Design and Analysis af Experiments with k Factors having p Levels

Henrik Spliid

Lecture notes in the Design and Analysis of Experiments

1st English edition 2002

Informatics and Mathematical Modelling

Technical University of Denmark, DK–2800 Lyngby, Denmark

Foreword

These notes have been prepared for use in the course 02411, Statistical Design of Experiments, at the Technical University of Denmark. The notes are concerned solely with experiments that have k factors, which all occur on p levels and are balanced. Such experiments are generally called p^k factorial experiments, and they are often used in the laboratory, where it is wanted to investigate many factors in a limited - perhaps as few as possible - number of single experiments.

Readers are expected to have a basic knowledge of the theory and practice of the design and analysis of factorial experiments, or, in other words, to be familiar with concepts and methods that are used in statistical experimental planning in general, including for example, analysis of variance technique, factorial experiments, block experiments, square experiments, confounding, balancing and randomisation as well as techniques for the calculation of the sums of squares and estimates on the basis of average values and contrasts.

The present version is a revised English edition, which in relation to the Danish has been improved as regards contents, layout, notation and, in part, organisation. Substantial parts of the text have been rewritten to improve readability and to make the various methods easier to apply. Finally, the examples on which the notes are largely based have been drawn up with a greater degree of detailing, and new examples have been added.

Since the present version is the first in English, errors in formulation an spelling may occur.

Henrik Spliid

IMM, March 2002

April 2002: Since the version of March 2002 a few corrections have been made on the pages 21, 25, 26, 40, 68 and 82.

Lecture notes for course 02411. IMM - DTU.

Contents

Tabels

1

1.1 Introduction

These lecture notes are concerned with the construction of experimental designs which are particularly suitable when it is wanted to examine a large number of factors and often under laboratory conditions.

The complexity of the problem can be illustrated with the fact that the number of possible factor combinations in a multi-factor experiment is the product of the levels of the single factors. If, for example, one considers 10 factors, each on only 2 levels, the number of possible different experiments is $2 \times 2 \times ... \times 2 = 2^k = 1024$. If it is wanted to investigate the factors on 3 levels, this number increases to $3^{10} = 59049$ single experiments. As can be seen, the number of single experiments rapidly increases with the number of factors and factor levels.

For practical experimental work, this implies two main problems. First, it quickly becomes impossible to perform all experiments in what is called a complete factor structure, and second, it is difficult to keep the experimental conditions unchanged during a large number of experiments. Doing the experiments, for example, necessarily takes a long time, uses large amounts of test material, uses a large number of experimental animals, or involves many people, all of which tend to increase the experimental uncertainty.

These notes will introduce general models for such multi-factor experiments where all factors are on p levels, and we will consider fundamental methods to reduce the experimental work very considerably in relation to the complete factorial experiment, and to group such experiments in small blocks. In this way, both savings in the experimental work and more accurate estimates are achieved.

An effort has been made to keep the notes as "non-mathematical" as possible, for example by showing the various techniques in typical examples and generalising on the basis of these. On the other hand, this has the disadvantage that the text is perhaps somewhat longer than a purely mathematical statistical run-through would need.

Generally, extensive numerical examples are not given nor examples of the design of experiments for specific problem complexes, but the whole discussion is kept on such a general level that experimental designers from different disciplines should have reasonable possibilities to benefit from the methods described. As mentioned in the foreword, it is assumed that the reader has a certain fundamental knowledge of experimental work and statistical experimental design.

Finally, I think that, on the basis of these notes, a person would be able to understand the idea in the experimental designs shown, and would also be able to draw up and analyse experimental designs that are suitable in given problem complexes. However, this must not prevent the designer of experiments from consulting the relevant specialist literature on the subject. Here can be found many numerical examples, both detailed and relevant, and in many cases, alternative analysis methods are suggested, which can be very useful in the interpretation of specific experiment results. Below, a few examples of "classical" literature in the field are mentioned.

1.2 Literature suggestions concerning the drawing up and analysis of factorial experiments

.

Box, G.E.P., Hunter, W.G. and Hunter, J.S.: Statistics for Experimenters, Wiley, 1978.

Chapter 10 introduces 2^k factorial experiments. Chapter 11 shows examples of their use and analysis. In particular, section 10.9 shows a method of analysing experiments with many effects, where one does not have an explicit estimate of uncertainty. The method uses the technique from the quantile diagram (Q-Q plot) and is both simple and illustrative for the user. A number of standard block experiments are given. Chapter 12 introduces fractional factorial designs and chapter 13 gives examples of applications. The book contains many examples that are completely calculated - although on the basis of quite modest amount of data. In general a highly recommendable book for experimenters.

Davies, O.L. and others: The Design and Analysis of Experiments, Oliver and Boyd, 1960 (1st edition 1954).

Chapters 7, 8, 9 and 10 deal with factorial experiments with special emphasis on 2^k and 3^k factorial experiments. A large number of practical examples are given based on real problems with a chemical/technical background. Even though the book is a little old, it is highly recommendable as a basis for conducting laboratory experiments. It also contains a good chapter (11) about experimental determination of optimal conditions where factorial experiments are used.

Fisher, R.A.: The Design of Experiments, Oliver and Boyd, 1960 (1st edition 1935)

A classic (perhaps "the classic"), written by one of the founders of statistics. Chapters 6, 7 and 8 introduce notation and methods for 2^k and 3^k factorial experiments. Very interesting book.

Johnson, N.L. and Leone, F.C,: Statistics and Experimental Design, Volume II, Wiley 1977.

Chapter 15 gives a practically orientated and quite condensed presentation of 2^k factorial experiments for use in engineering. With Volume I, this is a good general book about engineering statistical methods.

Kempthorne, O.: The Design and Analysis of Experiments, Wiley 1973 (1st edition 1952).

This contains the mathematical and statistical basis for p^k factorial experiments with which these notes are concerned (chapter 17). In addition it deals with a number of specific problems relevant for multi-factorial experiments, for example experiments with factors on both 2 and 3 levels (chapter 18). It is based on agricultural experiments in particular, but is actually completely general and highly recommended.

Montgomery, D.C.: Design and Analysis of Experiments, Wiley 1997 (1st edition 1976).

The latest edition (5th) is considerably improved in relation to the first editions. The book gives a good, thorough and relevant run-through of many experimental designs and methods for analysing experimental results. Chapters 7, 8 and 9 deal with 2^k factorial experiments and chapter 10 deals with 3^k factorial experiments. An excellent manual and, up to a point, suitable for self-tuition.

2 2^k–factorial experiment

Chapter 2 discusses some fundamental experimental structures for multi-factor experiments. Here, for the sake of simplicity, we consider only experiments where all factors occur on only 2 levels. These levels for example can be "low"/"high" for an amount of additive or "not present"/"present" for a catalyst.

A special notation is introduced and a number of terms and methods, which are generally applicable in planning experiments with many factors. This chapter should thus be seen as an introduction to the more general treatment of the subject that follows later.

2.1 Complete 2^k factorial experiments

2.1.1 Factors

The name, 2^k factorial experiments, refers to experiments in which it is wished to study k factors and where each factor can occur on only 2 levels. The number of possible different factor combinations is precisely 2^k, and if one chooses to do the experiment so that all these combinations are gone through in a randomised design, the experiment is called a complete 2^k factorial experiment.

In this section, the main purpose is to introduce a general notation, so we will only consider an experiment with two factors, each having two levels. This experiment is thus called a 2^2 factorial experiment.

The factors in the experiment are called A and B, and it is practical, not to say required, always to use these names, even if it could perhaps be wished to use, for example, T for temperature or V for volume for mnemonic reasons.

In addition, the factors are organised so that A is always the first factor and B is the second factor.

2.1.2 Design

For each combination of the two factors, we imagine that a number (r) of measurements are made. The random error is called (generally) E. The result of a single experiment with a certain factor combination is often called the **response**, and this terminology is also used for the sum of the results obtained for the given factor combination.

This design is as follows where there are r repetitions per factor combination in a completely randomised setup:

	B = 0	B = 1
A = 0	Y_{001} : Y_{00r}	Y_{011} : Y_{01r}
A = 1	Y_{101} : Y_{10r}	Y_{111} : Y_{11r}

If for example we investigate how the output from a process depends on pressure and temperature, the two levels of factor A can represent two values of pressure while the two levels of factor B represent two temperatures. The measured value, $Y_{ij\nu}$, then gives the result of the ν'th measurement with the factor combination (A_i,B_j).

2.1.3 Model for response, parametrisation

It is assumed, as mentioned, that the experiment is done as a completely randomised experiment, that is, that the $2 \times 2 \times r$ observations are made, for example, in completely random order or randomly distributed over the experimental material which may be used in the experiment.

The mathematical model for the yield of this experiment (the response) is, in that factor A is still the first factor and factor B is the second factor:

$$Y_{ij\nu} = \mu + A_i + B_j + AB_{ij} + E_{ij\nu} \text{ , where } i = (0, 1), j = (0, 1), \nu = (1, 2, .., r)$$

where the ususal restrictions apply

$$\sum_{i=0}^{1} A_i = 0 \quad , \quad \sum_{j=0}^{1} B_j = 0 \quad , \quad \sum_{i=0}^{1} AB_{ij} = 0 \quad , \quad \sum_{j=0}^{1} AB_{ij} = 0$$

These restrictions imply that

$$A_0 = -A_1 \quad , \quad B_0 = -B_1 \quad , \quad AB_{00} = -AB_{10} = -AB_{01} = +AB_{11}$$

Therefore, in reality, there are only 4 parameters in this model, namely the experiment level μ and the factor parameters A_1, $B_1$and AB_{11}, if one, (as usual) refers to the "high" levels of the factors.

2.1.4 Effects in 2^k–factor experiments

In a 2-level factorial experiment, one often speaks of the "effects" of the factors. By this is understood in this special case the mean change of the response that is obtained by changing a factor from its "low" to its "high" level.

The effects in an experiment where the factors have precisely 2 levels are therefore defined in the following manner:

$$A = A_1 - A_0 = 2A_1 \text{ , and likewise } B = 2B_1 \quad , \quad AB = 2AB_{11}$$

In other factorial experiments, one often speaks more generally about factor effects as expressions of the action of the factors on the response, without thereby referring to a definite parameter form.

2.1.5 Standard notation for single experiments

In the theoretical treatment of this experiment, it is practical to introduce a standard notation for the experimental results in the same way as for the effects in the mathematical model.

For the experiments that are done for example with the factor combination (A_1, B_0), the sum of the results of the experiment is needed. This sum is called a, that is

$$a = \sum_{\nu=1}^{r} Y_{10\nu}$$

where this sum is the sum of all data with factor A on the high level and the other factors on the low level. As mentioned, a is also called the response of the factor combination in question.

In the same way, the sum for the experiments with the factor combination (A_0, B_1) is called b, while the sum for (A_1, B_1) is called "ab". Finally, the sum for (A_0, B_0) is called "(1)".

In the design above, cell sums are thus found as in the following table

	B = 0	B = 1
A = 0	(1)	b
A = 1	a	ab

Some presentations use names that directly refer to the factor levels as for example:

	B = 0	B = 1
A = 0	00	01
A = 1	10	11

When one works with these cell sums, they are most practically shown in the so-called standard order for the 2^2 experiment:

$$(1), a, b, ab$$

It is important to keep strictly to the introduced notation, i.e. upper-case letter for parameters in the model and lower-case letters for cell sums, and that the order of parameters as well as data, is kept as shown. If not, there is a considerable risk of making a mess of it.

2.1.6 Parameter estimates

We can now formulate the analysis of the experiment in more general terms.

We find the following estimates for the parameters of the model:

$$\hat{\mu} = [(1) + a + b + ab]/(4 \cdot r) = [(1) + a + b + ab]/(2^k \cdot r)$$

where $k = 2$, as mentioned, gives the number of factors in the design and r is the number of repetitions of the single experiments.

Further we find:

$$\widehat{A}_1 = -\widehat{A}_0 = [-(1) + a - b + ab]/(2^k \cdot r)$$

$$\widehat{B}_1 = -\widehat{B}_0 = [-(1) - a + b + ab]/(2^k \cdot r)$$

$$\widehat{AB}_{11} = -\widehat{AB}_{10} = -\widehat{AB}_{01} = \widehat{AB}_{00} = [(1) - a - b + ab]/(2^k \cdot r)$$

If we also want to estimate for example the A-**effect**, i.e. the change in response when factor A is changed from low (i = 0) to high (i = 1) level, we find

$$\widehat{A} = \widehat{A}_1 - \widehat{A}_0 = 2\widehat{A}_1 = [-(1) + a - b + ab]/(2^{k-1} \cdot r)$$

The parenthesis $[-(1) + a - b + ab]$ gives the total increase in response, which was found by changing the factor A from its low level to its high level. This amount is called the A-**contrast**, and is called $[A]$. Therefore, in the case of the factor A, we have in summary the equations:

$$[A] = [-(1) + a - b + ab] \quad , \quad \widehat{A}_1 = -\widehat{A}_0 = [A]/(2^k \cdot r) \quad , \quad \widehat{A} = 2\widehat{A}_1$$

and correspondingly for the other terms in the model. Specifically for the total sum of observations, the notation $[I] = [(1) + a + b + ab]$ is used. This quantity can be called the **pseudo-contrast**.

2.1.7 Sums of squares

Further, we can derive the sums of squares for all terms in the model. This can be done with ordinary analysis of variance technique. For example, this gives in the case of factor A:

$$\mathrm{SSQ}_A = [A]^2/(2^k \cdot r)$$

Corresponding expressions apply for all the other factor effects in the model.

The sums of squares for these factor effects all have 1 degree of freedom.

If there are repeated measurements for the single factor combinations, i.e. $r > 1$, we can find the residual variation as the variation within the single cells in the design in the usual manner:

$$\mathrm{SSQ}_{resid} = \sum_{i=0}^{1}\sum_{j=0}^{1}([\sum_{\nu=1}^{r} Y_{ij\nu}^2] - T_{ij\cdot}^2/r) \quad ,$$

where

$$T_{ij\cdot} = \sum_{\nu=1}^{r} Y_{ij\nu}$$

is the sum (the total) in cell (i, j).

We can summarise these considerations in an analysis of variance table:

Source of variation	Sum of squares = SSQ	Degrees of freedom = f	S^2 =SSQ/f	F-value
A	$[A]^2/(2^k \cdot r)$	1	S_A^2	$F_A = S_A^2/S_{resid}^2$
B	$[B]^2/(2^k \cdot r)$	1	S_B^2	$F_B = S_B^2/S_{resid}^2$
AB	$[AB]^2/(2^k \cdot r)$	1	S_{AB}^2	$F_{AB} = S_{AB}^2/S_{resid}^2$
Residual	SSQ_{resid}	$2^k \cdot (r-1)$	S_{resid}^2	
Totalt	SSQ_{tot}	$r \cdot 2^k - 1$		

In the table, for example, F_A is compared with an F distribution with $(1, 2^k \cdot (r-1))$ degrees of freedom.

2.1.8 Calculation methods for contrasts

The salient point in the above analysis is the calculation of the contrasts. Various methods, some more practical than others, can be given to solve this problem.

Mathematically, the contrasts can be calculated by the following matrix equation:

$$\begin{bmatrix} I \\ A \\ B \\ AB \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} (1) \\ a \\ b \\ ab \end{bmatrix}$$

One notes that both contrasts and cell sums are given in standard order. In addition it can be seen that the row for example for the A-contrast contains +1 for a and ab, where factor A is at its high level, but -1 for (1) and b, where factor A is at its low level. Finally, it is noticed that the row for AB found by multiplying the rows for A and B by each other.

In some presentations, the matrix expresssion shown is given just as + and - signs in a table:

	(1)	a	b	ab
I	+	+	+	+
A	−	+	−	+
B	−	−	+	+
AB	+	−	−	+

2.1.9 Yates' algorithm

Finally we give a calculation algorithm which is named after the English statistician Frank Yates and is called Yates' algorithm. Data, i.e. the cell sums, are arranged in standard order in a column. Then these are taken in pairs and summed, and after that the same values are subtracted from each other. The sums are put at the top of the next column followed by the differences. When forming the differences, the uppermost value is subtracted from the bottom one (mnemonic rule: As complicated as possible). The operation is repeated as many times as there are factors. Here this would be $k = 2$ times:

Cell sums	1st time	2nd time	=	Contrasts	Sum of Sq.
(1)	$(1) + a$	$(1) + a + b + ab$	=	$[\ I\ ]$	$[I]^2/(2^k \cdot r)$
a	$b + ab$	$-(1) + a - b + ab$	=	$[\ A\ ]$	$[A]^2/(2^k \cdot r)$
b	$-(1) + a$	$-(1) - a + b + ab$	=	$[\ B\ ]$	$[B]^2/(2^k \cdot r)$
ab	$-b + ab$	$(1) - a - b + ab$	=	$[\ AB\ ]$	$[AB]^2/(2^k \cdot r)$

We give a numerical example where the data are shown in the following table:

	B=0	B=1
A=0	12.1	19.8
	14.3	21.0
A=1	17.9	24.3
	19.1	23.4

One finds $(1) = 12.1 + 14.3 = 26.4$, $a = 17.9 + 19.1 = 37.0$, $b = 19.8 + 21.0 = 40.8$ and $ab = 24.3 + 23.4 = 47.7$.

Yates= algorithm now gives

Cell sums	1st time	2nd time	=	Contrasts	Sums of Squares
$(1) = 26.4$	63.4	151.9	=	$[\ I\ ]$	$[I]^2/(2^k \cdot r) = 2884.20$
$a = 37.0$	88.5	17.5	=	$[\ A\ ]$	$[A]^2/(2^k \cdot r) = 38.28$
$b = 40.8$	10.6	25.1	=	$[\ B\ ]$	$[B]^2/(2^k \cdot r) = 78.75$
$ab = 47.7$	6.9	-3.7	=	$[\ AB\ ]$	$[AB]^2/(2^k \cdot r) = 1.71$

In this experiment $r = 2$, and SSQ_{resid} can be found as the sum of squares within the single factor combinations.

$$\begin{aligned}\mathrm{SSQ}_{resid} = &(12.1^2 + 14.3^2 - (12.1 + 14.3)^2/2) \\ &+(17.9^2 + 19.1^2 - (17.9 + 19.1)^2/2) \\ &+(19.8^2 + 21.0^2 - (19.8 + 21.0)^2/2) \\ &+(24.3^2 + 23.4^2 - (24.3 + 23.4)^2/2) \quad = 2.42 + 0.72 + 0.72 + 0.41 = 4.27\end{aligned}$$

ANOVA				
Source of variation	SSQ	df	s^2	F-value
A main effect	38.28	$2-1 = 1$	38.28	35.75
B main effect	78.75	$2-1 = 1$	78.75	73.60
AB interaction	1.71	$(2-1)(2-1) = 1$	1.71	1.60
Residual variation	4.27	$4(2-1) = 4$	1.07	
Totalt	123.01	$8-1 = 7$		

As we shall see, Yates' algorithm is generally applicable to all 2^k factorial experiments and for example can be easily programmed on a calculator. The algorithm also appears in signal analysis under the name “fast Fourier transform”.

The last column in the algorithm gives the contrasts that are used for the estimation as well as the calculation of the sums of squares for the factor effects.

2.1.10 Replications or repetitions

Before we move on to experiments with 3 or more factors, let us look at the following experiment

	B=0	B=1
A=0	Y_{001}	Y_{011}
A=1	Y_{101}	Y_{111}
Day no. 1		

,

	B=0	B=1
A=0	Y_{002}	Y_{012}
A=1	Y_{102}	Y_{112}
Day no. 2		

, $\cdots$,

	B=0	B=1
A=0	Y_{00R}	Y_{01R}
A=1	Y_{10R}	Y_{11R}
Day no. R		

that is, a 2×2, replicated R times. The mathematical model for this experiment is not identical with the model presented on page 8 the beginning of this chapter. The experiment is not completely randomised in that randomisation is done **within days**.

An experimental collection of single experiments that can be regarded as homogeneous with respect to uncertainty, such as the days in the example, is generally called a **block**.

If it is assumed that the contribution from the days can be described by an additive effect, corresponding to a general increase or reduction of the response on the single days (**block effect**), a reasonable mathematical model would be:

$$Y_{ij\tau} = \mu + A_i + B_j + AB_{ij} + D_\tau + F_{ij\tau} \quad , \quad i = (0,1), j = (0,1), \tau = (1,2,\ldots,R),$$

where D_τ gives the contribution from the τ'th dag, and $F_{ij\tau}$ gives the purely random error within days.

We will say that the 2^2 experiment is replicated R times.

This is essentially different from the case where for example $2 \times 2 \times r$ measurements are made in a completely randomised design as on page 8.

If one is in the practical situation of having to choose between the two designs, and it is assumed that both experiments (because of the time needed) must extend over several days, the latter design is preferable. In the first design the randomisation is done across days with r **repetitions**, and the experimental uncertainty, $E_{ij\nu}$ will also contain the variation between days.

One can regard D_τ, i.e. the effect from the τ'th day, as a randomly varying amount with the variance σ_D^2, while $F_{ij\tau}$, i.e. the experimental error within one day, is assumed to have the variance σ_F^2. From this can be derived that $E_{ij\nu}$, i.e. the total experimental error in a completely randomised design over several days, has the variance

$$\sigma_E^2 = \sigma_D^2 + \sigma_F^2$$

The example illustrates the advantage of dividing one's experiment into smaller homogeneous blocks as distinct from complete randomisation. It also shows that there is a fundamental difference between the analysis of an experiment with r repetitions in a completely randomised design and a randomised design replicated R times.

2.1.11 2^3 factorial design

We now state the described terms for the 2^3 factorial experiment with a minimum of comments.

The factors are now A, B, and C with indices i, j and k, respectively. The factors are again ordered so A is the first factor, B the second and C the third factor.

The mathematical model with r repetitions per cell in a completely randomised design is:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{ij} + C_k + AC_{ik} + BC_{jk} + ABC_{ijk} + E_{ijk\nu}$$

where $i, j, k = (0, 1)$ and $\nu = (1, .., r)$.

The usual restrictions are:

$$\sum_{i=0}^{1} A_i = \sum_{j=0}^{1} B_j = \sum_{i=0}^{1} AB_{ij} = \sum_{j=0}^{1} AB_{ij} = \sum_{k=0}^{1} C_k = \cdots = \sum_{k=0}^{1} ABC_{ijk} = 0$$

which implies that

$$A_1 = -A_0 \quad , \quad B_1 = -B_0 \quad , \quad AB_{11} = -AB_{10} = -AB_{01} = AB_{00} \quad ,$$
$$C_1 = -C_0 \quad , \quad \cdots \quad , \quad \text{(and further on until)}$$
$$ABC_{000} = -ABC_{100} = -ABC_{010} = ABC_{110} = -ABC_{001} = ABC_{101} = ABC_{011} = -ABC_{111}$$

The effects of the experiment (which give the difference in response when a factor is changed from "low" level to "high" level, cf. page 9) are

$$A = 2A_1 \quad , \quad B = 2B_1 \quad , \quad AB = 2AB_{11} \quad , \quad C = 2C_1 \ , \ \cdots \ , \ ABC = 2ABC_{111}$$

The standard order for the $2^3 = 8$ different experimental conditions (factor combinations) is:

$$(1) \quad , \quad a \quad , \quad b \quad , \quad ab \quad , \quad c \quad , \quad ac \quad , \quad bc \quad , \quad abc$$

where the introduction of the factor C is done by multiplying c onto the terms for the 2^2 experiment and adding the resulting terms to the sequence: $(1), a, b, ab, ((1), a, b, ab)c = (1), a, b, ab, c, ac, bc, abc$.

$$
\begin{aligned}
[I] &= [+(1) + a + b + ab + c + ac + bc + abc] \\
[A] &= [-(1) + a - b + ab - c + ac - bc + abc] \\
[B] &= [-(1) - a + b + ab - c - ac + bc + abc] \\
[AB] &= [+(1) - a - b + ab + c - ac - bc + abc] \\
[C] &= [-(1) - a - b - ab + c + ac + bc + abc] \\
[AC] &= [+(1) - a + b - ab - c + ac - bc + abc] \\
[BC] &= [+(1) + a - b - ab - c - ac + bc + abc] \\
[ABC] &= [-(1) + a + b - ab + c - ac - bc + abc]
\end{aligned}
$$

or in matrix formulation

$$
\begin{bmatrix} I \\ A \\ B \\ AB \\ C \\ AC \\ BC \\ ABC \end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
-1 & 1 & 1 & -1 & 1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix} (1) \\ a \\ b \\ ab \\ c \\ ac \\ bc \\ abc \end{bmatrix}
$$

Yates' algorithm is performed as above, but the operation on the columns should now be done 3 times as there are 3 factors. If one writes in detail what happens, one gets:

response	1st time	2nd time	3rd time	contrasts
(1)	$(1)+a$	$(1)+a+b+ab$	$(1)+a+b+ab+c+ac+bc+abc$	$[I]$
a	$b+ab$	$c+ac+bc+abc$	$-(1)+a-b+ab-c+ac-bc+abc$	$[A]$
b	$c+ac$	$-(1)+a-b+ab$	$-(1)-a+b+ab-c-ac+bc+abc$	$[B]$
ab	$bc+abc$	$-c+ac-bc+abc$	$(1)-a-b+ab+c-ac-bc+abc$	$[AB]$
c	$-(1)+a$	$-(1)-a+b+ab$	$-(1)-a-b-ab+c+ac+ab+abc$	$[C]$
ac	$-b+ab$	$-c-ac+bc+abc$	$(1)-a+b-ab-c+ac-bc+abc$	$[AC]$
bc	$-c+ac$	$(1)-a-b+ab$	$(1)+a-b-ab-c-ac+bc+abc$	$[BC]$
abc	$-bc+abc$	$c-ac-bc+abc$	$-(1)+a+b-ab+c-ac-bc+abc$	$[ABC]$

Parameter estimates are, with $k = 3$:

$$
\hat{\mu} = \frac{[I]}{2^k \cdot r} \ , \ \widehat{A}_1 = \frac{[A]}{2^k \cdot r} \ , \ \widehat{B}_1 = \frac{[B]}{2^k \cdot r} \ , \ \ldots \ , \ \widehat{ABC}_{111} = \frac{[ABC]}{2^k \cdot r}
$$

Correspondingly, the effect estimates are:

$$\widehat{A} = 2\widehat{A}_1 \quad , \quad \widehat{B} = 2\widehat{B}_1 \quad , \quad \cdots \quad , \quad \widehat{ABC} = 2\widehat{ABC}_{111}$$

The sums of squares are, for example:

$$SSQ_A = \frac{[A]^2}{2^k \cdot r} \ , \ SSQ_B = \frac{[B]^2}{2^k \cdot r} \ , \ SSQ_{ABC} = \frac{[ABC]^2}{2^k \cdot r}$$

The variances of the contrasts are found, with $[A]$ as example, as

$$\mathrm{Var}\{[A]\} = \mathrm{Var}\{-(1) + a - b + ab - c + ac - bc + abc\} = 2^k \cdot r \cdot \sigma^2 \ ,$$

where $k = 3$ here.

