Modelling and design of experiments

Chapter 4: Classical designs

Dr Jean-Marie Fürbringer

École Polytechnique Fédérale de Lausanne

Fall 2024

4.1 The Plackett Burman design

4.1.1 Introduction to Plackett-Burman Design

- **Purpose and Origin** : Developed by Robin Plackett and John Burman in 1946 to efficiently screen large numbers of factors with limited experimental runs.
- **Efficiency and Assumptions** : A two-level factorial design focusing on main effects ; interactions are typically assumed to be negligible in early-stage experimentation.
- **Applications and Benefits** : Widely used in fields like chemistry, engineering, and biotechnology, where it identifies key factors before committing to more complex experimental designs.

Example : Fermentation Process Optimization

- **Challenge** : Optimizing factors in fermentation, such as pH, temperature, nutrient concentration, and aeration, traditionally required numerous experiments.
- **Solution with Plackett-Burman** : By applying the design, companies could efficiently screen multiple factors, saving time and resources.
- **Impact** : Reduced time and cost, with quicker identification of key variables that influence yield and efficiency, helping companies like Merck streamline production.

Example : NASA Materials Engineering

- **Challenge** : Identifying critical material properties affecting durability and performance in harsh space environments would traditionally require numerous tests.
- **Solution with Plackett-Burman** : NASA used the design to screen multiple factors simultaneously, allowing efficient narrowing down of essential properties.
- **Impact** : Enabled quicker optimization of materials for spacecraft, resulting in robust material choices with fewer resources compared to standard methods.

4.1.2 Plackett Burman designs (Hadamard)

- Efficient estimation of the main effects of a system without (important) interactions : $y = a_o + \sum a_i x_i + \epsilon$
- N runs only necessary to analyze till N-1 factors
- Essay matrix composed of '1' y '-1' : some corner points of the parallellepipedic domain
- Usage : for screening
- Model matrix generated with a Matlab function : hadamard(N)
- How to generate these matrices ?
  - from the recursive relation
  - from a generator that can be found in the literature

4.1.3 Recursive construction for $N = 2^L$

The Hadamard matrices of order $N = 2^L$ can be generated from the recursive relation

$$H_{k+1} = \begin{bmatrix} H_k & H_k \\ H_k & -H_k \end{bmatrix}$$

with $H_0 = 1$. For example, when $N = 4$, we have

$$H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

4.1.4 Generators of Hadamard matrices

- 8 runs, 7 factors max :
  $+ + + - + - -$ or $- - - + - + +$
- 12 runs, 11 factors max :
  $+ + - + + + - - - + -$ or ...
- 20 runs, 19 factors max :
  $+ + - - + + + + - + - + - - - - + + -$ or ...
- 24 runs, 23 factors max :
  $+ + + + + - + - + + - - + + - - + - + - - - -$ or ...

4.1.5 Building a Plackett Burman

First line= generator	$- - - + - + +$	1
(N-2) next liges : circular permutations	$+ - - - + - +$	2
	$+ + - - - + -$	3
	$- + + - - - +$	4
	$+ - + + - - -$	5
	$- + - + + - -$	6
	$- - + - + + -$	7
Last line de + [1]	$+ + + + + + +$	8

When using the routine hadamard() with Matlab, the order of the experiments is different and the column of '+1' corresponding to the constant is also provided systematically

1. the number of + or of − must be the equal in each column

4.1.6 LSF estimates for orthogonal systems

The system $Y = X\alpha + \epsilon$ has for solution

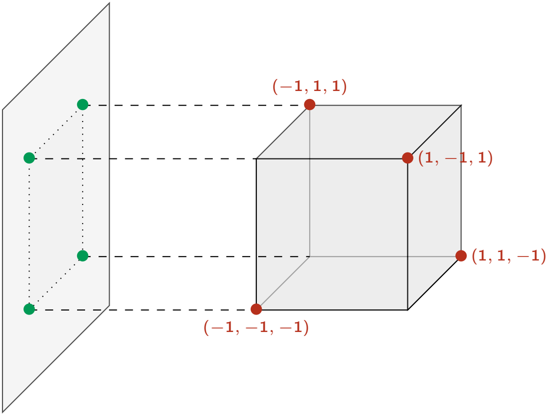
$$\begin{cases} \hat{\alpha} = (X'X)^{-1}\ X'Y \\ \hat{\epsilon} = Y - X\hat{\alpha} \end{cases}$$

However $(X'X)^{-1} = \frac{1}{N_{exp}}\ I_{N_{exp}}$, thus $\hat{\alpha} = \frac{1}{N_{exp}}\ X'Y$

Example : H_4

$$(H_4' H_4) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

4.1.7 Projectivity



When an Hadamard matrix is projected on a reference plane $x_i = 0$, the projection constitutes a factorial design.

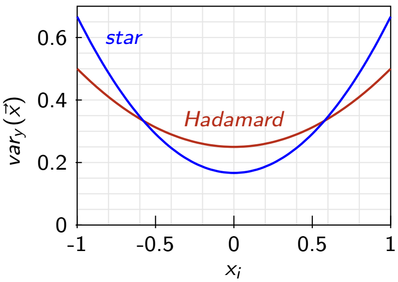
4.1.8 The VIF of an Hadamard design are optimal

$$D_{H4} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

$$VIF = diag\left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1}\right] = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

The design is orthogonal (the vectors constituting the matrix of the model are orthogonal , each one in reference to the others)

4.1.9 Variance Function



The variance function of the Hadamard design is prefered because, even if it is higher at the center of the domain, it increases less when towards the limit of the domain : the information is better spead in

Model :

$$y = a_o + \sum a_i x_i$$

Variance function :

$$var_y(\vec{x}) = f'(\vec{x}) \cdot (X'X)^{-1} \cdot f(\vec{x})$$

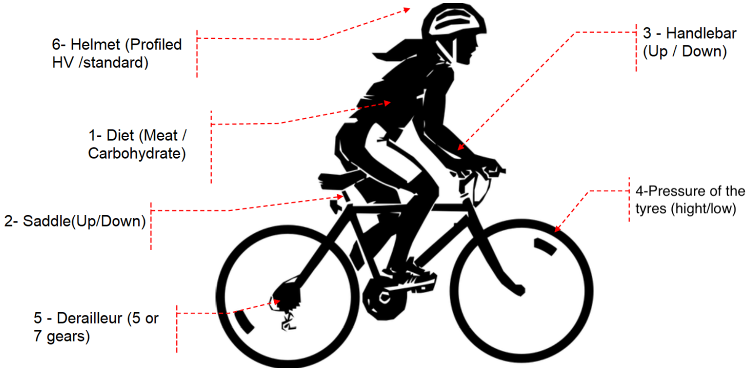
For the star design :

$$var_y(\vec{x}) = \frac{1}{6} + \frac{1}{2} \sum_i x_i^2$$

For the Hadamard design :

$$var_y(\vec{x}) = \frac{1}{4} + \frac{1}{4} \sum_i x_i^2$$

4.1.10 Optimizing for the bicycle race



4.1.11 Mind map

Bicycle

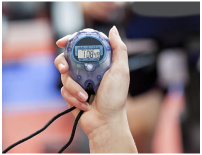
- **Objective**: Optimization of a bicycle to maximize the chance to win the race
- **Type of investigation**: Experiment (factors can be manipulated)
- **Factors**:
  - Continuous :
    1. Height of the saddle
    2. Height of the handle bar
    3. Pressure of the tyres
  - Discreet :
    4. Breakfeast (Cereals/Beacon & eggs)
    5. Jacket (standard/hightec
    6. Gear (5-speed/7/speed)
- **Response(s)**: Time to perform a given tour
- **Model(s)**: Empirical linear model $Y = a_o + \sum_i a_i x_1$
- **Strategies**: Plackett-Burman design : 8 runs

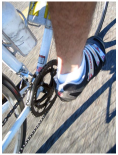
4.1.12 Hadamard Design H_8

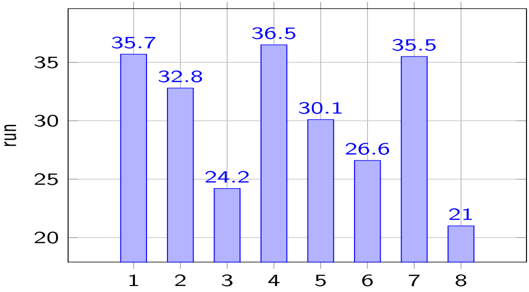
Table – Hadamard design of 8 runs

	Regime	Saddle	Handlebar	Pressure	Gear	Helmet	-
+	meat	High	high	high	7 speeds	profiled	-
−	pasta	Low	low	low	5 speeds	standard	-
1	1	1	1	1	1	1	1
2	-1	1	-1	1	-1	1	-1
3	1	-1	-1	1	1	-1	-1
4	-1	-1	1	1	-1	-1	1
5	1	1	1	-1	-1	-1	-1
6	-1	1	-1	-1	1	-1	1
7	1	-1	-1	-1	-1	1	1
8	-1	-1	1	-1	1	1	-1

