Experimental Design

Unit 8

Werner Müller

Institute of Applied Statistics

May 7th 2025

11.1 Introduction to two-level factorial experiments: basics

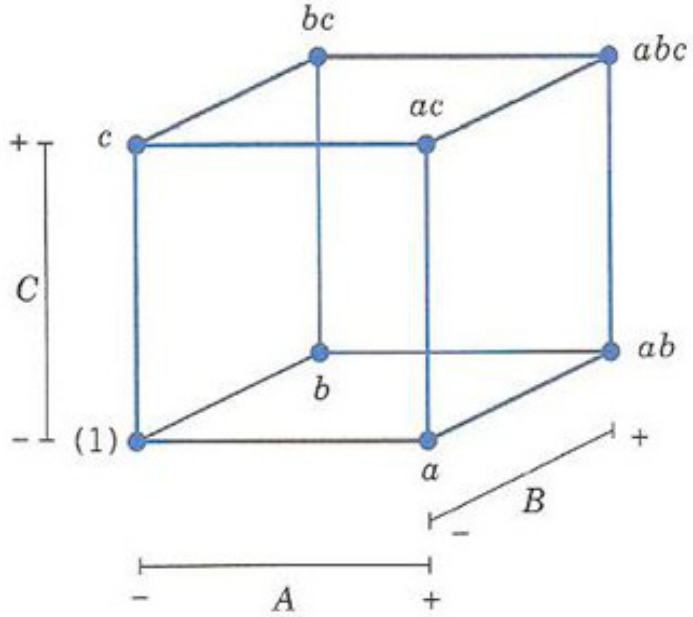
In many applications, experiments designed to examine the effects of two-level factors are especially common.

Restricting factors to two levels minimizes the number of treatments that must be considered for a given number of factors and maximizes the number of factors that can be examined in a factorial experiment of a given number of treatments.

If we consider f two-level factors, each of the 2^f treatments can be completely identified by an ordered string or vector of f binary „bits“ each symbolized by $(0, 1)$, $(-, +)$ or („low“, „high“)

Two-level experimental designs are often depicted „spatially“ by representing treatments as the corners of a square ($f = 2$), cube ($f = 3$), or hyper-cube ($f > 3$), where each spatial dimension is associated with a factor.

2^3 factorial design



Run	A	B	C	Label
1	−	−	−	(1)
2	+	−	−	a
3	−	+	−	b
4	+	+	−	ab
5	−	−	+	c
6	+	−	+	ac
7	−	+	+	bc
8	+	+	+	abc

geometric view

design matrix

A factorial effects model of overparameterized form can be written as:

$$y_{ijkt} = \mu + \dot{\alpha}_1 + \dot{\beta}_j + \dot{\gamma}_k + (\dot{\alpha\beta})_{ij} + (\dot{\alpha\gamma})_{ik} + (\dot{\beta\gamma})_{jk} + (\dot{\alpha\beta\gamma})_{ijk} + \varepsilon_{ijkt}$$

2^3 factorial design (cont.)

In each main effect and interaction of the overparameterized model there is just one parameter that is linearly independent of the former parameters. So for a full-rank parameterization we need just one matrix $\mathbf{F} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

With $\mathbf{f}_i$ denoting the i-th row of $\mathbf{F}$ we get

$$y_{ijkt} = \mu + \mathbf{f}_i\boldsymbol{\alpha} + \mathbf{f}_j\boldsymbol{\beta} + \mathbf{f}_k\boldsymbol{\gamma} + \mathbf{f}_i\mathbf{f}_j(\boldsymbol{\alpha\beta}) + \mathbf{f}_i\mathbf{f}_k(\boldsymbol{\alpha\gamma}) + \mathbf{f}_j\mathbf{f}_k(\boldsymbol{\beta\gamma}) + \mathbf{f}_i\mathbf{f}_j\mathbf{f}_k(\boldsymbol{\alpha\beta\gamma}) + \varepsilon_{ijkt}$$

The full-rank parameterization requires a total of 2^f parameters, the same as the number of cell means.

The above model equation has the form of a regression model with $\mathbf{f}$'s that are always either -1 or $+1$.

2^3 factorial design (cont.)

In matrix form, the data model for the complete, unreplicated 2^3 factorial design (with obvious extension to the general 2^f case) can be written as:

$$\mathbf{y} = \begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{pmatrix} = \begin{pmatrix} + & - & - & - & + & + & + & - \\ + & - & - & + & + & - & - & + \\ + & - & + & - & - & + & - & + \\ + & - & + & + & - & - & + & - \\ + & + & - & - & - & - & + & + \\ + & + & - & + & - & + & - & - \\ + & + & + & - & + & - & - & - \\ + & + & + & + & + & + & + & + \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{\gamma} \\ (\boldsymbol{\alpha\beta}) \\ (\boldsymbol{\alpha\gamma}) \\ (\boldsymbol{\beta\gamma}) \\ (\boldsymbol{\alpha\beta\gamma}) \end{pmatrix} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

The design matrix $\mathbf{X}$ has orthogonal columns of length $\sqrt{n}$:

$$\mathbf{X}^T\mathbf{X} = n\mathbf{I}_{2^f} \qquad \hat{\boldsymbol{\theta}} = \frac{1}{n}\mathbf{X}^T\mathbf{y}$$

2^3 factorial design (cont.)

