Experimental Design

Unit 5

Werner Müller

Institute of Applied Statistics

May 8th 2024

8.1 Introduction to random block effects

Up to now block effects have been seen as unknown, fixed model parameters. But there are also experimental situations in which it is reasonable to think of block effects as random, suggesting that a mixed effects model may be more appropriate.

If the random block effects assumption is reasonable, it can lead to additional analysis options.

In some designs, such as the split-plot designs (see later), a full analysis of experimental treatments cannot be made unless block effects can be treated as random.

8.2 Inter- and intra-block analysis

Suppose an experiment with b blocks of equal size k (so $n = b\,k$) of the form:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon} \qquad \mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \qquad \mathsf{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

- $\boldsymbol{\beta}$ is the b-vector of random effects with $\mathsf{E}(\boldsymbol{\beta}) = \mu_\beta\mathbf{1}$ and $\mathsf{Var}(\boldsymbol{\beta}) = \sigma^2_\beta\mathbf{I}$
-

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_k & \mathbf{0}_k & \cdots & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{1}_k & \cdots & \mathbf{0}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_k & \mathbf{0}_k & \cdots & \mathbf{1}_k \end{pmatrix}$$

If $\boldsymbol{\beta}$ is independent of $\boldsymbol{\varepsilon}$, it follows

$$\mathsf{E}(\mathbf{y}) = \mu_\beta\mathbf{1} + \mathbf{X}_2\boldsymbol{\tau} \qquad \mathsf{Var}(\mathbf{y}) = \sigma^2_\beta\mathbf{X}_1\mathbf{X}_1^T + \sigma^2\mathbf{I}$$

Inter- and intra-block analysis (cont.)

$\mathsf{Var}(\mathbf{y}) = \sigma_\beta^2 \mathbf{X}_1 \mathbf{X}_1^T + \sigma^2 \mathbf{I}$:

- each observation has the same variance $\sigma_\beta^2 + \sigma^2$
- pairs of observations associated with the same block have covariance σ_β^2

So the best linear unbiased estimator of $\boldsymbol{\tau}$ is the *generalized least-squares estimate*, i.e. any solution of the normal equations:

$$\mathbf{X}_2^T (\sigma_\beta^2 \mathbf{X}_1 \mathbf{X}_1^T + \sigma^2 \mathbf{I})^{-1} \mathbf{X}_2 \hat{\boldsymbol{\tau}} = \mathbf{X}_2^T (\sigma_\beta^2 \mathbf{X}_1 \mathbf{X}_1^T + \sigma^2 \mathbf{I})^{-1} \mathbf{y}$$

with $\rho = \frac{\sigma_\beta}{\sigma}$ we get

$$\mathbf{X}_2^T (\rho^2 \mathbf{X}_1 \mathbf{X}_1^T + \mathbf{I})^{-1} \mathbf{X}_2 \hat{\boldsymbol{\tau}} = \mathbf{X}_2^T (\rho^2 \mathbf{X}_1 \mathbf{X}_1^T + \mathbf{I})^{-1} \mathbf{y}$$

ρ is not known, so $\hat{\boldsymbol{\tau}}$ can only be determined using *iteratively reweighted least-squares* procedures:

- $\boldsymbol{\tau}$ is estimated using the normal equations with an estimated value of ρ^2 in place of the true variance ratio.
- ρ is estimated treating an estimated value of $\boldsymbol{\tau}$ as the true parameter vector.

Inter- and intra-block analysis (cont.)

We now generate two uncorrelated linear transformations of the data vector $\mathbf{y}$

$$\mathbf{y}_1 = \mathbf{U}^T\mathbf{y} \qquad\qquad \mathbf{y}_2 = \mathbf{X}_1^T\mathbf{y}$$

with $\mathbf{U}$ being any $(n \times l)$ matrix with arbitrary l such that $\mathbf{X}_1^T\mathbf{U} = \mathbf{0}$. $\mathbf{y}_1$ is an l-element vector.
Note that $\mathbf{y}_2$ is the b-element vector of block totals.

$$\mathbf{y}_1 = \mathbf{U}^T\mathbf{X}_1\boldsymbol{\beta} + \mathbf{U}^T\mathbf{X}_2\boldsymbol{\tau} + \mathbf{U}^T\boldsymbol{\varepsilon} = \mathbf{U}^T\mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}_1$$

$$\text{with} \qquad \mathsf{E}(\boldsymbol{\varepsilon}_1) = \mathbf{0} \qquad \text{and} \qquad \mathsf{Var}(\boldsymbol{\varepsilon}_1) = \sigma^2\mathbf{U}^T\mathbf{U}$$

$$\mathbf{y}_2 = \mathbf{X}_1^T\mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_1^T\mathbf{X}_2\boldsymbol{\tau} + \mathbf{X}_1^T\boldsymbol{\varepsilon} = k\,\mu_\beta\mathbf{1} + \mathbf{X}_1^T\mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}_2$$

$$\text{with} \qquad \mathsf{E}(\boldsymbol{\varepsilon}_2) = \mathbf{0} \qquad \text{and} \qquad \mathsf{Var}(\boldsymbol{\varepsilon}_2) = (k^2\sigma_\beta^2 + k\sigma^2)\mathbf{I}$$

Hence the transformed data can be represented as two linear models, each containing only a single random element.

Inter- and intra-block analysis (cont.)

$\mathbf{y}_1$ and $\mathbf{y}_2$ are uncorrelated:

$$\mathsf{Cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{U}^T(\sigma^2_\beta \mathbf{X}_1\mathbf{X}_1^T + \sigma^2\mathbf{I})\mathbf{X}_1 = \mathbf{0}$$

so if all random elements are normally distributed, analyses based on the two transformed models are statistically independent.