4.1.13 Measurements





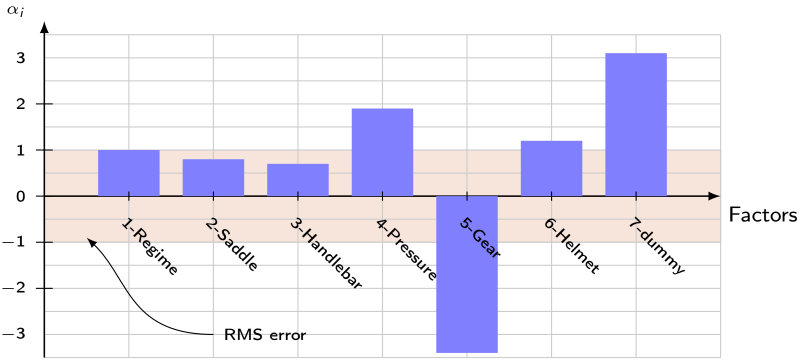


4.1.14 Determination of the effects

$$\hat{\alpha} = \left(X'X\right)^{-1} X'Y = \frac{1}{N} X'Y$$

$$\hat{\alpha} = \frac{1}{8} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}^T \begin{pmatrix} 35.7 \\ 32.8 \\ 24.2 \\ 36.5 \\ 30.1 \\ 26.0 \\ 35.5 \\ 22.1 \end{pmatrix} = \begin{pmatrix} 30.4 \\ 1.0 \\ 0.8 \\ 0.7 \\ 1.9 \\ -3.4 \\ 1.2 \\ 3.1 \end{pmatrix}$$

4.1.15 Half-Effects



4.1.16 Alias between main effects and interactions

- linear part of the model
  $y_1 = a_o + \sum_i a_i x_i = X_1 \alpha$
- interactions part of the model
  $y_2 = \sum_{i<j} a_{ij} x_i x_j = X_2 \alpha_2$
- Matrix of aliases between parts
  $A = (X_1' X_1)^{-1}\ X_1' X_2 = \frac{1}{8} X_1' X_2$

$$
\begin{aligned}
l_1 &= a_1 + a_{23} + a_{45} \\
l_2 &= a_2 + a_{13} + a_{46} \\
l_3 &= a_3 + a_{12} + a_{56} \\
l_4 &= a_4 + a_{15} + a_{26} \\
l_5 &= a_5 + a_{14} + a_{36} \\
l_6 &= a_6 + a_{24} + a_{35} \\
l_7 &= a_{16} + a_{25} + a_{34}
\end{aligned}
$$

	12	13	14	15	16	23	24	25	26	34	35	36	45	46	56
a_o	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a_1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
a_2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
a_3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
a_4	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
a_5	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
a_6	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
a_7	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0

4.1.17 Full foldover

Design	Essay matrix	Model matrix
Hadamard	E	$\begin{bmatrix}1 & E\end{bmatrix}$
Foldover	$-E$	$\begin{bmatrix}1 & -E\end{bmatrix}$
Full foldover	$\begin{bmatrix}E \\ -E\end{bmatrix}$	$\begin{bmatrix}1 & E \\ 1 & -E\end{bmatrix}$

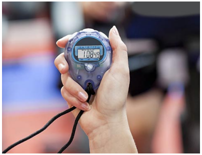
4.1.18 Deliasing with the *foldover*

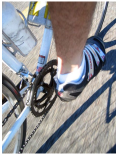
$$[1 \quad E] \qquad\qquad [1 \quad -E]$$

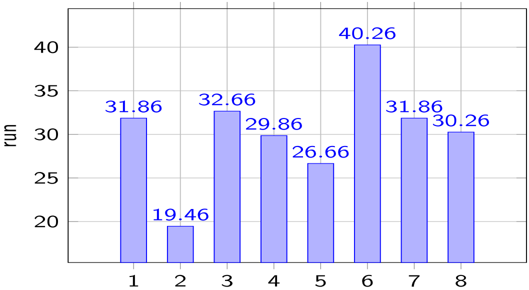
$$
\begin{aligned}
l_1 &= a_1 + a_{23} + a_{45} & \qquad l_1 &= a_1 - a_{23} - a_{45} \\
l_2 &= a_2 + a_{13} + a_{46} & l_2 &= a_2 - a_{13} - a_{46} \\
l_3 &= a_3 + a_{12} + a_{56} & l_3 &= a_3 - a_{12} - a_{56} \\
l_4 &= a_4 + a_{15} + a_{26} & l_4 &= a_4 - a_{15} - a_{26} \\
l_5 &= a_5 + a_{14} + a_{36} & l_5 &= a_5 - a_{14} - a_{36} \\
l_6 &= a_6 + a_{24} + a_{35} & l_6 &= a_6 - a_{24} - a_{35} \\
l_7 &= a_{16} + a_{25} + a_{34} & l_7 &= -a_{16} - a_{25} - a_{34}
\end{aligned}
$$

The foldover design changes the sign of the linear combination of the effects and then doubles the rank of the system

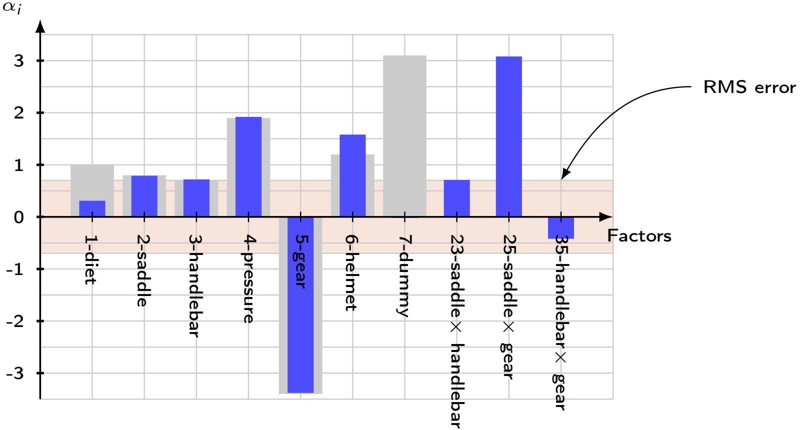
4.1.19 New Measurement







4.1.20 New Half-Effects



$$\hat{y} = a_0 + \sum a_i x_i + a_{23} x_2 x_3 + a_{25} x_2 x_5 + a_{35} x_3 x_5$$

4.1.21 Interaction coefficients (bicycle case)

In the present case it is finally possible

- to identify existing interactions $a_{23}, a_{25} eta_{35}$
- to see that they were not aliased in the full foldover
- to add columns x_2x_3, x_2x_5 and x_3x_5 to the model matrix
- to estimate the coefficients a_{23}, a_{25} and a_{35}

This aspect is used systematically in the Tagushi method

4.1.22 An example of a model with aliases

- Let's consider a model $y = 10 + 12\ x_1 + 16\ x_2 + 6\ x_3 + 4\ x_1 x_2$
- Let's choose a design that aliases a_{12} with a_3

$$E = \begin{pmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

- The third column is created from the product of columns '1' and '2', the generator being $3 = 12$
- The results of the experiments are then the following

$$\begin{aligned}
10 + 12 \times (-1) + 16 \times (-1) + 6 \times (+1) + 4 \times (-1) \times (-1) + \epsilon_1 &= -8 + \epsilon_1 \\
10 + 12 \times (-1) + 16 \times (+1) + 6 \times (-1) + 4 \times (-1) \times (+1) + \epsilon_2 &= 4 + \epsilon_2 \\
10 + 12 \times (+1) + 16 \times (-1) + 6 \times (-1) + 4 \times (+1) \times (-1) + \epsilon_3 &= -4 + \epsilon_3 \\
10 + 12 \times (+1) + 16 \times (+1) + 6 \times (+1) + 4 \times (+1) \times (+1) + \epsilon_4 &= 157 + \epsilon_4
\end{aligned}$$

4.1.23 Effect calculation

- The design has 4 runs, then 4 numbers (contrasts) can be computed
- If a model matrix with 5 columns is built :

$$X = \begin{pmatrix} 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The two last columns are identical : they will the produce the same result

$$\alpha^T = (10 \quad 12 \quad 16 \quad 10 \quad 10)^T$$

- Only $l_o = a_o, l_1 = a_1$ and $l_2 = a_2$ have been identified *correctly*.
- $l_3 = l_4 = a_3 + a_{12}$
- The matrix of the model can have only 4 columns

4.1.24 Matrix of alias

The alias matrix of the linear model versus the interaction 12 confirms that the alias is precisely the expected one.