Including the case of replications, i.e. if $r \geq 1$ we have

$$\hat{\boldsymbol{\theta}} = 2^{-f}\mathbf{M}^T\bar{\mathbf{y}}$$

Note that $\mathbf{M}$ is the above $(2^f \times 2^f)$-design matrix of the unreplicated design and $\bar{\mathbf{y}}$ is the 2^f-vector of treatment-specific averages.

Any linear combination of elements of $\boldsymbol{\theta}$ is estimable because $\mathbf{M}$ is square and of full rank, i.e. any $\mathbf{c}$ of dimension 2^f can be expressed as a linear combination of the rows of $\mathbf{M}$.

We just have one nuisance parameter, the intercept μ. The partitioned model equation is

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_2\boldsymbol{\phi} + \boldsymbol{\varepsilon}$$

Since $\mathbf{X}_2$ is orthogonal to $\mathbf{1}$, we have $\mathbf{X}_{2|1} = \mathbf{X}_2$.

This means that „correction for the mean“ or the regression model intercept is automatic in this parameterization. Hence we get

$$\mathcal{I}^{-1} = \frac{1}{n}\mathbf{I}_{2^f-1} \qquad \hat{\boldsymbol{\theta}} = \frac{1}{n}\mathbf{X}_2^T\mathbf{y}$$

11.4 Estimation of treatment contrasts

11.4.1 Full model

We order the treatment mean vector and the corresponding data average vector lexicographically by their indexes:

$$\mathbf{y}^T = \begin{pmatrix} y_{1\ldots11} & y_{1\ldots12} & y_{1\ldots21} & y_{1\ldots22} & \cdots & y_{2\ldots22} \end{pmatrix}$$

and $\mathsf{E}(\bar{\mathbf{y}}) = \boldsymbol{\mu} = \mathbf{M}\boldsymbol{\theta}$. $\mathbf{M}$ is orthogonal, so there is a one-to-one relationship between $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$.

We can unbiasedly estimate any linear combinations of the elements of $\boldsymbol{\mu}$

$$\widehat{\mathbf{c}^T\boldsymbol{\mu}} = \widehat{\mathbf{c}^T\mathbf{M}\boldsymbol{\theta}} = \mathbf{c}^T\mathbf{M}\hat{\boldsymbol{\theta}} = \mathbf{c}^T\mathbf{M}\,2^{-f}\mathbf{M}^T\bar{\mathbf{y}} = \mathbf{c}^T\bar{\mathbf{y}}$$

with

$$\mathsf{E}(\mathbf{c}^T\bar{\mathbf{y}}) = \mathbf{c}^T\boldsymbol{\mu} \qquad \text{and} \qquad \mathsf{Var}(\mathbf{c}^T\bar{\mathbf{y}}) = \frac{\sigma^2}{r}\,\mathbf{c}^T\mathbf{c}$$

11.4.2 Reduced model

Return to the $f = 3$ example, and suppose we assume that interactions involving γ don't exist, i.e.

$$(\alpha\gamma) = (\beta\gamma) = (\alpha\beta\gamma) = \mathbf{0}$$

We partition the model:

$$\boldsymbol{\mu} = \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2$$

where $\boldsymbol{\theta}_2$ is the p_2-vector of parameters assumed to be zero and $\boldsymbol{\theta}_1$ contains the p_1 parameters parameters remaining in the model.

If our assumptions are correct, we have

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\theta}_1 + \boldsymbol{\varepsilon} \qquad \text{and} \qquad \hat{\boldsymbol{\theta}}_1 = \frac{1}{n}\mathbf{X}_1^T\mathbf{y} = 2^{-f}\mathbf{M}_1^T\bar{\mathbf{y}}$$

where $\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 \end{pmatrix}$

Reduced model (cont.)

According to our assumptions

$$\widehat{\mathbf{c}^T\boldsymbol{\mu}} = \mathbf{c}^T\widehat{\mathbf{M}_1\boldsymbol{\theta}_1} = \mathbf{c}^T\mathbf{M}_1\hat{\boldsymbol{\theta}}_1 = 2^{-f}\mathbf{c}^T\mathbf{M}_1\mathbf{M}_1^T\bar{\mathbf{y}}$$

this is not necessarily the same thing as $\mathbf{c}^T\bar{\mathbf{y}}$, as follows when the full model is used.

We only have $\mathsf{E}(\widehat{\mathbf{c}^T\boldsymbol{\mu}}) = \mathbf{c}^T\boldsymbol{\mu}$ if the reduced model is correct. Otherwise we get a biased estimate!