With this structure, the treatments should be assigned such that the interesting linear combinations $\mathbf{c}^T\boldsymbol{\tau}$ use a vector $\mathbf{c}$ with

- $\mathbf{c} = \mathbf{c}_1^T(\mathbf{U}^T\mathbf{X}_2)$, with an l-vector $\mathbf{c}_1$ so that $\mathbf{c}^T\tau$ is estimable based on the analysis of $\mathbf{y}_1$, or
- $\mathbf{c} = \mathbf{c}_2^T(\mathbf{I}_b - \frac{1}{b}\mathbf{J}_b)(\mathbf{X}_1^T\mathbf{X}_2)$ with a b-vector $\mathbf{c}_2$ so that $\mathbf{c}^T\tau$ is estimable based on the analysis of $\mathbf{y}_2$.

Inter- and intra-block analysis (cont.)

$\mathbf{U}$ should be such that $\mathbf{X}_1^T\mathbf{U} = \mathbf{0}$, so one possible choice is $\mathbf{U} = \mathbf{I} - \mathbf{H}_1$, the projection matrix associated with the compliment of the column space of $\mathbf{X}_1$. With this choice the model for $\mathbf{y}_1$ can be rewritten as:

$$\mathbf{y}_1 = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}_1 \qquad \mathsf{Var}(\boldsymbol{\varepsilon}_1) = \sigma^2(\mathbf{I} - \mathbf{H}_1)$$

Analysis of this model leads to normal equations of the form:

$$\mathbf{X}_2^T(\mathbf{I}-\mathbf{H}_1)\left(\sigma^2(\mathbf{I}-\mathbf{H}_1)\right)^-(\mathbf{I}-\mathbf{H}_1)\mathbf{X}_2\hat{\boldsymbol{\tau}}_{intra} = \mathbf{X}_2^T(\mathbf{I}-\mathbf{H}_1)\left(\sigma^2(\mathbf{I}-\mathbf{H}_1)\right)^-\mathbf{y}_1$$

eliminating σ^{-2} from each side and noting that the identity matrix is a generalized inverse of $(\mathbf{I} - \mathbf{H}_1)$ we may rewrite the reduced normal equations as:

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

This is exactly the same reduced normal equations we have seen for the fixed-block scenario.

Inter- and intra-block analysis (cont.)

$\hat{\boldsymbol{\tau}}_{intra}$ is called the *intra-block estimate* of $\boldsymbol{\tau}$, because it relies on the data only through linear combinations that are contrasts within each block (since $\mathbf{U}^T\mathbf{X}_1 = \mathbf{0}$).

The model for $\mathbf{y}_2$ can be written as:

$$\mathbf{y}_2 = k\,\mu_\beta \mathbf{1} + \mathbf{X}_1^T\mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}_2 \qquad \mathsf{Var}(\boldsymbol{\varepsilon}_2) = k(k\sigma_\beta^2 + \sigma^2)\mathbf{I}$$

Analysis of this model is a regression of the vector of block totals $\mathbf{y}_2$ on a design matrix $\left(k\,\mathbf{1}_n | \mathbf{X}_1^T\mathbf{X}_2\right)$ and a parameter vector $\left(\mu_\beta, \boldsymbol{\tau}^T\right)^T$, resulting in the *inter-block estimate* $\hat{\boldsymbol{\tau}}_{inter}$ (via reduced normal equations corrected for μ_β):

$$\mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{I}_b - \frac{1}{b}\mathbf{J}_b\right)\mathbf{X}_1^T\mathbf{X}_2\hat{\boldsymbol{\tau}}_{inter} = \mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{I}_b - \frac{1}{b}\mathbf{J}_b\right)\mathbf{y}_2$$

Inter- and intra-block analysis (cont.)

The analysis is based on two different statistical models, so we can think of the design as having two different design information matrices. The intra-block analysis information matrix is as we have previously defined it for fixed-block analysis:

$$\mathcal{I}_{intra} = \mathbf{X}_2^T(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2$$

The „left side“ matrix of the inter-block reduced normal equations is

$$\mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{I}_b - \frac{1}{b}\mathbf{J}_b\right)\mathbf{X}_1^T\mathbf{X}_2$$

which might be regarded as the design information matrix for this analysis. The variance of estimable functions $\widehat{\mathbf{c}^T\boldsymbol{\tau}}$ and noncentrality parameters each depend on both the information matrix *and* a variance, specifically

$$\mathsf{Var}(\boldsymbol{\tau}) = \sigma^2\mathbf{c}^T\mathcal{I}^-\mathbf{c} \qquad \text{and} \qquad \frac{1}{\sigma^2}\boldsymbol{\tau}^T\mathcal{I}\boldsymbol{\tau}$$

Inter- and intra-block analysis (cont.)

$$\mathsf{Var}(\boldsymbol{\tau}) = \sigma^2 \mathbf{c}^T \mathcal{I}^- \mathbf{c} \qquad \text{and} \qquad \frac{1}{\sigma^2} \boldsymbol{\tau}^T \mathcal{I} \boldsymbol{\tau}$$

Hence simultaneously multiplying $\mathcal{I}$ and also the variance element with $\frac{1}{k}$ doesn't change the result. So we define the inter-block information matrix to be

$$\mathcal{I}_{inter} = \frac{1}{k} \mathbf{X}_2^T \mathbf{X}_1 \left(\mathbf{I}_b - \frac{1}{b} \mathbf{J}_b \right) \mathbf{X}_1^T \mathbf{X}_2$$

relative to the variance element

$$\frac{1}{k} \left(k^2 \sigma_\beta^2 + k\sigma^2 \right) = k\sigma_\beta^2 + \sigma^2$$

This adjustment yields inter- and intra-block variance elements $(k\sigma_\beta^2 + \sigma^2)$ for $\mathbf{y}_2$ and σ^2 for $\mathbf{y}_1$ that are directly comparable, because the coefficient of σ^2 is 1 in each case.