The result is seen by noting that there are 2^k terms, which all have the same variance, which for example is

$$\mathrm{Var}\{ab\} = \mathrm{Var}\{\sum_{\nu=1}^{r} Y_{110\nu}\} = r \cdot \sigma^2$$

Further, it is now found, that

$$\mathrm{Var}\{\widehat{A}_1\} = \mathrm{Var}\{[A]/(2^k \cdot r)\} = \sigma^2/(2^k \cdot r)$$

$$\mathrm{Var}\{\widehat{A}\} = \mathrm{Var}\{2\widehat{A}_1\} = \sigma^2/(2^{k-2} \cdot r)$$

2.1.12 2^k factorial experiment

The stated equations are generalised directly to factorial experiments with k factors, each on 2 levels, with r repetitions in a randomised design. Writing up the mathematical model, names for cell sums, calculation of contrasts etc. are done in exactly the same way as described above. For estimates and sums of squares, then generally

$$\text{Parameter estimate} = (\text{Contrast})/(2^k \cdot r)$$

$$\text{Effect estimate} = 2 \times \text{Parameter estimate}$$

$$\text{Sum of squares (SSQ)} = (\text{Contrast})^2/(2^k \cdot r)$$

Regarding the construction of confidence intervals for the parameters and effects, the variance of the estimates can be derived. One finds

$$\begin{aligned}
\text{Var}\{\text{ Contrast }\} &= \sigma^2 \cdot 2^k \cdot r \\
\text{Var}\{\text{ Parameter estimate }\} &= \text{Var}\{\text{ Contrast}\} \,/(2^k \cdot r)^2 = \sigma^2/(2^k \cdot r) \\
\text{Var}\{\text{ Effect estimate }\} &= 2^2\sigma^2/(2^k \cdot r) = \sigma^2/(2^{k-2} \cdot r)
\end{aligned}$$

The confidence intervals for parameters or effects can be constructed if one has an estimate of σ^2. Suppose that one has such an estimate, $\hat{\sigma}^2 = s^2$, and that it has f degrees of freedom. If $(1-\alpha)$ confidence intervals are wanted, one thereby gets

$$\begin{aligned}
I_{1-\alpha}(\text{parameter}) &= \text{Parameter estimate} \pm s \cdot t(f)_{(1-\alpha/2)}/\sqrt{2^k \cdot r} \\
I_{1-\alpha}(\text{effekt}) &= \text{Effect estimate} \pm 2 \cdot s \cdot t(f)_{(1-\alpha/2)}/\sqrt{2^k \cdot r}
\end{aligned}$$

where $t(f)_{(1-\alpha/2)}$ denotes the $(1-\alpha/2)$-fractile in the t-distribution with f degrees of freedom.

2.2 Block confounded 2^k factorial experiment

In experiments with many factors, the number of single experiments quickly becomes very large. For practical experimental work, this means that it can be difficult to ensure homogeneous experimental conditions for all the single experiments.

A generally occurring problem is that in a series of experiments, raw material is used that typically comes in the form of batches, i.e. homogeneous shipments. As long as we perform the experiments on raw material from the same batch, the experiments will give homogeneous results, while results of experiments done on material from different batches will be more non-homogeneous. The batches of raw material in this way constitute blocks.

In the same way, it will often be the case that experiments done close together in time are more uniform than experiments done with a long time between them.

In a series of experiments one will try to do experiments that are to be compared on the most uniform basis possible, since that gives the most exact evaluation of the treatments that are being studied. For example, one will try to do the experiment on the same batch and within as short a space of time as possible. But this of course is a problem when the number of single experiments is large.

Let us imagine that we want to do a 2^3 factorial experiment, i.e. an experiment with 8 single experiments, corresponding to the 8 different factor combinations. Suppose further that it is not possible to do all these 8 single experiments on the same day, but perhaps only four per day.

An obvious way to distribute the 8 single experiments over the two days could be to draw lots. We imagine that this drawing lots results in the following design:

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

For this design, we get for example the **A-contrast**:

$$[A] = [-(1) + a - b + ab - c + ac - bc + abc]$$

As long as the two days give results with exactly the same mean response, this estimate will, in principle, be just as good as if the experiments had been done on the same day. (however the **variance** is generally increased when experiments are done over two days instead of on one day).

But if on the other hand there is a certain unavoidable difference in the mean response on the two days, we obviously have a risk that this affects the estimates. As a simple model for such a difference in the days, we can assume that the response on day 1 is $1g$ under the ideal, while it is $2g$ over the ideal on day 2. An effect of this type is a block effect, and the days constitute the blocks. One says that the experiment is laid out in two blocks each with 4 single experiments.

For the A-contrast, it is shown below how these unintentional, but unavoidable, effects on the experimental results from the days will affect the estimation, as $1g$ is subtracted from all the results from day 1 and $2g$ is added to all the results from day 2:

$$\begin{aligned}
[A] &= [-((1)-1g)+(a-1g)-(b+2g)+(ab+2g)-(c-1g)+(ac+2g)-(bc+2g)+(abc-1g)] \\
&= [-(1) + a - b + ab - c + ac - bc + abc] + [1 - 1 - 2 + 2 + 1 + 2 - 2 - 1]g \\
&= [-(1) + a - b + ab - c + ac - bc + abc]
\end{aligned}$$

Thus, a difference in level on the results from the two days (blocks) will not have any effect on the estimate for the main effect of factor A. In other words, factor A is in **balance** with the blocks (the days).

If we repeat the procedure for the main effect of factor B, we get

$$[B] = [-((1)-1g)-(a-1g)+(b+2g)+(ab+2g)-(c-1g)-(ac+2g)+(bc+2g)+(abc-1g)]$$

$$= [-(1) - a + b + ab - c - ac + bc + abc] + [1 + 1 + 2 + 2 + 1 - 2 + 2 - 1]g$$
$$= [-(1) + a - b + ab - c + ac - bc + abc] + 6g$$

The estimate for the B effect (i.e. the difference in response when B is changed from low to high level) is thereby on average $(6g/4) = 1.5g$ higher than the ideal estimate.

If we look back at the design, this is because factor B was mainly at "high level" on day 2, where the response on average is a little above the ideal.

The same does not apply in the case of factor A. This has been at "high level" two times each day and likewise at "low level" two times each day. The same applies for factor C.

Thus factors A and C are in balance in relation to the blocks (the days), while factor B is not in balance.

An overall evaluation of the effect of the blocks (the days) on the experiment can be seen from the following matrix equation

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} (1) - 1g \\ a - 1g \\ b + 2g \\ ab + 2g \\ c - 1g \\ ac + 2g \\ bc + 2g \\ abc - 1g \end{bmatrix} = \begin{bmatrix} I + 4g \\ A \\ B + 6g \\ AB - 6g \\ C \\ AC \\ BC - 6g \\ ABC - 6g \end{bmatrix}$$

It can be seen that all contrasts that only concern factors A and C are found correctly, because the two factors are in balance in relation to the blocks in the design, while all contrasts that also concern B are affected by the (unintentional, but unavoidable) effect from the blocks.

What we now can ask is whether it is possible to find a distribution over the two days so that the influence from these is eliminated to the greatest possible extent.

We can note that it is the **difference** between the days that is important for the estimates of the effects of the factors, while general level of the days is absorbed in the common average for all data.

If we once more regard the calculation of the contrast $[A]$, we can draw up the following table, which shows how the influence of the days is weighted in the estimate:

Contrast [A]	Response	(1)	a	b	ab	c	ac	bc	abc
	Weight	−	+	−	+	−	+	−	+
	Day	1	1	2	2	1	2	2	1

We note that day 1 enters an equal number of times with + and with −, and day 2 as well. If we look at one of the contrasts where the days do not cancel, e.g. $[B]$, we get a table like the following:

Contrast [B]	Response	(1)	a	b	ab	c	ac	bc	abc
	Weight	−	−	+	+	−	−	+	+
	Day	1	1	2	2	1	2	2	1

where the balance is obviously not present.

The condition that is necessary so that an effect is not influenced by the days is obviously that there is a balance as described. The possibilities for creating such a balance are linked to the matrix of ones in the estimation:

$$\begin{bmatrix} I \\ A \\ B \\ AB \\ C \\ AC \\ BC \\ ABC \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} (1) \\ a \\ b \\ ab \\ c \\ ac \\ bc \\ abc \end{bmatrix}$$

This matrix has the special characteristic that the product sum of any two rows is zero. If one for example takes the rows for $[A]$ and $[B]$, one gets (-1)(-1) + (+1)(-1) + ... + (+1)(+1) = 0. The two contrasts $[A]$ and $[B]$ are thus orthogonal contrasts (linearly independent).

If one therefore chooses for example a design where the days follow factor B, it is absolutely certain that in any case factor A will be in balance in relation to the days. This design would be:

day 1	day 2
$(1)\quad a\quad c\quad ac$	$b\quad ab\quad bc\quad abc$

The influence from the days can now be calculated by adding $-1g$ to all data from day 1 and adding $+2g$ to all data from day 2:

$$
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
-1 & 1 & 1 & -1 & 1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
(1) - 1g \\ a - 1g \\ b + 2g \\ ab + 2g \\ c - 1g \\ ac - 1g \\ bc + 2g \\ abc + 2g
\end{bmatrix}
=
\begin{bmatrix}
I + 4g \\ A \\ B + 12g \\ AB \\ C \\ AC \\ BC \\ ABC
\end{bmatrix}
$$

One can see that now, because of the described attribute of the matrix, it is only the B contrast and the average that are affected by the distribution over the two days.

Of course this design is not very useful if we also want to estimate the effect of factor B, as we cannot unequivocally conclude whether a B-effect found comes from factor B or from differences in the blocks (the days). On the other hand, all the other effects are clearly free from the block effect (the effect of the days).

One says that main effect of factor B is **confounded** with the effect of the blocks (the word "confound" is from Latin and means to "mix up").

The last example shows how we (by following the $+1$ and -1 variation for the corresponding contrast) can distribute the 8 single experiments over the two days so that precisely one of the effects of the model is confounded with blocks, and no more than the one chosen. One can show that this can always just be done.

If, for example, we choose to distribute according to the three-factor interaction ABC, it can be seen that the row for $[ABC]$ has $+1$ for a, b, c og abc, but -1 for (1), ab, ac og bc. One can also follow the $+$ and $-$ signs in the following table :

	(1)	a	b	ab	c	ac	bc	abc
I	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
A	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$
B	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$+$
AB	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$
C	$-$	$-$	$-$	$-$	$+$	$+$	$+$	$+$
AC	$+$	$-$	$+$	$-$	$-$	$+$	$-$	$+$
BC	$+$	$+$	$-$	$-$	$-$	$-$	$+$	$+$
ABC	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$

This gives the following distribution, as we now in general designate the days as blocks and let these have the numbers 0 and 1:

block 0				block 1			
(1)	ab	ac	bc	a	b	c	abc

The block that contains the single experiment (1) is called the **principal block**. The practical meaning of this is that one can make a start in this block when constructing the design.

2.2.1 Construction of a confounded block experiment

The experiment described above is called a block confounded (or just confounded) 2^3 factorial experiment. The chosen confounding is given with the experiment's

$$\textbf{defining relation : } I = ABC$$

And in this connection ABC is called the **defining contrast**.

An easy way to carry out the design construction is to see if the single experiments have an even or an uneven number of letters in common with the defining contrast. Experiments with an even number in common should be placed in the one block and experiments with an uneven number in common should go in the other block.

Alternatively one may use the following **tabular method** where the column for 'Block' is found by **multiplying** the A, B and C columns:

A	B	C	code	Block = ABC
-1	-1	-1	(1)	-1
$+1$	-1	-1	a	$+1$
-1	$+1$	-1	b	$+1$
$+1$	$+1$	-1	ab	-1
-1	-1	$+1$	c	$+1$
$+1$	-1	$+1$	ac	-1
-1	$+1$	$+1$	bc	-1
$+1$	$+1$	$+1$	abc	$+1$

The experiment is analysed exactly as an ordinary 2^3 factorial experiment, but with the exception that the contrast $[ABC]$ cannot unambiguously be attributed to the factors in the model, but is confounded with the block effect.

One can ask whether it is possible to do the experiment in 4 blocks of 2 single experiments in a reasonable way. This has general relevance, since precisely the block size 2 (which naturally is the smallest imaginable) occurs frequently in practical investigations.

One could imagine that the 8 observations were put into blocks according to two criteria, i.e. by choosing two defining relations that for example could be:

$I_1 =$	(1) ab	ac bc		
ABC	c abc	a b		
	$I_2 = AB$			

block (0,0)∼ 1	block (0,1)∼ 2
block (1,0)∼ 3	block (1,1)∼ 4

One notices for example that the experiments in block (0,1) have an even number of letters in common with ABC and an uneven number of letters in common with AB.

The tabular method gives

A	B	C	code	B1 = ABC	B2 = AB	Block no.
-1	-1	-1	(1)	-1	$+1$	$1 \sim (0,0)$
$+1$	-1	-1	a	$+1$	-1	$4 \sim (1,1)$
-1	$+1$	-1	b	$+1$	-1	$4 \sim (1,1)$
$+1$	$+1$	-1	ab	-1	$+1$	$1 \sim (0,0)$
-1	-1	$+1$	c	$+1$	$+1$	$3 \sim (1,0)$
$+1$	-1	$+1$	ac	-1	-1	$2 \sim (0,1)$
-1	$+1$	$+1$	bc	-1	-1	$2 \sim (0,1)$
$+1$	$+1$	$+1$	abc	$+1$	$+1$	$3 \sim (1,0)$

In the figure, there is a 2×2 block system, corresponding to the grouping according to ABC and AB. One can note that the factors A and B are both on "high" as well as "low" level in all 4 blocks. These factors are obviously in balance in relation to the blocks.

However, this does not apply to factor C. It is at "high" level in two of the blocks and at "low" level in the other two. If it is so unfortunate that the two blocks designated (0,0) and (1,1) together result in a higher response than the other two blocks, we will get an undervaluation of the effect of factor C. Thus factor C is confounded with blocks.

To be able to foresee this, one can perceive ABC and AB as factors and then with a formal calculation find the interaction between them:

$$\text{Block effect} = \text{Block level} + ABC + AB + (ABC \times AB)$$

For the effect thus calculated $(ABC \times AB) = A^2B^2C$, the arithmetic rule is introduced that in the 2^k experiment, the exponents are reduced modulo 2. Thus $(ABC \times AB) = A^2B^2C \longrightarrow A^0B^0C \longrightarrow C$. Thereby one gets the formal expression for the block confounding:

$$\text{Block effect} = \text{Block level} + ABC + AB + C$$

which tells us that it is precisely the three effects ABC, AB and C that become confounded with the blocks in the given design.

If one wants to estimate the main effect of C, this design is therefore unfortunate. A better design would be:

	$I_2 = AB$						
$I_1 =$	(1)	abc	ac	b		block 0,0	block 0,1
AC	c	ab	a	bc		block 1,0	block 1,1

Since $(AC \times AB) = A^2BC = BC$, the influence of the blocks in the design is formally given by

$$\text{Blocks} = \text{Block level} + AC + AB + BC$$

and the defining relation: $I = AB = AC = BC$.

The three effects AB, AC and BC are confounded with blocks. All other effects can be estimated without influence from the blocks. Take special note that the main effects A, B and C all appear at both high and low levels in all 4 blocks. The three factors are thus all in balance in relation to the blocks.

The design shown is the best existing design for estimating the main effects of 3 factors in minimal blocks, that is, with 2 experiments in each. Since minimal blocks at the same time result in the most accurate experiments, the design is particularly important.

The design does not give the possibility of estimating the two-factor interactions AB, AC and BC.

2.2.2 A one-factor-at-a-time experiment

It could be interesting to compare the design shown with the following one-factor-at-a-time experiment, which is also carried out in blocks of size of 2:

$$\boxed{(1)\ a}\ \boxed{(1)\ b}\ \boxed{(1)\ c}$$

that is 3 blocks, where the factors are investigated each in one block.

The experiment could be a weighing experiment, where one wants to determine the weight of three items, A, B and C. The measurement (1) corresponds to the zero point reading, while a gives the reading when item A is (alone) on the weight and correspondingly for b and c.

In this design, an estimate for example of the A effect is found as

$$\widehat{A} = [-(1) + a] \;\; \text{with variance } 2\sigma^2$$

where it is here assumed that $r = 1$. In the previous 2^3 design in 2×2 blocks, it was found

$$\widehat{A} = [-(1) + a - ... + abc]/(2^{3-1}) \;\; \text{with variance } \sigma^2/2$$

If one is to achieve an accuracy as good as the "optimal" design with repeated use of the one-factor-at-a-time design, it has to be repeated $(2 \cdot \sigma^2/(\sigma^2/2) = 4$ times . Thus, there will be a total of $4 \times 6 = 24$ single experiments in contrast to the 8 that are used in the "optimal" design.

Another one-factor-at-a-time in 2 blocks of 2 single experiments is the following experiment:

$$\underset{\text{block 0}}{\boxed{(1) \quad a}} \quad \underset{\text{block 1}}{\boxed{b \quad c}}$$

Why is this a hopeless experiment? What can one estimate from the experiment?

2.3 Partially confounded 2^k factorial experiment

We will again consider the 2×2 experiment with the two factors A and B:

	B = 0	B = 1
A = 0	(1)	b
A = 1	a	ab

Suppose that this experiment is to be done in blocks of the size 2. The blocks can correspond for example to batches of raw material that are no larger than at most 2 experiments per batch can be done. By choosing the defining contrast as I=AB, the following block grouping is obtained:

$$\text{Experiment 1 :} \qquad \underset{\text{batch 1}}{\boxed{(1)_1 \quad ab_1}} \quad \underset{\text{batch 2}}{\boxed{a_2 \quad b_2}} \qquad I = AB$$

The mathematical model of the experiment is the following:

$$Y_{ij\nu} = \mu + A_i + B_j + AB_{ij} + E_{ij\nu} \text{ , where } i = (0,1), j = (0,1), \nu = 1$$

but the AB interaction effect is confounded with blocks.

Suppose now that we further want to estimate and/or test the interaction contribution AB_{ij}. This of course can only be done by doing yet another experiment, in which AB is not confounded with blocks. There are two such experiments, one where A is confounded with blocks and one where B is confounded with blocks. As example we choose the latter:

$$\text{Experiment 2 :} \qquad \underset{\text{batch 3}}{\boxed{(1)_3 \quad a_3}} \quad \underset{\text{batch 4}}{\boxed{b_4 \quad ab_4}} \qquad I = B$$

From the two experiments shown (each with two blocks and two single experiments in each block) we will now estimate the various effects. The main effect for factor A can be estimated both in the two first blocks and in the two last blocks and a total A contrast is found as:

$$[A]_{\text{total}} = [A]_1 + [A]_2 \quad ,$$

that is, the sum of the A contrasts in both the two experimental parts:

$$[A]_1 = -(1)_1 + a_2 - b_2 + ab_1 \quad \text{(from experiment 1)}$$
$$[A]_2 = -(1)_3 + a_3 - b_4 + ab_4 \quad \text{(from experiment 2)}$$

as the index of the 8 single experiments corresponds to the block (batch) in which the single experiments were made. The index of the contrast gives whether it is the first or the second experimental part it is calculated in.

Further, we can now find a contrast for the main effect B, but only from the first experiment:

$$[B]_1 = -(1)_1 - a_2 + b_2 + ab_1$$

Finally a contrast for the interaction effect AB is found, but now from the other experiment where it is not confounded with blocks:

$$[AB]_2 = +(1)_3 - a_3 - b_4 + ab_4$$

Since the two A contrasts are both free of block effects, in addition to their sum we can find their difference:

$$[A]_{\text{difference}} = [A]_1 - [A]_2$$

This amount measures the difference between the A estimates in the two parts of the experiment. This difference, as the experiment is laid out, can only be due to experimental uncertainty, and can thus be interpreted as an expression of the experimental uncertainty, that is, the residual variation.

The two contrasts

$$\begin{aligned}
[B]_2 &= -(1)_3 - a_3 + b_4 + ab_4 \quad \text{(from experiment 2)} \\
[AB]_1 &= +(1)_1 - a_2 - b_2 + ab_1 \quad \text{(from experiment 1)}
\end{aligned}$$

are both confounded with blocks.

As expressions of the experimental levels in the two parts of the experiment we find

$$\begin{aligned}
[I]_1 &= +(1)_1 + a_2 + b_2 + ab_1 \quad \text{(from experiment 1)} \\
[I]_2 &= +(1)_3 + a_3 + b_4 + ab_4 \quad \text{(from experiment 2)}
\end{aligned}$$

This results in

$$\begin{aligned}
[I]_{\text{total}} &= [I]_1 + [I]_2 \\
[I]_{\text{difference}} &= [I]_1 - [I]_2
\end{aligned}$$

The quantity $[I]_{\text{total}}$ and the contrast $[I]_{\text{difference}}$ measure the level of the whole experiment and the difference in level between the first and second part of the experiment, respectively.

One can investigate whether the quantities drawn up are orthogonal contrasts by looking at the following matrix expression:

$$\begin{bmatrix} [I] \\ [A]_{\text{total}} \\ [B]_1 \\ [AB]_2 \\ [A]_1 - [A]_2 \\ [B]_2 \\ [AB]_1 \\ [I]_1 - [I]_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} (1)_1 \\ a_2 \\ b_2 \\ ab_1 \\ (1)_3 \\ a_3 \\ b_4 \\ ab_4 \end{bmatrix}$$

One sees that all the contrasts are mutually orthogonal and that there are exactly 7 contrasts and the sum (the pseudo contrast). They can therefore as a whole describe all the variation between the 8 single experiments that have been carried out.

Bt means of the general formula for the sums of squares and for contrasts in particular, we can then find all the sums of squares.

$$
\begin{array}{lcll}
\text{SSQ}_A & = & \frac{[A]^2_{\text{total}}}{R_A \cdot 2^k \cdot r} & ,\quad \text{where } R_A = 2 \text{ and } k = 2 \\
\text{SSQ}_B & = & \frac{[B]^2_1}{R_B \cdot 2^k \cdot r} & ,\quad \text{where } R_B = 1 \text{ and } k = 2 \\
\text{SSQ}_{AB} & = & \frac{[AB]^2_2}{R_{AB} \cdot 2^k \cdot r} & ,\quad \text{where } R_{AB} = 1 \text{ and } k = 2 \\
\text{SSQ}_{A,\text{uncertainty}} & = & \frac{[A]^2_1 + [A]^2_2}{2^k \cdot r} - \frac{[A]^2_{\text{total}}}{R_A \cdot 2^k \cdot r} & ,\quad \text{where } R_A = 2 \\
\text{SSQ}_{B,\text{blocks}} & = & \frac{[B]^2_2}{R_{B,\text{blocks}} \cdot 2^k \cdot r} & ,\quad \text{where } R_{B,\text{blocks}} = 1 \\
\text{SSQ}_{AB,\text{blocks}} & = & \frac{[AB]^2_1}{R_{AB,\text{blocks}} \cdot 2^k \cdot r} & ,\quad \text{where } R_{AB,\text{blocks}} = 1 \\
\text{SSQ}_{\text{level difference}} & = & \frac{[I]^2_1 + [I]^2_2}{2^k \cdot r} - \frac{[I]^2_{\text{total}}}{R \cdot 2^k \cdot r} & ,\quad \text{where } R = 2
\end{array}
$$

All these sums of squares have 1 degree of freedom, and we can now draw up an analysis of variance table based on:

$$
\begin{array}{lll}
\text{SSQ}_A & ,\quad f_A = 1 & ,\quad (\text{precision}=1) \\
\text{SSQ}_B & ,\quad f_B = 1 & ,\quad (\text{precision}=1/2) \\
\text{SSQ}_{AB} & ,\quad f_{AB} = 1 & ,\quad (\text{precision}=1/2)
\end{array}
$$

$$\text{SSQ}_{\text{blocks}} = \text{SSQ}_{B,\text{blocks}} + \text{SSQ}_{AB,\text{blocks}} + \text{SSQ}_{\text{level difference}} \quad , \quad f_{\text{blocks}} = 3$$

$$\text{SSQ}_{\text{uncertainty}} = \text{SSQ}_{A,\text{uncertainty}} \quad , \quad f_{\text{uncertainty}} = 1$$

2.3.1 Some generalisations

As shown, all the drawn up contrasts are mutually orthogonal.

This means that the sum of squares for the 7 contrasts make up the total sum of squares. This is also synonymous with the statement that the effects we have in the experiment are all mutually balanced.

One can also note that the number of degrees of freedom is precisely $8 - 1 = 7$, namely one for each of the contrasts, to which comes 1 for the total sum $[I]$.

One can then test the individual effects of the model against the estimate of uncertainty. In the example, this estimate has only 1 degree of freedom, and of course this does not give a reasonably strong test.

The fact that one seemingly comes to the same F test for the effect of A, B as well as AB is not synonymous with the statement that the A, B and AB effects are estimated equally precise. For example this can be seen by calculating the variances in the parameter estimates:

$$\mathrm{Var}(\widehat{A}_1) = \mathrm{Var}\{\frac{[A]}{R_A \cdot 2^k \cdot r}\} = \sigma^2 \cdot \frac{R_A \cdot 2^k \cdot r}{(R_A \cdot 2^k \cdot r)^2} = \frac{\sigma^2}{R_A \cdot 2^k \cdot r} \quad , \quad R_A = 2$$

$$\mathrm{Var}(\widehat{B}_1) = \mathrm{Var}\{\frac{[B]_1}{R_B \cdot 2^k \cdot r}\} = \sigma^2 \cdot \frac{R_B \cdot 2^k \cdot r}{(R_B \cdot 2^k \cdot r)^2} = \frac{\sigma^2}{R_B \cdot 2^k \cdot r} \quad , \quad R_B = 1$$

$$\mathrm{Var}(\widehat{AB}_{11}) = \mathrm{Var}\{\frac{[AB]_2}{R_{AB} \cdot 2^k \cdot r}\} = \sigma^2 \cdot \frac{R_{AB} \cdot 2^k \cdot r}{(R_{AB} \cdot 2^k \cdot r)^2} = \frac{\sigma^2}{R_{AB} \cdot 2^k \cdot r} \quad , \quad R_{AB} = 1$$

so that the variances of the B and AB estimates are double the variance of the A estimate. This of course is due to the fact that the A estimate is based on twice as many observations as the other estimates ($R_A/R_B = 2$ and $R_A/R_{AB} = 2$).

The difference between the tests of the three effects is their power. The test of the A effect has greater power than the other two tests (for the same test level α).

One can generally write

$$\mathrm{Var}\{\text{Parameter estimate}\} = \mathrm{Var}\{[\text{Contrast}]/(R \cdot 2^k \cdot r)\} = \sigma^2/(R \cdot 2^k \cdot r) \quad ,$$

$$\mathrm{Var}\{\text{Effect estimate}\} = \mathrm{Var}\{2 \cdot [\text{Contrast}]/(R \cdot 2^k \cdot r)\} = \sigma^2 \cdot 4/(R \cdot 2^k \cdot r) \quad ,$$

where R gives the number of 2^k factorial experiments on which the estimate is based, and r gives the number of repetitions for the single factor combinations in these factorial experiments. For the A effect in the example, $R = 2$, while $R = 1$ for both the B and the AB effect.

If one has repeated a 2^k factorial experiment R times, where an effect can be estimated, one can find the variation between these estimates in a similar way as shown for the A effect in the example. If we suppose, for the sake of simplicity, that it is the A effect that can be estimated in these R different 2^k experiments with r repetitions per factor combination, we can generally find an estimate of uncertainty as the square sum:

$$SSQ_{A,\text{uncertainty}} = \frac{[A]_1^2 + [A]_2^2 + \cdots + [A]_R^2}{2^k \cdot r} - \frac{([A]_1 + [A]_2 + \cdots + [A]_R)^2}{R \cdot 2^k \cdot r}$$

which will have $R - 1$ degrees of freedom. The amount $[A]_\tau$ gives the A contrast in the τ'th factorial experiment.

One notes that this sum of squares is exactly the variation between the R estimates for the A effect. If one has several effects which in this way are estimated several times, all their uncertainty contributions can be summed up in a common uncertainty estimate, which can be used for testing.

Estimation af block effects

In some connections, it can be of interest to estimate specific block differences. If we again take the two confounded 2^k experiments that form the basis for this section, we could for example be interested in estimating the difference between block 0 and block 1. An estimate for this difference can be derived by remembering that the difference between blocks 0 and 1 is confounded with the AB effect, and that the pure AB effect can be estimated in the second experimental half. In other words, we can draw up the contrast

$$[AB]_1 - [AB]_2 = [(1)_1 - a_2 - b_2 + ab_1] - [(1)_3 - a_3 - b_4 + ab_4]$$

This quantity has $2\times$(difference between block 0 and block 1) as its expected value, and one can therefore use the estimate $X_1 = ([AB]_1 - [AB]_2)/2$ as the estimate for the block difference Block_1-Block_2.