On the other hand the variance of the reduced model estimate is

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\mu}}) = \mathbf{c}^T\mathbf{M}_1\mathsf{Var}(\hat{\boldsymbol{\theta}}_1)\mathbf{M}_1^T\mathbf{c} = \frac{\sigma^2}{r}2^{-f}\,\mathbf{c}^T\mathbf{M}_1\mathbf{M}_1^T\mathbf{c}$$

which in any case is not bigger than the variance of the full model estimate.

So estimates based on a reduced model have variance no greater than, and sometimes less than, estimates of the same quantity derived under the full model.

11.4.3 Examples

The following examples demonstrate the variance reduction of different estimates using a reduced model. We have to compare $\mathbf{c}^T\mathbf{M}_1\mathbf{M}_1^T\mathbf{c}$ (reduced model) with $\mathbf{c}^T\mathbf{M}\mathbf{M}^T\mathbf{c}$ (full model).

Single cell mean:

$\mathbf{c}^T\boldsymbol{\mu}$ is a single element of $\boldsymbol{\mu}$, i.e. $\mathbf{c}^T = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$

$$\mathbf{c}^T\mathbf{M}_1\mathbf{M}_1^T\mathbf{c} = p_1 \qquad \mathbf{c}^T\mathbf{M}\mathbf{M}^T\mathbf{c} = 2^f \qquad \text{variance reduction: } \frac{p_1}{2^f}$$

Treatment contrast corresponding to a reduced model effect:

$\mathbf{c}$ is a column of $\mathbf{M}_1$ and $\mathbf{c}^T\mathbf{M}_1 = \begin{pmatrix} 0 & \cdots & 0 & 2^f & 0 & \cdots & 0 \end{pmatrix}$

$$\mathbf{c}^T\mathbf{M}_1\mathbf{M}_1^T\mathbf{c} = 4^f \qquad \mathbf{c}^T\mathbf{M}\mathbf{M}^T\mathbf{c} = 4^f \qquad \text{variance reduction: } 1$$

Examples (cont.)

Treatment contrast corresponding to an effect removed from the full model:

$\mathbf{c}$ is a column of $\mathbf{M}_2$ and $\mathbf{c}^T\mathbf{M}_1 = \mathbf{0}$, i.e. the sample variance under the reduced model is zero!

$$\mathbf{c}^T\mathbf{M}_1\mathbf{M}_1^T\mathbf{c} = 0 \qquad \mathbf{c}^T\mathbf{M}\mathbf{M}^T\mathbf{c} = 4^f \qquad \text{variance reduction: } 0$$

The potential benefit of using a reduced model is improved precision associated with some treatment contrasts - depending on the contrast of interest.

But there is also risk in adopting a reduced model, specifically that the estimates will be biased and tests potentially invalid if the omitted effects are actually present.

11.5 Testing factorial effects

11.5.1 Individual model terms, experiments with replication

Suppose we wish to test

$$H_0 : (\diamondsuit) = 0$$

where $(\diamondsuit)$ represents any factorial effect, e.g., α or $(\alpha\beta\gamma)$.

The least square estimate of $(\diamondsuit)$ and its variance are

$$\widehat{(\diamondsuit)} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{m}_{(\diamondsuit)}^T\bar{\mathbf{y}} = 2^{-f}\,\mathbf{m}_{(\diamondsuit)}^T\bar{\mathbf{y}}$$

$$\mathsf{Var}\left(\widehat{(\diamondsuit)}\right) = 2^{-2f}\mathbf{m}_{(\diamondsuit)}^T\,\mathsf{Var}(\bar{\mathbf{y}})\,\mathbf{m}_{(\diamondsuit)} = 2^{-2f}\frac{\sigma^2}{r}\,\mathbf{m}_{(\diamondsuit)}^T\mathbf{m}_{(\diamondsuit)} = \frac{\sigma^2}{n}$$

where $\mathbf{m}_{(\diamondsuit)}$ is the column from $\mathbf{M}$ that corresponds to the parameter $(\diamondsuit)$.

Individual model terms, experiments with replication

The unbiased estimate of σ^2 is the mean squared error (MSE)

$$\hat{\sigma}^2 = s^2_{\text{pooled}} = \sum_{ij\ldots} \frac{s^2_{ij\ldots}}{2^f}$$

where $s^2_{ij\ldots}$ is the sample variance of the r data values collected under the treatment identified by $(ij\ldots)$. $\hat{\sigma}^2$ and $\widehat{(\diamondsuit)}$ are independent, so under H_0 the test statistic

$$\mathsf{t} = \widehat{(\diamondsuit)} \sqrt{\frac{n}{s^2_{\text{pooled}}}}$$

follows a t-distribution with $2^f(r-1)$ degrees of freedom.

11.5.2 Multiple model terms, experiments with replication

Suppose we wish to test $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$ or equivalently

$$H_0 : \boldsymbol{\mu} = \mathbf{X}_1\boldsymbol{\theta}_1 \qquad \text{against} \qquad H_1 : \boldsymbol{\mu} = \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2$$

where the subvectors $\boldsymbol{\theta}_i$ and the submatrices $\mathbf{X}_i$ correspond to the partitioned model above. We have to test all the parameters in $\boldsymbol{\theta}_2$ simultaneously, i.e. we have to perform an F-test.