8.3 Complete block designs (CBDs) and augmented CBDs

Consider a randomized CBD, with e.g. $k = t = 4$ treatments in each block and $b = 8$ blocks. The information matrix would be

$$\mathcal{I} = \mathbf{X}_2^T(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2 = b\left(\mathbf{I} - \frac{1}{t}\mathbf{J}\right) = 8\left(\mathbf{I}_4 - \frac{1}{4}\mathbf{J}_4\right)$$

Sometimes it might be reasonable to think of block-to-block differences as being random, i.e., there is a second source of unexplainable random „noise“.

In the above example the intra-block information matrix would be the same as the total information matrix $\mathcal{I}_{intra} = \mathcal{I}$ and there is no additional information due to the inter-blocks model.

CBDs and augmented CBDs (cont.)

The matrix of „regressors“ associated with $\boldsymbol{\tau}$ in the model for block totals (inter-block model), is $\mathbf{X}_1^T\mathbf{X}_2 = \mathbf{J}_{(8\times 4)}$, so that the matrices in both sides of the reduced normal equations:

$$\mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{I}_8 - \frac{1}{8}\mathbf{J}_8\right)\mathbf{X}_1^T\mathbf{X}_2\hat{\boldsymbol{\tau}}_{inter} = \mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{I}_8 - \frac{1}{8}\mathbf{J}_8\right)\mathbf{y}_2$$

have only zero elements. This would always be true for complete block designs in which each treatment is applied to one unit in each block.

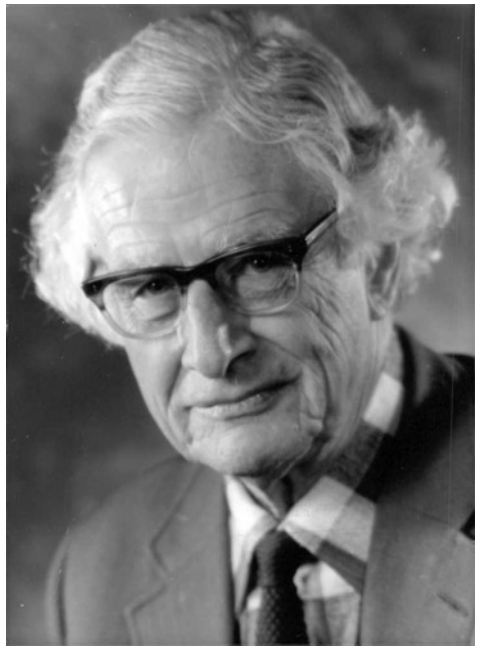
In this case no informative inter-block estimator is available. A minimal requirement for recovery of some inter-block information is that *the pattern of treatment assignments not be the same in each block.*

8.4 Balanced incomplete block designs (BIBDs)

Yates (1940) first demonstrated how inter-block analysis can increase the information available in BIBDs with random blocks.

Inter-block estimates can be of more practical value in BIBDs, when there are many blocks and blocks are small relative to the number of treatments.

The value of an inter-block analysis will always be limited, regardless of the design, if σ_β is large relative to σ.



Frank Yates (1902 - 1994)

Recall that BIBDs are characterized by five related design „parameters“:
b . . . # blocks, t . . . # treatments, $k < t$ . . . # units in each block, r . . . # units allocated to each treatment, and π . . . # blocks in which any two treatments are both applied to units.

BIBDs (cont.)

The intra-block analysis for BIBDs was developed in Chapter 7 (unit 4), the reduced normal equations are:

$$\frac{\pi\, t}{k}\left(\mathbf{I}-\frac{1}{t}\,\mathbf{J}\right)\hat{\boldsymbol{\tau}}_{intra}=\mathbf{T}-\frac{1}{k}\,\mathbf{X}_2^T\mathbf{X}_1\mathbf{B}$$

where $\mathbf{T}$ is the t-vector of treatment totals and $\mathbf{B}$ is the b-vector of block totals. The associated design information matrix is:

$$\mathcal{I}_{intra}=\frac{\pi\, t}{k}\left(\mathbf{I}-\frac{1}{t}\mathbf{J}\right)$$

Inter-block analysis again depends fundamentally on the matrix $\mathbf{X}_1^T\mathbf{X}_2$, which for general BIBDs has

- only elements of 0 and 1
- r ones and $(b-r)$ zeros in each column, and
- π rows in which both entries in any two columns are 1.

BIBDs (cont.)

Some algebra is needed to see that the reduced normal equations for inter-block analysis are

$$(r-\pi)\left(\mathbf{I}-\frac{1}{t}\mathbf{J}\right)\hat{\boldsymbol{\tau}}_{inter}=\mathbf{X}_2^T\mathbf{X}_1(\mathbf{B}-\bar{B}\,\mathbf{1})$$

where $\bar{B}$ is the average block total.

The corresponding design information matrix for the inter-block analysis is

$$\mathcal{I}_{inter}=\frac{1}{k}\mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{I}-\frac{1}{b}\mathbf{J}\right)\mathbf{X}_1^T\mathbf{X}_2=\frac{r-\pi}{k}\left(\mathbf{I}-\frac{1}{t}\mathbf{J}\right)$$

8.5 Combined estimator

Because intra- and inter-block estimators are uncorrelated, a weighted average of the two, with weights proportional to the inverse of their respective variances, is their optimal (minimum variance) linear combination.

$$\widehat{\mathbf{c}^T\boldsymbol{\tau}} = w\,\widehat{\mathbf{c}^T\boldsymbol{\tau}_{intra}} + (1-w)\,\widehat{\mathbf{c}^T\boldsymbol{\tau}_{inter}}$$

To get the weights we have to minimize

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}}) = w^2\,\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}_{intra}}) + (1-w)^2\,\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}_{inter}})$$

with respect to w. For BIBD the estimator variances are proportional to

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}_{intra}}) \propto \sigma^2\frac{k}{\pi\, t} \qquad \text{and} \qquad \mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}_{inter}}) \propto (k\sigma_\beta^2+\sigma^2)\frac{k}{r-\pi}$$

$$\text{Hence we get} \qquad w = \frac{\frac{k\sigma_\beta^2+\sigma^2}{r-\pi}}{\frac{k\sigma_\beta^2+\sigma^2}{r-\pi} + \frac{\sigma^2}{\pi\, t}}$$

Combined estimator (cont.)