In this estimation of block effects, the principle is the simple one that one estimates the effects that the blocks are confounded with and then breaks the confounding with these estimates.

We will not go further here with these ideas, but only point out the general possibilities that lie in using partial confounding.

It becomes possible to test and estimate all factor effects in factorial experiments with small blocks, just as it becomes possible to extract block effects with the outlined estimation technique.

2.4 Fractional 2^k factorial design

In this section, we will introduce a special and very important type of experiment, which under certain assumptions can help to reduce experimental work greatly in comparison with complete factorial experiments.

Example 2.1 : A simple weighing experiment with 3 items

Suppose we want to determine the weight of three items, A, B and C. A weighing result can be designated in the same way as described above. For example "a" designates the result of the weighing where item A is on the weight alone, while (1) designates weighing without any item being on the weight, i.e., the zero point adjustment.

The simplest experiment consists in doing the following 4 single experiments:

$$\boxed{(1) \quad a \quad b \quad c}$$

that is, that one measurement obtained without any item on the weight is obtained first, and the three items are weighed separately.

The estimates for the weight of the three items are:

$$\widehat{A} = [-(1) + a] \quad , \quad \widehat{B} = [-(1) + b] \quad , \quad \widehat{C} = [-(1) + c]$$

This kind of design is probably frequently (but unfortunately) used in practice. It can be briefly characterised as "one-factor-at-a-time".

One can directly find the variance in the estimates:

$$\mathrm{Var}\{\widehat{A}\} = \mathrm{Var}\{\widehat{B}\} = \mathrm{Var}\{\widehat{C}\} = 2 \cdot \sigma^2$$

A basic characteristic of good experimental designs is that all data are used in estimates for all effects. This is seen **not** to be the case here, and one can ask if one could possibly find an experimental design that is more "efficient" than the one shown.

The experiments that can be carried out are:

$$\boxed{(1) \quad a \quad b \quad ab \quad c \quad ac \quad bc \quad abc}$$

The complete factorial experiment of course consists of doing all 8 single experiments, and the estimates for the effects are found as previously shown. In the case of factor A, we get the effect estimate:

$$\widehat{A} = [-(1) + a - b + ab - c + ac - bc + abc]/4$$

which is thus the estimate for the weight of item A. The variance for this estimate is $\sigma^2/2$ (namely $8\sigma^2/4^2$).

One can easily convince oneself that the weight of item B and item C are balanced out of the estimate for A. The same applies to a possible zero point error in the scale of the weight (μ).

As an alternative to these two obvious experiments, we can consider the following experiment:

$$\boxed{(1) \quad ab \quad bc \quad ac}$$

The experiment thus consists of weighing the items together two by two. For example the estimate for the weight of A is:

$$\widehat{A} = [A]/2 = [-(1) + ab + ac - bc]/2$$

$$\mathrm{Var}\{\widehat{A}\} = \sigma^2$$

Note that the zero point μ as well as the weights of items B and C are eliminated in this estimate.

One also notes that in relation to the primitive "one-factor-at-a-time" experiment, in this design we can use all 4 observations to estimate the A effect, that is, the weight of item A. The same obviously applies to the estimates for the B and C effects. In addition, the variance of the estimate here is only half the variance of the estimates in the "one-factor-at-a-time" experiment. The experiment is therefore appreciably better than the "one-factor-at-a-time" experiment.

The experiment is called a $\frac{1}{2} \times 2^3$ factorial experiment or a 2^{3-1} factorial experiment, as it consists precisely of half the complete 2^3 factorial experiment.

Finally a small numerical example:

$$(1) = 6.78 \text{ g} \quad ab = 28.84 \text{ g} \quad ac = 20.66 \text{ g} \quad bc = 18.12 \text{ g}$$

$$\begin{array}{lclcr}
\widehat{A} & = & (-6.78 + 28.84 + 20.66 - 18.12)/2 & = & 12.30 \text{ g} \\
\widehat{B} & = & (-6.78 + 28.84 - 20.66 + 18.12)/2 & = & 9.76 \text{ g} \\
\widehat{C} & = & (-6.78 - 28.84 + 20.66 + 18.12)/2 & = & 1.58 \text{ g}
\end{array}$$

Let us suppose that the manufacturer has stated that the weight has an accuracy corresponding to the standard deviation $\sigma = 0.02g$. With this is found $\text{Var}\{\widehat{A}\} = 4 \times 0.02^2/2^2 = 0.02^2 \text{ g}^2$. The standard deviation of the estimated A weight is thus 0.02g. The same standard deviation is found for the weights of B and C.

A 95% confidence interval for the weight of A is 12.30 $\pm$ 2$\times$ 0.02 g = [12.26 , 12.34] g.

End of example 2.1

We will now discuss what can generally be estimated in an experiment as described in the above example. If one can assume that it is only the main effects that are important in the experiments, there are no problems estimating these. In the example, one can take it that the weight of the two items is exactly the sum of the weights of the two items, which corresponds to saying that there is no interaction.

Alternatively, we now imagine that the following general model applies for the described experiment with the three factors, A, B and C:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{ij} + C_k + AC_{ik} + BC_{jk} + ABC_{ijk} + E_{ijk\nu}$$

where $i, j, k = (0, 1)$ and $\nu = (1, \ldots, r)$ with the usual restrictions. Complete randomisation is assumed.

The quantity $E_{ijk\nu}$ designates the experimental error in the ν'th repetition of the single experiment indexed by (i, j, k).

For the single experiment "(1)" in the described experiment, all indices are on level "0", and its expected value is:

$$E\{(1)\} = \mu + A_0 + B_0 + AB_{00} + C_0 + AC_{00} + BC_{00} + ABC_{000}$$

By using the fact that $A_0 = -A_1$ and correspondingly for the other terms of the model, we find

$$E\{(1)\} = \mu - A_1 - B_1 + AB_{11} - C_1 + AC_{11} + BC_{11} - ABC_{111}$$

$$E\{ab\} = \mu + A_1 + B_1 + AB_{11} - C_1 - AC_{11} - BC_{11} - ABC_{111}$$

$$E\{ac\} = \mu + A_1 - B_1 - AB_{11} + C_1 + AC_{11} - BC_{11} - ABC_{111}$$

$$E\{bc\} = \mu - A_1 + B_1 - AB_{11} + C_1 - AC_{11} + BC_{11} - ABC_{111}$$

In this way, for the A contrast we can now find

$$E\{[A]\} = E\{-(1) + ab + ac - bc\} = 4(A_1 - BC_{11})$$

This means that if the factors B and C interact, so $BC_{11} \neq 0$, the estimate for the main effect of factor A will be affected in this half experiment. The effects A and BC are therefore confounded in the experiment. It holds true generally in this experiment that the **effects are confounded in groups of two**.

This is formally expressed through the **alias relation** "$A = BC$". The relation expresses that the effects A and BC act synchronously in the experiment and that they therefore are **confounded**. The A and BC effects cannot be destinguished from each other in the experiment.

The alias relations for the whole experiment are

$$\begin{array}{ccc} I & = & ABC \\ A & = & BC \\ B & = & AC \\ C & = & AB \end{array}$$

where the first relation, $I = ABC$, is called the **defining relation** of the experiment and ABC called the defining contrast - in the same way as in the construction of a confounded block experiment (cf. page 23). This expresses that the three-factor- interaction ABC does not vary in the experiment, but has the same level in all the single experiments (namely $-ABC_{111}$).

The other alias relations are simply derived by multiplying both sides of the defining relation with the effects of interest, and then reducing the exponents modulo 2. For example, the alias relation for the A effect is found as $A \times I = A \times ABC$ i.e. $A = A^2BC \longrightarrow BC$, where "$I$" is here treated as a "one" and the 2-exponent in A^2BC is reduced to 0 (modulo 2 reduction).

If we recall the confounded block experiment, where a complete 2^3 factorial experiment could be laid out in two blocks according to the defining relation $I = ABC$, we see that our experiment is precisely the principal block in that experiment. If it is a case of a $\frac{1}{2} \times 2^k$ factorial experiment, the fraction that contains "(1)" can be called the **principal fraction**.

We can check whether from the other half of the complete experiment one could find estimates that are just as good as in the half we treated in our example. The experiment is

$$\boxed{a \quad b \quad c \quad abc}$$

$$[A] = [a - b - c + abc]$$

$$E\{[A]\} = E\{a - b - c + abc\} = 4(A_1 + BC_{11})$$

Note that the confounding has the opposite sign compared with earlier. If one adds the two contrasts, that is

$$[-(1) + ab + ac - bc] + [a - b - c + abc]$$

one finds precisely the A contrast for the complete experiment, while subtracting them, that is

$$-[-(1) + ab + ac - bc] + [a - b - c + abc]$$

finds precisely the BC contrast.

The two alternative half experiments are called **complementary** fractional factorials, as together they form the complete factorial experiment.

We will now show how one chooses for example a $\frac{1}{2} \times 2^3$ factorial experiment in practice.

We note that a $\frac{1}{2} \times 2^3$ factorial experiment consists of 2^2 measurements. The experiment that is to be derived can therefore be understood as a 2^2 experiment with an extra factor put in. Let us therefore consider the complete 2^2 experiment with the factors A and B. The mathematical model for this experiment is:

$$Y_{ij\nu} = \mu + A_i + B_j + AB_{ij} + E_{ij\nu} \quad , \quad i = (0,1) \ , \ j = (0,1) \ , \ \nu = (1, 2, \ldots, r)$$

If we suspect that all 4 parameters in this model can be important, further factors cannot be put into the experiment, but if we assume that the interaction AB is negligible, as in the weighing experiment, we can introduce factor C, so that it is confounded with just AB.

We therefore choose to confound C with AB, that is using the alias relation $C = AB$. This alias relation can (only) be derived from the **defining relation** $I = ABC$, which can be seen by multiplying the alias relation $C = AB$ on both sides with C (or AB for that matter) and reducing all exponents modulo 2.

$$\boxed{C = AB} \implies \begin{array}{rcll} I &=& ABC & \text{(the defining relation)} \\ A &=& BC & \\ B &=& AC & \\ AB &=& C & \text{(the generator equation)} \end{array}$$

We shall call $C = AB$ the **generator equation** since it is the alias relation from which the design is generated.

The principal fraction is made up of all single experiments that have an **even** number of letters in common with ABC, i.e., the experiments (1), ab, ac, bc. Alternatively, the complementary fraction could be chosen, which contains all single experiments that have an **uneven** number of letters in common with ABC, i.e., a, b, c and abc.

With this last method, where the starting point is the complete factorial experiment for the two (first) factors A and B, it is said that these form an **underlying complete factorial** *for the fractional factorial design. We will return to this important concept later.*

Let us now suppose that we choose the experiment corresponding to “uneven”:

$$\boxed{a \quad b \quad c \quad abc}$$

To find the sign for the confoundings, it is enough to consider one of the alias relations, for example $C = AB$ and compare this with one of the experiments that is to be done, for example the experiment “a”.

For the experiment ”a”, the effect C has the value C_0 (since factor C is on 0 level), and the effect AB has the value AB_{10}. The confounding is therefore $C_0 = AB_{10}$. Since we calculate on the basis of the “high” levels C_1 and AB_{11}, these are put in.

Since $C_0 = -C_1$ and $AB_{10} = -AB_{11}$, we finally get that the alias relation is $C_1 = AB_{11}$. The rest of the alias relations get the same sign when they are expressed in the high levels. For example one gets $A_1 = BC_{11}$.

One writes for example

$$\boxed{+C = +AB} \implies \begin{array}{rcll} +I &=& +ABC & \text{(the defining relation)} \\ +A &=& +BC & \\ +B &=& +AC & \\ +AB &=& +C & \text{(the generator equation)} \end{array}$$

Whether the constructed experiment is a suitable experiment depends on whether the alias relations together give satisfactory possibilities for estimating the effects, which, a priori, are considered interesting.

Example 2.2 : A 1/4×2^5 factorial experiment .

We finish this section by showing how, with the help of the introduced ideas, one can construct a $1/4 \times 2^5$ factorial experiment, i.e., an experiment that consists only of $2^3 = 8$ measurements, but includes 5 factors. These are called (always) A, B, C, D and E (for 1st, 2nd, 3rd, 4th and 5th factor).

The complete factorial experiment with 3 factors contains in addition to the level μ the effects A, B, AB, C, AC, BC, and ABC.

Suppose now that it can be assumed that factors B and C do not interact, i.e. that BC=0. A reasonable inference from this could be that also ABC = 0. Thereby it would be natural to choose two **generator equations**, namely $D = BC$ and $E = ABC$. These give $I_1{=}BCD$ and $I_2{=}ABCE$, respectively. The principal fraction consists of the single experiments that have an even number of letters in common with both the defining contrasts BCD and $ABCE$. These single experiments are:

(1)	*ae*	*bde*	*abd*	*cde*	*acd*	*bc*	*abce*

A direct and easy method to construct this experiment is to write out a table as follows:

A	B	C	$D = -BC$	$E = ABC$	Code
-1	-1	-1	-1	-1	(1)
$+1$	-1	-1	-1	$+1$	*a e*
-1	$+1$	-1	$+1$	$+1$	*b de*
$+1$	$+1$	-1	$+1$	-1	*ab d*
-1	-1	$+1$	$+1$	$+1$	*c de*
$+1$	-1	$+1$	$+1$	-1	*ac d*
-1	$+1$	$+1$	-1	-1	*bc*
$+1$	$+1$	$+1$	-1	$+1$	*abc e*

Note that for the factors A, B and C the ordering of the levels correspond to the standard order: (1), a, b, ab, c, ac, bc, abc, as used in Yates algorithm, for example.

The minus sign in $D = -BC$ ensures that the experiment (1) is obtained as the first one, if the principal fraction is wanted.

This **tabular method** of writing out the experiment can be used quite generally as will be demonstrated in the following. A further advantage is that the signs of the confoundings are obtained directly.

An alternative experiment is found by constructing one of the other "fractions". If for example one wants an experiment that contains the single experiment "a", the corresponding fraction can be found by multiplying the principal fraction through with "a" and reducing the exponents modulo 2. In this way one gets:

a	e	$abde$	bd	$acde$	cd	abc	bce

The same experiment would have been obtained by changing the sign in $E = ABC$ so that $E = -ABC$ is used in the above tabular method.

Now, to find the alias relations in the experiment, we will again use the two defining relations.

The interaction of the two defining contrasts is found by multiplying them together and again reducing all exponents modulo 2:

$$D = BC,\ E = ABC \Longrightarrow I_1{=}BCD,\ I_2{=}ABCE \ \text{ and } \ I_3{=}I_1 \times I_2{=}AB^2\ C^2DE \rightarrow ADE$$

so that the defining relation and the alias relations (without signs) of the experiment are:

$$\begin{array}{lllllll}
I & = & BCD & = & ABCE & = & ADE \\
A & = & ABCD & = & BCE & = & DE \\
B & = & CD & = & ACE & = & ABDE \\
AB & = & ACD & = & CE & = & BDE \\
C & = & BD & = & ABE & = & ACDE \\
AC & = & ABD & = & BE & = & CDE \\
BC & = & D & = & AE & = & ABCDE \\
ABC & = & AD & = & E & = & BCDE
\end{array}$$

Roughly speaking, the experiment is only a good experiment if one can assume that the interactions are negligible (in relation to the main effects).

The signs for the confoundings can again be found by considering an alias relation, e.g. $A = ABCD = BCE = DE$, together with one of the single experiments that are part of the chosen experimental design.

For example "a" is in the experiment and it corresponds to a single experiment with indices (1,0,0,0,0) for the factors A, B, C, D and E, respectively. Thus

$$A_1 = ABCD_{1000} = BCE_{000} = DE_{00} \Longleftrightarrow +A_1 = -ABCD_{1111} = -BCE_{111} = +DE_{11}$$

This sign pattern is repeated in all the alias relations:

$$\begin{array}{lllllll}
+I & = & -BCD & = & -ABCE & = & +ADE \\
+A & = & -ABCD & = & -BCE & = & +DE \\
+B & = & -CD & = & -ACE & = & +ABDE \\
+AB & = & -ACD & = & -CE & = & +BDE \\
+C & = & -BD & = & -ABE & = & +ACDE \\
+AC & = & -ABD & = & -BE & = & +CDE \\
+BC & = & -D & = & -AE & = & +ABCDE \\
+ABC & = & -AD & = & -E & = & +BCDE
\end{array}$$

We need now to find estimates and sums of squares. This can be done by again using the fact that the experiment is formed on the basis of the complete underlying factorial structure composed of factors A, B and C. In this structure we now estimate all the effects corresponding to the three factors.

In order to subsequently find the D effect we only need to look up the BC row, where the D effect appears with the opposite sign. Data are grouped in standard order according to the factors A, B and C. This is done by ignoring "*d*" and "*e*". Then the contrasts can be calculated, with the use of Yates' algorithm, for example. One gets:

$$\begin{bmatrix}
I & = & -BCD & = & -ABCE & = & ADE \\
A & = & -ABCD & = & -BCE & = & DE \\
B & = & -CD & = & -ACE & = & ABDE \\
AB & = & -ACD & = & -CE & = & BDE \\
C & = & -BD & = & -ABE & = & ACDE \\
AC & = & -ABD & = & -BE & = & CDE \\
BC & = & -D & = & -AE & = & ABCDE \\
ABC & = & -AD & = & -E & = & BCDE
\end{bmatrix}$$

$$= \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
-1 & 1 & 1 & -1 & 1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
e \\ a \\ bd \\ abde \\ cd \\ acde \\ bce \\ abc
\end{bmatrix}$$

Note that the row $e, a, bd, abde, cd, acde, bce, abc$, becomes the row $(1), a, b, ab, c, ac, bc, abc$, if one leaves out d and e, i.e. the standard order for the complete 2^3 factorial experiment for A, B and C.

The experiment which we find by ignoring the factors D and E, i.e. the complete 2^3 factorial experiment including A, B and C, is again an underlying complete factorial experiment and A, B and C constitute an underlying complete factorial structure.

We can easily check that for example the ABC effect is confounded with -E. One way to do this is to consider the ABC contrast:

$$[ABC] = -e + a + bd - abde + cd - acde - bce + abc$$

where we note that all data with E at the high level, i.e., e, $abde$, $acde$ and bce, appear with -1 as coefficient, while the remainder, i.e. a, bd, cd and abc appear with $+1$. The contrast therefore contains a contribution of $-4(E_1) + 4(-E_1) = -8E_1$ from the factor E.

The suggested experiment could be done in two blocks of 4 by for example confounding the AB interaction with blocks. That would give the grouping:

block 0				block 1			
abc	e	cd	$abde$	a	bce	bd	$acde$

The confoundings in this experiment would be:

$$\begin{array}{lllllllll}
\mathrm{I} & = & -BCD & = & -ABCE & = & ADE & & \\
A & = & -ABCD & = & -BC\mathrm{E} & = & DE & & \\
B & = & -CD & = & -ACE & = & ABDE & & \\
AB & = & -ACD & = & -CE & = & BDE & = & \text{Blocks} \\
C & = & -BD & = & -ABE & = & ACDE & & \\
AC & = & -ABD & = & -BE & = & CDE & & \\
BC & = & -D & = & -AE & = & ABCDE & & \\
ABC & = & -AD & = & -E & = & BCDE & &
\end{array}$$

where the contrasts are calculated as previously, but where the contrast that appears in the AB row now contains possible factor effects as well as the block effects.

End of example 2.2

2.5 Factors on 2 and 4 levels

In many cases where several factors are analysed, it can be desirable and perhaps even necessary for single factors that they can appear on 3 or perhaps 4 levels together with the 2 levels of the other factors. In case there is a need for a mixture of 2 and 3 levels, it is difficult to construct good experimental designs, but in the textbook by Oscar Kempthorne (1952): The Design and Analysis of Experiments, Wiley, New York, there are however some suggestions for this.

If 4 levels are used, one can use the procedure below, which is demonstrated with the help of two examples, so the presentation is not too complicated.

Example 2.3 : A 2×4 experiment in 2 blocks

Suppose that in a factorial experiment two factors are to be analysed, namely a factor A that appears on 2 levels and a factor G that appears on 4 levels. The mathematical model for the experiment is

$$Y_{il\nu} = \mu + A_i + G_\ell + AG_{i\ell} + E_{i\ell\nu}$$

where $i = (0, 1)$, $\ell = (0, 1, 2, 3)$ and $\nu = (1, 2, \ldots, r)$.

Usual parameter restrictions are used

$$\sum_{i=0}^{1} A_i = \sum_{\ell=0}^{3} G_\ell = \sum_{i=0}^{1} AG_{i\ell} = \sum_{\ell=0}^{3} AG_{i\ell} = 0$$

To reformulate the model to a 2^k factorial structure, two new factors are introduced, B and C, as replacements for G.

G = 0	G = 1	G = 2	G = 3
B = 0 , C = 0	B = 1 , C = 0	B = 0 , C = 1	B = 1 , C = 1

$$Y_{ij\nu} = \mu + A_i + B_j + AB_{ij} + C_k + AC_{ik} + BC_{jk} + ABC_{ijk} + E_{ijk\nu}$$

where the index $j =$ remainder of $(\ell/2)$ and $k =$ integer part of $(\ell/2)$. Inversely, $\ell = j + 2k$.

The correspondence between factor G and the two artificial factors B and C is that

$$G_\ell = B_j + C_k + BC_{jk} \quad , \quad \ell = j + 2k$$

Suppose now that one wants to do a complete factorial experiment with the two factors A and G, i.e. a 2×4 experiment, or a total of 8 single experiments.

Suppose further that one wants to do the experiment in 2 blocks with 4 single experiments. In the reformulated model, where factor G is replaced by B and C, we see that the main effect of factor G is given as B+C+BC. The effects B, C and BC must therefore be estimated and cannot be used as **defining contrast** when dividing into blocks.

For the interaction between factor A and factor G, it holds true that

$$AG_{i\ell} = AB_{ij} + AC_{ik} + ABC_{ijk} \quad , \quad \ell = j + 2k$$

and one of these three effects (it does not matter which) can be reasonably used as defining contrast. We choose for example the AB effect.

The experiment thereby is

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

or converted to factors A and G:

Block 1				Block 2			
(1)	ab	c	abc	a	b	ac	bc
A = 0	A = 1	A = 0	A = 1	A = 1	A = 0	A = 1	A = 0
G = 0	G = 1	G = 2	G = 3	G = 0	G = 1	G = 2	G = 3

The experiment can be analysed with Yates' algorithm, and one gets a table of analysis of variance which (in outline) is built up as the following:

Source of variation	Sum of Squares	Degrees of dom	S^2	F-value
A	SSQ_A	1	S^2_A	
G	$\mathrm{SSQ}_B + \mathrm{SSQ}_C + \mathrm{SSQ}_{BC}$	3	S^2_G	
AG-unconfounded	$\mathrm{SSQ}_{AC} + \mathrm{SSQ}_{ABC}$	2	S^2_{AG}	
Blocks + AG	SSQ_{AB}	1	$S^2_{AG+\text{blocks}}$	
Possible residual from previous exp.				
Total				

One sees that some of the variation arising from the AG interaction can be taken out and tested, while the remaining part is confounded with blocks. On the other hand, one cannot estimate specific AG interaction effects, since the part described by the AB part cannot be estimated (AG=AB+AC+ABC).

End of example 2.3

Example 2.4 : A fractional 2×2×4 factorial design

If for example one wants to evaluate three factors A, B and G with 2, 2, and 4 levels respectively, two new artificial factors are introduced, C and D, so that G=C+D+CD, and in this case one must keep the three effects C, D and CD clear of confoundings. It could be wished to do such a $2 \times 2 \times 4$ design with a total of 16 possible single experiments as a $\frac{1}{2} \times 2^4$ experiment, using only 8 single experiments.

If it is assumed that the three factors A, B and G do not interact, the experiment can be constructed so that the effects A, B, C, D and CD can be found (as G=C+D+CD).

One can use the following defining contrast and alias relations:

$$\begin{array}{lcl}
\underline{I} & = & ABCD \\
\underline{A} & = & BCD \\
\underline{B} & = & ACD \\
AB & = & \underline{CD} \\
\underline{C} & = & ABD \\
AC & = & BD \\
BC & = & AD \\
ABC & = & \underline{D}
\end{array}$$

where effects that are interesting are underlined, while effects considered to be without interest are written normally.

The relation between the factors is that

$$\begin{array}{lcl}
G = C + D + CD & , & AG = AC + AD + ACD \quad , \\
BG = BC + BD + BCD & \text{and} & ABG = ABC + ABD + ABCD
\end{array}$$

The experiment wanted could be the following (try to construct it yourself!):

$$\boxed{a \quad b \quad c \quad abc \quad d \quad abd \quad acd \quad bcd}$$

or converted to the levels of the factors, as the index for the factor G is $k + 2\ell$, where k is the index for C while ℓ is the index for D:

a	b	c	abc	d	abd	acd	bcd
A = 1	A = 0	A = 0	A = 1	A = 0	A = 1	A = 1	A = 0
B = 0	B = 1	B = 0	B = 1	B = 0	B = 1	B = 0	B = 1
G = 0	G = 0	G = 1	G = 1	G = 2	G = 2	G = 3	G = 3

Of course, one could also have used the complementary experiment as the starting point:

$$\boxed{(1) \quad ab \quad ac \quad bc \quad ad \quad bd \quad cd \quad abcd}$$

Try to write the corresponding experiment out in factors A, B and G.

If the constructed experiment should be laid out in two blocks of 4 single experiments, one could use either AC or BC as defining contrast. If AC is used, one gets the design:

Block 1			
b	abc	d	acd
A = 0	A = 1	A = 0	A = 1
B = 1	B = 1	B = 0	B = 0
G = 0	G = 1	G = 2	G = 3

Block 2			
a	c	abd	bcd
A = 1	A = 0	A = 1	A = 0
B = 0	B = 0	B = 1	B = 1
G = 0	G = 1	G = 2	G = 3

Note, for example, that all three factors A, B and G are balanced within both blocks.

End of example 2.4

3 General methods for p^k-factorial designs

In this chapter we will introduce general methods for factorial experiments in which there are k factors, each on p levels. The purpose of this is to generalize the concepts and methods that were discussed in the previous chapter, where we considered k factors of which each was on only $p = 2$ levels.

In particular, we will look at experiments with many factors that have to be evaluated on 2 or 3 levels, which are most relevant in practice.

In general, no proofs are given, but the subject is presented through examples and direct demonstration in specific cases.

The method we will deal with is often called Kempthorne's method, and the interested reader is referred to the text book by Oscar Kempthorne (1952): The Design and Analysis of Experiments, Wiley, New York. This book has a somewhat more mathematical review of the experimental structures and models that we will deal with here. In fairness, it should be said that it was actually R. A. Fischer and others who, around 1935, formulated important parts of the basis for Kempthorne's presentation.

3.1 Complete p^k factorial experiments

We now consider experiments with k factors each on p levels, where p is everywhere assumed to be a prime number. In addition, complete randomisation is generally assumed. In cases where experiments are discussed in which there is used blocking, complete randomisation is assumed within blocks.

The factors are always called A, B, C, etc. Factor A is the first factor, B the second factor etc. In addition (to the greatest possible extent) we use the indices i, j, k, etc. for the factors A, B, C, etc., respectively.

The experiment is generally called a p^k factorial experiment, and the number of possible different factor combinations is precisely $p \times p \times \ldots \times p = p^k$.

For an experiment with 3 factors, A, B, and C, the standard mathematical model is:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i,j} + C_k + AC_{i,k} + BC_{j,k} + ABC_{i,j,k} + E_{ijk\nu}$$

where $i, j, k = (0, 1, .., p-1)$ and $\nu = (1, 2, .., r)$.

The index $\nu = (1, 2, .., r)$ gives the number of repetitions of each single experiment in the experiment. The other indices assume the values $(0,\ 1,\ 2,..,\ p-1)$. It should be noted that the index always runs from 0 up to and including $p - 1$.