For the full-rank factorial model with $\pm$ „coding“ the residual sum of squares are

$$\mathsf{SSE}_{H_1} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}}) = \mathbf{y}^T\mathbf{y} - r\,2^f\,\hat{\boldsymbol{\theta}}^T\hat{\boldsymbol{\theta}}$$

So the treatment sum of squares associated with $\boldsymbol{\theta}_2$ are

$$\begin{aligned}\mathsf{SST}_{\boldsymbol{\theta}_2} &= \mathsf{SSE}_{H_0} - \mathsf{SSE}_{H_1} = \mathbf{y}^T\mathbf{y} - n\,\hat{\boldsymbol{\theta}}_1^T\hat{\boldsymbol{\theta}}_1 - \mathbf{y}^T\mathbf{y} + n\,\hat{\boldsymbol{\theta}}^T\hat{\boldsymbol{\theta}} \\ &= n(\hat{\boldsymbol{\theta}}^T\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_1^T\hat{\boldsymbol{\theta}}_1)\end{aligned}$$

Multiple model terms, experiments with replication (cont.)

Since the columns of $\mathbf{M}$ are orthogonal, *in this case* omitting some parameters from the model does not change the estimates of the remaining parameters. So we have

$$\mathsf{SST}_{\boldsymbol{\theta}_2} = n(\hat{\boldsymbol{\theta}}^T\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_1^T\hat{\boldsymbol{\theta}}_1) = n\,\hat{\boldsymbol{\theta}}_2^T\hat{\boldsymbol{\theta}}_2$$

With $\mathsf{df}_{\boldsymbol{\theta}_2} = p_2$ we can compute $\mathsf{MST}_{\boldsymbol{\theta}_2}$ which is the denominator of the F-test-statistic.

$$\mathsf{F} = \frac{\frac{n\,\hat{\boldsymbol{\theta}}_2^T\hat{\boldsymbol{\theta}}_2}{p_2}}{\frac{\mathsf{SSE}_{H_1}}{2^f(r-1)}} \qquad \text{under } H_0 \text{ we have } \mathsf{F} \sim \mathsf{F}\left(p_2; 2^f(r-1)\right)$$

Because the information matrix for any subset of model parameters is $\mathcal{I} = n\mathbf{I}$, the noncentrality parameter for the test is $\lambda = \frac{n}{\sigma^2}\hat{\boldsymbol{\theta}}_2^T\hat{\boldsymbol{\theta}}_2$ and we may easily compute the power of the test.

11.5.3 Experiments without replication

If the full model is correct and ε is normally distributed, the effect estimates are independently distributed as:

$$\widehat{(\diamondsuit)} \sim \mathbf{N}\left((\diamondsuit); \frac{\sigma^2}{n}\right)$$

The estimates are approximately normal distributed even if ε is not normal.

If *most* effects are actually zero (*effect sparsity*), procedures developed to detect outliers can be used to identify the relatively few effects that appear to be „real“.

Graphical procedures like *normal plots* and *half-normal plots* of the effect estimates are used to detect nonzero effects as „outliers“.

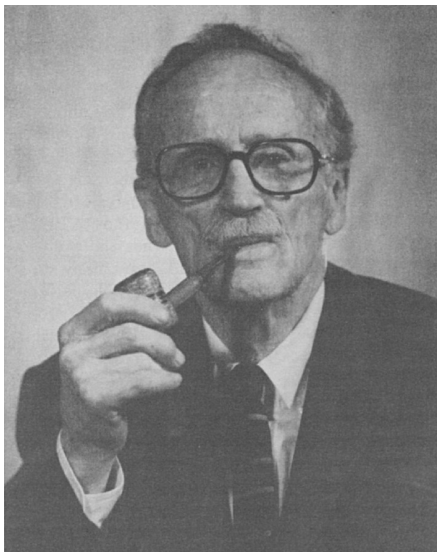
Graphical methods

For any group of effects, $(\diamondsuit)_i$, $i = 1, \ldots p$ (usually excluding μ) order the corresponding estimates from least to greatest, and refer to them as
$\widehat{(\diamondsuit)}_{[1]} \leq \widehat{(\diamondsuit)}_{[2]} \leq \cdots \leq \widehat{(\diamondsuit)}_{[p]}$

Then a normal plot is constructed by plotting the $\widehat{(\diamondsuit)}_{[i]}$ versus quantiles from the standard normal distribution, $\Phi^{-1}\left(\frac{i-0.5}{p}\right); i = 1, \ldots, p$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cdf.

Half-normal plots introduced by Daniel (1959), are constructed by plotting the sorted absolute values of effect estimates, $|\widehat{(\diamondsuit)}|_{[i]}$, versus quantiles from the „positive half“ of the standard normal distribution, $\Phi^{-1}\left(\frac{1}{2} + \frac{i-0.5}{2p}\right); i = 1, \ldots, p$.