In practice the values of the variance components are not known and so the weights have to be estimated.

σ^2 can be estimated unbiasedly by MSE_{intra}, the mean square error for the fixed-block analysis of $\mathbf{y}_1$.

$k\sigma_\beta^2 + \sigma^2$ can be estimated by MSE_{inter}, the mean square error for the regression analysis of block totals $\mathbf{y}_2$, divided by the block size k.

Substituting these estimates for their parameter counterparts leads to the estimated weight:

$$\hat{w} = \frac{\frac{\mathsf{MSE}_{inter}}{r-\pi}}{\frac{\mathsf{MSE}_{inter}}{r-\pi} + \frac{\mathsf{MSE}_{intra}}{\pi\, t}}$$

8.6 Why can information be „recovered“?

Why should an assumption of random block effects permit the recovery of more information about τ?.

An increase in information that leads to decreases in expected standard errors and more powerful tests requires an increase in the strength of assumptions (for a fixed amount of data).

the random-β assumption is actually „stronger“ than a fixed-β assumption. Saying that block effects are „fixed“ is really saying nothing at all about them.

The intra-block analysis is not affected by what the values of β might be because it relies only on linear combinations of data that are contrasts within each block, so that the elements of β „cancel out“ in the expectations of these contrasts.

A random-blocks assumption can lead to more informative inference if it is justified, but to invalid inference if it is not.

8.7 CBD reprise

The assumption of random block effects does not yield additional information with a CBD.

There is an additional advantage to assuming random blocks with an CBD if we add *treatment-by-block* interactions to the model:

$$y_{ij} = \alpha + \beta_i + \tau_j + \theta_{ij} + \varepsilon_{ij}$$

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

$$\beta_i \text{ iid with } \quad \mathsf{E}(\beta_i) = 0 \quad \text{and} \quad \mathsf{Var}(\beta_i) = \sigma^2_\beta$$

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

This model is exactly the same as a random-blocks model without interactions (just apply $\varepsilon^*_{ij} = \theta_{ij} + \varepsilon_{ij}$). So using random-blocks models covers even the situation of treatment-by-block interactions.

9.1 Introduction to factorial treatment structure

Up to now our mathematical description and handling of treatments have included no assumptions about relationships between any pair of them.

Many experiments are designed to compare treatments defined by selecting a *level* related to each of a collection of *factors*.

We describe any experiment in which the treatments have factorial structure as a *factorial experiment*.
In general, if f factors are used to define a treatment, and the ith of these factors has l_i levels, the number of different treatments is $t = \prod_{i=1}^{f} l_i$.

We restrict our attention to situations in which all possible combinations of factor levels are meaningful. In this chapter we will consider the structure of *full factorial experiments*, i.e., those in which all possible combinations of factor levels are represented.

In factorial experiments, the most interesting experimental questions are framed in a way that do not treat relationships between treatments symmetrically.

9.2 An overparameterized model

It is possible to frame the analysis of data from a factorial experiment within the context of unstructured linear models.
Let (i, j, k) be a set of 3 indices for 3 factors representing the levels of each factor, i.e. $i = 1, 2, \ldots, l_1; j = 1, 2, \ldots, l_2$ and $k = 1, 2, \ldots, l_f$. Then, a cell means model and an effects model for an unblocked experiment in which each treatment is replicated r times ($t = 1, 2, \ldots, r$) can be written as:

$$y_{ijkt} = \mu_{ijk} + \varepsilon_{ijkt} \qquad \text{or} \qquad y_{ijkt} = \alpha + \tau_{ijk} + \varepsilon_{ijkt}$$

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

Here a collection of f=3 subscripts is used to identify the treatment through the selected factor levels.

In most applications, however, a mathematically equivalent factorial model is more useful.

An overparameterized model (cont.)

A factorial model (3 factors) for a CRD might be written as:

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

In this parameterization, the collection of l_1 parameters $\dot{\alpha}_1, \dot{\alpha}_2, \ldots, \dot{\alpha}_{l_1}$ describe the main effect associated with the first factor.
$(\dot{\alpha\beta})_{ij}$ is the two-factor interaction associated with levels i and j of factors 1 and 2, respectively, and represents the nonadditive component of the joint contribution of these two factors etc. etc.

In factorial models with f factors the interactions of highest order are f-factor interactions, and there are f groups of $(f-1)$-factor interactions etc.

Let e.g. $l_1 = 2$, $l_2 = 3$ and $l_3 = 4$. Then we have to estimate the expected response for $t = 24$ different treatments. But the above model has 60 parameters which is a heavy over-parametrization.

9.2.1 Graphical logic

A major benefit of full factorial experimentation is that it is *fully efficient* for investigating the effects associated with each factor.

An n-run full factorial experiment in f factors provides the same information about main effects as f single-factor experiments, each of size n. But *further*, the factorial experiment provides information on how the factors affect the response in combination through interactions, something that cannot be learned from *one-factor-at-a-time* (OFAT) studies.

Suppose we do an unreplicated ($r = 1$) factorial experiment as an experiment designed to estimate the effects associated with changing factor 1 across its l_1 levels.
We could think of the experiment as a CBD with $t = l_1$ treatments (the levels of factor 1), and $b = l_2 \cdot l_3 \cdot \cdots \cdot l_f$ complete blocks (within which the levels of factors 2 through f are fixed).
As with CBD, we might then propose plotting block-corrected values for each level of factor 1.

9.2.2 Matrix development for the overparameterized model

The above model is severely overparameterized, i.e., there are far more parameters than experimental treatments. Hence the set of solutions to the reduced normal equations is especially ambiguous, and without external constraints offers little insight to the structure of the cell means.