For such experiments we introduce a standard notation for the single experiments in the same way as with the 2^k experiment. In the case where $p = 3$ and $k = 3$ we have the following table, which shows all the single experiments in the complete 3^3 factorial experiment:

	A (C=0)			A (C=1)			A (C=2)		
	0	1	2	0	1	2	0	1	2
B=0	(1)	a	a^2	c	ac	a^2c	c^2	ac^2	a^2c^2
B=1	b	ab	a^2b	bc	abc	a^2bc	bc^2	abc^2	a^2bc^2
B=2	b^2	ab^2	a^2b^2	b^2c	ab^2c	a^2b^2c	b^2c^2	ab^2c^2	$a^2b^2c^2$

As previously, we use one of these expressions as the term for a certain "treatment" or factor combination in a single experiment, as well as for the total response from the single experiments done with this factor combination. Thus, for example

$$ab^2c = \sum_{\nu=1}^{r} Y_{121\nu} = T_{121\cdot} \quad \text{or just} \quad T_{121}$$

For the 2^k factorial experiment, we arranged these terms in what was called a standard order. We can also do this for the p^k experiment in general. These standard orders are:

$2^k : (1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, \ldots$
$3^k : (1), a, a^2, b, ab, a^2b, b^2, ab^2, a^2b^2, c, ac, a^2c, .., a^2b^2c^2, d, \ldots$
$5^k : (1), a, a^2, a^3, a^4, b, ab, a^2b, a^3b, a^4b, b^2, ab^2, a^2b^2, a^3b^2, a^4b^2, \ldots, a^4b^4,$
$c, ac, a^2c, \ldots, a^4b^4c^4, d, ad, \ldots$
$7^k : (1), a, a^2, \ldots, a^6, b, ab, a^2b, \ldots, a^6b^6, c, ac, \ldots, a^6b^6c^6, d, ad, \ldots$

For example in the 3^k factorial experiment, a new factor is added by multiplying all the terms until now with the factor in the first power and in the second power and adding both these new rows to the original order.

These terms, of course, can perfectly well be used as names for factor combinations in completely general factorial experiments, but the results we will show are only generally applicable to experiments that can be formulated as p^k factorial experiments where p is a prime number.

Before we continue, it would be useful to look more closely at a 3^2 factorial experiment and show how the total variation in this experiment can be described and found with the help of a Graeco-Latin square. In addition we will introduce some mnemonic terms for new artificial effects, which will later prove to be practical in the construction of more sophisticated experimental designs.

Example 3.1 : Making a Graeco-Latin square in a 3^2 factorial experiment

The experiment has $3 \times 3 = 9$ different single experiments:

	A=0	A=1	A=2
B=0	(1)	a	a^2
B=1	b	ab	a^2b
B=2	b^2	ab^2	a^2b^2

The mathematical model for the experiment is

$$Y_{ij\nu} = \mu + A_i + B_j + AB_{i,j} + E_{ij\nu} \ , \ \ i = (0, 1, 2) \ , \ \ j = (0, 1, 2) \ , \ \ \nu = (1, \ldots, r)$$

$$\sum_{i=0}^{2} A_i = \sum_{j=0}^{2} B_j = \sum_{i=0}^{2} AB_{i,j} = \sum_{j=0}^{2} AB_{i,j} = 0$$

In this experiment we can introduce two artificial factors, which we can call X and Z. We let these factors have indices s and t, respectively, which we determine with

$$s = (i + j)_3 \ \text{ and } \ t = (i + 2j)_3$$

where the designation $(.)_3$ now stands for "modulo 3", i.e. "remainder of $(.)$ after division by 3".

We will now see how the indices s and t for the defined new effects X and Z vary throughout the experiment with the indices i and j of the two original factors A and B.

This is shown in the table below, as $i + j$ and $i + 2j$ are still calculated "modulo 3".

$$\underset{A_i}{\overset{i}{\boxed{\begin{array}{ccc} 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{array}}}} \quad \underset{B_j}{\overset{j}{\boxed{\begin{array}{ccc} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{array}}}} \quad \underset{X_{i+j}}{\overset{s=(i+j)_3}{\boxed{\begin{array}{ccc} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{array}}}} \quad \underset{Z_{i+2j}}{\overset{t=(i+2j)_3}{\boxed{\begin{array}{ccc} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{array}}}}$$

We note that if we fix one of the levels for one of the 4 indices, each of the other 3 indices appears precisely with the values 0, 1 and 2 within this level. As an example of this, we consider the single experiments where Z's index $t = (i + 2j)_3 = 1$:

$$\underset{A_i}{\overset{i}{\boxed{\begin{array}{ccc} & 1 & \\ & & 2 \\ 0 & & \end{array}}}} \quad \underset{B_j}{\overset{j}{\boxed{\begin{array}{ccc} & 0 & \\ & & 1 \\ 2 & & \end{array}}}} \quad \underset{X_{i+j}}{\overset{s=(i+j)_3}{\boxed{\begin{array}{ccc} & 1 & \\ & & 0 \\ 2 & & \end{array}}}} \quad \underset{Z_{i+2j}}{\overset{t=(i+2j)_3}{\boxed{\begin{array}{ccc} & 1 & \\ & & 1 \\ 1 & & \end{array}}}}$$

In the design shown, the 4 factors A, B, X and Z are obviously in balance in relation to each other. The design is a **Graeco-Latin square** with the introduced factors X and Z inside the square and with the factors A and B at the sides. Since the variation between the 9 single experiments or "treatments" in the experiment has a total of 9-1 degrees of freedom, and the 4 factors are in balance, as described, it can be shown that these 4 factors can describe the whole variation between the single experiments.

A and B are identical with the original main effects, and it can be shown that X and Z together precisely make up the interaction term $AB_{i,j}$ in the "natural" mathematical model of the experiment.

We will not prove this result, but only illustrate it with an example, where we imagine that a 3^2 experiment with one observation per cell has resulted in the following data:

	A=0	A=1	A=2	sum-B		sum-X		sum-Z
B=0	(1)=10	a=15	a^2=18	43	X=0	35	Z=0	33
B=1	b=8	ab=12	a^2b=16	36	X=1	34	Z=1	36
B=2	b^2=5	ab^2=9	a^2b^2=11	25	X=2	35	Z=2	35
sum-A	23	36	45	104		104		104

The usual two-sided analysis of variance for these data with the factors A and B gives:

ANOVA

Source of variation	Sum of squares	Degrees of freedom	F-value
A	81.556	(3-1)=2	
B	54.889	(3-1)=2	
AB	1.778	(3-1)(3-1)=4	
Residual	0.000	0	
Total	138.223	(9-1)=8	

To the right of the data table are sums for the two artificial factors X and Z. For example the (X = 0) sum is found as 10 + 16 + 9 = 35, i.e. the sum of the data where the index $(i + j)_3 = 0$ as X has the index $= (i + j)_3$.

From this we find the following sums of squares and degrees of freedom, where the 4 factors A, B, X and Z constitute a Graeco-Latin square:

SSQ(treatments)	=	$10^2 + 15^2 + 18^2 + 8^2 + .. + 11^2 - 104^2/9$	=	138.222 ,	f=9-1
SSQ(A)	=	$(23^2 + 36^2 + 45^2)/3 - 104^2/9$	=	81.556 ,	f=2
SSQ(B)	=	$(25^2 + 36^2 + 43^2)/3 - 104^2/9$	=	54.889 ,	f=2
SSQ(X)	=	$(35^2 + 34^2 + 35^2)/3 - 104^2/9$	=	0.222 ,	f=2
SSQ(Z)	=	$(33^2 + 36^2 + 35^2)/3 - 104^2/9$	=	1.556 ,	f=2
SSQ(A) + SSQ(B) + SSQ(X) + SSQ(Z)			=	138.223 ,	f=8

It is seen (except for the rounding) that for the interaction AB and corresponding degrees of freedom we have:

$$\text{SSQ}(AB\text{-interaction}) = \text{SSQ(X)} + \text{SSQ(Z)} \text{ , and } f(AB\text{-interaction}) = f(\text{X}) + f(\text{Z})$$

Further, it can be generally shown that for the interaction term it applies that

$$AB_{i,j} = X_{i+j} + Z_{i+2j} \quad , \quad i = (0,1,2) \ , \ j = (0,1,2)$$

This can be illustrated by finding the estimates for the interaction terms as well as for the artificial effects X and Z. As an example we can find the interaction estimate for (A=1,B=2), i.e. $AB_{1,2}$.

$$\hat{\mu} = 104/9 = 11.556 \quad , \quad \widehat{A}_1 = 36/3 - 104/9 = 0.444 \quad , \quad \widehat{B}_2 = 25/3 - 104/9 = -3.222$$

$$\Longrightarrow \widehat{AB}_{1,2} = Y_{1,2} - \hat{\mu} - \widehat{A}_1 - \widehat{B}_2 = 9.000 - 11.556 - 0.444 - (-3.222) = 0.222$$

$$\widehat{X}_{1+2} = \widehat{X}_3 \rightarrow \widehat{X}_0 = 35/3 - 104/9 = 0.111$$
$$\widehat{Z}_{1+2\cdot 2} = \widehat{Z}_5 \rightarrow \widehat{Z}_2 = 35/3 - 104/9 = 0.111$$

so that $\widehat{X}_{1+2} + \widehat{Z}_{1+2\cdot 2} = \widehat{AB}_{1,2}$, as postulated (remember that indices are still calculated "modulo 3"). Try to work out whether it is correct that $AB_{2,2} = X_{2+2} + Z_{2+2\cdot 2}$, when these are estimated.

In order to use the results of the example generally, it is practical to introduce some more mnemonic names for the two introduced effects X and Z. We thus set

$$X_{i+j} = AB_{i+j} \text{ and } Z_{i+2j} = AB^2_{i+2j}$$

Correspondingly, we write the original model on the form

$$Y_{ij\nu} = \mu + A_i + B_j + AB_{i+j} + AB^2_{i+2j} + E_{ij\nu}$$

where

$$AB_{i+j} + AB^2_{i+2j} = AB_{i,j} \quad , \quad i = (0,1,2) \ , \ j = (0,1,2)$$

It applies that with this new formal notation:

$$\sum_{i=0}^{2} A_i = \sum_{j=0}^{2} B_j = \sum_{r=0}^{2} AB_r = \sum_{s=0}^{2} AB^2_s = 0 \quad , \quad r = (i+j)_3 \ , \ s = (i+2j)_3$$

where all indices are still calculated "modulo 3".

The two effects AB_{i+j} and AB^2_{i+2j} in this way designate the artificially introduced effects, which enable a decomposition of the usual interaction term $AB_{i,j}$ from the traditional model formulation. The exponent "2" on AB^2 should only be considered as a mnemonic help and not as an expression of raising to a power of 2.

Finally note how the indices and names for the introduced artificial effects match.

End of example 3.1

Such artificial effects can be defined for general p^k factorial experiments. To be able to keep the effects in order, we introduce a "standard order" for effects. For an experiment where all factors have p levels, the artificial effects will likewise all have p levels:

$$2^k : I, A, B, AB, C, AC, BC, ABC, D, AD, BD, ABD, CD, ACD, BCD, ABCD, E, AE, ..$$

$$3^k : I, A, B, AB, AB^2, C, AC, AC^2, BC, BC^2, ABC, ABC^2, AB^2C, AB^2C^2, D, AD, AD^2, BD, BD^2, \ldots, AB^2C^2D^2, E, \ldots$$

$$5^k : I, A, B, AB, AB^2, AB^3, AB^4, C, AC, AC^2, \ldots, BC, \ldots, AB^4C, \ldots, AB^4C^4, D, \ldots$$

These effects have indices according to the same rules that were used in the previous example. That is, for example, that in the 5^k experiment with factors A, B and C, each with 5 levels, the effect AB^3C has index $= (i + 3j + k)_5$, i.e. $(i + 3j + k)$ **modulo 5**. Factor A is the first factor, B the second factor and C the third factor.

Note that this standard order can be derived from the standard order for single experiments by changing to upper-case letters and leaving out the terms where the exponent on the first factor in the effect is greater than 1. For example, AB^3C should be included, while for example B^2CD should be left out.

Example 3.2 : Latin cubes in 3^3 experiments

Let there be a completely randomised 3^3 experiment with r repetitions of each single experiment. We have in the usual model formulation:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i,j} + C_k + AC_{i,k} + BC_{j,k} + ABC_{i,j,k} + E_{ijk\nu}$$
$$\text{where} \quad i = (0,1,2) \ , \ j = (0,1,2) \ , \ k = (0,1,2) \ \text{ and } \ \nu = (1,2,\ldots,r)$$

The cells of the experiment or single experiments make up a cube, the length of its edge being 3.

It thus looks like this:

C=0

	A=0	A=1	A=2
B=0	(1)	a	a^2
B=1	b	ab	a^2b
B=2	b^2	ab^2	a^2b^2

C=1

	A=0	A=1	A=2
B=0	c	ac	a^2c
B=1	bc	abc	a^2bc
B=2	b^2c	ab^2c	a^2b^2c

C=2

	A=0	A=1	A=2
B=0	c^2	ac^2	a^2c^2
B=1	bc^2	abc^2	a^2bc^2
B=2	b^2c^2	ab^2c^2	$a^2b^2c^2$

With the stated arithmetic rules for indices used on the standard order for the introduced artificial effects for a 3^3 factorial experiment, we can now find the index values for all the effects. The index values are found as shown in the following tables:

index for A_i

	k=0: i=0	i=1	i=2	k=1: i=0	i=1	i=2	k=2: i=0	i=1	i=2
j=0	0	1	2	0	1	2	0	1	2
j=1	0	1	2	0	1	2	0	1	2
j=2	0	1	2	0	1	2	0	1	2

index for B_j

	k=0: i=0	i=1	i=2	k=1: i=0	i=1	i=2	k=2: i=0	i=1	i=2
j=0	0	0	0	0	0	0	0	0	0
j=1	1	1	1	1	1	1	1	1	1
j=2	2	2	2	2	2	2	2	2	2

index for AB_{i+j}

	k=0: i=0	i=1	i=2	k=1: i=0	i=1	i=2	k=2: i=0	i=1	i=2
j=0	0	1	2	0	1	2	0	1	2
j=1	1	2	0	1	2	0	1	2	0
j=2	2	0	1	2	0	1	2	0	1

index for AB^2_{i+2j}

	k=0: i=0	i=1	i=2	k=1: i=0	i=1	i=2	k=2: i=0	i=1	i=2
j=0	0	1	2	0	1	2	0	1	2
j=1	2	0	1	2	0	1	2	0	1
j=2	1	2	0	1	2	0	1	2	0

index for C_k

	k=0: i=0	i=1	i=2	k=1: i=0	i=1	i=2	k=2: i=0	i=1	i=2
j=0	0	0	0	1	1	1	2	2	2
j=1	0	0	0	1	1	1	2	2	2
j=2	0	0	0	1	1	1	2	2	2

index for AC_{i+k}

	k=0: i=0	i=1	i=2	k=1: i=0	i=1	i=2	k=2: i=0	i=1	i=2
j=0	0	1	2	1	2	0	2	0	1
j=1	0	1	2	1	2	0	2	0	1
j=2	0	1	2	1	2	0	2	0	1

	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0	0	1	2	2	0	1	1	2	0
j=1	0	1	2	2	0	1	1	2	0
j=2	0	1	2	2	0	1	1	2	0
		k=0			k=1			k=2	

index for AC^2_{i+2k}

	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0	0	0	0	1	1	1	2	2	2
j=1	1	1	1	2	2	2	0	0	0
j=2	2	2	2	0	0	0	1	1	1
		k=0			k=1			k=2	

index for BC_{j+k}

	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0	0	0	0	2	2	2	1	1	1
j=1	1	1	1	0	0	0	2	2	2
j=2	2	2	2	1	1	1	0	0	0
		k=0			k=1			k=2	

index for BC^2_{j+2k}

	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0	0	1	2	1	2	0	2	0	1
j=1	1	2	0	2	0	1	0	1	2
j=2	2	0	1	0	1	2	1	2	0
		k=0			k=1			k=2	

index for ABC_{i+j+k}

	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0	0	1	2	2	0	1	1	2	0
j=1	1	2	0	0	1	2	2	0	1
j=2	2	0	1	1	2	0	0	1	2
		k=0			k=1			k=2	

index for ABC^2_{i+j+2k}

	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0	0	1	2	1	2	0	2	0	1
j=1	2	0	1	0	1	2	1	2	0
j=2	1	2	0	2	0	1	0	1	2
		k=0			k=1			k=2	

index for AB^2C_{i+2j+k}

	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0	0	1	2	2	0	1	1	2	0
j=1	2	0	1	1	2	0	0	1	2
j=2	1	2	0	0	1	2	2	0	1
		k=0			k=1			k=2	

index for $AB^2C^2_{i+2j+2k}$

For example, we can look at the term BC_{j+k} and note where it has the index value 1. This is stated below:

	k=0			k=1			k=2		
	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0				1	1	1			
j=1	1	1	1						
j=2							1	1	1

index for BC_{j+k}

For any other arbitrarily chosen term, there will be an equal number of 0's, 1's and 2's in these 9 places. As an example, we take the term ABC^2, where the corresponding places are shown below:

	k=0			k=1			k=2		
	i=0	i=1	i=2	i=0	i=1	i=2	i=0	i=1	i=2
j=0				2	0	1			
j=1	1	2	0						
j=2							0	1	2

index for ABC^2_{i+j+2k}

We have thus constructed 13 effects, each with 3 levels (0, 1, and 2), which are in balance with each other in the same way as with the Graeco-Latin square in the example mentioned earlier.

It can be derived from this that we can re-write our original model with the help of the new artificial effects:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i+j} + AB^2_{i+2j} + C_k + AC_{i+k} + AC^2_{i+2k} + BC_{j+k}$$
$$+BC^2_{j+2k} + ABC_{i+j+k} + ABC^2_{i+j+2k} + AB^2C_{i+2j+k} + AB^2C^2_{i+2j+2k} + E_{ijk\nu}$$

where the terms in the model are decompositions of the conventional model terms:

$$\begin{array}{lcl}
A_i & \Rightarrow & A_i \\
B_j & \Rightarrow & B_j \\
AB_{i,j} & \Rightarrow & AB_{i+j} + AB^2_{i+2j} \\
C_k & \Rightarrow & C_k \\
AC_{i,k} & \Rightarrow & AC_{i+k} + AC^2_{i+2k} \\
BC_{j,k} & \Rightarrow & BC_{j+k} + BC^2_{j+2k} \\
ABC_{i,j,k} & \Rightarrow & ABC_{i+j+k} + ABC^2_{i+j+2k} + AB^2C_{i+2j+k} + AB^2C^2_{i+2j+2k}
\end{array}$$

with the usual meaning to the left and the artificial effects to the right.

In the 3^3 experiment, there are 27 cells or single experiments. To describe the mean values in these cells, 27 parameters should be used, of which one is μ, so that there should be $27 - 1 = 26$ degrees of freedom for factor effect.

The 13 terms in the standard order all have 3 levels, which sum up to 0. Thus $3 - 1 = 2$ free parameters (degrees of freedom) are connected to each of the 13 terms, or a total of $13 \times (3 - 1) = 26$ free parameters (degrees of freedom).

It is further seen that, because of the balance, all parameters are estimated by forming the average in the same way as for the main effects and correcting with the total average.

For example:

$$\widehat{AC_0^2} = \frac{1}{3^{3-1}} \sum_i \sum_j \sum_k \bar{Y}_{ijk.} \times \delta_{i+2k,0} - \bar{Y}_{....} \ ,$$

where

$$\delta_{r,s} = \begin{cases} 1 & \text{for} \quad r = s \\ 0 & \text{for} \quad r \neq s \end{cases}$$

as the indicator $\delta_{i+2k,0}$ points out the data where the index for AC^2 is 0 (zero), i.e. $(i + 2k)_3{=}0$, while $\bar{Y}_{ijk.}$ gives the average response in cells (i, j, k), and $\bar{Y}_{....}$ gives the average response for the whole experiment.

Thus, in order to estimate AC_0^2 the cells where the corresponding index, namely, $i + 2k$ modulo 3 is 0 (zero) are included. Correspondingly, $i + 2k$ has to be 1, respectively 2 to go into the estimates for AC_1^2 and AC_2^2, respectively:

$$\widehat{AC_0^2} = \frac{1}{9}(\bar{Y}_{000.} + \bar{Y}_{010.} + \bar{Y}_{020.} + \bar{Y}_{101.} + \bar{Y}_{111.} + \bar{Y}_{121.} + \bar{Y}_{202.} + \bar{Y}_{212.} + \bar{Y}_{222.}) - \bar{Y}_{....}$$

$$\widehat{AC_1^2} = \frac{1}{9}(\bar{Y}_{100.} + \bar{Y}_{110.} + \bar{Y}_{120.} + \bar{Y}_{201.} + \bar{Y}_{211.} + \bar{Y}_{221.} + \bar{Y}_{002.} + \bar{Y}_{012.} + \bar{Y}_{022.}) - \bar{Y}_{....}$$

$$\widehat{AC_2^2} = \frac{1}{9}(\bar{Y}_{200.} + \bar{Y}_{210.} + \bar{Y}_{220.} + \bar{Y}_{001.} + \bar{Y}_{011.} + \bar{Y}_{021.} + \bar{Y}_{102.} + \bar{Y}_{112.} + \bar{Y}_{122.}) - \bar{Y}_{....}$$

End of example 3.2

3.2 Calculations based on Kempthorne's method

We have seen with examples that the introduced new effects/parameters, which obviously do not refer directly for example to certain treatments (apart from the main effects), give rise to a **mathematical decomposition** of the interactions. This procedure and the methods derived from it are generally called "Kempthorne's" method, after the name of the statistician to whom its origin is often ascribed, and who has described it (cf. the list of literature suggestions at the beginning of these notes).

We have illustrated that the corresponding estimates are independent because of the described balance, and finally we saw in example 3.1, page 48, that we can also form sums of square, which are independent and sum to the total sum of squares.

We now consider an arbitrarily chosen effect in the standard order. We generally call this effect F:

$$F_t = A^{\alpha} B^{\beta} \dots C_t^{\gamma}$$

where the index is

$$t = i \cdot \alpha + j \cdot \beta + \ldots + k \cdot \gamma \quad , \qquad \text{modulo } p$$

With the notation introduced, where $Y_{ij\ldots k\nu}$ gives the response in the single experiment no. ν with the factor combination $(ij \ldots k)$, we have that

$$T_{ij\ldots k} = a^i b^j \ldots c^k = \sum_{\nu=1}^{r} Y_{ij\ldots k\nu}$$

and estimates are:

$$\widehat{F}_l = \frac{\sum_{ij\ldots k} T_{ij\ldots k} \times \delta_{l,t}}{N/p} - \frac{\sum_{ij\ldots k} T_{ij\ldots k}}{N}$$

$$\text{for} \quad l = (0,1,\ldots,p-1) \;\; , \;\; \text{where} \quad t = (i \cdot \alpha + j \cdot \beta + \ldots + k \cdot \gamma)_p \quad \text{and} \quad N = r \cdot p^k$$

$$\mathrm{SSQ}(F) = \frac{\sum_{l=0}^{p-1} \left(\sum_{ij\ldots k} T_{ij\ldots k} \times \delta_{l,t} \right)^2}{N/p} - \frac{\left(\sum_{ij\ldots k} T_{ij\ldots k} \right)^2}{N} = (N/p) \cdot \sum_{l=0}^{p-1} \widehat{F}_l^2$$

where we still use the indicator

$$\delta_{r,s} = \begin{cases} 1 & \text{for} \quad r = s \\ 0 & \text{for} \quad r \neq s \end{cases} \qquad ,$$

The effect estimate can be expressed in words

$$\widehat{F}_l = \frac{\text{sum of data, where } t = l}{\text{number of data, where } t = l} - \text{average af all data} \;\; , \;\; l = (0,1,\ldots,p-1)$$

Example 3.3 : Estimation and SSQ in the 3^2-factorial experiment

If we again consider the 3^2 experiment in example 3.1 page 48, where we let T_{ij} give the sum of data in cells (i, j) for example, we find:

$$\widehat{A}_i = \frac{T_{i0} + T_{i1} + T_{i2}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\text{SSQ}(A) = r \cdot 3^{2-1} \sum_{i=0}^{2} \widehat{A}_i^2$$

$$\widehat{B}_j = \frac{T_{0j} + T_{1j} + T_{2j}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\widehat{AB}_0 = \frac{T_{00} + T_{21} + T_{12}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\widehat{AB}_1 = \frac{T_{10} + T_{01} + T_{22}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\widehat{AB}_2 = \frac{T_{20} + T_{11} + T_{02}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\text{SSQ}(AB) = r \cdot 3((\widehat{AB}_0)^2 + (\widehat{AB}_1)^2 + (\widehat{AB}_2)^2)$$

$$\widehat{AB^2}_0 = \frac{T_{00} + T_{11} + T_{22}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\widehat{AB^2}_1 = \frac{T_{10} + T_{21} + T_{02}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\widehat{AB^2}_2 = \frac{T_{20} + T_{01} + T_{12}}{r \cdot 3^{2-1}} - \frac{T..}{r \cdot 3^2}$$

$$\text{SSQ}(AB^2) = r \cdot 3((\widehat{AB^2}_0)^2 + (\widehat{AB^2}_1)^2 + (\widehat{AB^2}_2)^2)$$

where the innermost 2 exponent is symbolic-mnemonic, while the outermost here is the usually squaring.

End of example 3.3

3.3 General formulation of interactions and artificial effects

Consider a p^k factorial experiment with factors A, B, ..., C, and p is a prime number.

The interaction between any factors is decomposed as we have seen in the previous examples:

$$AB_{i,j} = AB_{i+j} + AB^2_{i+2j} + \ldots + AB^{p-1}_{i+(p-1)j}$$

A general notation can be introduced for the interaction effects in a factor structure by introducing an operator "$\times$" in the following way

$$A \times B = AB + AB^2 + \ldots + AB^{p-1}$$

and $A \times I = A$. Later we will need the further arithmetic rule that

$$(A + B)^\alpha = A^\alpha + B^\alpha$$

so that for example

$$(A \times B)^\alpha = (AB + AB^2 + \ldots + AB^{p-1})^\alpha = (AB)^\alpha + (AB^2)^\alpha + \ldots + (AB^{p-1})^\alpha$$

In addition an even more general operator "$*$" can be introduced, working in the following way:

$$A * B = A + B + A \times B$$

In this way the operator "$*$" generates all the terms in the standard order for the factors on which it works.

For any complete p^k factorial experiment, the factor model can then be written

$$\begin{aligned} Y &= \mu + A * B * \ldots * C + E \\ &= \mu + A + B + A \times B + \ldots + C + A \times C + B \times C + \ldots + A \times B \times \ldots \times C + E \end{aligned}$$

If $p = 3$, the decomposition is as previously. For example

$$\begin{aligned} A \times B &= AB + AB^2 \\ A \times B \times C &= (AB + AB^2) \times C = ABC + ABC^2 + AB^2C + AB^2C^2 \end{aligned}$$

Suppose one begins with B as the first factor. In the 3^2 case this would give the decomposition

$$BA_{j,i} = B \times A = BA_{j+i} + BA^2_{j+2i}$$

but then it emerges that the indices for this set of artificial effects vary synchronously with the indices for the effects AB_{i+j} and AB^2_{i+2j}.

We can illustrate this with the following

Example 3.4 : Index variation with inversion of the factor order

Take a 3^2 experiment and consider the following table:

A	B	AB	AB^2	BA	BA^2
i	j	$i+j$	$i+2j$	$j+i$	$j+2i$
0	0	0	0	0	0
1	0	1	1	1	2
2	0	2	2	2	1
0	1	1	2	1	1
1	1	2	0	2	0
2	1	0	1	0	2
0	2	2	1	2	2
1	2	0	2	0	1
2	2	1	0	1	0

Note that for the indices of the two terms AB^2 and BA^2, it applies that

$$(i+2j)_3 = 0 \Longleftrightarrow (j+2i)_3 = (2i+j)_3 = 0$$
$$(i+2j)_3 = 1 \Longleftrightarrow (j+2i)_3 = (2i+j)_3 = 2$$
$$(i+2j)_3 = 2 \Longleftrightarrow (j+2i)_3 = (2i+j)_3 = 1$$

as we still calculate "modulo 3".