If all parameter estimates actually have expectation zero, the plotted points should lie approximately along a straight line. Any „real“ effects (those that are not zero) tend to appear as „outliers“.

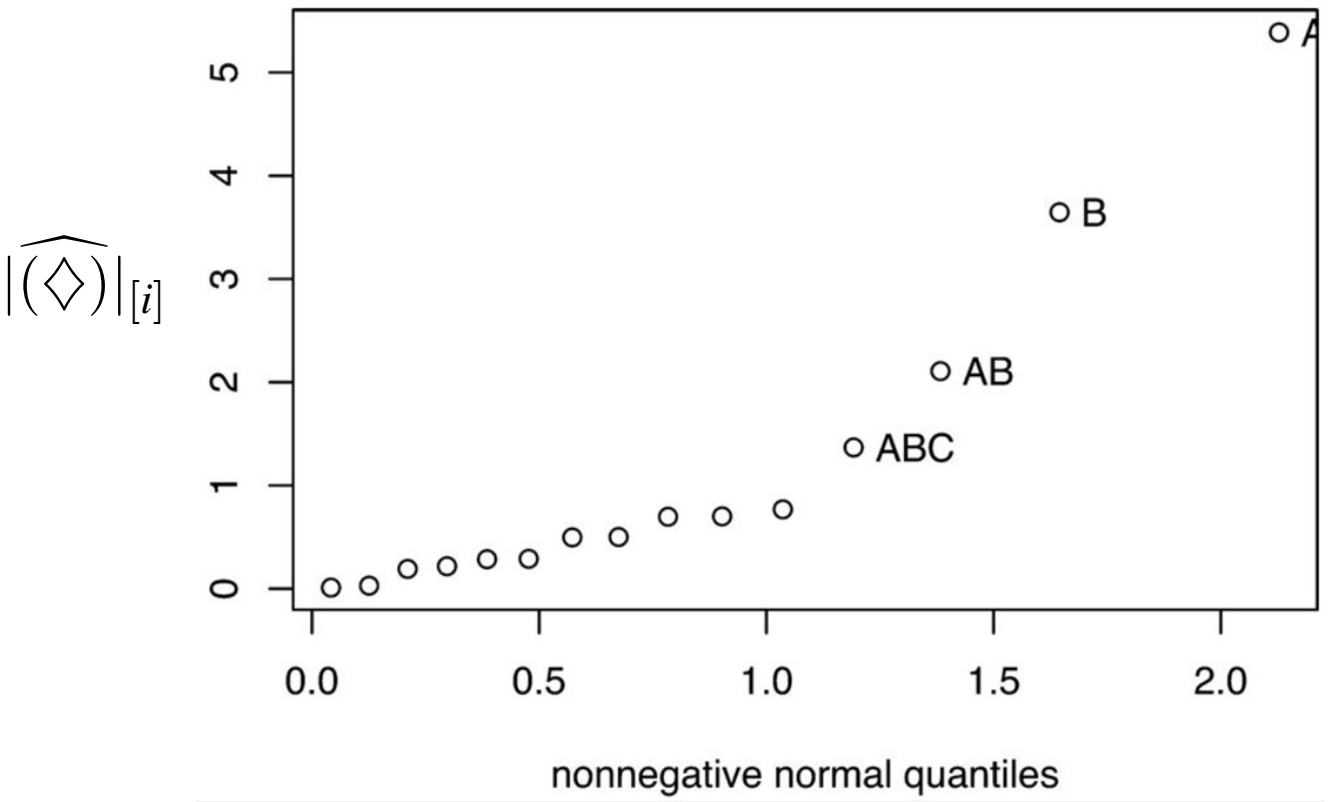


Cuthbert Daniel (1904 - 1997)

Graphical methods (cont.)

To demonstrate the use of a half-normal plot, data were simulated for an unreplicated 2^4 factorial experiment, where all factorial effects are actually zero except:

$$\alpha = 6 \quad \beta = -4 \quad (\alpha\beta) = 2 \qquad \text{and } \sigma = 3$$



Lenth's method

Lenth proposed an algorithm for performing the half-normal plot analysis in a more automatic and objective way:

- let $\mathcal{B}$ the set of absolute values of estimated coefficients of interest
- compute $s_0 = 1.5 \cdot \mathsf{median}(\mathcal{B})$ as an initial robust estimate of $\frac{\sigma}{\sqrt{n}}$
- let $\mathcal{B}^* = \{|\hat{\theta}| \in \mathcal{B} : |\hat{\theta}| < 2.5 \cdot s_0\}$, i.e. remove the clearly nonzero elements from $\mathcal{B}$
- compute the so-called „pseudo standard error" $\mathsf{PSE} = 1.5 \cdot \mathsf{median}(\mathcal{B}^*)$ as a refined estimate of $\frac{\sigma}{\sqrt{n}}$
- any estimated effect is significant if it is greater than $t \cdot \mathsf{PSE}$

Russ Lenth (*1948)

A table of critical values of t for $\alpha = 0.05$ and $\mathcal{B}$ of at least moderate size was published by Lenth 1989. t is between 2 and 2.5.