In the design matrix of the three-factor example above, there would be groups of columns in $\mathbf{X}_2$ corresponding to each main effect group and each interaction group. A matrix model for the entire experiment could then be written as

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

$$= \mathbf{1}\,\mu + \left(\mathbf{X}_{\dot\alpha} \,\middle|\, \mathbf{X}_{\dot\beta} \,\middle|\, \mathbf{X}_{\dot\gamma} \,\middle|\, \mathbf{X}_{(\dot{\alpha\beta})} \,\middle|\, \cdots \,\middle|\, \mathbf{X}_{(\dot{\alpha\beta\gamma})} \right) \begin{pmatrix} \dot{\boldsymbol{\alpha}} \\ \dot{\boldsymbol{\beta}} \\ \dot{\boldsymbol{\gamma}} \\ (\dot{\boldsymbol{\alpha\beta}}) \\ \vdots \\ (\dot{\boldsymbol{\alpha\beta\gamma}}) \end{pmatrix} + \boldsymbol{\varepsilon}$$

$$\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \qquad \mathsf{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

Matrix development for the overparameterized model (cont)

In the above model $\dot{\phi}$ is used for the parameter vector instead of τ to reflect a parameterization motivated by factorial structure, rather than the unstructured treatment coding employed in previous chapters.

The over-dots are used here to distinguish this parameterization from a different one for the full rank model (later).

The design matrix is partitioned into blocks corresponding to the parameter groups in $\dot{\phi}$ and we will develop a systematic characterization of the model matrix and the various submatrices of $\mathbf{X}_2$ and $\mathbf{X}_2^T\mathbf{X}_2$ based on the use of the *Kronecker product* $\otimes$.

Matrix development for the overparameterized model (cont)

If $\mathbf{A}$ is an $(m \times n)$-matrix and $\mathbf{B}$ is a $(p \times q)$-matrix, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $(mp \times nq)$ block matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}$$

Kronecker products have many interesting properties, we need here just:

- $\otimes$ is non-commutative, i.e. in general $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$
- transposition is distributive over $\otimes$, i.e. $(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T$
- mixed-product property of $\otimes$: $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$

Matrix development for the overparameterized model (cont)

We will show the development of the submatrices of $\mathbf{X}_2$ on the basis of the above 3-factor example with $l_1 = 2$, $l_2 = 3$ and $l_3 = 4$. Let the response vector be ordered, s.t. the index for the first factor changes most slowly, and that for the third factor changes most quickly, in lexicographical order:

$$\mathbf{y}^T = (y_{1111} \cdots y_{111r}, y_{1121} \cdots y_{112r}, y_{1211} \cdots y_{121r}, \cdots, y_{2341} \cdots y_{234r})$$

Given this ordering, the submatrices of $\mathbf{X}_2$ can be written as:

$$\mathbf{X}_{\dot{\alpha}} = \mathbf{1}_r \otimes \mathbf{1}_4 \otimes \mathbf{1}_3 \otimes \mathbf{I}_2 \qquad \mathbf{X}_{\dot{\beta}} = \mathbf{1}_r \otimes \mathbf{1}_4 \otimes \mathbf{I}_3 \otimes \mathbf{1}_2$$

$$\mathbf{X}_{\dot{\gamma}} = \mathbf{1}_r \otimes \mathbf{I}_4 \otimes \mathbf{1}_3 \otimes \mathbf{1}_2 \qquad \mathbf{X}_{(\dot{\alpha\beta})} = \mathbf{1}_r \otimes \mathbf{1}_4 \otimes \mathbf{I}_3 \otimes \mathbf{I}_2$$

$$\mathbf{X}_{(\dot{\alpha\gamma})} = \mathbf{1}_r \otimes \mathbf{I}_4 \otimes \mathbf{1}_3 \otimes \mathbf{I}_2 \qquad \mathbf{X}_{(\dot{\beta\gamma})} = \mathbf{1}_r \otimes \mathbf{I}_4 \otimes \mathbf{I}_3 \otimes \mathbf{1}_2$$

$$\mathbf{X}_{(\dot{\alpha\beta\gamma})} = \mathbf{1}_r \otimes \mathbf{I}_4 \otimes \mathbf{I}_3 \otimes \mathbf{I}_2$$

Matrix development for the overparameterized model (cont)

Given these representations, the diagonal blocks in $\mathbf{X}_2^T\mathbf{X}_2$ corresponding to any parameter group are multiples of the identity matrix. For our example:

$$\mathbf{X}_{\dot\alpha}^T\mathbf{X}_{\dot\alpha} = l_2 l_3 r\,\mathbf{I}_{l_1} = 12r\,\mathbf{I}_2 \qquad \mathbf{X}_{\dot\beta}^T\mathbf{X}_{\dot\beta} = l_1 l_3 r\,\mathbf{I}_{l_2} = 8r\,\mathbf{I}_3$$

$$\mathbf{X}_{\dot\gamma}^T\mathbf{X}_{\dot\gamma} = l_1 l_2 r\,\mathbf{I}_{l_3} = 6r\,\mathbf{I}_4 \qquad \mathbf{X}_{(\alpha\dot\beta)}^T\mathbf{X}_{(\alpha\dot\beta)} = l_3 r\,\mathbf{I}_{l_1 l_2} = 4r\,\mathbf{I}_6$$

$$\mathbf{X}_{(\alpha\dot\gamma)}^T\mathbf{X}_{(\alpha\dot\gamma)} = l_2 r\,\mathbf{I}_{l_1 l_3} = 3r\,\mathbf{I}_8 \qquad \mathbf{X}_{(\beta\dot\gamma)}^T\mathbf{X}_{(\beta\dot\gamma)} = l_1 r\,\mathbf{I}_{l_2 l_3} = 2r\,\mathbf{I}_{12}$$

$$\mathbf{X}_{(\alpha\dot\beta\gamma)}^T\mathbf{X}_{(\alpha\dot\beta\gamma)} = r\,\mathbf{I}_{l_1 l_2 l_3} = r\,\mathbf{I}_{24}$$

Off-diagonal blocks of $\mathbf{X}_2^T\mathbf{X}_2$ can be calculated in a similar way (exercise).