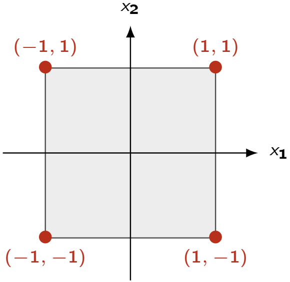
$$A = \frac{1}{4}\begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

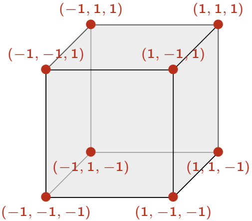
4.1.25 Summary - Plans de Hadamard

- Best possible estimates of the main effects : $var(\alpha_i) = \frac{\sigma^2}{N_{exp}}$
- The aliases are different from an Hadamard matrix to another
- Possibility to dealiase the main effects from the first order interactions with a foldover
- Possibility to estimate the interactions if they are only a few and identified
- Plackett-Burman designs are not the only screening designs... it exists a few others such as the hyper-saturated designs

4.2 Full factorial design 2^n

4.2.1 Full factorial design 2^n

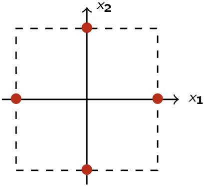




$$E = \begin{pmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$

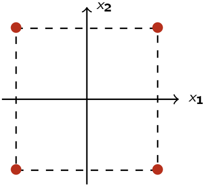
$$E = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

4.2.2 OFAT vs factorial design



$$\text{var(effet)} = 2\sigma^2$$

si $\frac{\sigma}{\mu} = 10\%$ $\rightarrow \frac{s}{m} = \frac{\sqrt{2}\ \sigma}{\mu} = 14\%$



$$\text{var(effet)} = \frac{\sigma^2}{N}$$

si $\frac{\sigma}{\mu} = 10\%$ $\rightarrow \frac{s}{m} = \frac{\sigma}{\sqrt{4}\ \mu} = 5\%$

4.2.3 What can we do with a 2^N factorial design ?

Identify the main effects and every interaction of n factors :

$$y = a_0 + \sum_{i}^{n} a_i x_i + \sum_{i<j}^{n} a_{ij} x_i x_j + \sum_{i<j<k}^{n} a_{ijk} x_i x_j x_k + \dots$$

The number of coefficients by level is

$$N_m = \begin{pmatrix} n \\ m \end{pmatrix} = \frac{n!}{m!(n-m)!}, \ m = 0 : n$$

And we know that $\sum_{m=0}^{n} \begin{pmatrix} n \\ m \end{pmatrix} = 2^n = N_{exp}$

In MATLAB use the function b=nchoosek(n,k) to calculate the binomial coefficient

4.2.4 2^N Factorial design on Matlab

Matlab

- Routine *fullfact()* generates a complete factorial plan depending on the number of indicated levels for each factor
- The levels begin at 1→ standardize the matrix
- *E=2*(fullfact([2 2 . . .])-1.5)*

- Routine *ff2n()* generates a complete factorial plan with two levels for each factor
- The levels begin at 0→ standardize the matrix
- *E=2*(ff2n(nfact)-0.5)*
- *E=fracfact('a b c d')*

4.2.5 Model matrix of a 2^n design

Matlab

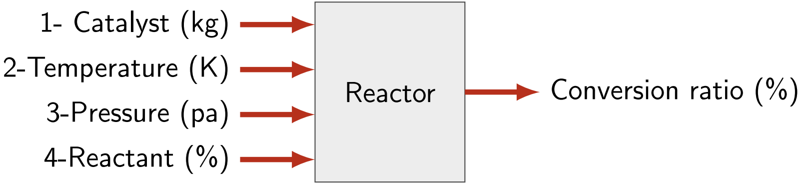
- *X=x2fx(E,modelspec)*
- If *modelspec* is the keyword *'interactions '* only the interactions of first order are computed
- To integrate more (or less) interaction coefficients $2 \times 2, 3 \times 3, \dots n \times n$, the model must be defined by a coefficient matrix :

$$\textbf{modelspec} = \begin{pmatrix} 0 & \dots & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \\ 1 & 1 & 1 & 0 & \dots \\ etc \end{pmatrix}$$

- Example :
modelspec=[zeros(1,n) ;eye(N) ;unique(perms([1 1 0 . . .],"rows")) ; . . .]

4.2.6 Case of a chemical reactor

4.2.7 Block diagram



4.2.8 Mind map

Chemical reactor

- **Objective**: Identification of the dominant factors and interactions managing yield of the chemical reaction in an industrial reactor
- **Type of investigation**: Experiment (factors can be manipulated)
- **Factors**: Continuous
  1. Catalyst charge $in[10lb, 15lb]$
  2. Temperature $in[220°C, 240°C]$
  3. Pressure $in[50psi, 80psi]$
  4. Concentration of reactant A $in[10\%, 12\%i]$
- **Response(s)**: Conversion ratio (%)
- **Model(s)**: Empirical linear model with interaction $Y = a_o + \sum_i a_i\ x_1 + \sum_{i<j} a_{ij}\ x_i\ x_j$
- **Strategies**: Full factorial design : $2^4 = 16$ runs

4.2.9 : Essay Matrix $2^4 \quad \rightarrow 16 \times 4$

Standardized matrix

x_1	x_2	x_3	x_4
-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

Matrix with laboratory values

Catalyst [lb]	Temperature [$^\circ C$]	Pressure [psi]	Reactant A [-]
10	220	50	10%
10	220	50	12%
10	220	80	10%
10	220	80	12%
10	240	50	10%
10	240	50	12%
10	240	80	10%
10	240	80	12%
15	220	50	10%
15	220	50	12%
15	220	80	10%
15	220	80	12%
15	240	50	10%
15	240	50	12%
15	240	80	10%
15	240	80	12%

Matlab

```
E=fracfact('a b c d')
```

4.2.10 Model Matrix (16 × 11)

$$y = a_0 + \sum a_i x_i + \sum_{i<j} a_{ij} x_i x_j + \epsilon \qquad (1)$$

I	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4
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	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	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

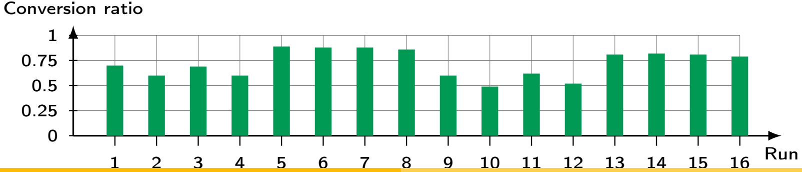
Matlab

```
X = x2fx(E,'interactions')
```

4.2.11 Measurement

4.2.11 Case 8 : Results of the measurement

Run	Order	Catalyst	Temp.	Pressure	Reactant	Conversion [%]
1	9	-1	-1	-1	-1	70
2	16	-1	-1	-1	1	60
3	15	-1	-1	1	-1	69
4	3	-1	-1	1	1	60
5	10	-1	1	-1	-1	89
6	11	-1	1	-1	1	88
7	1	-1	1	1	-1	88
8	6	-1	1	1	1	86
9	2	1	-1	-1	-1	60
10	5	1	-1	-1	1	49
11	9	1	-1	1	-1	62
12	12	1	-1	1	1	52
13	4	1	1	-1	-1	81
14	14	1	1	-1	1	82
15	13	1	1	1	-1	81
16	7	1	1	1	1	79

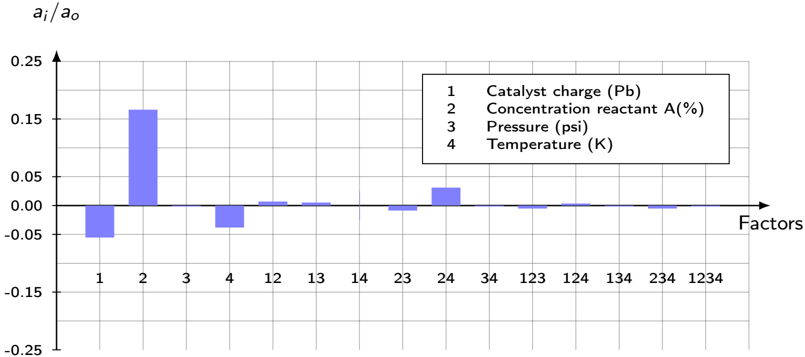


4.2.12 Effect inference

$$\hat{\alpha} = \frac{1}{16} X'Y$$

Coefficient	Estimator	Estimate
a_o	I	72.25
a_1	1	-4.00
a_2	2	12.00
a_3	3	-0.13
a_4	4	-2.75
a_{12}	12	0.50
a_{13}	13	0.37
a_{14}	14	0.00
a_{23}	23	-0.62
a_{24}	24	2.20
a_{34}	34	-0.13
a_{123}	123	-0.38
a_{124}	124	0.25
a_{134}	134	-0.13
a_{234}	234	-0.38
a_{1234}	1234	-0.13

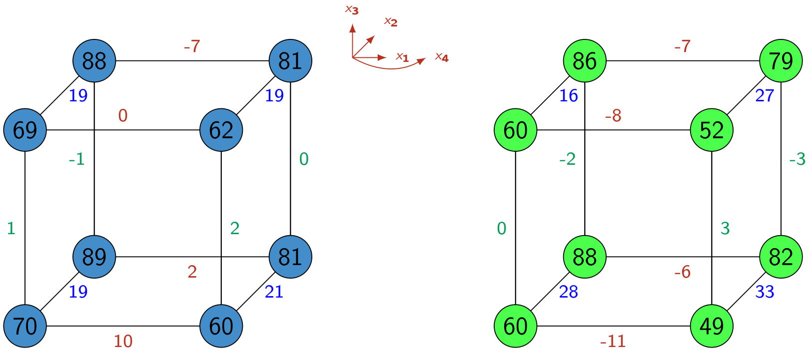
4.2.13 Relative half-effects



Matlab

```
alpha = X' * Y / Nexp
alpha_relative = alpha / alpha(1)
```

4.2.13 Visualization of the Effects



4.3.0 Methods for selecting significant effects

- Several methods exist
- Two are presented here :
  - The reference distribution
  - Normal plot
  - Replicated runs and analysis of variance

