed as either a CRD or a CBD and sometimes CBD has to be chosen because of operational requirements. But especially for small designs the CRD can be expected to yield more precise confidence intervals even if $\sigma^2_{CBD} < \sigma^2_{CRD}$.

For any particular estimable contrast $\mathbf{c}^T\boldsymbol{\tau}$, the expected squared length of the associated confidence intervals with fixed n are

$$4\, \mathrm{t}_{1-\frac{\alpha}{2}}(n-t)\, \sigma^2_{CRD}\, \mathbf{c}^T\mathbf{c}\, \frac{t}{n} \qquad \text{and}$$

$$4\, \mathrm{t}_{1-\frac{\alpha}{2}}(n-b-t+1)\, \sigma^2_{CBD}\, \mathbf{c}^T\mathbf{c}\, \frac{t}{n}$$

respectively. So the CBD can only be expected to yield more precise intervals if

$$\frac{\sigma^2_{CBD}}{\sigma^2_{CRD}} < \frac{\mathrm{t}_{1-\frac{\alpha}{2}}(n-t)}{\mathrm{t}_{1-\frac{\alpha}{2}}(n-b-t+1)}$$

4.4.1 Experiment size

for any estimable function $\mathbf{c}^T\boldsymbol{\tau}$, the variance of the estimate is

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}}) = \frac{\sigma^2}{b}\sum_{j=1}^{t} c_j^2 = \frac{\sigma^2\mathbf{c}^T\mathbf{c}}{b}$$

For any treatment contrast the value of

$$\boldsymbol{\Psi} = \frac{\mathbf{c}^T\boldsymbol{\tau}}{\sqrt{\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}})}} = \frac{\mathbf{c}^T\boldsymbol{\tau}\sqrt{b}}{\sigma\sqrt{\mathbf{c}^T\mathbf{c}}} = \psi\sqrt{\frac{b}{\mathbf{c}^T\mathbf{c}}}$$

should be acceptably large. Notice that the true signal-to-noise ratio $\psi = \frac{\mathbf{c}^T\boldsymbol{\tau}}{\sigma}$ is unknown. For a given desired value of $\boldsymbol{\Psi}$ and signal-to-noise ratio is ψ, we can solve for the required number of blocks

$$b \geq \frac{\boldsymbol{\Psi}^2}{\psi^2}\mathbf{c}^T\mathbf{c}$$

4.5 Influence of design on hypothesis testing

The variance decomposition for this model is

$$\mathsf{TSS} = \mathsf{SST} + \mathsf{SSB} + \mathsf{SSE}$$

where SSB is the sum of squares associated with the blocks. In detail we have

$$\mathsf{SST} = \sum_{j=1}^{t} b(\bar{y}_j - \bar{y})^2 \qquad \mathsf{SSB} = \sum_{i=1}^{b} t(\bar{y}_i - \bar{y})^2 \qquad \mathsf{SSE} = \sum_{i,j} (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$$

The associated degrees of freedom are

$$\mathsf{df}_{\mathsf{SST}} = t - 1 \qquad \mathsf{df}_{\mathsf{SSB}} = b - 1 \qquad \mathsf{df}_{\mathsf{SSE}} = n - t - b + 1$$

So the entire analysis of data from a CBD cannot be carried out ignoring blocks although the form of the least-squares estimates of estimable functions of treatment effects is the same for CRDs and CBDs. The reason is in the different error sum of squares: $\mathsf{SSE}_{CRD} > \mathsf{SSE}_{CBD}$!

Influence of design on hypothesis testing (cont.)