One has to distinguish between off-diagonal blocks for which the parameter groups associated with rows and columns do reference common factors (e.g. $\mathbf{X}_{\dot\alpha}^T\mathbf{X}_{(\alpha\dot\beta)}$) and such which do not reference common factors (e.g. $\mathbf{X}_{\dot\gamma}^T\mathbf{X}_{(\alpha\dot\beta)}$).

9.3 An equivalent full-rank model

Given the above structure, an overwhelming variety of solutions to the least-squares problem exist. Some are of simple form while others are quite complicated, reflecting the particular generalized inverse selected for $\mathbf{X}^T\mathbf{X}$.

An alternative full-rank parameterization that eliminates many of these complications, at the cost of some interpretative simplicity will be presented now.

We now consider independent variables other than binary-coded indicator variables, and write:

$$\mathsf{E}(y) = \mu + \mathbf{x}\boldsymbol{\phi} + \varepsilon \qquad \text{or in matrix form} \qquad \mathsf{E}(\mathbf{y}) = \mathbf{1}\mu + \mathbf{X}_2\boldsymbol{\phi} + \boldsymbol{\varepsilon}$$

where the row vector $\mathbf{x}$ and model matrix $\mathbf{X}_2$ can contain elements other than 0 and 1, and $\boldsymbol{\phi}$ represents a set of treatment-related parameters that are related to, but not the same as, those denoted by $\dot{\boldsymbol{\phi}}$.

An equivalent full-rank model (cont.)

We need to define a set of $l_i - 1$ „regressors“ for the i-th factor, $i = 1, 2, \ldots, f$:

For example, the levels of a three-level factor can be „coded“ using x_1 and x_2 defined as beside.

factor level	x_1	x_2
1	$-\sqrt{\frac{3}{2}}$	$-\frac{1}{\sqrt{2}}$
2	0	$\sqrt{2}$
3	$\sqrt{\frac{3}{2}}$	$-\frac{1}{\sqrt{2}}$

The new regressors x_1 and x_2 have 3 important characteristcs:

- x_1 and x_2 are orthogonal, i.e. $x_1^T x_2 = 0$ **MOST IMPORTANT!**
- the length of x_1 and x_2 equals the number of factor levels, i.e. $x_1^T x_1 = x_2^T x_2 = l_i$
- the sum of the elements of x_1 and x_2 is zero, i.e. $\mathbf{1}^T x_1 = \mathbf{1}^T x_2 = 0$

Similar codings can be easily constructed for factors with any number of levels.

An equivalent full-rank model(cont.)

Suppose the tabulated coding was for the second factor of our above example. We could combine the new regressors to the matrix $\mathbf{F}^\beta$ and denote the i-th row of $\mathbf{F}^\beta$ by $\mathbf{f}_i^\beta$. In the same way we get matrices for the first and second factor:

$$\mathbf{F}^\alpha = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad \mathbf{F}^\beta = \begin{pmatrix} -\sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \sqrt{2} \\ \sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \qquad \mathbf{F}^\gamma = \begin{pmatrix} -\frac{3}{\sqrt{5}} & 1 & -\frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & -1 & \frac{3}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -1 & -\frac{3}{\sqrt{5}} \\ \frac{3}{\sqrt{5}} & 1 & \frac{1}{\sqrt{5}} \end{pmatrix}$$

With this, we may write a model for data from a three-factor experiment as:

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

An equivalent full-rank model(cont.)

Note that the parameter groups for main effects and interactions now contain fewer (only linear independent) parameters.

A main effect associated with factor i (l_i levels) corresponds to a group of $(l_i - 1)$ parameters, an interaction between factors i and j corresponds to a group of $(l_i - 1)(l_j - 1)$ parameters etc.

The (overall) number of model parameters used to describe the mean structure, including μ, is now exactly the same as the number of treatments.

the interpretation of the parameters of this full-rank model is sometimes a little tricky but there is a one-to-one linear relationship between the t cell means μ_{ijk} of and the t elements of the parameter vector $(\mu, \boldsymbol{\phi}^T)^T$.

9.3.1 Matrix development for the full-rank model

The full-rank model may be written as:

$$\mathbf{y} = \mathbf{X} \begin{pmatrix} \mu \\ \phi \end{pmatrix} + \varepsilon = \mathbf{1}\mu + \mathbf{X}_2\phi + \varepsilon$$

$$\mathsf{E}(\varepsilon) = \mathbf{0} \qquad \mathsf{Var}(\varepsilon) = \sigma^2\mathbf{I}$$

where $\mathbf{X}$ contains $n = r \cdot \prod_{i=1}^{f} l_i$ rows and $t = \prod_{i=1}^{f} l_i$ columns.

We now connect the matrices $\mathbf{F}$ associated with the factors to the design matrix $\mathbf{X}$:

For each factor we define a square matrix appending a vector of ones to the associated $\mathbf{F}$, e.g.

$$\mathbf{G}^{\alpha} = (\mathbf{1}|\mathbf{F}^{\alpha})$$

This matrices are orthogonal, i.e.

$$\mathbf{G}^{\alpha T}\mathbf{G}^{\alpha} = l_1\,\mathbf{I}_{l_1}$$

Matrix development for the full-rank model (cont.)