This means that the indices for AB^2 and BA^2 vary synchronously so that $AB^2_0 \equiv BA^2_0$, $AB^2_1 \equiv BA^2_2$, and $AB^2_2 \equiv BA^2_1$ or said in another way: In order to extract the proper sum of squares we only need one of them, of which we have chosen AB^2.

End of example 3.4

The example illustrates the rule that we should only include effects where the exponent on the first factor in the effect equals 1. In certain situations, however, one can come to effects that do not fulfil this condition. But fortunately it is easy to find the effect from the standard order with which it can be replaced.

3.4 Standardisation of general effects

We consider a general non-standardised effect. As example we can take an effect such as $CA^2B^4D^3$ from a 3^k factorial experiment. To find the effect from the standard order with an index variation that varies synchronously with this, one proceeds as follows:

1. Arrange the factors in the factor order A, B, C, D, ... etc
$$CA^2B^4D^3 \longrightarrow A^2B^4CD^3$$
2. Reduce all exponents modulo p
$$A^2B^4CD^3 \longrightarrow A^2BCD^0 \longrightarrow A^2BC \ (p\text{=3 her})$$
3. If the exponent on the first factor in the effect (here A) is 1, we are finished. Otherwise, the whole effect is lifted to the second power and the exponents are again reduced modulo p:
$$A^2BC \longrightarrow A^4B^2C^2 \longrightarrow AB^2C^2$$
4. Step 3 is repeated until the first factor in the effect has the exponent 1.

With the help of this algorithm, one can always make the exponent on the first factor in a Kempthorne effect be 1, and by using a fixed order of factors, one gets an unequivocal standard order. For example we see that the index for the effect $CA^2B^4D^3$ varies synchronously with the index for the effect AB^2C^2 - calculated modulo 3.

Example 3.5 : Generalised interactions and standardisation

If we have two effects in a 3^5 experiment, for example ABC^2 and $ABDE$, we find their generalised interaction as (where $p = 3$)

$$ABC^2 \times ABDE = ABC^2(ABDE) + ABC^2(ABDE)^2 = ABCD^2E^2 + CDE$$

where we have also used the above-mentioned squaring method to get the exponent 1 on the first factor in the effect.

In this connection it can be useful to remember that in a square experiment, one factor can be moved "out of the square" and out to the edge, whereby the edge in question "moves into the square". From the 3^2 experiment dealt with in example 3.1 page 48:

$$
\begin{array}{cc}
i \rightarrow & \\
j & 0 \\
\downarrow & 1 \\
 & 2 \\
 &
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{array}} \\
A_i
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{array}} \\
B_j
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{array}} \\
AB_{i+j}
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{array}} \\
AB^2_{i+2j}
\end{array}
$$

The main effects A_i and B_j constitute the edges in the square and AB_{i+j} and AB^2_{i+2j} are "inside the square". If we now move AB_{i+j} and AB^2_{i+2j} out to the sides, we see that A_i and B_j move "into the square":

$$
\begin{array}{cc}
i+j \rightarrow & \\
i+2j & 0 \\
\downarrow & 1 \\
 & 2 \\
 &
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{array}} \\
AB_{i+j}
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{array}} \\
AB^2_{i+2j}
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 2 \end{array}} \\
A_i
\end{array}
\quad
\begin{array}{c}
0 \;\; 1 \;\; 2 \\
\boxed{\begin{array}{ccc} 0 & 2 & 1 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{array}} \\
B_j
\end{array}
$$

Examples: $(i + j = 1 \text{ and } i + 2j = 2) \Rightarrow (i = 0 \text{ and } j = 1)$, $(i + j = 2 \text{ and } i + 2j = 0) \Rightarrow (i = 1 \text{ and } j = 1)$.

$$AB \times AB^2 = AB(AB^2) + AB(AB^2)^2 = A^2B^3 + A^3B^5 = A^2 + B^2 = A + B$$

Therefore, by using the rules above, we could have foreseen that A and B would come into the square as generalised interactions for the artificial effects AB and AB^2.

The four effects A, B, AB and AB^2 together form the elements in what is called a group. In brief, it distinguishes itself by the fact that with the introduced arithmetic rules we can create new elements from other elements, and all elements created will belong to the group.

End of example 3.5

Example 3.6 : Latin squares in 2^3 factorial experiments and Yates' algorithm

In chapter 2 we went through the 2^k experiment, while here - exemplified with the 3^k experiment - we have introduced more general p^k experiments. We will now show briefly how the introduced more general methods look in a 2^k experiment.

With three factors, A, B, and C, the mathematical model in the introduced formulation, with $p = 2$, is:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i+j} + C_k + AC_{i+k} + BC_{j+k} + ABC_{i+j+k} + E_{ijk\nu}$$

where $i, j, k = (0, 1)$ and $\nu = (1, .., r)$.

The usual restrictions are:

$$\sum_{i=0}^{1} A_i = \sum_{j=0}^{1} B_j = \sum_{i+j=0}^{1} AB_{i+j} = \sum_{k=0}^{1} C_k = \sum_{i+k=0}^{1} AC_{i+k} = \sum_{j+k=0}^{1} BC_{j+k} = \sum_{i+j+k=0}^{1} ABC_{i+j+k} = 0$$

The connection between the traditional model formulation and the formulation introduced here is (as also shown on page 54 for the 3^k experiment) that (with the usual formulation to the left and the new formulation to the right):

$$\begin{array}{lcl}
A_i & \Rightarrow & A_i \\
B_j & \Rightarrow & B_j \\
AB_{i,j} & \Rightarrow & AB_{i+j} \\
C_k & \Rightarrow & C_k \\
AC_{i,k} & \Rightarrow & AC_{i+k} \\
BC_{j,k} & \Rightarrow & BC_{j+k} \\
ABC_{i,j,k} & \Rightarrow & ABC_{i+j+k}
\end{array}$$

If data are analysed with the help of Yates' algorithm, one must ensure that the effect estimates get the correct sign. Yates' algorithm always gives estimates corresponding to the level where all factors in the parameter are on level "1". For the ABC interaction, Yates' algorithm gives that

$$\widehat{ABC}_{1,1,1} = [ABC \text{ kontrast}]/(2^k \cdot r)$$

If the data are analysed according to the introduced model, it is found that

$$\widehat{ABC}_{1,1,1} \Rightarrow \widehat{ABC}_{1+1+1} \rightarrow \widehat{ABC}_3 \rightarrow \widehat{ABC}_1$$

that is, that the algorithm finds the ABC_{i+j+k}-parameter level "1".

If on the other hand, the interaction AB is considered, Yates' algorithm finds

$$\widehat{AB}_{1,1} \Rightarrow \widehat{AB}_{1+1} \rightarrow \widehat{AB}_2 \rightarrow \widehat{AB}_0 = -\widehat{AB}_1$$

that is plus AB_{i+j} parameter level "0" or minus its level "1".

It thus generally applies that Yates' algorithm used for a 2^k factorial experiment for the introduced general effects with an **uneven** number of factors gives the parameters' level "1", while the algorithm for parameters with an **eve**n number of factors gives the parameters' level "0".

End of example 3.6

3.5 Block-confounded p^k factorial experiment

In this section we will generalise the methods that were introduced in section 2.2 page 18, and we start with the following

Example 3.7 : 2^3 **factorial experiment in 2 blocks of 4 single experiments**

We consider a 2^3 factorial experiment with factors A, B and C and with r repetitions per factor combination.

The traditional mathematical model for this experiment is

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i,j} + C_k + AC_{i,k} + BC_{j,k} + ABC_{i,j,k} + E_{ijk\nu}$$

where $i, j, k = (0, 1)$ and $\nu = (1, .., r)$

We have previously seen that such an experiment can be laid out in two blocks by choosing to confound one of the factor effects with blocks, and we have seen that this is formalised by choosing a defining contrast. The effect corresponding to this will be confounded with blocks. In order to use the introduced method for analysis of p^k factorial experiments, we will write the model on the general form, which for $p = 2$ is:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i+j} + C_k + AC_{i+k} + BC_{j+k} + ABC_{i+j+k} + E_{ijk\nu}$$

where $i, j, k, = (0, 1)$ and $\nu = (1, ..., r)$

To divide the experiment into two parts, we now choose a

$$\text{Defining relation : } I = ABC$$

where, as an example, we choose to confound the 3-factor interaction ABC with blocks.

This effect has index $= (i + j + k)_2$, which thus takes the values 0 or 1. We let the block number follow this index, i.e. that in Block 0 are placed the experiments where it applies that $(i + j + k)_2 = 0$. Correspondingly, experiments where $(i + j + k)_2 = 1$ are put in block 1. To find the principal block, we must in other words find all the solutions to the equation:

$$(i+j+k) \ \ (\text{modulo } 2) \ \ = 0$$

We try :

$$\begin{aligned} i=0 \ , \ j=0 &\Longrightarrow k=0 : \text{ experiment } = (1) \\ i=1 \ , \ j=0 &\Longrightarrow k=1 : \text{ experiment } = ac \\ i=0 \ , \ j=1 &\Longrightarrow k=1 : \text{ experiment } = bc \\ i=1 \ , \ j=1 &\Longrightarrow k=0 : \text{ experiment } = ab \end{aligned}$$

The last solution could be found by adding the two previous solutions to each other:

$$\begin{array}{ccccccccc|c} i & & j & & k & & & & & \\ \hline 1 & + & 0 & + & 1 & = & 2 & \rightarrow & 0 & ac \\ 0 & + & 1 & + & 1 & = & 2 & \rightarrow & 0 & bc \\ \hline 1 & + & 1 & + & 2 \rightarrow 0 & = & 2 & \rightarrow & 0 & ac \times bc = ab \end{array}$$

One notes that this index addition corresponds to "multiplying" the two solutions ac and bc by each other.

The other block is constructed by finding the solutions to $(i+j+k)_2 = 1$. These solutions are (a, b, c, abc).

In this way the blocking is found:

Block 0				Block 1			
(1)	ab	ac	bc	a	b	c	abc

We note that this solution is exactly the same as the one we found in section 2.2 using for example the **tabular method**.

End of example 3.7

We have now seen a simple example of the use of Kempthorne's method to make block experiments. The principle is still that we let the block variable vary synchronously with the levels for the factor effect that we will confound with blocks.

Example 3.8 : 3^2 **factorial experiment in 3 blocks**

Suppose we have the experiment

	A=0	A=1	A=2
B=0	(1)	a	a^2
B=1	b	ab	a^2b
B=2	b^2	ab^2	a^2b^2

where we again write the model on the general form, which for $p = 3$ is :

$$Y_{ij\nu} = \mu + A_i + B_j + AB_{i+j} + AB^2_{i+2j} + E_{ij\nu}$$

We now want to carry out the experiment be in 3 blocks, each with 3 single experiments. For that purpose we can let the block index follow the index for the artificial effect AB_{i+j}, whereby it is still possible to estimate the two main effects A and B:

$$\text{Defining relation :} I = AB$$

$$\text{Index} = i + j : \quad \begin{array}{r|ccc|} & \text{i=0} & \text{i=1} & \text{i=2} \\ \cline{2-4} j=0 & 0 & 1 & 2 \\ j=1 & 1 & 2 & 0 \\ j=2 & 2 & 0 & 1 \\ \cline{2-4} \end{array}$$

Block 0	Block 1	Block 2
(1) ab^2 a^2b	a b a^2b^2	a^2 b^2 ab

Block 0 is given with all solutions to the equation $(i + j)_3 = 0$. The other two blocks are given with $(i + j)_3 = 1$ and $(i + j)_3 = 2$ respectively.

The design could have been computed directly using the following **tabular method**:

i	j	code	Block=$(i + j)_3$
0	0	(1)	0
1	0	a	1
2	0	a^2	2
0	1	b	1
1	1	ab	2
2	1	a^2b	0
0	2	b^2	2
1	2	ab^2	0
2	2	a^2b^2	1

If we only wanted the principal block we can use the method shown in the previous example, which consists of solving the equation:

$$\begin{array}{llllll} \multicolumn{4}{l}{(i + j) \text{ modulo } 3 = 0} & & \\ \hline i = 0 & \Rightarrow & j = 0 & & & \text{Experiment : } (1) \\ i = 1 & \Rightarrow & j = -1 \rightarrow -1 + 3 & = 2 & & \text{Experiment : } ab^2 \\ i = 2 & \Rightarrow & j = -2 \rightarrow -2 + 3 & = 1 & & \text{Experiment : } a^2b \end{array}$$

It is seen that the principal block has the appearance:

$$\begin{array}{c}\text{Block 0}\\ \boxed{(1)\quad x\quad x^2}\end{array}\qquad\qquad \text{where } x = ab^2 \Rightarrow x^2 = (ab^2)^2 = a^2b^4 = a^2b$$

where x can represent any chosen solution to $(i+j)_3 = 0$ except $(i,j) = (0,0)$.

The other two blocks are can be found by the above tabular method or, equivalently, by solving the equations $(i+j)_3 = 1$ and $(i+j)_3 = 2$ respectively. For example, the term a is the solution to $(i+j)_3 = 1$ in that $i = 1$, and $j = 0$ correspond to a.

The block that contains the experiment a can be constructed by "multiplying" a on the principal block found:

$$a\times \begin{array}{c}\text{Block 0}\\ \boxed{(1)\quad ab^2\quad a^2b}\\ \text{principal block}\end{array} \Longrightarrow \begin{array}{c}\text{Block 1}\\ \boxed{a\quad a^2b^2\quad b}\\ \end{array}$$

The last block is constructed by finding a solution to the equation $(i+j)_3 = 2$, for example a^2 and multiplying this on the principal block :

$$a^2\times \begin{array}{c}\text{Block 0}\\ \boxed{(1)\quad ab^2\quad a^2b}\\ \text{principal block}\end{array} \Longrightarrow \begin{array}{c}\text{Block}\\ \boxed{a^2\quad b^2\quad ab}\\ \end{array}$$

It is easy to show that with this blocking, the only effect in our model that is confounded with the block effect is precisely the AB_{i+j} effect.

If we wanted an alternative block grouping, where the AB^2 effect was confounded with blocks, we would use the defining relation I $= AB^2$ and determine the principal block by solving the index equation (i + 2j) = 0. One solution is ab, and the principal block is therefore $\boxed{(1),\ ab,\ (ab)^2} = \boxed{(1)\ ,\ ab,\ a^2b^2}$. After this one finds the blocking

$$\begin{array}{c}\text{Block 0}\\ \boxed{(1)\quad ab\quad a^2b^2}\\ \text{principal block}\\ (i+2j)_3 = 0\end{array}\quad \begin{array}{c}\text{Block 1}\\ \boxed{a\quad a^2b\quad b^2}\\ \\ \end{array}\quad \begin{array}{c}\text{Block 2}\\ \boxed{a^2\quad b\quad ab^2}\\ \\ \end{array}$$

Try it yourself!

End of example 3.8

We have seen how a p^k factorial experiment can be divided into p blocks so that an effect chosen in advance is confounded with blocks.

We can generalise this method to a division into p^q blocks, where $q < k$. To do this, we start by dividing a 2^3 experiment into $2 \times 2 = 2^2 = 4$ blocks.

Example 3.9 : Division of a 2^3 factorial experiment into 2^2 blocks

Let there be a 2^3 factorial experiment with factors A, B and C. With the introduced formulation the model is, in that $p = 2$:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i+j} + C_k + AC_{i+k} + BC_{j+k} + ABC_{i+j+k} + E_{ijk\nu}$$

where all indices $i, j, k = (0, 1)$ and $\nu = (1, .., r)$

To define 4 (= 2×2) blocks, we use 2 defining relations, for example

$$I_1 = AB \quad \text{and} \quad I_2 = AC$$

as previously shown on page 25.

The structure of the 4 blocks can be illustrated

		$I_1 = AB$	
		$i+j=0$	$i+j=1$
$I_2 = AC$	$i+k=0$	Block (0,0)	Block (1,0)
	$i+k=1$	Block (0,1)	Block (1,1)

If the index for both AB_{i+j} and AC_{i+k} is 0 for example, the single experiments are placed in block (0,0).

In this way, or by using the **tabular method**, one finds the blocking :

		$I_1 = AB$	
		$i+j=0$	$i+j=1$
$I_2 = AC$	$i+k=0$	(1) abc	b ac
	$i+k=1$	ab c	a bc

If the block effects are modelled as a 2×2 design, we can write that the blocks contribute with

$$\text{Blocks} = \xi + F_f + G_g + FG_{f+g} \ , \quad \text{where } f = (i+j)_2 \text{ and } g = (i+k)_2$$

It is clear that the effect AB varies synchronously with F and that the two effects are confounded. Correspondingly, AC is confounded with G. That part of the block variation, which is here called FG, has the index $(f+g)_2 = ((i+j)+(i+k))_2 = (j+k)_2$, which is precisely the index for the term BC in the model for the response of the experiment.

Therefore it can be concluded that the effect BC will also be confounded with blocks, which can also be seen from the following table, where the index of the BC effect is 0 on one diagonal and 1 on the other one:

		$I_1 = AB$	
		$i+j=0$	$i+j=1$
$I_2 = AC$	$i+k=0$	$j+k=0$	$j+k=1$
	$i+k=1$	$j+k=1$	$j+k=0$

More formally we can write:

$$\text{Blocks} = AB + AC + AB \times AC = AB + AC + BC$$

End of example 3.9

3.6 Generalisation of the division into blocks with several defining relations

Let $I_1 = A^\alpha B^\beta \ldots C^\gamma$ denote a defining relation, that divides a p^k factorial experiment into p blocks. Further, let $I_2 = A^a B^b ... C^c$ denote a defining relation that likewise divides the p^k experiment into p blocks.

In the division of the experiment into $p \times p$ blocks on the basis of these defining relations, both effects

$$I_1 = A^\alpha\ B^\beta ...\ C^\gamma \text{ and } I_2 = A^a\ B^b ...\ C^c$$

will be confounded with blocks. In addition, their generalised interaction will be confounded with blocks so that besides I_1 and I_2 the effects given in the expression:

$$I_1 \times I_2 = (A^\alpha\ B^\beta ...\ C^\gamma) \times (A^a\ B^b ...\ C^c)$$

will be confounded with blocks. All terms in the expression:

$$I_1 * I_2 = I_1 + I_2 + I_1 \times I_2 = I_1 + I_2 + I_1\ I_2 + I_1\ (I_2)^2 + ... + + I_1\ (I_2)^{(p-1)}$$

are confounded with blocks.

We can generally write up the confoundings for any division of a p^k factorial experiment in p^q blocks.

If we have the corresponding defining relations given by I_1, I_2, I_3,.., I_q, all the effects in the equation

$$I_1 * I_2 * I_3 * ... * I_q = I_1 + I_2 + I_1 \times I_2 + I_3 + ... + I_1 \times I_2 \times I_3 \times ... \times I_q$$

will be confounded with blocks. The operators "$*$" and "$\times$" work as stated in section 3.3 page 58

Example 3.10 : Dividing a 3^3 factorial experiment into 9 blocks

Let there be a 3^3 factorial experiment with factors A, B and C.

As an example we divide the experiment into 3×3 blocks using

$$I_1 = ABC^2 \text{ and } I_2 = AC$$

Thereby, ABC^2 and AC together with their generalised interaction are confounded with blocks, that is, all the effects in the expression (where $p = 3$):

$$ABC^2 * AC = ABC^2 + AC + ABC^2 \times AC$$

$$= ABC^2 + AC + ABC^2(AC) + ABC^2(AC)^2 = ABC^2 + AC + AB^2 + BC$$

In the analysis of variance table, the sums of squares for ABC^2, AC, AB^2 and BC therefore also contain possible block effects and thus they cannot be interpreted as expressing factor effects alone.

The principal block in this experiment is found by solving the equations ($p = 3$):

$$(i + j + 2k)_3 = 0 \text{ and } (i + k)_3 = 0$$

One finds for example $i = 1 \Rightarrow k = 2 \Rightarrow j = 1$, that is abc^2. The principal block contains $3^3/3^2 = 3$ single experiments. This means that it satisfies to find one solution in order to determine the block. If this solution is called "x", the principal block experiments are (1), x and x^2.

In our case, we then for $x = abc^2$ get the three experiments (1), abc^2 and $(abc^2)^2 = a^2b^2c$. One can check that $(i = 2, j = 2, c = 1)$ is also a solution to the two index equations.

The other blocks are found by finding the solutions to the index equations for the right-hand sides equal to (0,1,2) in the case of both equations, i.e. a total of 9 different cases, corresponding to the 3×3 blocks.

For a any one of these blocks, it applies that they can be found when just one experiment is found in the block. By multiplying this experiment on the principal block, the whole block is determined.

End of example 3.10

Example 3.11 : Division of a 2^5 experiment into 2^3 blocks

Let the factors be A, B, C, D and E, which all appear on 2 levels. To divide the experiment into $2 \times 2 \times 2$ blocks, 3 defining relations are used, f.ex.

$$I_1 = ABC,\ I_2 = BDE \text{ and } I_3 = ABE$$

Thereby all effects in the following expression are confounded with blocks:

$$I_1 * I_2 * I_3 = I_1 + I_2 + I_1 \times I_2 + I_3 + I_1 \times I_3 + I_2 \times I_3 + I_1 \times I_2 \times I_3$$

That is, in addition to ABC, BDE and ABE, the following terms (since $p = 2$):

$$\begin{array}{lclcl}
(I_1 \times I_2) & = & ABC\ BDE & = & ACDE \\
(I_1 \times I_3) & = & ABC\ ABE & = & CE \\
(I_2 \times I_3\ ) & = & BDE\ ABE & = & AD \\
(I_1 \times I_2 \times I_3) & = & ABC\ BDE\ ABE & = & BCD
\end{array}$$

The design can be found by the **tabular method** (se f.ex. page 23).

If only the principal block is wanted we can solve the equations

$$(i + j + k)_2 = 0 \quad , \quad (j + l + m)_2 = 0 \quad , \quad (i + j + m)_2 = 0$$

One block contains $2^5/2^3 = 2^2 = 4$ single experiments. Therefore 2 solutions have to be found. If these solutions are called x and y, the principal block is: $\boxed{(1)\ ,\ x\ ,\ y \text{ and } xy}$

An ingenious way to find these solutions is to try with ($i = 1$,$j = 0$) and ($i = 0$,$j = 1$), which correspond to the elements a and b in the factor structure for factors A and B. The method works if the main effects A and B are **not** confounded with each other or with blocks.

We find

$$\begin{array}{llllll} (i=1, j=0) & \Rightarrow & k=1, & m=1, & l=1, & \text{the experiment is } x = acde \\ (i=0, j=1) & \Rightarrow & k=1, & m=1, & l=0, & \text{the experiment is } y = bce \end{array}$$

The principal block is therefore

$$\begin{array}{ccc} \text{Principal block} & & \text{Block(0,0,0)} \\ \boxed{(1) \quad x \quad y \quad xy} & = & \boxed{(1) \quad acde \quad bce \quad abd} \end{array}$$

Note that all experiments in the principal block have an **even** number of letters in common with the 3 defining contrasts, ABC, BDE and ABE. The remaining blocks can now be found by multiplying with elements that are not in the principal block.

For the block corresponding to the equations

$$(i+j+k)_2 = 1 \quad , \quad (j+l+m)_2 = 0 \quad , \quad (i+j+m)_2 = 0$$

that is block (1,0,0), there is a solution: $(i, j, k, l, m) = (1, 0, 0, 1, 1) = ade$ (start with $(i, j) = (1, 0)$, which is the easiest method). The rest of the block is found by multiplying this solution onto the principal block:

$$\begin{array}{cccc} & \text{Principal block} & & \text{Block(1,0,0)} \\ ade\times & \boxed{(1) \quad acde \quad bce \quad abd} & \Longrightarrow & \boxed{ade \quad c \quad abcd \quad be} \end{array}$$

The remaining 6 blocks can be found by setting the right-hand sides of the equations to (0,1,0), (1,1,0), (0,0,1), (1,0,1), (0,1,1) and (1,1,1), respectively.

End of example 3.11

Example 3.12 : Division of 3^k experiments into 3^3 blocks

Let there be a 3^k factorial experiment and suppose that I_1, I_2 and I_3 define a division of the experiment into $3 \times 3 \times 3 = 27$ blocks.

The confounding is thereby given with

$$I_1 * I_2 * I_3 = I_1 + I_2 + I_1 \times I_2 + I_3 + I_1 \times I_3 + I_2 \times I_3 + I_1 \times I_2 \times I_3$$
$$= I_1 + I_2 + I_1\, I_2 + I_1\, I_2^2 + I_3 + I_1\, I_3 + I_1\, I_3^2 + I_2\, I_3 + I_2\, I_3^2 + I_1\, (I_2 \times I_3) + I_1\, (I_2 \times I_3)^2$$

where exponents are reduced modulo 3. For the two last terms we have:

$$I_1(I_2 \times I_3) = I_1(I_2\ I_3 + I_2\ I_3^2) = I_1\ I_2\ I_3 + I_1\ I_2\ I_3^2$$

and

$$I_1\ (I_2 \times I_3)^2 = I_1\ (I_2\ I_3 + I_2\ I_3^2)^2 = I_1\ I_2^2\ I_3^2\ + I_1\ I_2^2\ I_3^4 = I_1\ I_2^2\ I_3^2\ + I_1\ I_2^2\ I_3$$

The terms found will all be confounded with blocks. Each of the corresponding effects has precisely 3 levels, i.e. the variation between these 3 levels has 2 degrees of freedom. A total of 13 terms with 2 degrees of freedom are found, i.e. a total of 26 degrees of freedom, which correspond exactly to the variation between 27 blocks.

The design can be found by the **tabular method** (se f.ex. page 65).

End of example 3.12

3.6.1 Construction of blocks in general

We have seen above that in a p^k factorial experiment, one defining relation, I_1, divides the experiment into p blocks, while q relations, for example I_1, ..., I_q, divide the experiment into p^q blocks each containing precisely p^{k-q} single experiments.

This corresponds to the fact that in one block are just as many single experiments as there are in a complete p^{k-q} experiment, that is, an experiment with $k-q$ factors each on p levels.

We first construct the principal block among these p^q blocks, and, on the basis of this, the remaining blocks can be determined, as is shown in the examples.

The method we will use is in brief:

1): : Let there be q defining contrasts I_1, I_2, ..., I_q, and again let all the single experiments be designated with $(1), a, a^2, \ldots, a^{p-1}, b, ab, \ldots, a^{p-1}b, b^2, ab^2, \ldots, a^{p-1}b^{p-1}, \ldots,$ etc.

2): Determine $k-q$ of these single experiments, which are in the principal block. For example they can be: f_1 ,f_2,...,f_{k-q}. It is required for these single experiments that the corresponding solutions to the index equations are **linearly independent**.

3): The principal block can now be constructed by using these single experiments as "basic experiments" and making a complete p^{k-q} factorial experiment, i.e. the standard order for the single experiments on the basis of f_1, ,f_2,...,f_{k-q}:

$$(1), f_1, {f_1}^2, \ldots, {f_1}^{p-1}, f_2, f_1f_2, {f_1}^2f_2, \ldots, {f_1}^{p-1}{f_2}^{p-1}, \ldots, {f_1}^{p-1}{f_2}^{p-1} \cdot \cdot {f_{k-q}}^{p-1}$$

This collection of single experiments then makes up the principal block. During the formation, all exponents are reduced modulo p.

4): Change the level for one of the index equations and thereby find a new single experiment that is not in the principal block and multiply the experiment on all the experiments in the principal block. In this way a new block is formed.