4.3.0 The reference distribution

- The method needs some replicates of at least one data point
- Let's consider g sets with N_i data in the set i
- The DF are $\nu_i = N_i - 1$ and the variances $s_i^2 = \frac{1}{\nu_i}\sum_j (y_{ij} - \mu_j)$
- The variance is

$$s_y^2 = \frac{\sum_i \nu_i \; s_i^2}{\sum_i \nu_i}$$

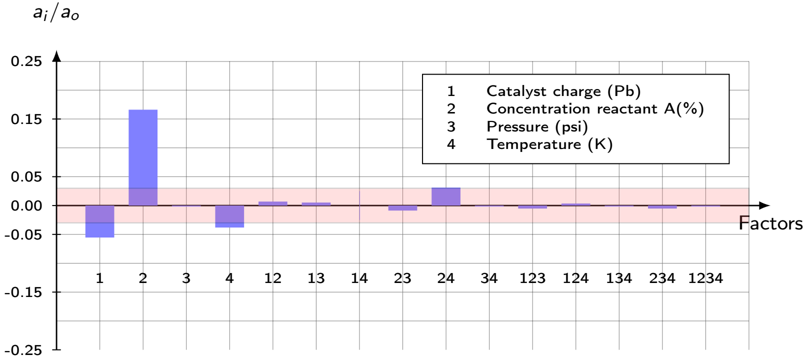
- If one experiment only has been duplicated :

$$s^2(y_1, y_2) = \frac{1}{2}(y_1 - y_2)^2 = \frac{\Delta^2}{2}$$

- with an estimate of s^2, it is possible to compute a confidence interval (see chap.2)

$$CI_{1-\alpha} = t_{\frac{\alpha}{2},\nu} \sqrt{\frac{s^2}{N}}$$

4.3.1 Discriminating half-effect with CI



Let's consider that run 1 has been replicated with results 69% and 71% $\rightarrow S^2 = 2$

$CI_{0.9} = t_{0.05,1}\sqrt{\frac{2}{17}} = \pm 2.2$ in relative range it gives 3%

4.3.2 Normal plot - rational (1)

- Each effect is a linear combination of random variables :

$$a_j = \sum_i x_{ij} Y_i$$

- If effect would result from an experimental noise only, they would follow a Normal distribution :

$$\text{if } y_i \sim N(\mu, \sigma^2) \quad \text{then } a_j \sim \ N\left(\sum x_{ij}\mu,\ \sum x_{ij}\sigma^2\right)$$

- Then effects that are significantly distinct of a Normal distribution have a large probability to be real.

4.3.2 Normal plot - rational (2)

- To check if the effects follow a Normal distribution, the cumulative distribution function of the effects

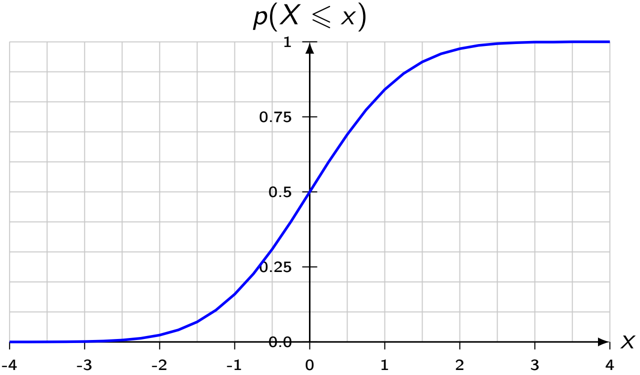
$$p(\alpha_i \leq x) = \int_{-\infty}^{x} p(x')dx'$$

is compared to the Normal cumulative distribution function (which is a sigmoid) :

$$p(X \leq x) = \Phi(\frac{x-\mu}{\sigma}) = \frac{1}{2}\left(1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})\right)$$

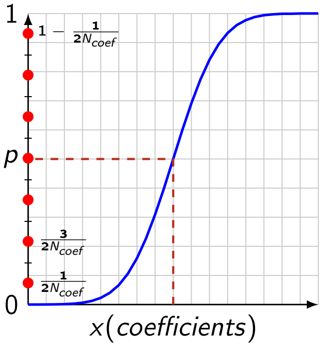
4.3.3 The cumulative Normal distribution

CFD of $N\,(\nu = 0, \sigma = 1)$



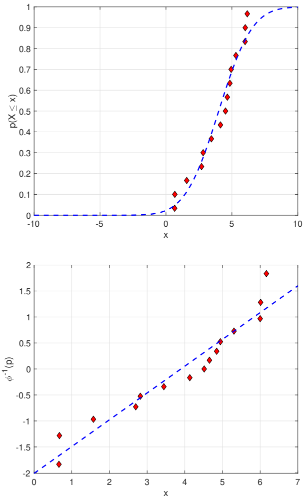
4.3.4 Normal plot - Reasoning(3)

- Each effect represents a fraction $\frac{1}{N_{coef}}$ of the considered population. N_{coef} is the number of coefficients without the constant
- The Y axis is divided in N_{coef} intervals
- The y-coordinate of the plot is placed at the middle of each interval
- The first y-coordinate is thus in $\frac{1}{2N_{coef}}$
- The following y-coordinates are at regular intervals $\frac{1}{N_{coef}}$



4.3.5 Normal plot - Reasoning(4)

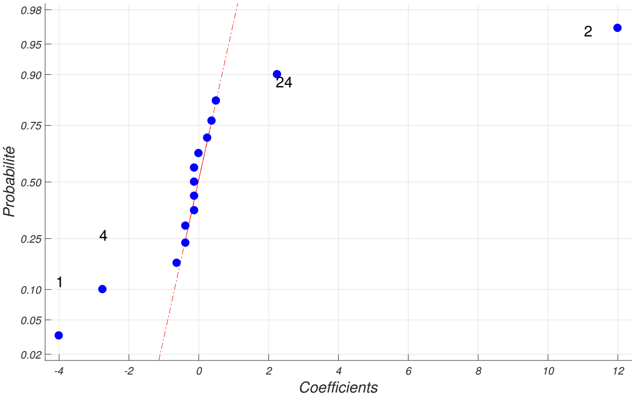
- Plotting $\{x = \alpha_{[i]}, y = p(i)\}$ produces a sigmoidal curve $f_p(\alpha)$
- $f_p(\alpha)$ is compared to $\Phi(x)$
- However it is difficult to visually compare curves !
- Then let's transform the sigmoid into a straight line $\Phi^{-1}(f_p(\alpha))$



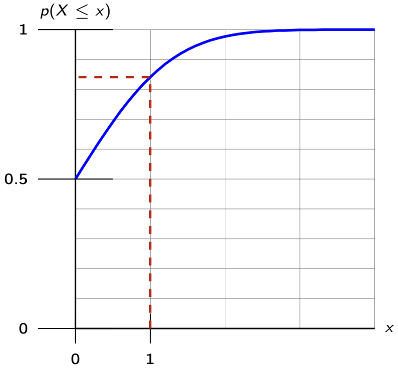
Matlab

normplot(x)
probplot(dist,x)

4.3.6 Normal plot for the case of the reactor



4.3.7 Function $\Phi(x)$ with $x > 0$

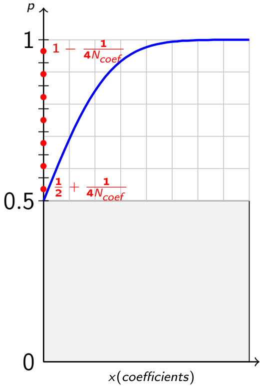


4.3.8 Half-normal plot

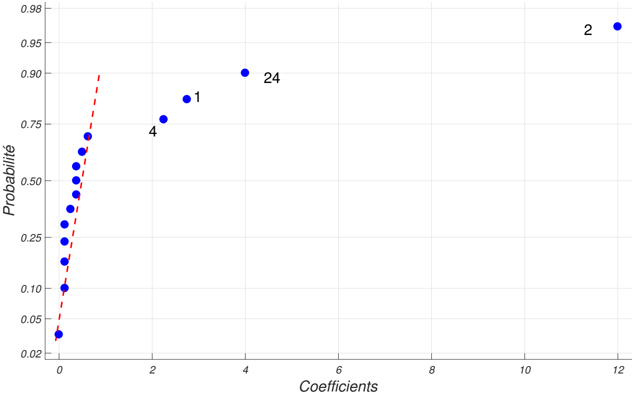
- The sign of the effects depends on the coding by -1 and 1
- Inverting the coding for the factors with negative effects, main effects are obtained
- Then for testing the normality of the effects the sign has no specific signification
- Then it is possible to work with half of the normal distribution and so increase the density of the sampling

Matlab

probplot('halfnormal',abs(x))



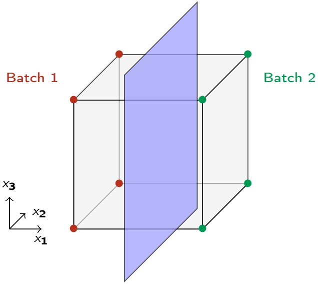
4.3.9 Half normal plot - reactor case



4.3.10 Blocking to mitigate expected source of error

- Sometimes it is not possible tontreatv all the experimental units the same manner :
  - If the process require a treatment that can not be applied in a single batch
  - the experiment are not done in the same time
  - If the measurement are shared between different operators or laboratories
- So what will be the impact of separating the experimental unit in , let's say, two batches ?