11.6 Additional guidelines for model editing

Half-normal plots and Lenth's method can help to find a reduced model form that conforms well to the data.

However, sometimes they lead to a suggested model that „makes no sense“: An example would be a 2^4 factorial model with only α and $(\beta\gamma\delta)$ as significant parameters.

There are two principles for „meaningful“ models

- *Effect Hierarchy Principle*: If an interaction involving a given set of factors is included in the model, all main effects and interactions involving subsets of these factors should also be included.
- *Effect Heredity Principle*: If an interaction involving a given set of factors is included in the model, at least one effect of the next smallest order involving a subset of these factors should also be included.

12.1 Introduction to Two-level factorial experiments: blocking

If n is so large that consistent experimental control cannot be exerted throughout all runs, or if several batches of experimental material must be used to complete the study, blocking the experiment into a few or several subexperiments may account for substantial uncontrolled variation.

For a two-level blocked experiment a full-rank effects model can be written as:

$$y_{mij\ldots} = \mu + \theta_m + [x_1\boldsymbol{\alpha} + x_2\boldsymbol{\beta} + \cdots + x_1x_2\cdots(\boldsymbol{\alpha\beta}\cdots)] + \varepsilon_{mij\ldots}$$

where θ_m denotes the additive effect associated with block m and $x_i = \pm 1$ indicates that factor i is at its high or low level.

Because the blocks form a partition of the experimental runs, an equivalent form for the full-rank parameterization is:

$$y_{mij\ldots} = \theta_m + [x_1\boldsymbol{\alpha} + x_2\boldsymbol{\beta} + \cdots + x_1x_2\cdots(\boldsymbol{\alpha\beta}\cdots)] + \varepsilon_{mij\ldots}$$

12.1.1 Models

A matrix form for the entire blocked experiment is:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\boldsymbol{\phi} + \boldsymbol{\varepsilon}$$

If each block contains k observations and the elements of $\mathbf{y}$ are ordered by block we have

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_k & \mathbf{0}_k & \cdots & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{1}_k & \cdots & \mathbf{0}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_k & \mathbf{0}_k & \ldots & \mathbf{1}_k \end{pmatrix} \qquad \mathbf{H}_1 = \frac{1}{k}\begin{pmatrix} \mathbf{J}_k & \mathbf{O}_k & \cdots & \mathbf{O}_k \\ \mathbf{O}_k & \mathbf{J}_k & \cdots & \mathbf{O}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_k & \mathbf{O}_k & \ldots & \mathbf{J}_k \end{pmatrix}$$

In a CBD $k = 2^f$ and the rows of $\mathbf{X}_2$ corresponding to each k-row section of $\mathbf{X}_1$ „code for“ all 2^f factorial treatments.

12.1.2 Notation

For two-level experiments a simpler notation can be used.

Factors are often denoted by upper-case letters: A, B, ... The corresponding columns of the design matrix are sometimes denoted the same way, e.g., ABC can refer to the column vector of $x_1x_2x_3$ values, the „regressor“ associated with the parameter $(\alpha\beta\gamma)$. The upper-case „I“, the identity element is used to represent the column of ones associated with μ.

A treatment is sometimes designated by listing lower-case letters associated with factors set to level 2, e.g.,

- „ac“ ... the treatment defined by setting factors A and C at level 2, and others at level 1
- „abcd“ ... the treatment defined by setting factors A to D at level 2, and others at level 1
- „(1)“ ... the treatment defined by setting all factors at level 1

12.2 Complete blocks

The most straightforward form of blocking in 2f experimentation is the CBD.

E.g. a CBD for $f = 2$ factors, arranged in r complete blocks, can be depicted as beside.

block 1	block 2	...	block r
(1)	(1)		(1)
a	a	...	a
b	b		b
ab	ab		ab

Let $\mathbf{M}_2$ be the $\left(2^f \times (2^f - 1)\right)$-design matrix associated with just the factorial effects (i.e., excluding μ) for an unreplicated 2^f experiment. If the elements of $\mathbf{y}$ are ordered by block $\mathbf{X}_2$ can be written as:

$$\mathbf{X}_2 = \begin{pmatrix} \mathbf{M}_2 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_2 \end{pmatrix} \qquad \text{With } \mathbf{H}_1 = \frac{1}{2^f} \begin{pmatrix} \mathbf{J}_{2^f} & \mathbf{O}_{2^f} & \cdots & \mathbf{O}_{2^f} \\ \mathbf{O}_{2^f} & \mathbf{J}_{2^f} & \cdots & \mathbf{O}_{2^f} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{2^f} & \mathbf{O}_{2^f} & \cdots & \mathbf{J}_{2^f} \end{pmatrix}$$

we get $\qquad \mathbf{X}_{2|1} = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2 = \mathbf{X}_2 - \mathbf{O} = \mathbf{X}_2$

Complete blocks (cont.)