The complete design matrix may be written as:

$$\mathbf{X} = \mathbf{1}_r \otimes \mathbf{G}^{\gamma} \otimes \mathbf{G}^{\beta} \otimes \mathbf{G}^{\alpha}$$

Because of the orthogonal structure of each $\mathbf{G}$ we get

$$\mathbf{X}^T\mathbf{X} = r \prod_{i=1}^{f} l_i \, \mathbf{I}_t = n \, \mathbf{I}_t$$

Because all pairs of columns in $\mathbf{X}$ are orthogonal, matrix forms associated with estimation are especially simple. $\mathbf{X}_2$ equals $\mathbf{X}$ without the first column of ones. In this case, $\mathbf{X}_{2|1} = \mathbf{X}_2$, and so the reduced normal equations are:

$$\mathbf{X}_2^T\mathbf{X}_2\hat{\phi} = \mathbf{X}_2^T\mathbf{y}$$

But since $\mathbf{X}_2^T\mathbf{X}_2 = n\,\mathbf{I}$, this leads immediately to the unique least-squares estimator:

$$\hat{\phi} = \frac{1}{n}\mathbf{X}_2^T\mathbf{y}$$

9.4 Estimation

The experiment offers equal information about each parameter in ϕ, as reflected by the fact that:

$$\mathcal{I} = \mathbf{X}_2^T \mathbf{X}_2 = n\, \mathbf{I}_{t-1}$$

Under the full-rank model questions associated with the „overall“ effect of e.g. factor 1 may be addressed by linear contrasts of form $\mathbf{c}^T \mathbf{F}^\alpha \boldsymbol{\alpha}$

The unique least-squares estimator is (note that $\mathbf{c}^T \mathbf{1} = 0$):

$$\widehat{\mathbf{c}^T \mathbf{F}^\alpha \boldsymbol{\alpha}} = \mathbf{c}^T \mathbf{F}^\alpha \hat{\boldsymbol{\alpha}} = \mathbf{c}^T \mathbf{F}^\alpha \frac{1}{n} (\mathbf{1}^T \otimes \mathbf{1}^T \otimes \mathbf{1}^T \otimes \mathbf{F}^{\alpha T}) \mathbf{y} =$$

$$= \frac{l_1}{n} \sum_{i=1}^{l_1} c_i y_{i\ldots} = \sum_{i=1}^{l_1} c_i \bar{y}_{i\ldots}$$

Since the information matrix for any subset of ϕ is $\mathcal{I} = n\mathbf{I}$ of appropriate dimension, the variance of the estimate is simple

$$\mathsf{Var}(\widehat{\mathbf{c}^T \mathbf{F}^\alpha \boldsymbol{\alpha}}) = \frac{l_1}{n}\, \sigma^2\, \mathbf{c}^T \mathbf{c}$$

9.5 Partitioning of variability and hypothesis testing

The ANOVA decomposition is especially simple in the case of the full-rank model, and facilitates the examination of variability that can be associated with each parameter group. The general from of the treatment sum of squares can be written as:

$$\mathsf{SST} = \mathbf{y}^T \mathbf{H}_{2|1} \mathbf{y} = \frac{1}{n} \mathbf{y}^T \mathbf{X}_2 \mathbf{X}_2^T \mathbf{y} = n \hat{\phi}^T \hat{\phi}$$

where $\mathbf{H}_{2|1}$ projects on the linear space spanned by $\mathbf{X}_{2|1}$ which in our case equals $\mathbf{X}_2$. This expression can be further reduced to individual sums of squares of estimates from each parameter group.
In our three-factor example we have

$$\mathsf{SST} = n \left(\hat{\boldsymbol{\alpha}}^T \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\gamma}}^T \hat{\boldsymbol{\gamma}} + (\widehat{\boldsymbol{\alpha\beta}})^T (\widehat{\boldsymbol{\alpha\beta}}) + (\widehat{\boldsymbol{\alpha\gamma}})^T (\widehat{\boldsymbol{\alpha\gamma}}) + \right.$$
$$\left. + (\widehat{\boldsymbol{\beta\gamma}})^T (\widehat{\boldsymbol{\beta\gamma}}) + (\widehat{\boldsymbol{\alpha\beta\gamma}})^T (\widehat{\boldsymbol{\alpha\beta\gamma}}) \right)$$

If ε has a multivariate normal distribution the seven terms in this sum are independent sums of squares because they are orthogonal contrasts in the data.

Partitioning of variability and hypothesis testing (cont.)

Because of the orthogonality of $\mathbf{X}_2$ the treatment sum of squares SST can be split up as sums of squares in ANOVA in the case of balanced data, e.g.

$$\mathsf{SST}_{\boldsymbol{\alpha}} = n\,\hat{\boldsymbol{\alpha}}^T\hat{\boldsymbol{\alpha}} = n\left(\frac{1}{n}\mathbf{X}_\alpha\mathbf{y}\right)^T\left(\frac{1}{n}\mathbf{X}_\alpha\mathbf{y}\right) = \frac{1}{n}\mathbf{y}^T\mathbf{X}_\alpha\mathbf{X}_\alpha^T\mathbf{y} =$$

$$= \frac{1}{n}\begin{pmatrix} y_{1\ldots} & y_{2\ldots} & \cdots & y_{l_1\ldots}\end{pmatrix}\mathbf{F}^\alpha\mathbf{F}^{\alpha T}\begin{pmatrix} y_{1\ldots}\\ y_{2\ldots}\\ \vdots\\ y_{l_1\ldots}\end{pmatrix} = \frac{1}{n}\mathbf{y}_1^T\mathbf{F}^\alpha\mathbf{F}^{\alpha T}\mathbf{y}_1$$

where $\mathbf{X}_\alpha$ is the submatrix of $\mathbf{X}_2$ corresponding to the parameter group $\boldsymbol{\alpha}$ and $y_{i\ldots}$ are the response totals of the i-th level of the factor corresponding to the parameter group.

So in the above example $\mathbf{y}_1$ is the vector of the data totals specific to the levels of factor 1.

Partitioning of variability and hypothesis testing (cont.)