However the form the noncentrality parameter for the F-test of

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$$

is identical to that of a CRD with equal numbers of units assigned to each treatment ($n_i = b$), just σ^2 differs from the CRD:

$$\lambda = \frac{1}{\sigma^2}\boldsymbol{\tau}^T\mathcal{I}\boldsymbol{\tau} = \frac{1}{\sigma^2}\boldsymbol{\tau}^T\left(b\mathbf{I} - \frac{b^2}{n}\mathbf{1}\mathbf{1}^T\right)\boldsymbol{\tau}$$

We may compare a CRD and a CBD of the same size in the context of the power $(1-\beta)$ of the above test:

$1-\beta_{\mathsf{CRD}} = Pr(W_{\mathsf{CRD}} > \mathsf{F}_{1-\alpha}(t-1; n-t))$ $\qquad W_{CRD} \sim \mathsf{F}(t-1; n-t; \frac{\boldsymbol{\tau}^T\mathcal{I}\boldsymbol{\tau}}{\sigma^2_{CRD}})$

$1-\beta_{\mathsf{CBD}} = Pr(W_{\mathsf{CBD}} > \mathsf{F}_{1-\alpha}(t-1; n-t-b+1))$

$W_{CBD} \sim \mathsf{F}(t-1; n-t-b+1; \frac{\boldsymbol{\tau}^T\mathcal{I}\boldsymbol{\tau}}{\sigma^2_{CBD}})$

Again, the trade-off is between the degrees of freedom (favoring the CRD) and the size of the noncentrality parameter

4.6 Orthogonality and „Condition E“

Normal equations and design information matrices for the CBD and CRD with $n_j = b$ are equivalent. This equivalence can be generalized if two conditions are satisfied - we will call them „Condition E“.

We consider two designs with t treatments and n runs each, which can be modeled with a partitioned linear model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}$.
$\mathbf{X}_1$ may have a different number of columns for the two designs.

Then the two designs satisfy *Condition E* if, for some ordering of rows

- $\mathbf{X}_2$ is the same matrix for each design, and
- $\mathbf{H}_1\mathbf{X}_2$ is the same matrix for each design

As a consequence $\mathbf{X}_{2|1} = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2$ is the same for the two designs.

Orthogonality and „Condition E“ (cont.)

For CRD and a CBD that assign the same number of units to each treatment, the estimable contrasts in the elements of τ can be estimated in a CBD ignoring blocks, just as they can be estimated in a CRD ignoring α.

We may express this by „*treatments are orthogonal to blocks*“ in a CBD.

But this is also true for any other blocked design which, together with a CRD, satisfies Condition E. CBD are not the only arrangements for which treatments and blocks are orthogonal.

Orthogonality and „Condition E“ (cont.)

Consider a design with b blocks, and the blocks are of size $m_1, m_2, \ldots, m_b$ units. Also the number of units to which any treatment is applied may be different in each block. We may express Condition E for such a design and a CRD of same size as follows

- the number of units associated with the various treatments $n_1, \ldots, n_t$ has to be the same in each design, and
- in the block design any specific treatment must be applied to the same proportion of units in each block (although these proportions do not need to be the same for each treatment).

Note that this condition cannot be met for all possible integer values of $t, m_1, \ldots, m_b$, however, it can be easily satisfied when all blocks contain the same number of units. If this common block size is greater than t, we call the design an *augmented complete block designs*.

Orthogonally blocked designs are attractive because they result in simple reduced normal equations.

Orthogonality and „Condition E“ (cont.)

Are *orthogonally blocked designs* really statistically superior to other blocked arrangements? YES!

Consider a CRD and a blocked design, each of which assigns n_j units to treatment $j = 1, \ldots, t$. Also the error variance shall be the same for both designs: $\mathsf{Var}(\mathbf{y}) = \sigma^2 \mathbf{I}$.

Then for the CRD the variance of $\widehat{\mathbf{c}^T \boldsymbol{\tau}}$ is

$$\mathsf{Var}(\widehat{\mathbf{c}^T \boldsymbol{\tau}}) = \sigma^2 \sum_{j=1}^{t} \frac{c_j^2}{n_j}$$

For a general blocked design we have

$$\mathsf{Var}(\widehat{\mathbf{c}^T \boldsymbol{\tau}}) = \sigma^2 \sum_{j=1}^{t} \frac{c_j^2}{n_j} + \mathbf{c}^T \mathbf{Q}\, \mathbf{c}$$

where $\mathbf{Q}$ is a positive semidefinite $(t \times t)$-matrix.

Hence, apart from its effect on the value of σ^2, blocking cannot improve the variance properties of a design, and can make them much worse.

On the other hand, the variance properties are not degraded for blocked designs which, together with a CRD of the same size, satisfy Condition E.