5): Continue with 4) until all blocks are formed.

In order to assure that all the single experiments in the principal block are different, we must require for the original $(k-q)$ solutions that they are linearly independent (where all are still calculated modulo p).

Otherwise the principal block will not be completely determined, and the same single experiments will be found several times when trying to find the experiments in the block.

With the same kind of argumentation, it can be shown how, on the basis of one experiment belonging to an alternative block, the rest of that block can be formed by multiplying it onto the principal block.

For example if, with the help of a spreadsheet or a computer program, one wants to find a block distribution, the simplest method is to run through all single experiments in standard order and for each single experiment calculate the value of the indices of the defining contrasts, that is to use the **tabular method**.

Example 3.13 : Dividing a 3^4 factorial experiment into 3^2 blocks

Let the notation be as usual. The single experiments are given by:

$$(1)\ ,\ a\ ,\ a^2\ ,\ b\ ,\ ab\ ,\ a^2b\ ,\ b^2\ ,\ ab^2\ ,\ a^2b^2\ ,\ \ldots\ ,\ a^2b^2c^2d^2$$

Take for example the defining relations :

$$I_1 = AB_{i+j} \text{ and } I_2 = BCD^2_{j+k+2l}$$

The principal block consists of the experiments where both $(i+j)_3 = 0$ and $(j+k+2l)_3 = 0$. In one block there are $3^{4-2} = 3^2$ single experiments.

The complete design can be constructed and written out by means of the tabular method (see page 65).

If we want the principal principal block, for example, we must just determine two "linearly independent" single experiments and from that form the rest as a 3^2 experiment.

Thus: Find two linearly independent solutions to:

$$i + j = 0 \quad \text{and} \quad j + k + 2l = 0$$

Try with $i = 0 \Rightarrow j = 0 \Rightarrow k + 2l = 0$ and choose $(k = 1, l = 1)$, for example, giving $(i, j, k, l) = (0, 0, 1, 1)$ as a usable solution. The experiment is cd.

Then try for example with $i = 1 \Rightarrow j = 2$, $(j + k + 2l) = 0 \Rightarrow (k + 2l)_3 = (-2)_3 = (-2 + 3)_3 = 1$ where we for example choose $l = 0$ and $k = 1$. Note that one can always add an arbitrary multiple of "3" to a (negative) number when one has to find "modulo 3" of the number. That is to say that generally $(x)_p = (x + kp)_p$ where $(.)_p$ here denotes "(.) modulo p".

Thus $(i, j, k, l) = (1, 2, 1, 0)$ is a usable combination and the experiment is "ab^2c".

Check the independence by verifying that $cd(ab^2c)^\lambda \neq (1)$ for all λ (the relevant λ's are 1 and 2): OK.

Now call $f_1 = cd$ and $f_2 = ab^2c$. The principal block then is

$$\boxed{\begin{array}{ccc} (1) & f_1 & f_1^2 \\ f_2 & f_1 f_2 & f_1^2 f_2 \\ f_2^2 & f_1 f_2^2 & f_1^2 f_2^2 \end{array}} = \boxed{\begin{array}{ccc} (1) & cd & (cd)^2 \\ ab^2c & cdab^2c & (cd)^2 ab^2c \\ (ab^2c)^2 & cd(ab^2c)^2 & (cd)^2(ab^2c)^2 \end{array}}$$

by ordering the elements, multiplying out and reducing all exponents modulo 3, the block is found:

$$\boxed{\begin{array}{ccc} (1) & cd & c^2d^2 \\ ab^2c & ab^2c^2d & ab^2d^2 \\ a^2bc^2 & a^2bd & a^2bcd^2 \end{array}}$$

To find an alternative block, we look for a single experiment that is not in the block already found. We can for example take "a".

The new block is then:

$$a\times \boxed{\begin{array}{ccc} (1) & cd & c^2d^2 \\ ab^2c & ab^2c^2d & ab^2d^2 \\ a^2bc^2 & a^2bd & a^2bcd^2 \end{array}} \Longrightarrow \boxed{\begin{array}{ccc} a & acd & ac^2d^2 \\ a^2b^2c & a^2b^2c^2d & a^2b^2d^2 \\ bc^2 & bd & bcd^2 \end{array}}$$

or by multiplying with b :

$$b\times \boxed{\begin{array}{ccc} (1) & cd & c^2d^2 \\ ab^2c & ab^2c^2d & ab^2d^2 \\ a^2bc^2 & a^2bd & a^2bcd^2 \end{array}} \Longrightarrow \boxed{\begin{array}{ccc} b & bcd & bc^2d^2 \\ ac & ac^2d & ad^2 \\ a^2b^2c^2 & a^2b^2d & a^2b^2cd^2 \end{array}}$$

End of example 3.13

Example 3.14 : Dividing a 5^3 factorial experiment into 5 blocks

A 5^3 experiment consists of a total of 125 single experiments. With the division into 5 blocks, there are 25 single experiments in each block.

The factors are A, B and C, and as defining relation we choose for example

$$I = ABC^3_{i+j+3k}$$

In the principal block, where $p = 5$, it applies that

$$i + j + 3k = 0 \quad \text{(modulo 5)}$$

Since the size of the block is $5 \times 5 = 5^2$, we have to find 2 linearly independent solutions to this equation. For example,

$$(i, j, k) = (1, 0, 3) \sim ac^3 \text{ and } (i, j, k) = (0, 1, 3) \sim bc^3$$

can be used. As a start, the principal block is thereby

(1)	ac^3	a^2c^6	a^3c^9	a^4c^{12}
bc^3	abc^6	a^2bc^9	a^3bc^{12}	a^4bc^{15}
b^2c^6	ab^2c^9	$a^2b^2c^{12}$	$a^3b^2c^{15}$	$a^4b^2c^{18}$
b^3c^9	ab^3c^{12}	$a^2b^3c^{15}$	$a^3b^3c^{18}$	$a^4b^3c^{21}$
b^4c^{12}	ab^4c^{15}	$a^2b^4c^{18}$	$a^3b^4c^{21}$	$a^4b^4c^{24}$

and after reduction of the exponents modulo 5, one finally gets

(1)	ac^3	a^2c	a^3c^4	a^4c^2
bc^3	abc	a^2bc^4	a^3bc^2	a^4b
b^2c^1	ab^2c^4	$a^2b^2c^2$	a^3b^2	$a^4b^2c^3$
b^3c^4	ab^3c^2	a^2b^3	$a^3b^3c^3$	a^4b^3c
b^4c^2	ab^4	$a^2b^4c^3$	$a^3b^4c^1$	$a^4b^4c^4$

It can be interesting to note that this is a 5×5 Latin square, which with C inside the square is:

	A=0	A=1	A=2	A=3	A=4
B=0	0	3	1	4	2
B=1	3	1	4	2	0
B=2	1	4	2	0	3
B=3	4	2	0	3	1
B=4	2	0	3	1	4

which for instance shows that the three factors are mutually balanced within the block found. The same will naturally apply within the other 4 blocks in the experiment. One of these blocks can be easily constructed for example by multiplying the principal block with an experiment that is not included in the principal block. By multiplying with a, for example, we find

a	a^2c^3	a^3c	a^4c^4	c^2
abc^3	a^2bc	a^3bc^4	a^4bc^2	b
ab^2c^1	$a^2b^2c^4$	$a^3b^2c^2$	a^4b^2	b^2c^3
ab^3c^4	$a^2b^3c^2$	a^3b^3	$a^4b^3c^3$	b^3c
ab^4c^2	a^2b^4	$a^3b^4c^3$	$a^4b^4c^1$	b^4c^4

which is thus also a Latin square.

The remaining blocks can be found in the same way, but naturally one can also let a program construct all the blocks by calculating the value of the index $(i + j + 3k)$ for all single experiments and placing the experiments according to whether $(i+j+3k)$ (modulo 5) is 0, 1, 2, 3 or 4, that is by the **tabular method**.

End of example 3.14

3.7 Partial confounding

Partial confounding in 2^k factorial experiments was introduced in section 2.3 page 26.

We will give another example of partial confounding in a 2^k experiment, where we now for the sake of illustration use Kempthorne's method to form the relevant blocks.

Example 3.15 : Partially confounded 2^3 factorial experiment

Again we consider an experiment with 3 factors A, B and C, each on 2 levels. We assume that the experiments can only be done in blocks which each contain 4 single experiments. To be able to estimate all the effects in the model

$$Y_{ijk} = \mu + A_i + B_j + AB_{i+j} + C_i + AC_{i+k} + BC_{j+k} + ABC_{i+j+k} + E$$

it is necessary to do a partially confounded factorial experiment.

Suppose that in the first experimental series we choose to confound the three-factor interaction ABC.

To divide the experiment into 2 blocks, we have to find 2 solutions to the index equation since the block size is $2^{3-1} = 2 \times 2$.

Therefore we have to find 2 solutions to the equation $(i + j + k)_2 = 0$.

By trial and error, we find for example $x = ac$ and $y = bc$.

The principal block is then

$$\boxed{(1) \quad x \quad y \quad xy} = \underset{(i+j+k)_2 = 0}{\overset{\text{block 1}}{\boxed{(1) \quad ac \quad bc \quad ab}}}$$

By multiplying with a, we get the other block, which of course consists of the remaining single experiments in the complete 2^3 factorial experiment:

$$a\times \boxed{(1) \quad x \quad y \quad xy} = \underset{(i+j+k)_2 = 1}{\overset{\text{block 2}}{\boxed{a \quad c \quad abc \quad b}}}$$

Analysis of this first block-confounded experiment can be done with Yates' algorithm, which gives a result that can also be expressed in matrix form in the usual way:

$$\begin{bmatrix} I_{(1)} \\ A_{(1)} \\ B_{(1)} \\ AB_{(1)} \\ C_{(1)} \\ AC_{(1)} \\ BC_{(1)} \\ ABC_{(1)} = \text{blocks} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} (1) \\ a \\ b \\ ab \\ c \\ ac \\ bc \\ abc \end{bmatrix}$$

The index $(.)_{(1)}$ on the contrasts refers to this first experiment.

We then do another experiment, but this time we choose to confound the effect AB. The blocking of the experiment is thus:

$$\underset{(i+j)_2 = 0}{\overset{\text{block 3}}{\boxed{(1) \quad ab \quad abc \quad c}}} \quad \text{and} \quad \underset{(i+j)_2 = 1}{\overset{\text{block 4}}{\boxed{a \quad b \quad ac \quad bc}}}$$

For this experiment we can find contrasts in the same way as in the first experiment. And finally we will combine the two experiments. We have the following sources of variation:

1)	Factor effects	which are not confounded
2)	Factor effects	which are partially confounded
3)	Blockeffects ,	i.e. variation between the totals of the 4 blocks
4)	Residual variation	

Analysis of the two experiments gives respectively:

$$\begin{bmatrix} I_{(1)} \\ A_{(1)} \\ B_{(1)} \\ AB_{(1)} \\ C_{(1)} \\ AC_{(1)} \\ BC_{(1)} \\ ABC_{(1)} = \text{blocks} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} I_{(2)} \\ A_{(2)} \\ B_{(2)} \\ AB_{(2)} = \text{blocks} \\ C_{(2)} \\ AC_{(2)} \\ BC_{(2)} \\ ABC_{(2)} \end{bmatrix}$$

where index $(.)_{(2)}$ corresponds to the second experiment.

We can find sums of squares and degrees of freedom corresponding to the four sources of variation:

1)	Unconfounded factor effects			
	SSQ_A	$=$	$([A_{(1)}]+[A_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	SSQ_B	$=$	$([B_{(1)}]+[B_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	SSQ_C	$=$	$([C_{(1)}]+[C_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	SSQ_{AC}	$=$	$([AC_{(1)}]+[AC_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	SSQ_{BC}	$=$	$([BC_{(1)}]+[BC_{(2)}])^2/(2\cdot 2^3)$,	f = 1
2)	Partially confounded factor effects			
	SSQ_{AB} (half precision)	$=$	$[AB_{(1)}]^2/(2^3)$,	f = 1
	SSQ_{ABC} (half precision)	$=$	$[ABC_{(2)}]^2/(2^3)$,	f = 1
3)	Block effects and confounded factor effects			
	Between experiments	$=$	$([I_{(1)}]-[I_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	$\text{SSQ}_{AB+\text{blocks(3-4)}}$	$=$	$[AB_{(2)}]^2/(2^3)$,	f = 1
	$\text{SSQ}_{ABC+\text{blocks(1-2)}}$	$=$	$[ABC_{(1)}]^2/(2^3)$,	f = 1
4)	Residual variation:			
	Between A-estimates ($\text{SSQ}_{\text{A,Uncert.}}$)	$=$	$([A_{(1)}]-[A_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	Between B-estimates ($\text{SSQ}_{\text{B,Uncert.}}$)	$=$	$([B_{(1)}]-[B_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	Between C-estimates ($\text{SSQ}_{\text{C,Uncert.}}$)	$=$	$([C_{(1)}]-[C_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	Between AC-estimates ($\text{SSQ}_{\text{AC,Uncert.}}$)	$=$	$([AC_{(1)}]-[AC_{(2)}])^2/(2\cdot 2^3)$,	f = 1
	Between BC-estimates ($\text{SSQ}_{\text{BC,Uncert.}}$)	$=$	$([BC_{(1)}]-[BC_{(2)}])^2/(2\cdot 2^3)$,	f = 1
5)	Total variation	$=$	SSQ_{tot} with degrees of freedom	f = 15

Generally, the variation can be calculated between for example R_A A estimates that are all based on contrasts from (R_A different) 2^k experiments with r repetitions (in which they are all unconfounded) with the expression:

$$\text{SSQ}_{\text{A,Uncert.}} = \frac{[A_{(1)}]^2+\ldots+[A_{(R_A)}]^2}{2^k\cdot r} - \frac{([A_{(1)}]+\ldots+[A_{(R_A)}])^2}{R_A\cdot 2^k\cdot r}$$

In the example, $R_A = 2$, $k = 3$ and $r = 1$. For other non-confounded estimates, naturally, corresponding expressions are found. See, too, page 29.

We have thereby accounted for all the variation in the two experiments collectively. Note that we have calculated sums of squares corresponding to a total of 15 sources of variance, each with one degree of freedom. This corresponds precisely to the total variation between the 16 single experiments, which gives rise to $(16-1) = 15$ degrees of freedom.

If there are r repetitions for each of the single experiments, all the SSQ's have to be divided by r. In this case, one can, of course, find variation within each factor combination (a total of 8 + 8 single experiments with $(r-1)$ degrees of freedom) and use them to calculate a separate estimate for the remainder variation. This estimate can, if necessary, be compared with the mentioned estimate, which was calculated above.

End of example 3.15

The example shown illustrates the principles for combining several experiments with different confoundings. The whole analysis can be summarised to the following. Suppose that, in all, experiments are made in R blocks med n_{block} single experiments in each block. The variation can then be decomposed in the following contributions, where $T_{block\ i}$ gives the total in the i'th block:

$$
\begin{aligned}
\text{SSQ}_{\text{blocks}} &= (T^2_{block\ 1} + T^2_{block\ 2} + \ldots + T^2_{block\ R})/n_{block} - (T^2_{tot})/(R \cdot n_{block}) \\
&= \text{variation between block totals} \\
\text{SSQ}_{\text{effects}} &= \text{SSQ for all factor effects based on experiments in which} \\
&\ \text{the effects are not er confounded with blocks} \\
\text{SSQ}_{\text{resid}} &= \text{variation between effect estimates from experiments, where} \\
&\ \text{the effects are not confounded} \\
\text{SSQ}_{\text{uncertainty}} &= \text{variation between repeated single experiments within blocks}
\end{aligned}
$$

The first contribution, $\text{SSQ}_{\text{blocks}}$ also contains, in addition to the total variation between blocks, the variation from confounded factor effects.

Example 3.16 : Partially confounded 3^2 factorial experiment

We finish this section by showing the principles for the construction and analysis of a partially confounded 3^2 factorial experiment. This experiment is possibly little used in practice, but it illustrates the general procedure well. And it shows how all the main effects and interactions in a 3×3 experiment can be determined, even though the size of the block is only 3.

<u>Experiment 1</u> : Divide the 3^2 experiment into 3 blocks of 3 according to $I = AB_{i+j}$

One finds:

$$
\begin{aligned}
&\text{block 1} \quad : \quad \boxed{(1)_{(1)} \quad ab^2_{(1)} \quad a^2b_{(1)}} \quad \text{total} = T_1 \\
&\text{block 2} \quad : \quad \boxed{a_{(1)} \quad a^2b^2_{(1)} \quad b_{(1)}} \quad \text{total} = T_2 \\
&\text{block 3} \quad : \quad \boxed{a^2_{(1)} \quad b^2_{(1)} \quad ab_{(1)}} \quad \text{total} = T_3
\end{aligned}
$$

Index (1) indicates that this is experiment 1.

Experiment 2 : Now divide according to $I = AB^2_{i+2j}$. This gives the blocking:

block 4	:	$(1)_{(2)}$ $ab_{(2)}$ $a^2b^2_{(2)}$		total $= T_4$
block 5	:	$a_{(2)}$ $a^2b_{(2)}$ $b^2_{(2)}$		total $= T_5$
block 6	:	$a^2_{(2)}$ $b_{(2)}$ $ab^2_{(2)}$		total $= T_6$

Index (2) indicates that this is experiment 2.

We can find the sum of squares between the 6 blocks, in that $T_{tot} = T_1 + T_2 + \ldots + T_6$:

$$\text{SSQ}_{blocks} = (T_1^2 + T_2^2 + \ldots + T_6^2)/3 - T_{tot}^2/18$$

We then have

$$T_{A_0} = (1)_{(1)} + b_{(1)} + b^2_{(1)} + (1)_{(2)} + b_{(2)} + b^2_{(2)}$$

$$T_{A_1} = a_{(1)} + ab_{(1)} + ab^2_{(1)} + a_{(2)} + ab_{(2)} + ab^2_{(2)}$$

$$T_{A_2} = a^2_{(1)} + a^2b_{(1)} + a^2b^2_{(1)} + a^2_{(2)} + a^2b_{(2)} + a^2b^2_{(2)}$$

That is, that for example T_{A_0} = the sum of all the measurements where factor A has been on level "0" in the R_A experiments where effect A is not confounded with blocks, and correspondingly for levels "1" and "2". With $T_{tot,A} = T_{A_0} + T_{A_1} + T_{A_0}$ we get quite generally:

$$\text{SSQ}_A = (T_{A_0}^2 + T_{A_1}^2 + T_{A_2}^2)/(R_A \times 3^{k-1}) - T_{tot,A}^2/(R_A \times 3^k)$$

with $f = (3-1) = 2$ degrees of freedom. In our example $R_A = 2$ and $k = 2$.

$$T_{B_0} = (1)_{(1)} + a_{(1)} + a^2_{(1)} + (1)_{(2)} + a_{(2)} + a^2_{(2)}$$

$$T_{B_1} = b_{(1)} + ab_{(1)} + a^2b_{(1)} + b_{(2)} + ab_{(2)} + a^2b_{(2)}$$

$$T_{B_2} = b^2_{(1)} + ab^2_{(1)} + a^2b^2_{(1)} + b^2_{(2)} + a^2b^2_{(2)} + a^2b^2_{(2)}$$

$$\text{SSQ}_B = (T^2_{B_0} + T^2_{B_1} + T^2_{B_2})/(R_B \times 3^{k-1}) - T^2_{tot,B}/(R_B \times 3^k)$$

with $f = (3-1) = 2$ degrees of freedom and $R_B = 2$ and $k = 2$.

$$\begin{aligned} T_{AB_0} &= (1)_{(2)} + ab^2_{(2)} + a^2b_{(2)} \\ T_{AB_1} &= a_{(2)} + a^2b^2_{(2)} + b_{(2)} \\ T_{AB_2} &= a^2_{(2)} + b^2_{(2)} + ab_{(2)} \end{aligned}$$

that is, sums from the R_{AB} experiments in which the artificial effect AB_{i+j} is not confounded with blocks, i.e. the experiment consisting of blocks 4, 5 and 6. With $T_{tot,AB} = T_4 + T_5 + T_6$, one finds

$$\text{SSQ}_{AB} = (T^2_{AB_0} + T^2_{AB_1} + T^2_{AB_2})/(R_{AB} \times 3^{k-1}) - T^2_{tot,AB}/(R_{AB} \times 3^k)$$

with $f = (3-1) = 2$ degrees of freedom and $R_{AB} = 1$ and $k = 2$.

Finally one finds

$$\begin{aligned} T_{AB^2_0} &= (1)_{(1)} + ab_{(1)} + a^2b^2_{(1)} \\ T_{AB^2_1} &= a_{(1)} + a^2b_{(1)} + b^2_{(1)} \\ T_{AB^2_2} &= a^2_{(1)} + b_{(1)} + ab^2_{(1)} \end{aligned}$$

that is, sums from the experiments in which the artificial effect AB^2_{i+j} is not confounded with blocks, i.e. the experiment consisting of blocks 1, 2 and 3. With $T_{tot,AB^2} = T_1+T_2+T_3$ one finds

$$\text{SSQ}_{AB^2} = (T^2_{AB^2{}_0} + T^2_{AB^2{}_1} + T^2_{AB^2{}_2})/(R_{AB^2} \times 3^{k-1}) - T^2_{tot,AB^2}/(R_{AB^2} \times 3^k)$$

with $f = (3-1) = 2$ degrees of freedom and $R_{AB^2} = 1$ and $k = 2$.

The residual variation is found as the variation between estimates for effects that are not confounded with blocks.

From the A effect one finds, where SSQ_A(both experiments) is the above calculated sum of squares for effect A, while SSQ_A(experiment 1) and SSQ_A(experiment 2) are the sums of squares for effect A calculated separately for the two experiments:

$$\text{SSQ}_{UA} = \text{SSQ}_A(\text{experiment 1}) + \text{SSQ}_A(\text{experiment 2}) \text{ - } \text{SSQ}_A(\text{both experiments})$$

with $f = 4 - 2 = 2$ degrees of freedom.

From the B effect one finds correspondingly

$$\text{SSQ}_{UB} = \text{SSQ}_B(\text{experiment 1}) + \text{SSQ}_B(\text{experiment 2}) \text{ - } \text{SSQ}_B(\text{both experiments})$$

likewise with $f = 4 - 2 = 2$ degrees of freedom.

Since the remaining effects AB_{i+j} and AB^2_{i+2j} are only "purely" estimated one time each, we do not get any contribution to the residual variation from these effects.

In summary, we get the following variance decomposition:

Blocks and/or confounded factor effects	SSQ_{blocks}	5	
Main effect A_i	SSQ_A	2	
Main effect B_j	SSQ_B	2	
Interaction AB_{i+j}	SSQ_{AB}	2	(half præcision)
Interaction AB^2_{i+j}	SSQ_{AB^2}	2	(half præcision)
Residual variation	SSQ_{UA} + SSQ_{UB}	2 + 2	
Total	SSQ_{total}	17	

Note that we have derived variation corresponding to $17 = 18 - 1$ degrees of freedom.

In conclusion we can give estimates for all the effects in this experiment:

$$\widehat{\sigma^2}_{\text{resid}} = (\text{SSQ}_{UA} + \text{SSQ}_{UB})/4$$

$$(\hat{A}_0, \hat{A}_1, \hat{A}_2) = (\frac{T_{A_0}}{6} - \frac{T_{tot}}{18}, \frac{T_{A_1}}{6} - \frac{T_{tot}}{18}, \frac{T_{A_2}}{6} - \frac{T_{tot}}{18})$$

and the main effect B is found correspondingly.

For the interaction, the estimates are found in the blocks where they are not confounded:

$$(\widehat{AB}_0, \widehat{AB}_1, \widehat{AB}_2) = (\frac{T_{AB_0}}{3} - \frac{T_{tot,2}}{9}, \frac{T_{AB_1}}{3} - \frac{T_{tot,2}}{9}, \frac{T_{AB_2}}{3} - \frac{T_{tot,2}}{9}) \quad \text{(from experiment 2)}$$

$$(\widehat{AB^2_0}, \widehat{AB^2_1}, \widehat{AB^2_2}) = (\frac{T_{AB^2_0}}{3} - \frac{T_{tot,1}}{9}, \frac{T_{AB^2_1}}{3} - \frac{T_{tot,1}}{9}, \frac{T_{AB^2_2}}{3} - \frac{T_{tot,1}}{9}) \quad \text{(from experiment 1)}$$

and with the help of the relation $AB_{i,j} = AB_{i+j} + AB^2_{i+2j}$ one can finally find the parameter estimates in the traditional model formula $Y_{i,j} = \mu + A_i + B_j + AB_{i,j} + E$.

End of example 3.16

3.8 Construction of a fractional factorial design

We will now concern ourselves with constructing designs where the factors form Latin squares/cubes. The presentation is a generalisation of the results in section 2.4, where we introduced fractional 2^k factorial designs. We will limit the discussion to experiments with factors on 2 or on 3 levels, since these are the experiments that are used most frequently in practice.

As before, mainly examples are used to show the different techniques.

Example 3.17 : Factor experiment done as a Latin square experiment

Let us assume that we have three factors, A, B and C and that we want to evaluate these each on 3 levels. We choose to make a Latin square experiment with the factor C inside the square. According to the same principle described in the previous section, we can for example let C have index $k = (i + j)_3$. This means that the main effect C will have the same index as the Kempthorne effect AB with index $(i + j)_3$.

The experimental design where the index for factor C is inside the square is thus:

	A=0	A=1	A=2
B=0	0	1	2
B=1	1	2	0
B=2	2	0	1

Index for C

or equivalently :

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

Instead of the $3 \times 3 \times 3$ single experiments in the complete factor experiment, we choose to do only the 3×3 single experiments in the Latin square.

Note that if $x = ac$ and $y = bc$ which both correspond to a solution to the index equation $k = (i + j)_3$, the experiment can be written:

(1)	x	x^2
y	xy	x^2y
y^2	xy^2	x^2y^2

Finally, we can also by means of the **tabular method** write out the design:

Experiment no.	Experiment sequence	A level i	B level j	C level $(i+j)_3$	code
1	3	0	0	0	(1)
2	7	1	0	1	ac
3	1	2	0	2	a^2c^2
4	9	0	1	1	bc
5	5	1	1	2	abc^2
6	6	2	1	0	a^2b
7	2	0	2	2	b^2c^2
8	4	1	2	0	ab^2
9	8	2	2	1	a^2b^2c

In the practical execution of the experiment, the order is randomised, as exemplified in the table.

The variation of the mean value due to the two factors A and B can be written

$$A_i + B_j + AB_{i,j} = A_i + B_j + AB_{i+j} + AB^2_{i+2j}$$

where the left-hand side is the conventional meaning, while the right-hand side is the formulation according to Kempthorne's method.

The introduction of factor C, as mentioned, was done by giving C the index value $k = (i + j)_3$. If C is purely additive, i.e. if C does not interact with the other factors, the following model will describe the response, where the influence from C is put in:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i+j} + AB^2_{i+2j} + C_{k=i+j} + E_{ijk\nu}$$

where index ν corresponds to possible repetitions of the 9 single experiments.

If there is interaction between the two factors A and B, that part of the interaction described by the artificial effect AB_{i+j} cannot be regarded as negligible (the same is true of course for AB^2_{i+2j}).

If we now try to estimate the C effect, we cannot avoid having the AB_{i+j} part of the AB interaction confounded the C estimate, precisely because we used $k = (i + j)_3$

The effects C_k and AB_{i+j} in other words are **confounded** in the experiment.