Noting that $\mathbf{F}^{\alpha}\mathbf{F}^{\alpha T} = \mathbf{G}^{\alpha}\mathbf{G}^{\alpha T} - \mathbf{1}\mathbf{1}^T = l_1\mathbf{I}_{l_1} - \mathbf{J}_{l_1}$

the above sum of squares may be written

$$\mathsf{SST}_{\boldsymbol{\alpha}} = \frac{1}{n}\mathbf{y}_1^T\mathbf{F}^{\alpha}\mathbf{F}^{\alpha T}\mathbf{y}_1 = \frac{1}{n}\mathbf{y}_1^T(l_1\mathbf{I}_{l_1} - \mathbf{J}_{l_1})\mathbf{y}_1 = \frac{l_1}{n}\sum_{i=1}^{l_1} y_{i\ldots}^2 - n\bar{y}_{\ldots\ldots}^2 =$$

$$= \frac{n}{l_1}\sum_{i=1}^{l_1}(\bar{y}_{i\ldots}^2 - \bar{y}_{\ldots\ldots}^2) = \frac{n}{l_1}\sum_{i=1}^{l_1}(\bar{y}_{i\ldots} - \bar{y}_{\ldots\ldots})^2$$

Independent sums of squares can be written for each of the $(2^f - 1)$ parameter groups, although tests for these groups are not independent because each relies on the same denominator sum of squares $\mathsf{SSE} = \sum_{ijkt}(y_{ijkt} - \bar{y}_{ijk.})^2$.

Partitioning of variability and hypothesis testing (cont.)

Testing the main effect parameters of factor 1, i.e. H_0: $\boldsymbol{\alpha} = \mathbf{0}$ in the above example assuming r replications of each treatment can be carried out by comparing

$$\frac{\mathsf{MST}_\alpha}{\mathsf{MSE}} = \frac{\frac{\mathsf{SST}_\alpha}{\mathsf{df}_\alpha}}{\frac{\mathsf{SSE}}{\mathsf{df}_{\mathsf{SSE}}}} = \frac{\frac{12r\sum_{i=1}^{2}(\bar{y}_{i\ldots}-\bar{y}_{\ldots.})^2}{2-1}}{\frac{\sum_{ijkt}(y_{ijkt}-\bar{y}_{ijk.})^2}{24(r-1)}} = \frac{288r(r-1)\sum_{i=1}^{2}(\bar{y}_{i\ldots}-\bar{y}_{\ldots.})^2}{\sum_{ijkt}(y_{ijkt}-\bar{y}_{ijk.})^2}$$

with the $(1-\alpha)$-quantile of $\mathbf{F}\left(1; 24(r-1)\right)$. Since the information matrix for $\boldsymbol{\alpha}$ is $24r\,\mathbf{I}_1$, the noncentrality parameter associated with this test is

$$\lambda = \frac{24r}{\sigma^2}\,\boldsymbol{\alpha}^T\boldsymbol{\alpha}$$

So for nonzero $\boldsymbol{\alpha}$ the power of the test performed at level 0.01 is

$$P\left(W > \mathbf{F}_{0.99}(1; 24(r-1))\right) \qquad \text{where} \qquad W \sim \mathbf{F}\left(1; 24(r-1); \frac{24r}{\sigma^2}\boldsymbol{\alpha}^T\boldsymbol{\alpha}\right)$$

9.6 Factorial experiments as CRDs, CBDs, LSDs, and BIBDs

The three fundamental experimental design plans (CRD, CBD, LSD) can be employed in factorial settings by simply ignoring the factorial structure.

Hence a three-factor treatment structure with $l_1 = 2$, $l_2 = 3$ and $l_3 = 4$ can be examined via a CRD with $24r$ unblocked units, a CBD in r blocks of 24 units each, or a Latin square design (LSD) in 576 units organized in 24 rows and 24 columns in which each treatment is applied r = 24 times.

The degrees of freedom available to estimate this residual variability is $24(r-1)$ for the CRD, $23(r-1)$ for the CBD and $22(r-1) = 506$ for the LSD.

Experiments for factorial structures can also be implemented using BIBD plans, again by ignoring the factors and using randomization of treatments to units as in the unstructured case.

9.7 Model reduction

The substantial and practical difficulty that often arises with factorial experiments is the potentially very large number of parameters.

However, in most practical situations it turns out that interactions of higher order are zero or close to zero. There are substantial statistical advantages to reducing the number of parameters in the model used in analyzing data from a factorial experiment.

We rewrite the full-rank model in further partitioned form as:

$$\mathsf{E}(\mathbf{y}) = \mathbf{1}\mu + \mathbf{X}_2\phi = \mathbf{1}\mu + \mathbf{W}_1\phi_1 + \mathbf{W}_2\phi_2$$

where the columns of $\mathbf{W}_1$ and $\mathbf{W}_2$ form a partition of those in $\mathbf{X}_2$, and ϕ_1 and ϕ_2 form the corresponding partition of ϕ.

Let's now assume that $\phi_2 = \mathbf{0}$. The least-squares estimate of ϕ_1 then is:

$$\hat{\phi}_1 = \frac{1}{n}\mathbf{W}_1^T\mathbf{y}$$

and these estimates are exactly the same functions of $\mathbf{y}$ as they would be under the full model.

Model reduction (cont.)

Now, any contrast in cell means is estimated by $\mathbf{l}^T\mathbf{X}_2\hat{\phi}$ under the full model, or $\mathbf{l}^T\mathbf{W}_1\hat{\phi}_1$ under the reduced model.

The estimation variance based on the reduced model can be no more than that based on the full model, and depending on the specific vector $\mathbf{l}$ of interest and partitioning of ϕ, may be much less.

The questions that must be answered are whether the full model can be reduced, and if so, which terms can be eliminated.

Model selection can be done by a procedure similar to backward elimination in classical ANOVA and requires a hierarchical model structure. This can lead to a remarkable reduction in the number of parameters in the model required to summarize the systematic differences in observed responses.
Reduced models provide more precise estimates of treatment contrasts, and more degrees of freedom for estimating error variance.

This is the main advantage of experiments with factorial treatment structure against unstructured treatment experiments.