$\mathbf{X}_{2|1} = \mathbf{X}_2$ because the sum of each column of $\mathbf{M}_2$ is zero.

Hence the estimates of factorial effects are computed as if the experiment were not blocked.

The standard assumption is that there is no block-by-treatment interaction.

In the ANOVA decomposition the treatment sum of squares SST with $\mathsf{df}_\mathsf{T} = 2^f - 1$ can be further decomposed into one-degree-of-freedom components for each factorial effect, e.g. for a 2^2 CBD with r blocks

$$\mathsf{SST} = \sum_{ij} r\,(\bar{y}_{.ij} - \bar{y}_{...})^2 = n\left(\hat{\alpha}^2 + \hat{\beta}^2 + \widehat{(\alpha\beta)}^2\right) \qquad \text{with } \mathsf{df}_\mathsf{T} = 3$$

the component associated with factor A is $\mathsf{SST}_\mathrm{A} = n\,\hat{\alpha}^2$ with $\mathsf{df}_\mathrm{A} = 1$.

12.3 Balanced incomplete block designs (BIBDs)

BIBD structures can also be taken in constructing blocked factorial designs.

See a BIBD with $t = 2^2$ treatments, organized in $b = 4$ blocks, each of size $k = 3$ beside.

block 1	block 2	block 3	block 4
(1)	(1)	(1)	a
a	a	b	b
b	ab	ab	ab

Analysis of data from a factorial experiment organized as a BIBD follows easily from the results of unstructured BIBDs. The factorial representation is just a reparameterization of the models we would use if we ignored the factorial structure of the experiment.

For example, we can arbitrarily assign the 2^2 treatments in the above experiment to:

treatment no.	1	2	3	4
„new notation“	(1)	a	b	ab

Balanced incomplete block designs (cont.)

The parameters of the treatment effects in either notation correspond to:

unstructured		factorial
τ_1	=	$-\alpha - \beta + (\alpha\beta)$
τ_2	=	$+\alpha - \beta - (\alpha\beta)$
τ_3	=	$-\alpha + \beta - (\alpha\beta)$
τ_4	=	$+\alpha + \beta + (\alpha\beta)$

The solution of this set of linear equations for α, β and $(\alpha\beta)$ is left to the reader.

The most commonly used blocking technique in 2^f experiments is fundamentally different.

In most factorial experiment settings, some factorial effects (e.g., main effects and low-order interactions) are of substantially more interest than others. The result is called „*regular* blocking scheme“.

12.4 Regular blocks of size 2^{f-1}

We start with „half-replicate“-blocking, i.e. one half of all possible treatments are included in each block of size 2^{f-1} so that each treatment appears exactly once in each *pair* of blocks.

Blocks are arranged so that for a selected effect $(\diamondsuit)$

- $(\diamondsuit)$ is multiplied by + in the effects model representation for all treatments included in one block, and
- $(\diamondsuit)$ is multiplied by - in the effects model representation for all treatments included in the paired block.

In this case $(\diamondsuit)$ is no longer estimable, so the highest-order interaction is often chosen for $(\diamondsuit)$.

We say $(\diamondsuit)$ is *confounded with* the block contrast. Let e.g. $(\diamondsuit)$ be the factorial effect ABC. Those treatments for which the entry in the ABC column is +1 are assigned in one block, while those for which it is -1 are placed in the other. Then the parameter $(\alpha\beta\gamma)$ is „confounded with“ the block contrast $\theta_2 - \theta_1$

Regular blocks of size 2^{f-1} (cont.)

To see how ABC „splits“ the treatments of a 2^3 experiments into blocks we list the treatments explicitly

treatment	I	A	B	C	AB	AC	BC	ABC	blocks
(1)	+	−	−	−	+	+	+	−	1
a	+	+	−	−	−	−	+	+	2
b	+	−	+	−	−	+	−	+	2
c	+	−	−	+	+	−	−	+	2
ab	+	+	+	−	+	−	−	−	1
ac	+	+	−	+	−	+	−	−	1
bc	+	−	+	+	−	−	+	−	1
abc	+	+	+	+	+	+	+	+	2

The paired blocks then have the treatments assigned as beside.

Other treatment estimates and sums of squares are unchanged by the introduction of blocks.

block 1	block 2
(1)	a
ab	b
ac	c
bc	abc

Regular blocks of size 2^{f-1} (cont.)

The matrix form of the model with r replicates, i.e. $2r$ blocks, is very similar to the complete block situation (slide 24).

We just have to partition the $\mathbf{M}_2$ matrix: with

$$\mathbf{M}_2 = \begin{pmatrix} \mathbf{M}_{2,1} \\ \mathbf{M}_{2,2} \end{pmatrix} \quad \text{we get} \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{M}_{2,1} \\ \mathbf{M}_{2,2} \\ \mathbf{M}_{2,1} \\ \mathbf{M}_{2,2} \\ \vdots \\ \mathbf{M}_{2,1} \\ \mathbf{M}_{2,2} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{2^f-1} \end{pmatrix}$$

$\mathbf{M}_{2,1}$ and $\mathbf{M}_{2,2}$ each have columns with zero sums except for the column corresponding to the confounded effect.