What would be the consequence of separating the measurement in two groups ?



4.3.11 Avoid Blocking aliased with a main effect

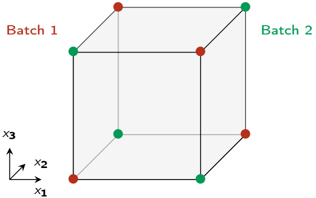
- The division into two batches according to the value of a factor would cause an alias of the main effect of this factor with a possible batch effect

$$y = a_o + (a_1 + a_B)\, x_1 + a_2 x_2 + \ldots$$

- This is precisely what must be avoid because it would degrade a part of the model that interests us in priority

$$E = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

4.3.12 Blocking aliased with an interaction



$$E = \begin{pmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{matrix} block1 \\ block2 \\ block2 \\ block1 \\ block2 \\ block1 \\ block1 \\ block2 \end{matrix}$$

4.3.13 Blocking vs randomisation

- Blocking is used to mitigate identified and predictable sources of error
- Randomization is used to mitigate random sources of error

4.3.14 Summary of full factorial design

- A 2^N full factorial design is used to explore the relationship between factors and their effect on a response variable. It involves testing all possible combinations of the factors at two levels each.
- The variance coefficients are $1/N$ which is the optimum.
- The number of experimental runs required is equal to 2^N, where N is the number of factors : It can also be resource-intensive.
- The data from a full factorial design is typically analyzed using ANOVA that is used to identify which factors have significant effects on the response variable.
- In case of genuine replicate normal plot or half-normal plot can be used to discriminate non significant effects.
- One limitation is that it assumes a linear relationship between the factors and the response variable.

4.4 Fractional factorial designs 2^{n-p}

4.4.1 Critics of the full factorial designs

- With the factorial designs in general and 2^n designs in particular, the number of runs rapidly becomes too large in comparison to the useful information that is gathered
- Let's take the example of the 2^4 design
  - 16 runs
  - 4 main effects
  - 6 first order interactions (2×2)
  - 4 second order interactions (3×3)
  - 1 third order interaction (4×4)
- So 5 coefficients for which the interest is low, not to mention that among the main effects and the first order interactions there can be non significant effects
- We have to find a better design !

4.4.2 Main Effects and First-Level Interactions

1. Efficiency in Experimentation

- **Resource Optimization :** Lower-order effects require fewer runs, saving time and resources while providing robust data.
- **Statistical Power :** Concentrating on main effects and two-way interactions increases precision in effect estimation by minimizing noise.
- **Noise Filtering :** Higher-order interactions often contribute minimal variation and may be indistinguishable from random noise.

2. Practical Relevance

- **Actionable Insights :** Main effects and two-way interactions are easier to interpret and apply in practical settings.
- **Feasibility in Application :** Simple interactions are more likely to be implementable in real-world engineering and design contexts.

4.4.3 Epistemology and Theory

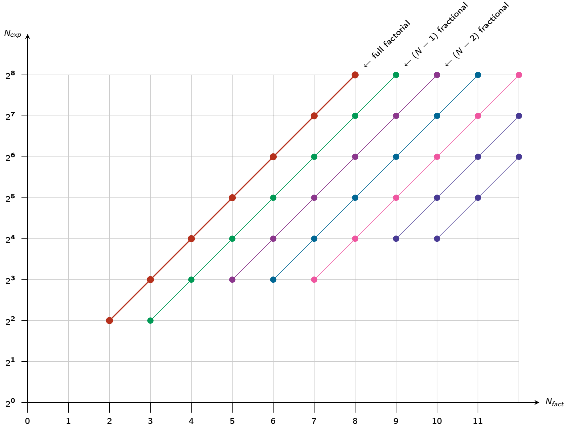
3. **Model Interpretability**
   - **Conceptual Clarity :** Lower-order interactions create more intuitive and accessible models.
   - **Ease of Communication :** Results are simpler to communicate to non-specialists, aiding decision-making.

4. **Scientific and Theoretical Alignment**
   - **Hierarchical Principle :** If higher-order effects exist, main and two-way interactions likely dominate, forming a stable model foundation.
   - **Mechanistic Understanding :** Main effects and two-way interactions reflect primary causal relationships, making theories based on them more meaningful.

5. **Theory Building and Simplified Models**
   - **Generalizability :** Simplified models based on main and two-way interactions are more likely to generalize across contexts.

4.4.4 Full factorial vs fractional

Fractional design allows to compensate the exponential expansion of the factorial design



4.4.5 Fractional factorial design $2^{(4-1)}$

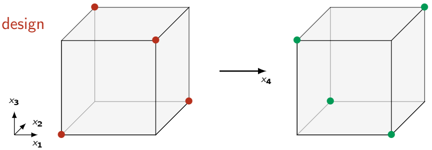
- Let's consider an experimental situation with 4 factors whose main effects and first order interactions has to be determined.
- In the situation with blocking(last chapter) the interaction with the highest order has been used to create batches,
- In the present example, it starts with 3 factors and then a fourth factor is introduced as previously the partition of the runs into two batches,
- Finally, it constitutes a design of 8 runs for 4 factors, but with aliases.
- It is named : 2^{4-1}

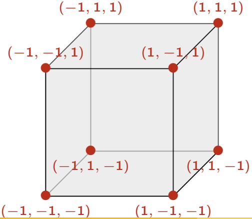
Fractional factorial plan 2^{4-1} obtained by adding a column **123** to a design 2^3

$$E = \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

4.4.6 Geometrical view of the $2^{(4-1)}$ design

Fractional 2^{4-1} factorial design





4.4.7 What can be done with a 2^{4-1} plan ?

- The model of interest has 11 coefficients

$$
\begin{aligned}
y = \; & a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + \\
& a_{12} x_1 x_2 + a_{13} x_1 x_3 + a_{14} x_1 x_4 + a_{23} x_2 x_3 + a_{24} x_2 x_4 + a_{34} x_3 x_4
\end{aligned}
$$

- They can not be all identified in only 8 runs
- What are the aliases ?
  - between main effects and interaction effects
  - between interaction effects

4.4.8 The group of columns of a factorial matrix

The set of 2^N columns of the model matrix of a factorial design and the multiplication operation term by term form a group of Galois :

- The result of the multiplication of two columns in a column,
- The set that contain a neutral element, the column I of the constant,
- Each column has an inverse, itself,
- The group is commutative.

4.4.9 The aliases of the 2^{4-1} design

- The fourth column of the essay matrix has been built with the product of the three first columns :

$$4 = 123$$

Which reveals the alias between a_4 and a_{123}

- We can multiply each side of the equation by 4 :

$$44 = I = 1234$$

So a_o is aliased with a_{1234},

- The relation $I = 1234$ is the *canonical form of the generator*,

- We can multiply the previous equation by 1 (or 2, or 3) :

$$1 = 11234 = 234$$

Each main effect is aliased by an interaction 3×3

- We can multiply each side of the equation by 2 (or 3, or 4)

$$12 = 34$$
$$13 = 24$$
$$14 = 23$$

Which reveals the aliases between the interactions 2×2.

4.4.10 The contrast table of the 2^{4-1} design

$$
\begin{aligned}
l_0 &= a_o + a_{1234} \\
l_1 &= a_1 + a_{234} \\
l_2 &= a_2 + a_{134} \\
l_3 &= a_3 + a_{124} \\
l_4 &= a_4 + a_{123} \\
l_5 &= a_{12} + a_{34} \\
l_6 &= a_{13} + a_{24} \\
l_7 &= a_{14} + a_{23}
\end{aligned}
$$

- With a 2^{4-1} design we can estimate height contrasts
- Neglecting 3×3 and 4×4 interactions, it brings estimates of the constant, the main effects and linear combinations of the 2×2 interactions.
- It can be sufficient in order to reveal the significant effects

- A *normalplot* allows to select the significant effects
- If all of the contrasts are significant, it is possible to *de-aliase* the system with a complementary plan and dispose of a complete 2^4 factorial plan.