We can also demonstrate this by direct calculation. Using $\bar{Y}_{tot}$ for the average of the 9 measurements, we have, since $(1) = Y_{0,0,0}$, $a^2b = Y_{2,1,0}$, and $ab^2 = Y_{1,2,0}$:

$$\widehat{C}_0 = (Y_{0,0,0} + Y_{2,1,0} + Y_{1,2,0})/3 - \bar{Y}_{tot}$$

which has the expected value

$$E\{\widehat{C}_0\} = E\{(Y_{0,0,0} + Y_{2,1,0} + Y_{1,2,0})/3 - \bar{Y}_{tot}\} = C_0 + AB_0$$

Further it is found

$$E\{\widehat{C}_1\} = E\{(Y_{0,1,1} + Y_{1,0,1} + Y_{2,2,1})/3 - \bar{Y}_{tot}\} = C_1 + AB_1$$

and

$$E\{\widehat{C}_2\} = E\{(Y_{0,2,2} + Y_{1,1,2} + Y_{2,0,2})/3 - \bar{Y}_{tot}\} = C_2 + AB_2$$

In this experiment we have a certain possibility of evaluating whether the AB interaction can be regarded as negligible, because we can examine the AB^2_{i+2j} effect. If this is negligible, one can perhaps allow oneself to conclude that the AB effect as a whole can be zero.

In summary, one can see that only if the factors A and B do not interact, is the experiment suitable for estimating C.

As we shall see below, we need to assume that all two-factor interactions are zero in order to estimate the main effects A, B and C in the described $(1/3) \times 3^3 = 3^{3-1}$ experiment.

End of example 3.17

We have seen above and previously in section 2.4 that it is not without problems to put further factors into an experiment in the form of a square experiment. But with appropriate assumptions about the lack of interactions, it can be done. In the following, we will try to show how it is done in practice.

Example 3.18 : Confoundings in a $3^{-1} \times 3^3$ factorial experiment, alias relations

Consider again the above example. If we should have done an ordinary 3^3 factorial experiment, it would have consisted of $3 \times 3 \times 3 = 27$ single experiments. The experiment we did is only 1/3 of this, namely a total of $3^3/3 = 3^{3-1} = 9$ single experiments.

If in general there are interactions between all the factors, the effects of the experiment will be confounded with each other in groups of 3 effects, analogous with for example the

2^{3-1} experiment, where they were confounded in groups of 2. The complete mathematical model for the 3^3 factorial experiment can be written, once again:

$$Y_{ijk\nu} = \mu + A_i + B_j + AB_{i+j} + AB^2_{i+2j} + C_k + AC_{i+k} + AC^2_{i+2k} + BC_{j+k}$$
$$+BC^2_{j+2k} + ABC_{i+j+k} + ABC^2_{i+j+2k} + AB^2C_{i+2j+k} + AB^2C^2_{i+2j+2k} + E_{ijk\nu}$$

For the 9 single experiments we considered in the previous example, we used $k = i + j$, corresponding to the confounding $C_k = AB_{i+j}$.

This **generator equation** can be changed to a **defining relation** by multiplying on both sides of the equation sign with C^2 , and then reorganising the expressions and reducing the exponents modulo 3. The result is $C^2C = C^2AB \longrightarrow I = ABC^2$. Thus we have the

$$\text{Defining relation : } \quad I = ABC^2$$

If one now wants which effects in the general model an arbitrary effect is confounded with, the defining relation can be used. It is multiplied with the effect in question in first and second power (because the factors are on 3 levels and according to the rules layed out in section 3.3). One finds for the A effect:

$$A \times (I = ABC^2) \longrightarrow A = (A)(ABC^2) = (A)^2(ABC^2)$$

Since now $A(ABC^2) = A^2BC^2 \to A^4B^2C^4 \to AB^2C$ and $(A)^2(ABC^2) = A^3BC^2 \to BC^2$, it is found that

$$A = AB^2C = BC^2$$

One can be convinced that indices for these three effects vary synchronously throughout the experiment, because it is required that $k = (i + j)_3$. "Modulo 3" calculation gives (try it yourself):

Effects	A	AB^2C	BC^2
Indices	i	$(i + 2j + k)_3$	$(j + 2k)_3$
	$0 \Rightarrow$	0	0
	$1 \Rightarrow$	2	2
	$2 \Rightarrow$	1	1

Of course the same calculations can be made for all effects in the experiment and one can be convinced that it will generally hold true that all effects are confounded in groups of 3.

The complete set of alias relations is found on the basis of the defining relation by multiplying with the effects in the underlying factor structure in first and second power:

Generator $C = AB \Longrightarrow$ Defining relation $I = ABC^2$				
Alias relations				
A	$=$	AB^2C	$=$	BC^2
B	$=$	AB^2C^2	$=$	AC^2
AB	$=$	ABC	$=$	C
AB^2	$=$	AC	$=$	BC

For example $AB^2 \times (I = ABC^2) \longrightarrow AB^2 = AC = BC$.

Remember again, that the factors A and B constitute an underlying complete factor structure and that factor C is introduced into this structure by means of the **generator equation** $C = AB$. This examplifies the general method of construction of fractional factorials.

The alias relations are most usefully written up with one relation per effect in the underlying factor structure and in standard order, as is shown in the table.

End of example 3.18

3.8.1 Resolution for fractional factorial designs

The term resolution describes which orders of effects are confounded with each other.

Corresponding to the example page 32 where fractional 2^k factorial designs were introduced, page 35 shows alias relations for a 2^{3-1} factorial experiment. It can be seen here that the main effects (first order effects) are confounded with 2-factor interactions (second order effects). Such an experiment is called a **resolution III** experiment. It should be noted that precisely **3** factors are present in the defining relation ($I = ABC$) for the experiment.

In the above example a 3^{3-1} factorial design is described with the defining relation $I = ABC^2$. This experiment too is called a resolution III experiment, since main effects are confounded with two-factor interactions (or higher). The defining relation involves at least **3** factors.

If all main effects in a fractional factorial design are confounded with effects of at least second order (2-factor interactions), the experiment is called a **resolution III** experiment. This corresponds to the fact that no effect in the defining relation of the experiment is of a lower order than 3.

If it holds true that no effect in the defining relation of the experiment is of a lower order than 4, the experiment is called a **resolution IV** experiment.

In a resolution IV experiment, the main effects are all confounded with effects of at least the third order, i.e. 3-factor interactions. Two-factor interactions will generally be confounded with other 2-factor interactions and/or interactions of a higher order in a resolution IV experiment.

In many practical circumstances, one can not assume in advance that the 2-factor interactions are unimportant compared with the main actions. One will therefore often want an experiment of resolution IV - at least .

In a resolution V experiment, the main effects are all confounded with effects of at least the fourth order, i.e. 4-factor interactions. Two-factor interactions will generally be confounded with 3-factor interactions and/or interactions of a higher order in a resolution V experiment.

As a rule, experiments with a higher resolution than V will not be needed to be done, if the factors involved are of a quantitative nature (temperature, pressure, concentration, time, density etc.) where main effects and 2-factor interactions are most frequently of considerably greater importance than the interactions of higher order.

3.8.2 Practical and general procedure

By means of the method outlined above, we can now construct arbitrary $1/p^q \times p^k$ factorial experiments and find the confoundings (the alias relations) in the experiment.

k factors are considered (A, B, C, ..., K) and these factors are ordered so that the first factors are a priori attributed the greatest importance. Meaning that the experimenter expects that factor A will prove to have the greatest importance (effect) on the response Y, and that B has the next greatest importance etc.

This ordering of the factors before the experiment is a great help both with regard to creating a suitable design and with regard to evaluating the results obtained.

In the review, in addition, one has to make a decision as to which factors that could be thought to interact and which ones that can be assumed to act additively. As the general rule, interactions between factors that have a large effect will be larger than interactions between factors with more moderate effects.

In addition, one will generally expect that interactions of a high order will be less important than interactions of a lower order.

In many cases one often allows oneself to assume that interactions of an order higher than 2 (i.e. 3-factor effects such as ABC, ABD, BCD etc. and effects of even higher order) are assumed to have considerably less importance than the main effects.

After this, the experiment is most simply constructed by starting with the complete factorial experiment, which is made up of the $(k-q)$ first (and expected to be most important) factors and putting the remaining q factors into this factor structure by confounding with effects regarded as negligible. The first $(k-q)$ and often most important factors will thereby form the underlying factor structure in the $1/p^q \times p^k$ factorial experiment wanted.

Example 3.19 : A $2^{-2} \times 2^5$ factorial experiment

Suppose that one considers 5 factors A, B, C, D and E, which one wants to evaluate each on 2 levels in a $1/2^2 \times 2^5$ factorial experiment, i.e. in $2^{5-2} = 2^3 = 8$ single experiments.

We imagine that a closer evaluation of the problem at hand indicates that factors A, B and C will have the greatest effect, and we thus let the underlying factor structure consist of precisely these factors.

The design is then generated by confounding factors D and E with effects in the underlying factor structure:

Generators

I		
A		
B		
AB		
C		
AC		
BC	$=$	E
ABC	$=$	D

$\Longrightarrow$

Aliasrelationer

I	$=$	$ABCD$	$=$	BCE	$=$	ADE
A	$=$	BCD	$=$	$ABCE$	$=$	DE
B	$=$	ACD	$=$	CE	$=$	$ABDE$
AB	$=$	CD	$=$	ACE	$=$	BDE
C	$=$	ABD	$=$	BE	$=$	$ACDE$
AC	$=$	BD	$=$	ABE	$=$	CDE
BC	$=$	AD	$=$	E	$=$	$ABCDE$
ABC	$=$	D	$=$	AE	$=$	$BCDE$

The experiment is a resolution III experiment.

The experiment can easily be written out using the tabular method as follows (if the principal fraction with the experiment "(1)" is chosen)

i	j	k	$l = (i+j+k)_2$	$m = (j+k)_2$	Code
0	0	0	0	0	(1)
1	0	0	1	0	*ad*
0	1	0	1	1	*bde*
1	1	0	0	1	*abe*
0	0	1	1	1	*cde*
1	0	1	0	1	*ace*
0	1	1	0	0	*bc*
1	1	1	1	0	*abcd*

which is probably the easiest way to find the experiment and at the same time time write out the whole plan, for example using a simple spreadsheet program.

We may write the plan as:

$D = ABC$, $E = -BC$ $l = (i+j+k)_2$, $m = (j+k)_2$							
(1)	ad	bde	abe	cde	ace	bc	$abcd$

where the indices are i, j, k, l and m for factors A, B, C, D and E, respectively, and the construction of the experiment is given with the previously introduced sign notation for 2^k experiments as well as with the index method that is used in the present chapter. See, too, the example on page 38.

There are 3 alternative possibilities, namely

$D = -ABC$, $E = -BC$ $l = (i+j+k+1)_2$, $m = (j+k)_2$							
d	a	be	$abde$	ce	$acde$	bcd	abc

$D = +ABC$, $E = +BC$ $l = (i+j+k)_2$, $m = (j+k+1)_2$							
e	ade	bd	ab	cd	ac	bce	$abcde$

$D = -ABC$, $E = +BC$ $l = (i+j+k+1)_2$, $m = (j+k+1)_2$							
de	ae	b	abd	c	acd	$bcde$	$abce$

A prerequisite for obtaining a "good" experiment by doing one of these experiments is that factors B and C do not interact with each other (BC and ABC unimportant), and that factors D and E do not interact with other factors at all or with each other. Factor A can be allowed to interact with the two factors B and C, i.e. AB and AC can differ from 0.

If these preconditions cannot be regarded as fulfilled to a reasonable degree, the experiment will not be appropriate to study the 5 factors simultaneously in a fractional factorial design with only 8 single experiments.

The alternatives will then be either to exclude one of the factors (by keeping it constant in the experiment) and being content with a $1/2 \times 2^4$ experiment or to extend the experiment to a $1/2 \times 2^5$ experiment, i.e. an experiment with 16 single experiments. The second alternative could reasonably be constructed by putting factor E into the factor structure consisting of factors A, B, C and D by the relation $ABCD = E$, with the following alias relations:

I	$=$	$ABCDE$
A	$=$	$BCDE$
B	$=$	$ACDE$
AB	$=$	CDE
C	$=$	$ABDE$
AC	$=$	BDE
BC	$=$	ADE
ABC	$=$	DE
D	$=$	$ABCE$
AD	$=$	BCE
ABD	$=$	CE
CD	$=$	ABE
ACD	$=$	BE
BCD	$=$	AE
$ABCD$	$=$	E

If it is assumed that all interactions of an order higher than 2 are unimportant, one gets **reduced alias relations** (the full defining relation is retained)

I	$=$	$ABCDE$
A	$=$	
B	$=$	
AB	$=$	
C	$=$	
AC	$=$	
BC	$=$	
	$=$	DE
D	$=$	
AD	$=$	
	$=$	CE
CD	$=$	
	$=$	BE
	$=$	AE
	$=$	E

One can see that all 2-factor interactions can be tested in this design. If some of these are reasonably small, their sums of squares could be pooled into a residual sum of squares and used to test higher order effects.

This experiment is a resolution V experiment.

One can choose one of the two following complementary experiments:

$E = -ABCD$ or $m = (i+j+k+l)_2$															
(1)	ae	be	ab	ce	ac	bc	$abce$	de	ad	bd	$abde$	cd	$acde$	$bcde$	$abcd$

$E = +ABCD$ or $m = (i+j+k+l+1)_2$															
e	a	b	abe	c	ace	bce	abc	d	ade	bde	abd	cde	acd	bcd	$abcde$

End of example 3.19

3.8.3 Alias relations with $1/p^q \times p^k$ experiments

When q factors are put into a factor structure consisting of $(k-q)$ factors, q generator equations are used. Each equation gives rise to one defining relation. That is, one finds q defining relations with q **defining contrasts**: I_1, I_2, ... , I_q. The complete defining relation can then symbolically be written as

$$I = I_1 * I_2 * \ldots * I_q$$

where the operator "$*$" is defined on page 58. By calculating the expression and replacing all "+" with "=", one finds the complete defining relation:

$$I = I_1 = I_2 = I_1 I_2 = I_1 I_2^2 = \ldots = I_1 I_2^{p-1} = \ldots = I_1 I_2^{p-1} \cdots I_q^{p-1}$$

corresponding to the "standard order" for q factors, called I_1, I_2, ... , I_q.

The alias relations of the experiment drawn up can be found for an arbitrary effect, F, by calculating the expression

$$F = F * I \Longrightarrow F = F * I_1 * I_2 * \ldots * I_q$$

and F and all the effects emerging on the right-hand side of the expression will be confounded with each other.

From calculation of the expression and replacement of "=" with "+" subsequently, one gets:

$$F = FI_1 = \ldots = FI_1^{p-1} = FI_2 = \ldots = FI_2^{p-1} = \ldots = F(I_1 I_2^{p-1} \cdots I_q^{p-1})^{p-1}$$

During the calculation of the single effects in the expression, it can be helpful to use the fact that for two arbitrary effects X and Y, it holds true that

$$XY^\alpha = X^\alpha Y$$

such that in the calculation of expressions with two effects, it is often easiest to lift up the simplest effect to the power in question. For example in a 3^2 factorial experiment, both $(AB^2C)^2(AB)$ and $(AB^2C)(AB)^2$ become BC. Test this.

Example 3.20 : Construction of $3^{-2} \times 3^5$ factorial experiment

Let there be 5 factors A, B, C, D and E each on 3 levels. One wants to do only 1/9 of the whole experiment, i.e. a total of $3^{5-2} = 3^3 = 27$ single experiments.

Again we start with a complete factor structure for three of the factors. And it is assumed that it is reasonable to choose A, B and C. In this factor structure, a further 2 factors are put in, namely D and E.

$$\begin{array}{c}
\text{Design generators} \\
\boxed{\begin{array}{ccc}
I & & \\
A & & \\
B & & \\
AB & & \\
AB^2 & & \\
C & & \\
AC & & \\
BC & & \\
ABC & & \\
AB^2C & & \\
AC^2 & & \\
BC^2 & = & E \\
ABC^2 & & \\
AB^2C^2 & = & D
\end{array}}
\end{array}
\Longrightarrow
\begin{array}{rcl}
I_1 & = & AB^2C^2D^2 \\
I_2 & = & BC^2E^2
\end{array}$$

Other alternatives can be chosen, for example to put both D and E into the 3-factor interaction ABC (which is decomposed in 4 parts each with 2 degrees of freedom) by for example $D = AB^2C^2$ and $E = ABC^2$. (Try to find the characteristics of this experiment (alias relations)).

With the confounding chosen in the table, one finds the defining relation

$$I = AB^2C^2D^2 = BC^2E^2 = (AB^2C^2D^2)(BC^2E^2) = (AB^2C^2D^2)(BC^2E^2)^2$$

which after reduction gives

$$I = AB^2C^2D^2 = BC^2E^2 = ACD^2E^2 = ABD^2E$$

The alias relations of the experiment are

I	$=$	$AB^2C^2D^2 = BC^2E^2 = ACD^2E^2 = ABD^2E =$ Defining relation
A	$=$	$ABCD = ABC^2E^2 = AC^2DE = AB^2DE^2 = BCD = AB^2CE = CD^2E^2 = BD^2E$
B	$=$	$AC^2D^2 = BCE = ABCD^2E^2 = AB^2D^2E = ABC^2D^2 = CE = AB^2CD^2E^2 = AD^2E$
AB	$=$	$ACD = AB^2C^2E^2 = AB^2C^2DE = ABDE^2 = BC^2D^2 = ACE = BC^2DE = DE^2$
AB^2	$=$	$AB^2CD = AC^2E^2 = ABC^2DE = ADE^2 = CD = ABCE = BCD^2E^2 = BDE^2$
C	$=$	$AB^2D^2 = BE^2 = AC^2D^2E^2 = ABCD^2E = AB^2CD^2 = BCE^2 = AD^2E^2 = ABC^2D^2E$
AC	$=$	$ABD = ABE^2 = ACDE = AB^2C^2DE^2 = BC^2D = AB^2C^2E = DE = BC^2D^2E$
BC	$=$	$AD^2 = BE = ABC^2D^2E^2 = AB^2CD^2E = ABCD^2 = CE^2 = AB^2D^2E^2 = AC^2D^2E$
ABC	$=$	$AD = AB^2E^2 = AB^2CDE = ABC^2DE^2 = BCD^2 = AC^2E = BDE = CDE^2$
AB^2C	$=$	$AB^2D = AE^2 = ABCDE = AC^2DE^2 = CD^2 = ABC^2E = BD^2E^2 = BCDE^2$
AC^2	$=$	$ABC^2D = ABCE^2 = ADE = AB^2CDE^2 = BD = AB^2E = CDE = BCD^2E$
BC^2	$=$	$ACD^2 = BC^2E = ABD^2E^2 = AB^2C^2D^2E = ABD^2 = E = AB^2C^2D^2E^2 = ACD^2E$
ABC^2	$=$	$AC^2D = AB^2CE^2 = AB^2DE = ABCDE^2 = BD^2 = AE = BCDE = CD^2E$
AB^2C^2	$=$	$AB^2C^2D = ACE^2 = ABDE = ACDE^2 = D = ABE = BC^2D^2E^2 = BC^2DE^2$

The alias relation for example of the main effect A is found with the help of:

$$A = A * I_1 * I_2 = A * (AB^2C^2D^2) * (BC^2E^2)$$

which gives

$$A = A(AB^2C^2D^2) = A(AB^2C^2D^2)^2 = A(BC^2E^2) = \ldots = A(AB^2C^2D^2)^2(BC^2E^2)^2$$

The expressions are organised in A-B-C-D-E order and the exponents reduced modulo 3. If necessary the exponent 1 on the first factor in the expressions is found by raising to the power of 2 and reducing modulo 3.

In the same way, the alias relations for each of the other effects are found in the underlying factor structure as shown in the table.

To elucidate the characteristics of the experimental design, all effects considered unimportant can be removed. This is the case for the BC effect and all other effects involving more than 2 factors. For the sake of clarity, the effects from the underlying factor structure are retained, but in parenthesis for assumingly unimportant effects.

In this way, the following table is found, which shows that 2-factor interactions are usually confounded with other 2-factor interactions or with main effects.

Reduced alias relations		
I	$=$	$AB^2C^2D^2 = BC^2E^2$
	$=$	$ACD^2E^2 = ABD^2E$
A	$=$	
B	$=$	CE
AB	$=$	DE^2
AB^2	$=$	CD
C	$=$	BE^2
AC	$=$	DE
(BC)	$=$	$AD^2 = BE = CE^2$
(ABC)	$=$	AD
(AB^2C)	$=$	$AE^2 = CD^2$
AC^2	$=$	BD
(BC^2)	$=$	E
(ABC^2)	$=$	$BD^2 = AE$
(AB^2C^2)	$=$	D

The experiment is a resolution III experiment. One can see that it is necessary to assume that a number of the 2-factor interactions are unimportant if the experiment is to be suitable.

If, for example, one can furthermore ignore interactions involving factors D and E, all else can be tested and estimated. This shows the usefulness of ordering the factors according to importance (i.e. main effects and thus interactions from D and E relatively small).

If it holds true that D and E have only additive effects and do not interact with the other factors, the alias relations can be reduced to the table shown on page 94, where the experiment was constructed.

There are $3 \times 3 = 9$ possibilities for implementing the experiment. If, for example, we want the fraction including "(1), the experiment will be given by the index restrictions $(i + 2j + 2k + 2l)_3 = 0$ and $(j + 2k + 2m)_3 = 0$.

If we want to write out a table of indices for the factors (that is the design), we use the **tabular method** and the generator equations $l = i + 2j + 2k$ and $m = j + 2k$ as follows:

Factors and levels					
A	B	C	D	E	Experiment
i	j	k	$l=(i+2j+2k)_3$	$m=(j+2k)_3$	code
0	0	0	0	0	(1)
1	0	0	1	0	ad
2	0	0	2	0	a^2d^2
0	1	0	2	1	bd^2e
1	1	0	0	1	abe
2	1	0	1	1	a^2bde
0	2	0	1	2	b^2de^2
1	2	0	2	2	$ab^2d^2e^2$
2	2	0	0	2	a^2bd^2e
0	0	1	2	2	cd^2e^2
1	0	1	0	2	ace^2
...	...	...	...	...	...
...	...	...	...	...	...
2	2	2	1	0	$a^2b^2c^2d$

The experiment is $3 \times 3 \times 3 = 27$ single experiments. One can also derive the single experiments by solving index equations. If 3 experiments which (independently) fulfil index equations are called x, y and z, the experiment will be:

(1)	x	x^2
y	xy	x^2y
y^2	xy^2	x^2y^2
z	xz	x^2z
yz	xyz	x^2yz
y^2z	xy^2z	x^2y^2z
z^2	xz^2	x^2z^2
yz^2	xyz^2	x^2yz^2
y^2z^2	xy^2z^2	$x^2y^2z^2$

In the underlying factor structure A, B and C, $x' = a$, $y' = b$ and $z' = c$ will be solutions, and the corresponding index sets are $(i,j,k) = (1,0,0)$, $(i,j,k) = (0,1,0)$ and $(i,j,k) = (0,0,1)$.

To find three solutions x, y and z, we therefore try with

$$\begin{array}{lcccl} x' : (i,j,k) = (1,0,0) & \Longrightarrow & (l,m) = (1,0) & \Longrightarrow & x = ad \\ y' : (i,j,k) = (0,1,0) & \Longrightarrow & (l,m) = (2,1) & \Longrightarrow & y = bd^2e \\ z' : (i,j,k) = (0,0,1) & \Longrightarrow & (l,m) = (2,2) & \Longrightarrow & z = cd^2e^2 \end{array}$$

The experiment then consists of the single experiments below:

(1)	ad	$(ad)^2$
bd^2e	$ad(bd^2e)$	$(ad)^2(bd^2e)$
$(bd^2e)^2$	$ad(bd^2e)^2$	$(ad)^2(bd^2e)^2$
cd^2e^2	$adcd^2e^2$	$(ad)^2cd^2e^2$
$(bd^2e)cd^2e^2$	$ad(bd^2e)cd^2e^2$	$(ad)^2(bd^2e)cd^2e^2$
$(bd^2e)^2cd^2e^2$	$ad(bd^2e)^2cd^2e^2$	$(ad)^2(bd^2e)^2cd^2e^2$
$(cd^2e^2)^2$	$ad(cd^2e^2)^2$	$(ad)^2(cd^2e^2)^2$
$(bd^2e)(cd^2e^2)^2$	$ad(bd^2e)(cd^2e^2)^2$	$(ad)^2(bd^2e)(cd^2e^2)^2$
$(bd^2e)^2(cd^2e^2)^2$	$ad(bd^2e)^2(cd^2e^2)^2$	$(ad)^2(bd^2e)^2(cd^2e^2)^2$

which are reorganised and the exponents reduced modulo 3:

(1)	ad	a^2d^2
bd^2e	abe	a^2bde
b^2de^2	$ab^2d^2e^2$	$a^2b^2e^2$
cd^2e^2	ace^2	a^2cde^2
bcd	$abcd^2$	a^2bc
b^2ce	ab^2cde	$a^2b^2cd^2e$
c^2de	ac^2d^2e	a^2c^2e
bc^2e^2	abc^2de^2	$a^2bc^2d^2e^2$
$b^2c^2d^2$	ab^2c^2	$a^2b^2c^2d$

In all, there are 9 different possibilities to construct the experiment, corresponding to the following table:

	$(i+2j+2k+2l)_3 = 0$	$(i+2j+2k+2l)_3 = 1$	$(i+2j+2k+2l)_3 = 2$
$(j+2k+2m)_3 = 0$	1 = the design shown	2	3
$(j+2k+2m)_3 = 1$	4	5	6
$(j+2k+2m)_3 = 2$	7	8	9

The three experiments "1", "4" and "7", for example, are complementary with regard to the generator equation $BC^2 = E$, i.e. the defining relation $I_2 = BC^2E^2$.

The same holds true for "2", "5" and "8", as well as for "3", "6" and "9".

If one carries out one of these sets of complementary experiments, one breaks the confoundings originating in the choice of $BC^2 = E$, and the whole experiment will then be a $1/3 \times 3^4$ experiment with the defining relation $I_1 = AB^2C^2D^2$ and the factors A, B, C and E in the underlying factor structure.

End of example 3.20

3.8.4 Estimation and testing in $1/p^q \times p^k$ factorial experiments

The important thing to realise is that $1/p^q \times p^k$ factorial experiments are constructed on the basis of a complete factor structure, i.e. the underlying complete factor structure. The analysis of the experiment is then done in the following steps:

1) For the fractional factorial design, the underlying complete factor structure is identified.
2) Data are arranged in accordance with this underlying structure, and the sums of squares are determined in the usual way for the factors and interactions in it.
3) The alias relations indicate how all the effects are confounded in the experiment. Thereby the sums of squares are found for effects that are not in the underlying factor structure.
4) By considering specific factor combinations in a single experiment, one can decide how the index relations are between the effects that are part of the same alias relation. In this way estimates are determined for the individual levels for the effects that are not in the underlying structure.

As an illustration of this, we consider the following.

Example 3.21 : Estimation in a $3^{-1} \times 3^3$-factorial experiment

We have factors A, B and C, all on 3 levels and we assume that A, B and C are purely additive, so that it is relevant to do a fractional factorial experiment instead of a complete factorial experiment.

As the underlying factor structure, the (A,B) structure is chosen.