Regular blocks of size 2^{f-1} (cont.)

The $\mathbf{H}_1$ matrix has now four times as many blocks as in the complete block situation and the block matrix is divided by 2^{f-1} instead of 2^f.
Each of the $\mathbf{O}_{2^f}$-blocks is quartered into four blocks of $\mathbf{O}_{2^{f-1}}$ and the $\mathbf{J}_{2^f}$-blocks are quartered as per

$$\mathbf{J}_{2^f} = \begin{pmatrix} \mathbf{J}_{2^{f-1}} & \mathbf{O}_{2^{f-1}} \\ \mathbf{O}_{2^{f-1}} & \mathbf{J}_{2^{f-1}} \end{pmatrix}$$

If $\mathbf{x}_{2^f-1}$ is the column in $\mathbf{X}_2$ corresponding to the factorial effect confounded with blocks, we get

$$\mathbf{X}_{2|1} = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2 = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{2^f-2} & 0 \end{pmatrix}$$

For $r = 1$, there are no degrees of freedom for residual sum of squares under the full model. If we apply a reduced model instead the model variance could be estimated again because degrees of freedom are „released“ from the treatment sum of squares.

12.4.1 Random blocks

An experiment divided into regular blocks can be analyzed as a split-plot experiment, with levels of ABC compared between blocks, and other factorial effects compared within blocks.

Here ABC is confounded with blocks in each replicate.

12.4.2 Partial confounding

Split-plot designs allow fully efficient estimation of all factorial effects *except* for the effect selected for confounding, for which formal inferences cannot be made if block effects are fixed, and for which generally less informative whole-plot inferences can be made if block effects are random.

If we cannot treat block effects as random and need to make inference about all effects in blocks of size 2^{f-1} we should try *partial confounding*.

Partial confounding (cont.)

Different factorial effects are confounded with blocks in each complete replicate. For example, with 2^3 treatments and blocks of size 4, in the first replicate the blocks might be confounded with the interaction effect ABC, in the second replicate the blocks might be confounded with the interaction effect BC and so on.

Estimates and sums of squares for nonconfounded effects are computed as usual, e.g., $n\,\hat{\alpha}^2$ if A is not selected as the confounding effect *in any* replicate.

Estimates and sums of squares for confounded effects are computed *using only* data from replicates in which they are *not confounded*.

12.5 Regular blocks of size 2^{f-2}

When blocks of size 2^{f-1} are too large, we split up these blocks into blocks of size 2^{f-2} by selecting a second factorial effect to confound with blocks.

Continuing the 2^3-design example above where ABC was chosen to generate the first treatment split, now add BC for the second. ABC and BC are confounded with blocks, so $(\alpha\beta\gamma)$ and $(\beta\gamma)$ are not estimable. So he degrees of freedom for the model is reduced by 2.

The $2^3 = 8$ treatments have to be distributed to 4 blocks of size 2. This means we need 3 additional degrees of freedom for the blocks.

This also means that a third factorial effect must be confounded with blocks too, i.e. the values of a third effect must also be constant within the blocks.

In this case it is the factorial effect A, which is the so-called *generalized interaction* between the confounded effects ABC and BC.

Regular blocks of size 2^{f-2} (cont.)

The effects ABC, BC and hence also A split the treatments into 4 blocks:

treatment	I	A	B	C	AB	AC	BC	ABC	blocks
(1)	+	−	−	−	+	+	+	−	1
a	+	+	−	−	−	−	+	+	2
b	+	−	+	−	−	+	−	+	3
c	+	−	−	+	+	−	−	+	3
ab	+	+	+	−	+	−	−	−	4
ac	+	+	−	+	−	+	−	−	4
bc	+	−	+	+	−	−	+	−	1
abc	+	+	+	+	+	+	+	+	2

block 1	block 2	block 3	block 4
(-)	a	b	ab
bc	abc	c	ac

A better choice for the effects confounded with blocks would be AB, AC and BC because then no main effect would be „sacrificed“.

12.6 Regular blocks: general case

The ideas already introduced for blocks of size 2^{f-1} and 2^{f-2} can be generalized to construct designs with smaller blocks by sequentially re-splitting treatments into groups s times, to obtain 2^s blocks of 2^{f-s} units each per replicate.

We can think in terms of the s factorial effects „independently“ selected to confound with blocks, one corresponding to each of the sequence of „splits“ of the 2^f treatments, with the understanding that new generalized interactions are also confounded with blocks at each split after the first.

As blocking schemes become more complicated, it becomes less obvious which treatments are applied to units in each block.

Model, parameter estimation and hypotheses testing is in line with regular blocks of size 2^{f-1}.