Effects	#
Constant	$\frac{4!}{4!\ 0!} = 1$
Main	$\frac{4!}{3!\ 1!} = 4$
2×2	$\frac{4!}{2!\ 1!} = 6$
3×3	$\frac{4!}{1!\ 2!} = 4$
4×4	$\frac{4!}{0!\ 4!} = 1$

4.4.11 Analyse 5 factors with 8 experiments

- A fifth column of the essay matrix can be introduced in the model matrix, for example :

$$5 = 12$$

- As previously, we can take the canonical form of the generator :

$$I = 1234 = 125$$

- This new generator creates a list of additional aliases :

$$1 = 25$$
$$2 = 15$$

- But also with the higher levels of interaction, for example the 3×3 interactions

$$\begin{aligned} 135 &= 23 \\ 145 &= 24 \\ 235 &= 13 \\ 245 &= 14 \\ 345 &= 1234 \end{aligned}$$

4.4.12 The table of the constrats of the 2^{5-2} design

$$
\begin{aligned}
l_0 &= a_o + a_{125} + a_{345} + \dots \\
l_1 &= a_1 + a_{25} + a_{234} + \dots \\
l_2 &= a_2 + a_{15} + a_{134} + \dots \\
l_3 &= a_3 + a_{45} + a_{124} + \dots \\
l_4 &= a_4 + a_{35} + a_{123} + \dots \\
l_5 &= a_5 + a_{12} + a_{34} + \dots \\
l_6 &= a_{13} + a_{24} + a_{145} + a_{235} + \dots \\
l_7 &= a_{14} + a_{23} + a_{135} + a_{245} + \dots
\end{aligned}
$$

Effects	#
Constant	$\frac{5!}{5!\ 0!} = 1$
Main	$\frac{5!}{4!\ 1!} = 5$
2×2	$\frac{5!}{3!\ 2!} = 10$
3×3	$\frac{5!}{2!\ 3!} = 10$
4×4	$\frac{5!}{1!\ 4!} = 5$
5×5	$\frac{5!}{0!\ 5!} = 1$

- Usually the 3×3 interactions and higher are negligible

4.4.13 Alias vs contrast

Alias An alias is a linear combination, which clarifies the confusion between the unknowns (coefficients) of the under-determined system.

Contrast A contrast is a number, often denoted l_i in this course, resulting from the resolution of an under-determined linear system.

Alias table The alias table is the list of correspondence between contrasts and aliases.

4.4.13 Contrasts of the 2^{5-2} design

$$
\begin{aligned}
l_0 &= a_o + \dots \\
l_1 &= a_1 + a_{25} + \dots \\
l_2 &= a_2 + a_{15} + \dots \\
l_3 &= a_3 + a_{45} + \dots \\
l_4 &= a_4 + a_{35} + \dots \\
l_5 &= a_5 + a_{12} + a_{34} + \dots \\
l_6 &= a_{13} + a_{24} + \dots \\
l_7 &= a_{14} + a_{23} + \dots
\end{aligned}
$$

- Once simplified, the table shows the aliases between main effects and first order interactions.
- It can be sufficient to reveal the significant effects.
- A *normalplot* allows to select the significant effects.
- The two generators share the 2^5 plan in four
- If all of the effects are significant, we can *de-aliaze* the plan with one of the three fractions left and obtain a 2^{5-1} plan.
- The other fractions are generated by the following combinations :

$$
\begin{aligned}
I &= 1234 = -125 \\
I &= -1234 = 125 \\
I &= -1234 = -125
\end{aligned}
$$

4.4.14 Generation of a fractional factorial matrix

MATLAB

- *[E,conf] = fracfact(gen,'MaxInt',m)*

  E : runs matrix
  conf : table of alias
  gen : generator in the form of a chain of characters such as 'a b c abc'
  m : maximal level of the interactions to consider in the table of alias

- *gen = fracfactgen(terms)*
- *gen = fracfactgen(terms,k)*
- *gen = fracfactgen(terms,k,R)*

  terms : factors in the form of a chain of characters such as 'a b c abc'
  k : exponent defining the number of runs ($N_{exp} = 2^k$)
  R : resolution of the fractional plan (III, IV ou V, but in Arabic numbers)

4.4.15 The table of 2^{n-p} designs

	3	4	5	6	7	8	9	10	11
4	2^{3-1}_{III} ±3=12								
8	2^3	2^{4-1}_{IV} ±4=123	2^{5-2}_{III} ±4=12 ±5=13	2^{6-3}_{III} ±4=12 ±5=13 ±6=23	2^{7-4}_{III} ±4=12 ±5=13 ±6=23 ±7=123				
16	2^3 2 fois	2^4	2^{5-1}_{V} ±5=1234	2^{6-2}_{IV} ±5=123 ±6=234	2^{7-3}_{IV} ±5=123 ±6=234 ±7=134	2^{8-4}_{IV} ±5=234 ±6=134 ±7=123 ±8=124	2^{9-5}_{III} ±5=123 ±6=234 ±7=134 ±8=124 ±9=1234	2^{10-6}_{III} ±5=123 ±6=234 ±7=134 ±8=124 ±9=1234 ±10=12	2^{11-7}_{III} ±5=123 ±6=234 ±7=134 ±8=124 ±9=1234 ±10=12 ±11=13
32	2^3 4 fois	2^4 2 fois	2^5	2^{6-1}_{VI} ±6=12345	2^{7-2}_{IV} ±6=1234 ±7=1245	2^{8-3}_{IV} ±6=123 ±7=124 ±8=2345	2^{9-4}_{IV} ±6=2345 ±7=1345 ±8=1245 ±9=1235	2^{10-5}_{IV} ±6=1234 ±7=1235 ±8=1245 ±9=1345 ±10=2345	2^{11-6}_{IV} ±6=123 ±7=234 ±8=345 ±9=134 ±10=145 ±11=245
64	2^3 8 fois	2^4 4 fois	2^5 2 fois	2^6	2^{7-1}_{VII} ±7=123456	2^{8-2}_{V} ±7=1234 ±8=1256	2^{9-3}_{IV} ±7=1234 ±8=1356 ±9=3456	2^{10-4}_{IV} ±7=2346 ±8=1346 ±9=1245 ±10=1235	2^{11-5}_{IV} ±7=345 ±8=1234 ±9=126 ±10=2456 ±11=1456
128	2^3 16 fois	2^4 8 fois	2^5 4 fois	2^6 2 fois	2^7	2^{8-1}_{VIII} ±8=1234567	2^{9-2}_{VI} ±8=13467 ±9=23567	2^{10-3}_{V} ±8=1237 ±9=2345 ±10=1346	2^{11-4}_{V} ±8=1237 ±9=2345 ±10=1346 ±11=1234567

4.4.16 The resolution of a design

R=III No alias between the main effects a_i,
Aliases of the main effects a_i with first order interactions a_{ij}.

R=IV No alias between the main effects a_i,
No alias between the main effects a_i and first order interactions a_{ij},
Aliases between the first order interactions a_{ij}.

R=V No alias between the main effects a_i,
No alias between the main effects a_i and first order interactions a_{ij},
No alias between the first order interactions a_{ij}.
Aliases of the first order interactions a_{ij} with second order interactions a_{ijk}.

4.4.17 Chemical reactor with 5 factors

Chemical reactor

Objective

Identification of the dominant factors and interactions managing yield of the chemical reaction in an industrial reactor

Type of investigation

Experiment (factors can be manipulated)

Factors

Continuous

1. Feeding $in[10l/min, 15l/min$
2. Catalyst charge $in[5lb, 7.5lb]$
3. Mixing $in[100rpm, 120rpm]$
4. Temperature $in[220^\circ C, 240^\circ C]$
5. Concentration of reactant A $in[10\%, 12\%i]$

Response(s)

Conversion ratio (%)

Model(s)

1. Linear model $Y = a_o + \sum_i a_i \, x_1$
2. Linear model with interactions $Y = a_o + \sum_i a_i \, x_1 + \sum_{i<j} a_{ij} \, x_i \, x_j$

Strategies

1. Full factorial design : $2^5 = 32$ runs
2. Fractional factorial design : $2^{5-1} = 16$ runs, R=V
3. Fractional factorial design : $2^{5-2} = 8$ runs, R=III

4.4.18 Runs and aliases

- We choose the 2_V^{5-1} plan
- We choose as generator 5=1234
- The plan is
- The plan is of resolution V and allows thus to evaluate the 5 main effects and the 10 first order interactions

Matlab

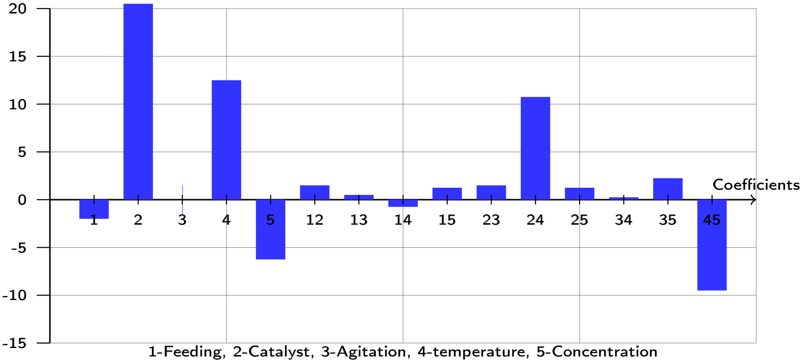
```
>> [X,conf]=fracfact('bcde b c d e','MaxInt',2)

X =

     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    -1
    -1     1     1    -1     1
    -1     1     1     1    -1
     1     1     1     1     1
```

```
conf =

  16×3 cell array

    'Term'       'Generator'    'Confounding'
    'X1'         'bcde'         'X1'
    'X2'         'b'            'X2'
    'X3'         'c'            'X3'
    'X4'         'd'            'X4'
    'X5'         'e'            'X5'
    'X1*X2'      'cde'          'X1*X2'
    'X1*X3'      'bde'          'X1*X3'
    'X1*X4'      'bce'          'X1*X4'
    'X1*X5'      'bcd'          'X1*X5'
    'X2*X3'      'bc'           'X2*X3'
    'X2*X4'      'bd'           'X2*X4'
    'X2*X5'      'be'           'X2*X5'
    'X3*X4'      'cd'           'X3*X4'
    'X3*X5'      'ce'           'X3*X5'
    'X4*X5'      'de'           'X4*X5'
```

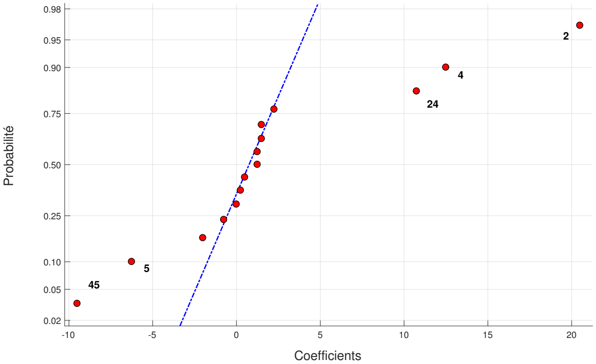
4.4.19 Experimental Results

	Feeding	Catalyst	Agitation	Temperature	Concentration	Rate
	l/min	kg	rpm	°C	%	%
−	10	5.0	100	220	10	
+	15	7.5	120	240	12	
1	1	-1	-1	-1	-1	53
2	-1	1	-1	-1	-1	63
3	-1	-1	1	-1	-1	53
4	1	1	1	-1	-1	61
5	-1	-1	-1	1	-1	69
6	1	1	-1	1	-1	93
7	1	-1	1	1	-1	60
8	-1	1	1	1	-1	95
9	-1	-1	-1	-1	1	56
10	1	1	-1	-1	1	65
11	1	-1	1	-1	1	55
12	-1	1	1	-1	1	67
13	1	-1	-1	1	1	45
14	-1	1	-1	1	1	78
15	-1	-1	1	1	1	49
16	1	1	1	1	1	82

4.4.20 Standardized effects

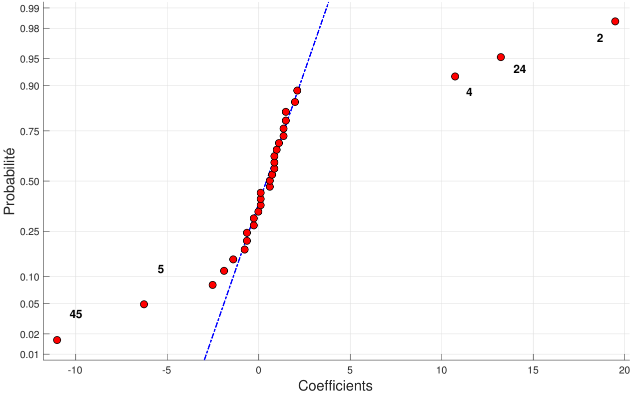


4.4.21 Normal plot of the 2^{5-1} design



1	Feeding
2	Catalyst
3	Agitation
4	Temperature
5	Concentration

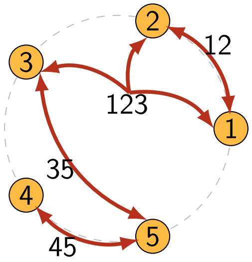
4.4.22 Normalplot of the 2^5



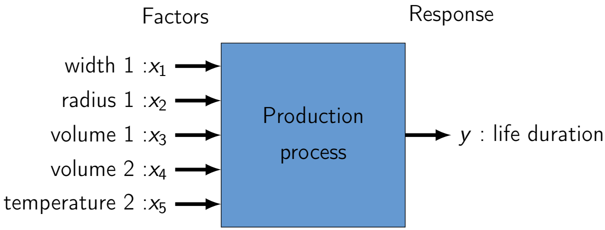
1 Feeding
2 Catalyst
3 Agitation
4 Temperature
5 Concentration

4.4.22 A smart Pareto analysis

How DOE helps to determine interactions



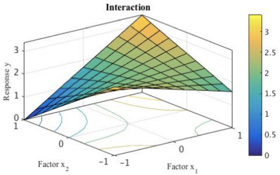
4.4.22 Case presentation



- During a technology validation phase in laboratory, the process of production is analysed to determine the dominant factors (and their potential interactions) on the life duration of a medical device.
- There are 5 factors, x_i
- There is one answer, y, the life duration of a device

4.4.22 Empirical model

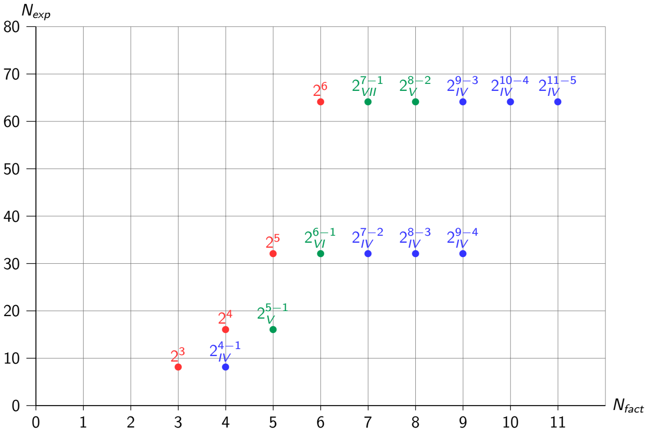
$$y(\vec{x}) = a_o + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 + \epsilon$$



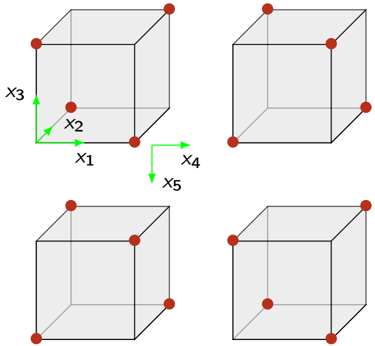
- In this case, the target model has 16 coefficients

$$y = a_o + \sum_{i=1}^{5} a_i x_i + \sum_{i>j}^{5} a_{ij} x_i x_j$$

4.4.22Fractional factorial designs

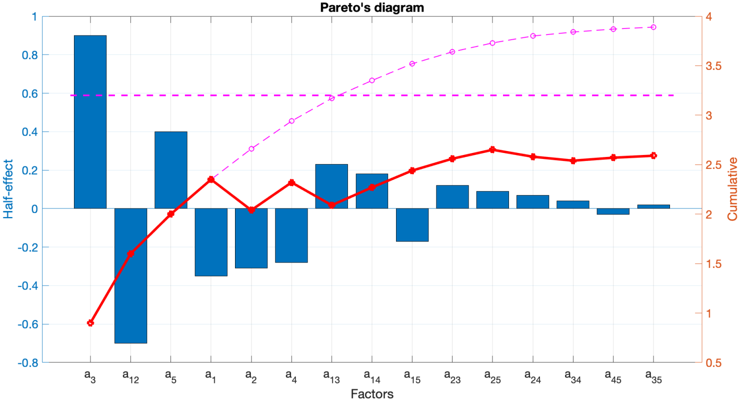


4.4.22 The fractional factorial design 2_V^{5-1}



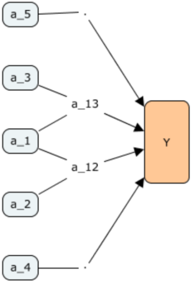
- To identify the 16 coefficients of the model, we can use the 2_V^{5-1} fractional factorial design
- It counts 16 runs
- The variance of the coefficients is $var(\beta) = \frac{\sigma^2}{16}$

4.4.22 Pareto Diagram

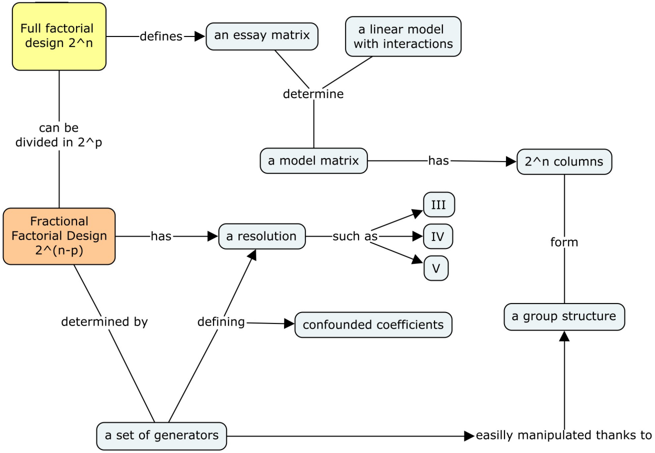


4.4.22Conclusion for this case

- Interactions are key information for managing a process
- In this case, the use of a DOE technique brings efficiency in term of determining the interactions and permitting a smart Pareto analysis.
- The choice of the adequate design let's us minimizing the number of experiments
- DOE expertise allows the experimenters to determine before the experiments which coefficients are calculable and also the size of their confidence interval in function of the precision of the measurement technique



4.4.23 Summary fractional factorial design



4.4.24 Summary fractional factorial design

- **Definition** : A fractional factorial design is a statistical technique used to study the effect of a subset of factors on a response variable. Unlike full factorial designs, fractional factorial designs do not test all possible combinations of the factors, but instead test a carefully chosen subset of them.
- **Design** : In a fractional factorial design, the number of experimental runs required is a fraction of the total number of runs required for a full factorial design. The design is carefully constructed to minimize the confounding of higher-order interactions between the independent variables.
- **Analysis** : The data from a fractional factorial design is typically analyzed using an ANOVA to determine the main effects of each factors and their interactions.
- **Advantages** : Fractional factorial designs are more efficient than full factorial designs, as they require fewer experimental runs while controlling for confounding between higher-order interactions.
- **Limitations** : fractional factorial designs may not be able to identify higher-order interactions between the independent variables, which may be important in some applications. Finally, the choice of the fraction used in the design may impact the accuracy of the estimates of the effects of the independent variables.

4.4.25 Learning outcomes of lesson on fractional factorial design

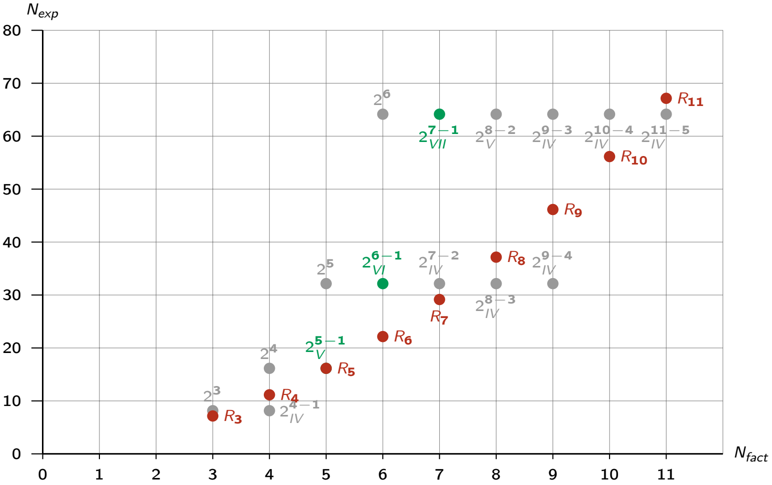
- At the end of this lesson, you must be able to select, use and analyze a fractional factorial design → table by Box
- to choose a generator
- to determine the aliases
- to build the essay matrix
- to perform the analysis of the results → Normalplot

4.5 Rechtschaffner's designs

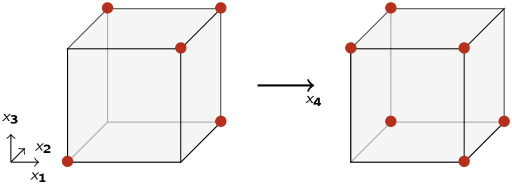
4.5.1 Rechtschaffner's designs

- Specialized design for estimating main effects and first-order interactions,
- Three types of lines (for n factors) :
  1. a line for the constant
     $[++\ldots+]$ ou $[--\ldots-]$
  2. n lines for the main effects
     $[-+\ldots+]$ ou $[+-\ldots-]$
  3. $\frac{n!}{2(n-2)!}$ lines for the interactions of first order
     $[--+\ldots+]$ ou $[++-\ldots-]$
- The most interesting designs are R_4,R_6, R_7 et R_8 their performances pass the factorial designs 2^4,2^{6-1}_{VI}, 2^{7-2}_{IV}, 2^{8-3}_{IV},
- But these designs are not orthogonal

4.5.2 Rechtschaffner VS factorial



4.5.3 Rechtschaffner design for 4 factors R_4



Essay matrix

$$E = \begin{pmatrix} -1 & -1 & -1 & -1 \\ \hline -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ \hline -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

4.5.4 Dispersion matrix (Rechtschaffner-4)

$$D = \frac{1}{144}\left(\begin{array}{c|cccc|cccccc}
14 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\hline
-1 & 20 & 2 & 2 & 2 & -2 & -2 & -2 & 7 & 7 & 7 \\
-1 & 2 & 20 & 2 & 2 & -2 & 7 & 7 & -2 & -2 & 7 \\
-1 & 2 & 2 & 20 & 2 & 7 & -2 & 7 & -2 & 7 & -2 \\
-1 & 2 & 2 & 2 & 20 & 7 & 7 & -2 & 7 & -2 & -2 \\
\hline
1 & -2 & -2 & 7 & 7 & 20 & 2 & 2 & 2 & 2 & -7 \\
1 & -2 & 7 & -2 & 7 & 2 & 20 & 2 & 2 & -7 & 2 \\
1 & -2 & 7 & 7 & -2 & 2 & 2 & 20 & -7 & 2 & 2 \\
1 & 7 & -2 & -2 & 7 & 2 & 2 & -7 & 20 & 2 & 2 \\
1 & 7 & -2 & 7 & -2 & 2 & -7 & 2 & 2 & 20 & 2 \\
1 & 7 & 7 & -2 & -2 & -7 & 2 & 2 & 2 & 2 & 20
\end{array}\right)$$

- Transfer of the experimental variance : 10%(a_o) and 14% (a_i, a_{ij}) to be compared with 6.25% for a 2^4 design that counts 5 more experiments,
- VIF : 1.1(a_o) and 1.5 (a_i, a_{ij}) to be compared to 1 for a 2^4 design.

4.5.5 Angles between the estimators

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \phi \quad \Rightarrow \quad \cos \phi_{ij} = \frac{1}{N_{exp}} X^T X$$

$$\phi_{ij} = \{75^o, 85^o, 95^o, 105^o\}$$

Alias with higher order interactions coefficients (a_{ijk} and a_{ijkl})

$$A = \frac{1}{3} \left(\begin{array}{cccc|c} -1 & -1 & -1 & -1 & 1 \\ \hline -1 & -1 & -1 & 2 & -2 \\ -1 & -1 & 2 & -1 & -2 \\ -1 & 2 & -1 & -1 & -2 \\ 2 & -1 & -1 & -1 & -2 \\ \hline 1 & 1 & -2 & -2 & -1 \\ 1 & -2 & 1 & -2 & -1 \\ -2 & 1 & 1 & -2 & -1 \\ 1 & -2 & -2 & 1 & -1 \\ -2 & 1 & -2 & 1 & -1 \\ -2 & -2 & 1 & 1 & -1 \end{array} \right)$$

4.5.6 Rechtschaffner - 6 factors

- Transfer of the experimental error : about 5% compared to 3.1% for the 2^{6-1}_{IV} design which has 10 more experiments,
- VIF : $1.2(a_o)$ and $1.15\ (a_i, a_{ij})$.

$$E = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 & -1 & 1 \\ \hline -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 \end{pmatrix}$$

4.5.7 Orthogonality vs. Resolution

- **Orthogonality** : Ensures that estimates of main effects are completely independent of each other, but often requires a larger number of runs.
- **Resolution** : Determines the level of confounding between effects. Higher-resolution designs (e.g., Resolution V) allow clear estimation of main effects and selected interactions, but achieving this with orthogonality may demand more experimental runs.
- **Trade-off** : Rechtschafner's designs prioritize higher resolution at the expense of full orthogonality, balancing interpretability with resource constraints.

4.5.8 The "Right" Trade-Off

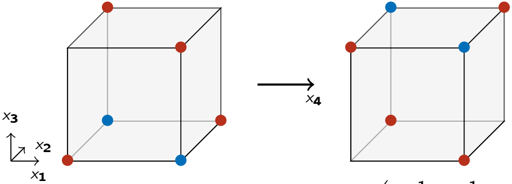
- **When to Prioritize Orthogonality :**
  - Main effects are the primary focus.
  - Interactions are negligible or assumed unimportant.
  - Ample resources to afford more experimental runs.

- **When to Prioritize Resolution :**
  - Interactions may significantly influence outcomes.
  - Limited resources (e.g., time, materials).
  - Partial confounding is acceptable, especially during screening.

4.5.9 Advantages of Rechtschafner's designs

- Ideal for **efficient screening** of factors in resource-constrained experiments.
- **Resolution V** separates main effects from two-factor interactions.
- Structured confounding allows for **interpretability** without requiring the extensive resources of fully orthogonal designs.

4.5.10 Design 3/4 for 4 factors



Essay matrix

$$E = \left(\begin{array}{ccc|c}
-1 & -1 & -1 & -1 \\
-1 & -1 & 1 & 1 \\
-1 & 1 & -1 & 1 \\
-1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 \\
1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 \\
\hline
-1 & 1 & -1 & -1 \\
-1 & 1 & 1 & 1 \\
1 & -1 & -1 & -1 \\
1 & -1 & 1 & 1
\end{array}\right)$$

4.5.11 Conclusions - Rechtschaffner's design

Feature	Three-Quarter Design	Rechtschafner Design
Levels per Factor	Two $(-1, +1)$	Two $(-1, +1)$
Number of Runs	$\frac{3}{4} \times 2^k$	$1 + k + \binom{k}{2}$
Resolution	Typically IV or higher	Typically V
Confounding	Reduced	Minimal
Orthogonality	Partial orthogonality	Structured confounding
Use Case	Sequential exp. or refinement	Screening experiments, min. runs

Table – Comparison of Three-Quarter and Rechtschafner Designs