The generalised (Kempthorne) effects in this structure are A, B, AB and AB^2. We choose to confound for example with AB, i.e. $AB_{i+j} = C_k$. This choice entails the following defining relation and alias relations:

I	$=$	ABC^2		
A	$=$	AB^2C	$=$	BC^2
B	$=$	AB^2C^2	$=$	AC^2
AB	$=$	ABC	$=$	C
AB^2	$=$	AC	$=$	BC

The index restriction on the principal fraction of the experiment, i.e. the fraction that contains "(1)", is $(i+j+2k)_3 = 0 \Leftrightarrow k = (i+j)_3$. Two linearly independent solutions have to be found for this and by starting in "a" and "b", one finds:

$$(i,j) = (1,0) \Longrightarrow k = (1+0) = 1 \;\;:\; \text{the experiment is } \; ac$$
$$(i,j) = (0,1) \Longrightarrow k = (0+1) = 1 \;\;:\; \text{the experiment is } \; bc$$

One possible experiment is the principal fraction in which "(1)" is a part:

$$\boxed{\begin{array}{ccc} (1) & ac & (ac)^2 \\ bc & acbc & (ac)^2bc \\ (bc)^2 & ac(bc)^2 & (ac)^2(bc)^2 \end{array}} \Longrightarrow \boxed{\begin{array}{ccc} (1) & ac & a^2c^2 \\ bc & abc^2 & a^2b \\ b^2c^2 & ab^2 & a^2b^2c \end{array}}$$

An alternative possibility is to carry out one of the (two) other fractions, for example the one of which the single experiment a is part. This fraction is determined by "multiplying" the principal fraction with a:

$$a \times \boxed{\begin{array}{ccc} (1) & ac & a^2c^2 \\ bc & abc^2 & a^2b \\ b^2c^2 & ab^2 & a^2b^2c \end{array}} \Longrightarrow \boxed{\begin{array}{ccc} a & a^2c & c^2 \\ abc & a^2bc^2 & b \\ ab^2c^2 & a^2b^2 & b^2c \end{array}}$$

This experiment has the index restriction $(i+j+2k)_3 = 1 \Leftrightarrow k = (i+j+2)_3$.

This experiment is chosen here and data are organised and analysed now in the usual way according to factors A and B (neglecting C):

$$\begin{array}{cccc} & \text{A=0} & \text{A=1} & \text{A=2} \\ \text{B=0} & c^2 & a & a^2c \\ \text{B=1} & b & abc & a^2bc^2 \\ \text{B=2} & b^2c & ab^2c^2 & a^2b^2 \end{array} \quad \begin{array}{c} \text{or} \\ \text{without } c: \end{array} \quad \boxed{\begin{array}{ccc} (1) & a & a^2 \\ b & ab & a^2b \\ b^2 & ab^2 & a^2b^2 \end{array}}$$

$$\begin{array}{lclclcl}
T_{A_0} = c^2 + b + b^2c & , & T_{A_1} = a + abc + ab^2c^2 & , & T_{A_2} = a^2c + a^2bc^2 + a^2b^2 \\
T_{B_0} = c^2 + a + a^2c & , & T_{B_1} = b + abc + a^2bc^2 & , & T_{B_2} = b^2c + ab^2c^2 + a^2b^2 \\
T_{AB_0} = c^2 + a^2bc^2 + ab^2c^2 & , & T_{AB_1} = a + b + a^2b^2 & , & T_{AB_2} = a^2c + abc + b^2c \\
T_{AB^2_0} = c^2 + abc + a^2b^2 & , & T_{AB^2_1} = a + a^2bc^2 + b^2c & , & T_{AB^2_2} = a^2c + b + ab^2c^2
\end{array}$$

$$
\begin{aligned}
\mathrm{SSQ}(A) &= \left([T_{A_0}]^2 + [T_{A_1}]^2 + [T_{A_2}]^2\right)/3r - [T_{tot}]^2/9r \quad , \quad f = 3-1 \\
\mathrm{SSQ}(B) &= \left([T_{B_0}]^2 + [T_{B_1}]^2 + [T_{B_2}]^2\right)/3r - [T_{tot}]^2/9r \quad , \quad f = 3-1 \\
\mathrm{SSQ}(AB) &= \left([T_{AB_0}]^2 + [T_{AB_1}]^2 + [T_{AB_2}]^2\right)/3r - [T_{tot}]^2/9r \quad , \quad f = 3-1 \\
\mathrm{SSQ}(AB^2) &= \left([T_{AB^2_0}]^2 + [T_{AB^2_1}]^2 + [T_{AB^2_2}]^2\right)/3r - [T_{tot}]^2/9r \quad , \quad f = 3-1
\end{aligned}
$$

where r indicates that a total of r single measurements could be made for each single experiment. In that case, it is assumed that these r repetitions are randomised over the whole experiment.

Finally, the effects can be estimated:

$$
\begin{array}{lll}
\widehat{A}_0 = T_{A_0}/3r - T_{tot}/9r \ , & \widehat{A}_1 = T_{A_1}/3r - T_{tot}/9r \ , & \widehat{A}_2 = T_{A_2}/3r - T_{tot}/9r \\
\widehat{B}_0 = T_{B_0}/3r - T_{tot}/9r \ , & \widehat{B}_1 = T_{B_1}/3r - T_{tot}/9r \ , & \widehat{B}_2 = T_{B_2}/3r - T_{tot}/9r \\
\widehat{AB}_0 = T_{AB_0}/3r - T_{tot}/9r \ , & \widehat{AB}_1 = T_{AB_1}/3r - T_{tot}/9r \ , & \widehat{AB}_2 = T_{AB_2}/3r - T_{tot}/9r \\
\widehat{AB^2}_0 = T_{AB^2_0}/3r - T_{tot}/9r \ , & \widehat{AB^2}_1 = T_{AB^2_1}/3r - T_{tot}/9r \ , & \widehat{AB^2}_2 = T_{AB^2_2}/3r - T_{tot}/9r
\end{array}
$$

To find the connection between C and the AB effect, the index relation is found from the specific experiment by considering two single experiments, for example c^2 $(i=0, j=0, k=2)$ and a $(i=1, j=0, k=0)$. One finds that it holds true that

	Index		
AB_{i+j}	0	1	2
C_k	2	0	1

where index $k = (i+j+2)_3$. Therefore, $\widehat{C}_0 = \widehat{AB}_1$, $\widehat{C}_1 = \widehat{AB}_2$ and $\widehat{C}_2 = \widehat{AB}_0$.

The mathematical model of the experiment could be written as

$$
Y_{ijk\nu} = \mu + A_i + B_j + C_{k=i+j+2} + E_{ijk\nu} \text{ , where } \ \nu = 1, 2 \ldots, r
$$

and, if $\nu > 1$, and there is used complete randomisation correctly, the residual sum of squares is

$$
\mathrm{SSQ}_{resid} = \sum_{i=1}^{3}\sum_{j=1}^{3}\left[\sum_{\nu=1}^{r} Y^2_{ijk\nu} - r \cdot \bar{Y}^2_{ijk\cdot}\right]
$$

Note that sums are made over indices i and j alone, since index k is of course given by i and j in this 3^{3-1} factorial experiment (which consists of 9 single experiments each repeated r times).

The precondition of additivity between the three factors A, B and C could be tested by testing the AB^2 effect against this sum of squares.

End of example 3.21

Example 3.22 : Two *SAS* examples

The calculations shown in the above example are relatively easy to program. A program can also be written for the statistical package *SAS* which will do the work. The following small example with data ($r = 1$) illustrates how analysis of variance can be done corresponding to factors A and B alone, i.e. the underlying factor structure.

	A=0	A=1	A=2
B=0	c^2=15.1	a=16.9	a^2c=23.0
B=1	b=9.8	abc=12.6	a^2bc^2=21.7
B=2	b^2c=5.0	ab^2c^2=10.0	a^2b^2=12.8

```
data exempel1;    input A B C Y;
AB  = mod(A+B,3);
AB2 = mod(A+B*2,3);
cards;
0 0 2 15.1
1 0 0 16.9
2 0 1 23.0
0 1 0 9.8
1 1 1 12.6
2 1 2 21.7
0 2 1 5.0
1 2 2 10.0
2 2 0 12.8
;
proc GLM;   class A B AB AB2;
  model       Y   = A B AB AB2;
  means             A B AB AB2;
run;
```

In the example starting on page 94 with data layout shown on page 96, a *SAS* job could look like the following:

```
data exempel2;    input A B C D E Y;
```

```
AB  = mod(A+B,3);      AB2 = mod(A+B*2,3);    AC = mod(A+C,3);
AC2 = mod(A+C*2,3);    BC  = mod(B+C,3);      BC2 = mod(B+C*2,3);
ABC = mod(A+B+C,3);    ABC2 = mod(A+B+C*2,3);
AB2C = mod(A+B*2+C,3); AB2C2 = mod(A+B*2+C*2,3);
cards;
0 0 0 0 0   31.0
1 0 0 1 0   16.0
2 0 0 2 0    4.0
0 1 0 0 1   23.8
1 1 0 1 1   23.6
2 1 0 2 1    9.7
.....
.....
.....
1 2 2 0 0   12.8
2 2 2 1 0   15.1
;
proc GLM ; class A B AB AB2 C AC AC2 BC BC2 ABC ABC2 AB2C AB2C2 ;
model       Y  = A B AB AB2 C AC AC2 BC BC2 ABC ABC2 AB2C AB2C2 ;
means            A B AB AB2 C AC AC2 BC BC2 ABC ABC2 AB2C AB2C2 ;
run;
```

And sums of squares and estimates of effects outside the underlying factor structure (A,B,C) can be directly found using the alias relations.

End of example 3.22

3.8.5 Fractional factorial design laid out in blocks

A fractional factorial design can be laid out in smaller blocks because of a wish to increase the accuracy in the experiment (different batches, groups of experimental animals, several days etc.). Other reasons could be that for the sake of saving time one wants to do the single experiments on parallel experimental facilities (several ovens, reactors, set-ups and such).

In the organisation of such an experiment, the fractional factorial design is first set up without regard to these possible blocks, since it is important first and foremost to have an overview of whether it is possible to construct a good fractional factorial design and how the factor effects of the experiment will be confounded.

When a suitable fractional factorial design has been constructed, a choice is made of which effect or effects would be suitable to confound with blocks, and a control is made that all block confoundings are sensible, perhaps the whole confounding table is reviewed. In the example on page 112, an example of this is shown.

Both during the construction of the fractional factorial design and in the subsequent formation of blocks for the experiment, the underlying factor structure is used, which most practically is composed of the most important factors, called A (first factor), B (second factor) etc.

In practice, one can naturally imagine a large number of variants of such experiments, but the following examples illustrate the technique rather generally.

Example 3.23 : A $3^{-2} \times 3^5$ factorial experiment in 3 blocks of 9 single experiments

Let us again consider an experiment in which there are 5 factors: A, B, C, D and E. A fractional factorial design consists of $3^3 = 27$ single experiments. We imagine that for practical reasons, it can be expedient to divide these 27 single experiments into 3 blocks of 9; for example it can be difficult to maintain uniform experimental conditions throughout all 27 single experiments.

The $1/3^2 \times 3^5$ factorial experiment wanted is found from two **generator equations**. As previously discussed, 2 factors, D and E, are introduced into a complete 3^3 factor structure for the factors A, B and C.

As in the example on page 94, we choose to put in D and E as in the following table:

$$\begin{array}{c}
\text{Design generators} \\
\boxed{\begin{array}{ccc}
I & & \\
A & & \\
B & & \\
AB & & \\
AB^2 & & \\
C & & \\
AC & & \\
BC & & \\
ABC & & \\
AB^2C & & \\
AC^2 & & \\
BC^2 & = & E \\
ABC^2 & & \\
AB^2C^2 & = & D
\end{array}}
\end{array}
\Longrightarrow
\begin{array}{ccc}
I_1 & = & AB^2C^2D^2 \\
I_2 & = & BC^2E^2
\end{array}$$

With this confounding, one gets (as previously) the **defining relation**

$$I = AB^2C^2D^2 = BC^2E^2 = ACD^2E^2 = ABD^2E$$

Construction of the experiment still follows the example on page 94, and one could perhaps again choose the **principal fraction** (see page 98):

(1)	ad	a^2d^2
bd^2e	abe	a^2bde
b^2de^2	$ab^2d^2e^2$	$a^2b^2e^2$
cd^2e^2	ace^2	a^2cde^2
bcd	$abcd^2$	a^2bc
b^2ce	ab^2cde	$a^2b^2cd^2e$
c^2de	ac^2d^2e	a^2c^2e
bc^2e^2	abc^2de^2	$a^2bc^2d^2e^2$
$b^2c^2d^2$	ab^2c^2	$a^2b^2c^2d$

To now divide this experiment consisting of the 27 single experiments into 3 blocks of 9, one chooses yet another generating relation which indicates how the blocks are formed.

When this relation is to be chosen, one again starts with the alias relations of the experiment in such a reduced form that one has an overview of how the main effects and/or interactions of interest are confounded. If we again follow the same example, these reduced alias relations could be as shown in the following table, in which we now also put in the blocks:

I	$=$	$AB^2C^2D^2 = BC^2E^2 = ACD^2E^2 = ABD^2E$
A	$=$	
B	$=$	CE
AB	$=$	DE^2
AB^2	$=$	CD
C	$=$	BE^2
AC	$=$	DE
(BC)	$=$	$AD^2 = BE = CE^2 =$ blocks
(ABC)	$=$	AD
(AB^2C)	$=$	$AE^2 = CD^2$
AC^2	$=$	BD
(BC^2)	$=$	E
(ABC^2)	$=$	$BD^2 = AE$
(AB^2C^2)	$=$	D

The easiest way to write out the this experiment is shown on page 108, however we will discuss the design a little in detail.

The choice of confounding with blocks means that all effects in the underlying factor structure that are not confounded with an effect of interest can be used. The effect BC could be such an effect (but, for example, not BC^2, why?).

The defining contrast BC has the index value $(j+k)_3$. The block division is then determined by whether $(j+k)_3 = 0$, 1 or 2.

To find the 3 blocks, one can again start with the underlying factor structure and it can be seen that the block division is solely determined by indices for the factors B and C, namely j and k.

As we saw in the previous example, the experiment, as described above, was also found on the basis of the underlying factor structure, and the block number corresponding to the single experiments is inserted in the following table:

Experiment	block	Experiment	block	Experiment	block
(1)	0	ad	0	a^2d^2	0
bd^2e	1	abe	1	a^2bde	1
b^2de^2	2	$ab^2d^2e^2$	2	$a^2b^2e^2$	2
cd^2e^2	1	ace^2	1	a^2cde^2	1
bcd	2	$abcd^2$	2	a^2bc	2
b^2ce	0	ab^2cde	0	$a^2b^2cd^2e$	0
c^2de	2	ac^2d^2e	2	a^2c^2e	2
bc^2e^2	0	abc^2de^2	0	$a^2bc^2d^2e^2$	0
$b^2c^2d^2$	1	ab^2c^2	1	$a^2b^2c^2d$	1

To find the three blocks, we could also solve the equations (modulo 3):

Generatorer:	Blocks $= BC$	$D = AB^2C^2$	$E = BC^2$
Block 0 :	$j+k=0$	$i+2j+2k+2l=0$	$j+2k+2m=0$
Block 1 :	$j+k=1$	$i+2j+2k+2l=0$	$j+2k+2m=0$
Block 2 :	$j+k=2$	$i+2j+2k+2l=0$	$j+2k+2m=0$

For example 2 solutions have to be found for "Block 0" which consists of $3 \times 3 = 9$ single experiments, and after that one further solution for each of the other two blocks.

The structure of block 0 can be illustrated

(1)	u	u^2
v	uv	u^2v
v^2	uv^2	u^2v^2

where u and v represent solutions to the equations for block 0.

For example, with $i = 1$ and $j = 0$, it is found from $j + k = 0$, that $k = 0$. Further, $i + 2j + 2k + 2l = 0$ indicates that $l = 1$, and from $j + 2k + 2m = 0$ is found that $m = 0$. A solution is thereby $u = ad$.

With $i = 0$, $j = 1$, it is found that $k = 2$, $l = 0$ and $m = 2$, from which $v = bc^2e^2$ is found.

(1)	$u = ad$	$u^2 = a^2d^2$
$v = bc^2e^2$	$uv = abc^2de^2$	$u^2v = a^2bc^2d^2e^2$
$v^2 = b^2ce$	$uv^2 = ab^2cde$	$u^2v^2 = a^2b^2cd^2e$

And it can be seen that this is precisely the block 0 found above.

To find block 1, one solution is derived for $j+k = 1$, $i+2j+2k+2l = 0$ and $j+2k+2m = 0$. Such a solution is $i = 0$, $j = 0$, $k = 1$, from which $l = 2$ and $m = 2$, corresponding to the single experiment cd^2e^2.

By "multiplying" cd^2e^2 on the already found block 0, block 1 is formed. Try it yourself.

Block 2 is found by solving the equations $j+k = 2$, $i+2j+2k+2l = 0$ and $j+2k+2m = 0$. A solution is $i = 0$, $j = 1$, $k = 1$, from which $l = 1$ and $m = 0$, corresponding to the single experiment bcd. This solution "is multiplied" on block 0, by which block 2 appears.

When the experiment is analysed, the block effect is reflected in the BC effect together with the other effects with which BC is confounded. In other words, the experiment is again analysed on the basis of the underlying factor structure determined by the factors A, B and C.

Finally the experiment could also be constructed directly on the basis of the generators that are chosen

I		
A		
B		
AB		
AB^2		
C		
AC		
BC	=	Blocks
ABC		
AB^2C		
AC^2		
BC^2	=	E
ABC^2		
AB^2C^2	=	D

and calculating the factor levels and block numbers as shown in the following table by means of the **tabular method**:

Experimental design $D = (A+2B+2C)_3$, $E = (B+2C)_3$ and $Block = (B+C)_3$							
No.	A	B	C	D	E	Block	Experiment
1	0	0	0	0	0	0	(1)
2	1	0	0	1	0	0	ad
3	2	0	0	2	0	0	a^2d^2
4	0	1	0	2	1	1	bd^2e
5	1	1	0	0	1	1	abe
6	2	1	0	1	1	1	a^2bde
7	0	2	0	1	2	2	b^2de^2
8	1	2	0	2	2	2	$ab^2d^2e^2$
9	2	2	0	0	2	2	$a^2b^2e^2$
10	0	0	1	2	2	1	cd^2e^2
11	1	0	1	0	2	1	ace^2
12	2	0	1	1	2	1	a^2cde^2
13	0	1	1	1	0	2	bcd
14	1	1	1	2	0	2	$abcd^2$
15	2	1	1	0	0	2	a^2bc
16	0	2	1	0	1	0	b^2ce
17	1	2	1	1	1	0	ab^2cde
18	2	2	1	2	1	0	$a^2b^2cd^2e$
19	0	0	2	1	1	2	c^2de
20	1	0	2	2	1	2	ac^2d^2e
21	2	0	2	0	1	2	a^2c^2e
22	0	1	2	0	2	0	bc^2e^2
23	1	1	2	1	2	0	abc^2de^2
24	2	1	2	2	2	0	$a^2bc^2d^2e^2$
25	0	2	2	2	0	1	$b^2c^2d^2$
26	1	2	2	0	0	1	ab^2c^2
27	2	2	2	1	0	1	$a^2b^2c^2d$

End of example 3.23

Finally two examples are given that illustrate the practical procedure in the construction of two resolution IV experiments for 8 and 7 factors respectively. These experiments are of great practical relevance, since they include relatively many factors in relatively few single experiments, namely only 16. At the same time, the examples show division into 2 and 4 blocks, enabling the advantages such blocking can have.

Example 3.24 : A $2^{-4} \times 2^8$ factorial in 2 blocks

The experiment could be done in connection with a study of the manufacturing process for a drug, for example.

We imagine that the given factors and their levels are circumstances which, during manufacture, one normally aims to keep constant, or at least within given limits. It is the effect of variation within these permitted limits that we want to study.

Eight factors are studied in a $2^{-4} \times 2^8$ factorial in two blocks. The 8 factors are 2 waiting times during two phases of the process, 3 temperatures, 2 pH values and the content of zinc in the finished product. The factors are ordered so that factor A is considered the most important, while B is the next most important etc.

The experiment is a *resolution IV* experiment. Under the assumption of negligible third order interactions, all main effects can be analysed in this design.

The experiment is randomised within two blocks, as it is assumed that it is done in two facilities (R_0 and R_1) in parallel experiments in completely random order.

The experiment is constructed as given in the following tables.

Factors and levels chosen		
Factor	Low level	High level
A: $\text{Tid}_{\text{opløsning 1 + filtrering}}$	(.) 70+30 min	(a) 30+70 min
B: $\text{T}_{\text{blanding 1}}$	(.) 20 ± 1 °C	(b) 27 ± 1 °C
C: $\text{Tid}_{\text{opløsning 2}}$	(.) 30 min	(c) 100 min
D: $\text{T}_{\text{opløsning 2}}$	(.) 5 ± 1 °C	(d) 17 ± 1 °C
E: T_{proces}	(.) 5 ± 1 °C	(e) 17 ± 1 °C
F: $\text{pH}_{\text{råprodukt 1}}$	(.) 2.65 ± 0.02	(f) 3.25 ± 0.02
G: $\text{Zink}_{\text{færdig mix}}$	(.) 20.0 μg/ml	(g) 26.0 μg/ml
H: $\text{pH}_{\text{færdig mix}}$	(.) 7.20 ± 0.02	(h) 7.40 ± 0.02

Confoundings		
I		
A		
B		
AB		
C		
AC		
BC		
ABC	=	H
D		
AD		
BD		
ABD	=	G
CD		
ACD	=	F
BCD	=	E
ABCD	=	Blocks

We use the **tabular method** for calculating the levels of the factors and the block number on the basis of the underlying complete factor structure consisting of factors A, B, C and D, as shown in the following table:

Experimental design $E=(B+C+D)_2$, $F=(A+C+D)_2$, $G=(A+B+D)_2$, $H=(A+B+C)_2$ and $Facility=(A+B+C+D)_2$												
No.	A	B	C	D	E	F	G	H	Experiment	Facility	Randomis.
1	0	0	0	0	0	0	0	0	(1)	R_0	9
2	1	0	0	0	0	1	1	1	a fgh	R_1	4
3	0	1	0	0	1	0	1	1	b egh	R_1	6
4	1	1	0	0	1	1	0	0	ab ef	R_0	11
5	0	0	1	0	1	1	0	1	c efh	R_1	16
6	1	0	1	0	1	0	1	0	ac eg	R_0	7
7	0	1	1	0	0	1	1	0	bc fg	R_0	5
8	1	1	1	0	0	0	0	1	abc h	R_1	10
9	0	0	0	1	1	1	1	0	d efg	R_1	12
10	1	0	0	1	1	0	0	1	ad eh	R_0	3
11	0	1	0	1	0	1	0	1	bd fh	R_0	1
12	1	1	0	1	0	0	1	0	abd g	R_1	14
13	0	0	1	1	0	0	1	1	cd gh	R_0	13
14	1	0	1	1	0	1	0	0	acd f	R_1	2
15	0	1	1	1	1	0	0	0	bcd e	R_1	8
16	1	1	1	1	1	1	1	1	abcd efgh	R_0	15

Prescriptions for the single experiments

Below are shown the factor settings for the two first single experiments and the two last ones.

Carried out on facility R_0 Testnr. FF– 1 X19 Experiment = bd(fh)	
Proces parameter	Level in experiment
A: $Tid_{opløsning\ 1\ +\ filtrering}$	(.) 70+30 min
B: $T_{blanding\ 1}$	(b) 27 ± 1 °C
C: $Tid_{opløsning\ 2}$	(.) 30 min
D: $T_{opløsning\ 2}$	(d) 17 ± 1 °C
E: T_{proces}	(.) 5 ± 1 °C
F: $pH_{råprodukt\ 1}$	(f) 3.25 ± 0.02
G: $Zink_{færdig\ mix}$	(.) 20.0 μg/ml
H: $pH_{færdig\ mix}$	(h) 7.40 ± 0.02

Carried out on facility R_1 Testnr. FF– 2 X19	
Experiment = acd(f)	
Proces parameter	Level in experiment
A: $\text{Tid}_{\text{opløsning 1 + filtrering}}$	(a) 30+70 min
B: $T_{\text{blanding 1}}$	(.) 20 ± 1 °C
C: $\text{Tid}_{\text{opløsning 2}}$	(c) 100 min
D: $T_{\text{opløsning 2}}$	(d) 17 ± 1 °C
E: T_{proces}	(.) 5 ± 1 °C
F: $\text{pH}_{\text{råprodukt 1}}$	(f) 3.25 ± 0.02
G: $\text{Zink}_{\text{færdig mix}}$	(.) 20.0 μg/ml
H: $\text{pH}_{\text{færdig mix}}$	(.) 7.20 ± 0.02

Carried out on facility R_0 Testnr. FF– 15 X19	
Experiment = abcd(efgh)	
Proces parameter	Level in experiment
A: $\text{Tid}_{\text{opløsning 1 + filtrering}}$	(a) 30+70 min
B: $T_{\text{blanding 1}}$	(b) 27 ± 1 °C
C: $\text{Tid}_{\text{opløsning 2}}$	(c) 100 min
D: $T_{\text{opløsning 2}}$	(d) 17 ± 1 °C
E: T_{proces}	(e) 17 ± 1 °C
F: $\text{pH}_{\text{råprodukt 1}}$	(f) 3.25 ± 0.02
G: $\text{Zink}_{\text{færdig mix}}$	(g) 26.0 μg/ml
H: $\text{pH}_{\text{færdig mix}}$	(h) 7.40 ± 0.02

Carried out on facility R_1 Testnr. FF– 16 X19	
Experiment = c(efh)	
Proces parameter	Level in experiment
A: $\text{Tid}_{\text{opløsning 1 + filtrering}}$	(.) 70+30 min
B: $T_{\text{blanding 1}}$	(.) 20 ± 1 °C
C: $\text{Tid}_{\text{opløsning 2}}$	(c) 100 min
D: $T_{\text{opløsning 2}}$	(.) 5 ± 1 °C
E: T_{proces}	(e) 17 ± 1 °C
F: $\text{pH}_{\text{råprodukt 1}}$	(f) 3.25 ± 0.02
G: $\text{Zink}_{\text{færdig mix}}$	(.) 20.0 μg/ml
H: $\text{pH}_{\text{færdig mix}}$	(h) 7.40 ± 0.02

End of example 3.24

Example 3.25 : A $2^{-3} \times 2^7$ factorial experiment in 4 blocks

Suppose that there are 7 factors which one wants studied in 16 single experiments. The first four factors, A, B, C and D are used as the underlying factor structure. The factors E, F and G are put into this according to the table below in a resolution IV experiment.

At the same time, one could want the 16 single experiments done in 4 blocks of 4 single experiments. Since $4 = 2\times2$ blocks have to be used, 2 defining equations for blocks have to be chosen. A suggestion for the construction of the experimental design could be:

Generators		
I		
A		
B		
AB		
C		
AC		
BC		
ABC	=	blocks
D		
AD		
BD		
ABD	=	G
CD		
ACD	=	F
BCD	=	E
ABCD	=	blocks

With these choices, the effects ABC and ABCD, but also the effect ABC×ABCD will be confounded with blocks. Now, since ABC×ABCD = D, this is not a good choice, because the main effect D is obviously confounded with blocks. A better choice could be:

Generators		
I		
A		
B		
AB		
C		
AC		
BC		
ABC	=	blocks
D		
AD		
BD		
ABD	=	G
CD		
ACD	=	F
BCD	=	blocks
ABCD	=	E

This choice will entail that ABC, BCD and ABC×BCD = AB will be confounded with blocks. The experimental design can be written out using the tabular method:

Design $E=(A+B+C+D)_2$, $F=(A+C+D)_2$, $G=(A+B+D)_2$ and $Block=(A+B+C)_2 + 2\times(B+C+D)_2$									
Nr	A	B	C	D	E	F	G	Experiment	Block
1	0	0	0	0	0	0	0	(1)	0
2	1	0	0	0	1	1	1	a efg	1
3	0	1	0	0	1	0	1	b eg	3
4	1	1	0	0	0	1	0	ab f	2
5	0	0	1	0	1	1	0	c ef	3
6	1	0	1	0	0	0	1	ac g	2
7	0	1	1	0	0	1	1	bc fg	0
8	1	1	1	0	1	0	0	abc e	1
9	0	0	0	1	1	1	1	d efg	2
10	1	0	0	1	0	0	0	ad	3
11	0	1	0	1	0	1	0	bd f	1
12	1	1	0	1	1	0	1	abd eg	0
13	0	0	1	1	0	0	1	cd g	1
14	1	0	1	1	1	1	0	acd ef	0
15	0	1	1	1	1	0	0	bcd e	2
16	1	1	1	1	0	1	1	abcd fg	3

End of example 3.25

Index

My own notes: