Optimal design of experiments

Session 2: Standard designs

Peter Goos

Universiteit Antwerpen

Outline

- quantitative variables
  - first-order design
  - second-order design
- qualitative or categorical variables
  - completely randomized designs
  - block designs

Experimental design textbooks

- quantitative experimental variables
  - Box & Draper (1987), Myers & Montgomery (1995)

→ regression designs, response surface designs

- qualitative experimental variables
  - Montgomery (1991), Wu & Hamada (2000)

→ anova

- mixture variables
  - Cornell (2002)
- optimal design
  - Atkinson, Donev & Tobias (2007)
- orthogonal arrays
  - Hedayat, Sloane & Stufken (1999)
- classical block designs
  - Cochran & Cox (1957), Cox (1958)

Two-level factorial design

- each experimental variable has 2 levels
- each combination of levels is tested
- number of combinations?

$$\begin{array}{rl} 2 & \text{options for } x_1 \\ \times\, 2 & x_2 \\ \vdots & \\ \times\, 2 & x_m \\ \hline 2^m & \text{for } m \text{ variables} \end{array}$$

→ 2^m factorial design

2^3 factorial design

- three variables ($m = 3$)

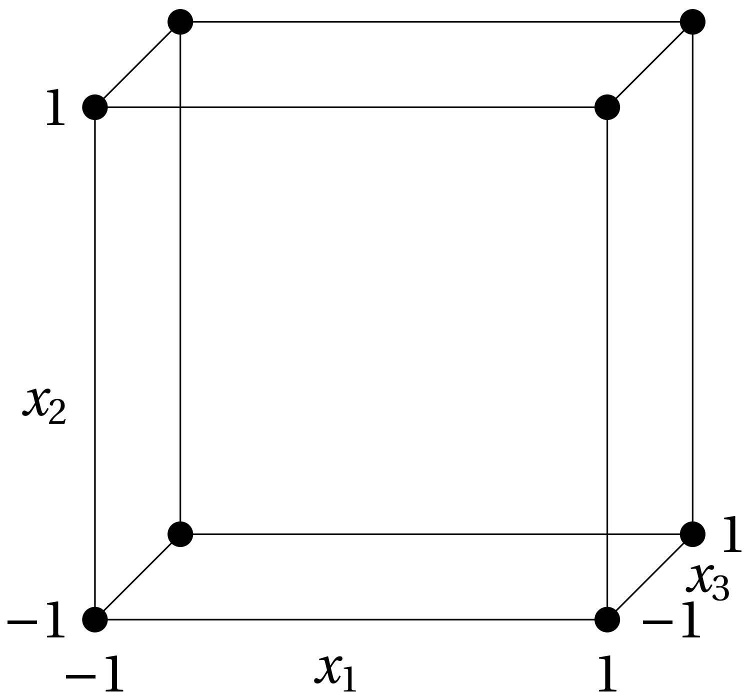
	x_1	x_2	x_3
1	−1	−1	−1
2	+1	−1	−1
3	−1	+1	−1
4	+1	+1	−1
5	−1	−1	+1
6	+1	−1	+1
7	−1	+1	+1
8	+1	+1	+1

- model?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 \\ + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

- n is power of 2

2^3 factorial design



Information matrix 2^3 factorial design

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix} = 8\mathbf{I}_8$$

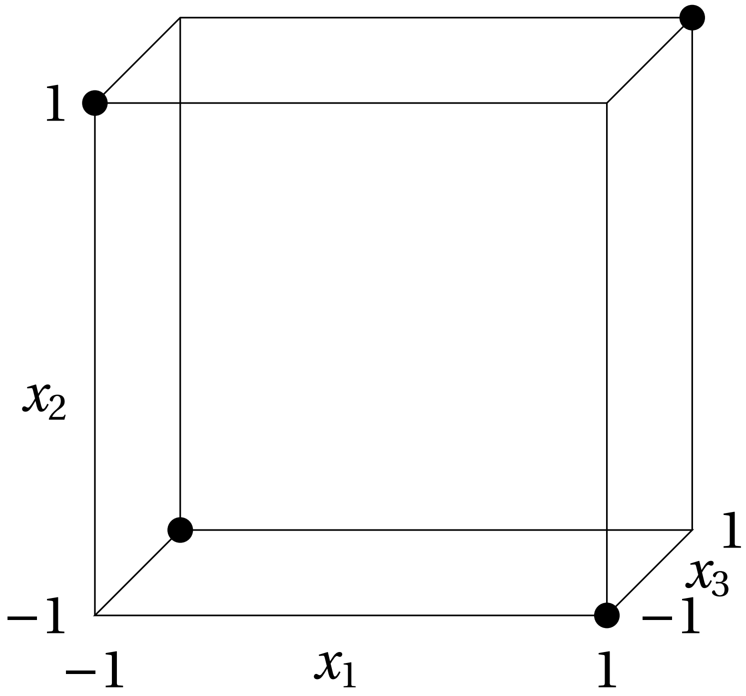
Other two-level designs

- two-level fractional factorial designs
  - 2^{m-f} fractional factorial design
    - not every combination is tested
    - screening purposes
    - full model cannot be estimated
  - n is a power of 2
- Plackett-Burman designs
  - only meant for estimating main-effects model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

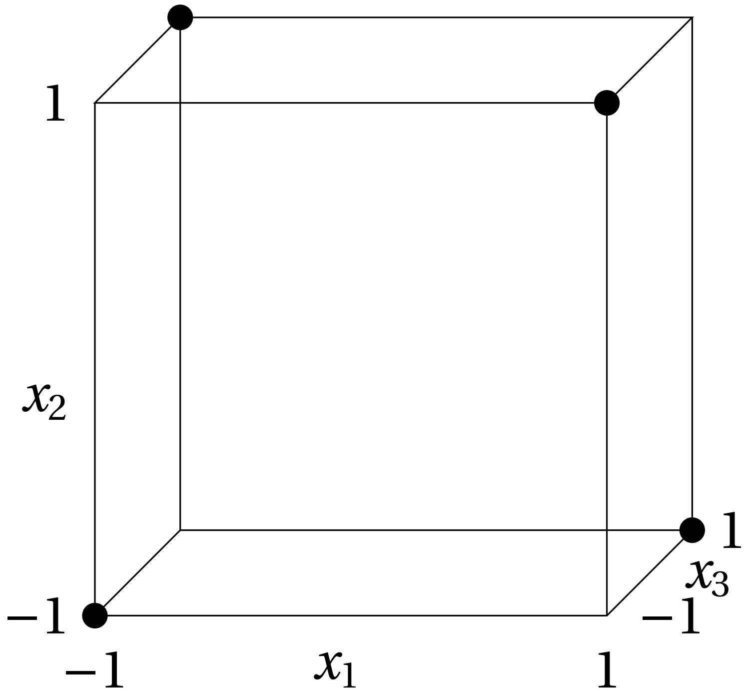
  - n is a multiple of 4
- Hadamard matrices

2^{3-1} factorial design



$$(+1 = x_1 x_2 x_3)$$

2^{3-1} factorial design



$$(-1 = x_1 x_2 x_3)$$

Plackett-Burman design

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
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

8-point Plackett-Burman design for $m = 7$

Hadamard matrices

$$H_1 = \begin{bmatrix} 1 \end{bmatrix}$$

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

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Two-level designs

- all these designs have excellent properties
- one of them is that

$$\text{var}(\hat{\beta}) = \sigma^2 \left(\mathbf{X}^T \mathbf{X}\right)^{-1}$$

  is a diagonal matrix (actually it is $\frac{\sigma^2}{n}\mathbf{I}$)
- all model parameters are estimated independently of each other
- all these designs can be used for quantitative and qualitative variables

Second-order designs

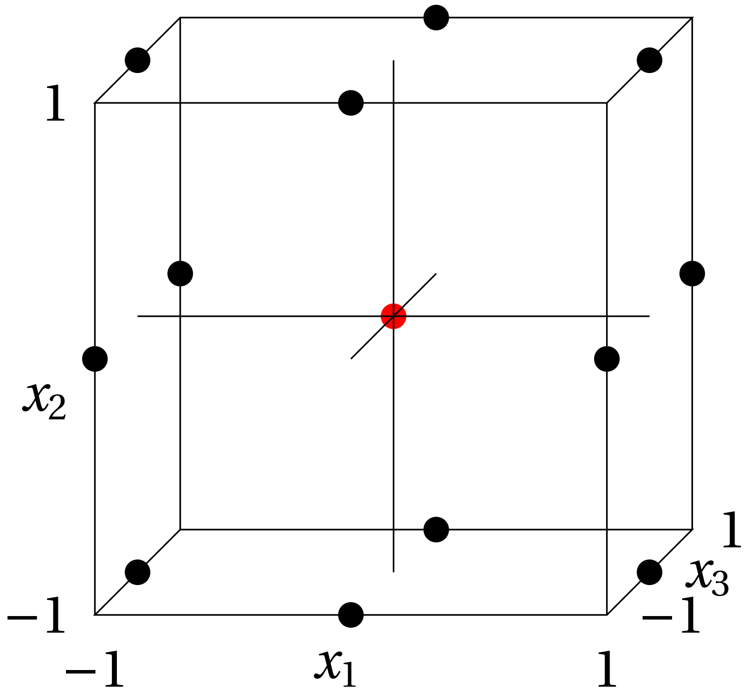
- designs for quantitative variables
  - Box-Behnken designs
  - central composite designs
- Box-Behnken designs (1960) combine

$$\begin{cases} \text{two-level factorial design} \\ \text{balanced incomplete block design} \end{cases}$$

- model for three variables

$$\begin{aligned} y = \beta_0 &+ \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &+ \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \\ &+ \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \epsilon \end{aligned}$$

Box-Behnken design



Box-Behnken design

run	x_1	x_2	x_3
1	−1	−1	0
2	−1	1	0
3	1	−1	0
4	1	1	0
5	0	−1	−1
6	0	−1	1
7	0	1	−1
8	0	1	1
9	−1	0	−1
10	1	0	−1
11	−1	0	1
12	1	0	1
13	0	0	0
14	0	0	0
15	0	0	0

Information matrix

$$\begin{bmatrix} 15 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 8 & 8 \\ 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 8 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 4 & 4 \\ 8 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 8 & 4 \\ 8 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 4 & 8 \end{bmatrix}$$

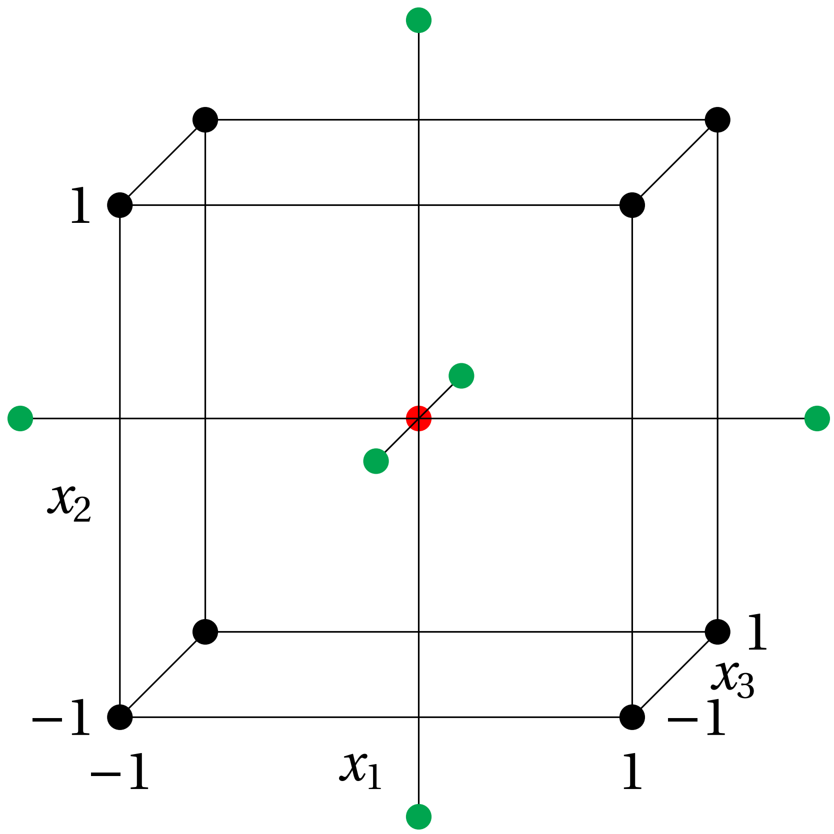
Central composite design

- Box & Wilson (1951)
- CCD
- 3 components:
  1. 2^m factorial (or 2^{m-f} fractional factorial) design
  2. $2m$ axial points (at distance α from center)
  3. at least one center point

  → $\alpha = 1$: face-centered CCD

  → $\alpha = \sqrt{m}$: spherical CCD

Central composite design ($\alpha > 1$)



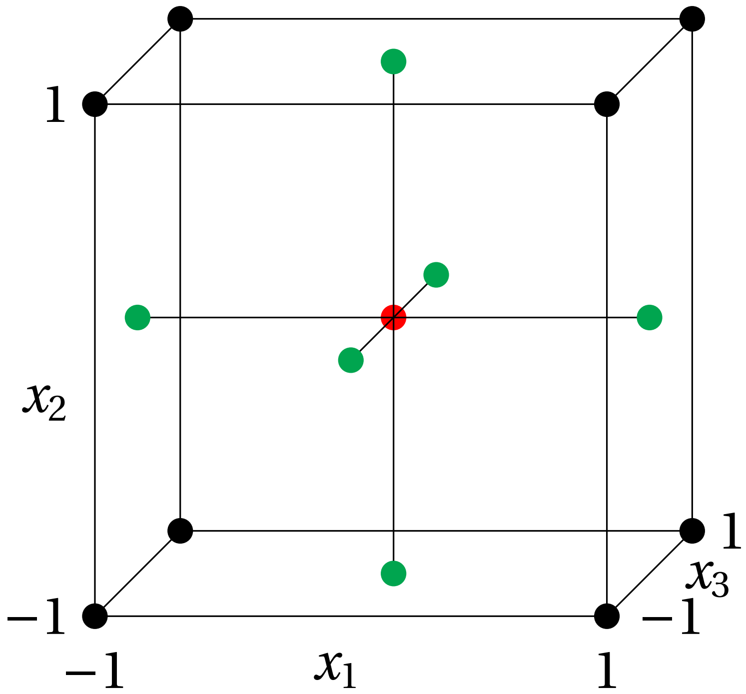
Central composite design

run	x_1	x_2	x_3
1	−1	−1	−1
2	−1	−1	1
3	−1	1	−1
4	−1	1	1
5	1	−1	−1
6	1	−1	1
7	1	1	−1
8	1	1	1
9	−1.68	0	0
10	1.68	0	0
11	0	−1.68	0
12	0	1.68	0
13	0	0	−1.68
14	0	0	1.68
15	0	0	0
16	0	0	0

Information matrix

$$
\begin{bmatrix}
16 & 0 & 0 & 0 & 0 & 0 & 0 & 13.66 & 13.66 & 13.66 \\
0 & 13.66 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 13.66 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 13.66 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\
13.66 & 0 & 0 & 0 & 0 & 0 & 0 & 24 & 8 & 8 \\
13.66 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 24 & 8 \\
13.66 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 8 & 24
\end{bmatrix}
$$

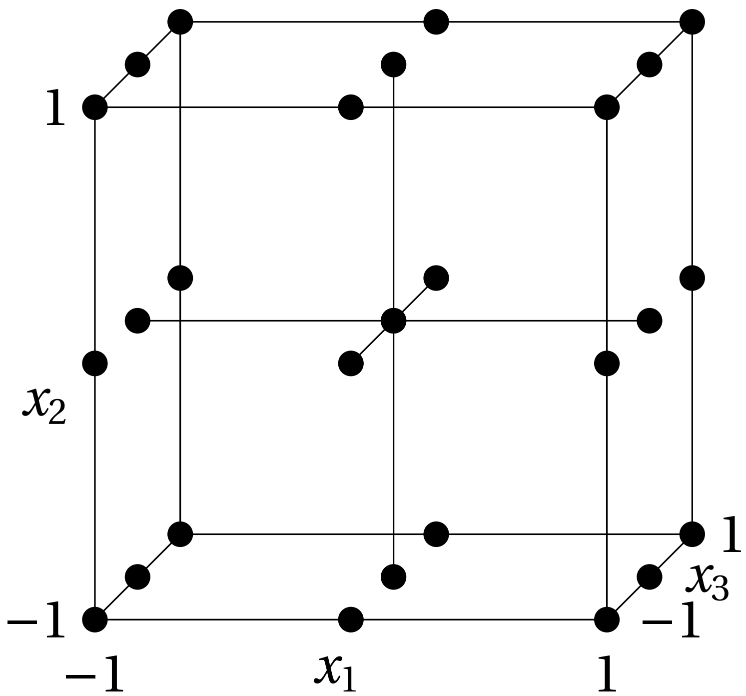
Face-centered central composite design



Three-level factorial design

- each experimental variable has 3 levels
- denoted by 3^m factorial design
- can be used for second-order models
- often used as candidate set for design construction algorithms
- 3^{m-f} factorial designs also exist

3^m factorial design



Other second-order designs

- many, many other second-order designs:
  - Hoke
  - Roquemore
  - Mee
  - small composite designs
  - ...
- all have at least three levels for each experimental variable to capture curvature

Note on design regions

- quantitative factors
- cuboidal design regions

$$\begin{cases} -1 \leq x_1 \leq +1 \\ \vdots \\ -1 \leq x_m \leq +1 \end{cases}$$

$$(x_1, x_2, \ldots, x_m) \in [-1, +1]^m$$

- spherical design regions

$$x_1^2 + x_2^2 + \cdots + x_m^2 \leq r^2$$

- design region is usually denoted by χ

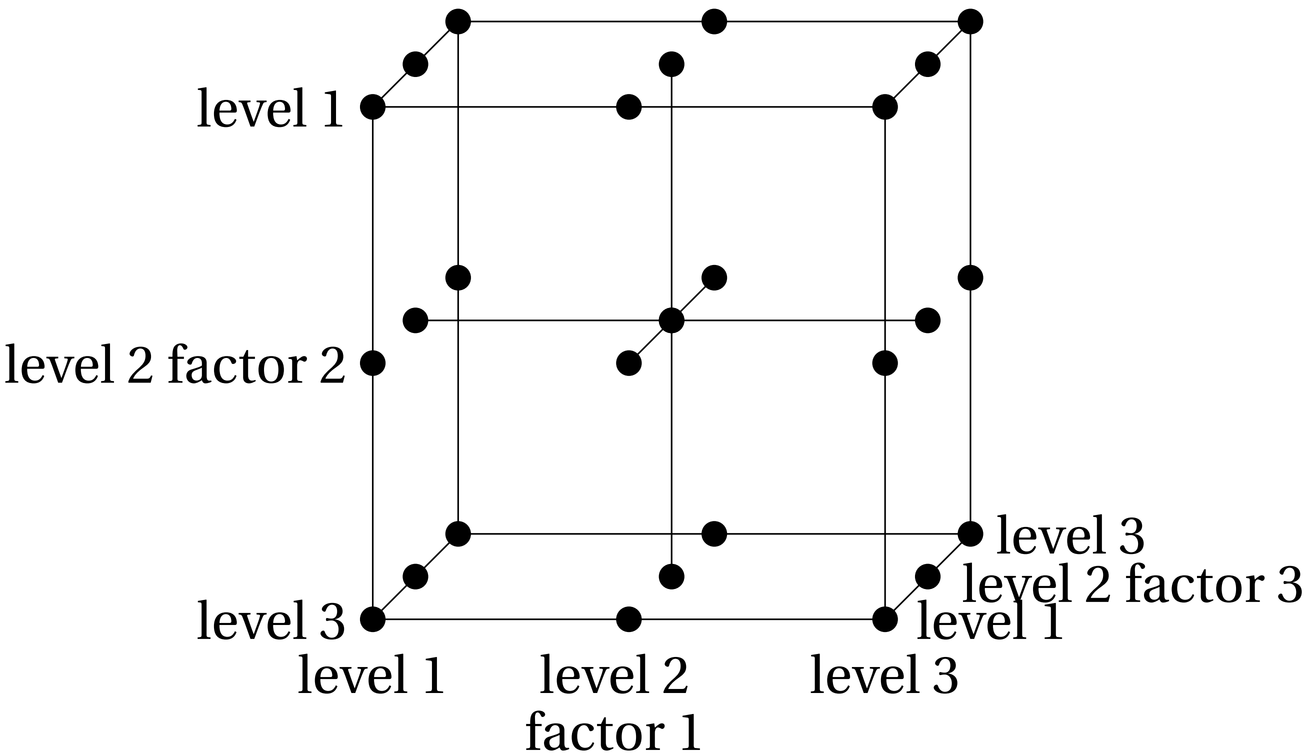
Categorical designs

- factorial designs with 2 or more levels
  - 3^m factorial design
  - $2 \times 3 \times 4$ factorial design
  - fractional factorial designs
  - orthogonal arrays
- model (3 variables)

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$$

- anova type of models

3^m factorial design



Orthogonal array

	Factor				
Run	A	B	C	D	E
1	0	0	0	0	0
2	0	1	1	1	1
3	1	0	1	0	1
4	1	1	0	1	0
5	2	0	0	1	1
6	2	1	1	0	0
7	3	0	1	1	0
8	3	1	0	0	1

Block designs

- useful when not all experimental tests can be done under homogeneous circumstances
- balanced incomplete block designs
  - BIBDs
  - 1 categorical experimental variable
    5 levels = treatments
    10 test persons = blocks
  - excellent properties
- partially balanced incomplete block designs
  - BIBDs often require large n
- latin square designs

Balanced incomplete block design

b	t		
1	1	2	3
2	1	2	4
3	1	2	5
4	1	3	4
5	1	3	5
6	1	4	5
7	2	3	4
8	2	3	5
9	2	4	5
10	3	4	5

5 treatments, 10 blocks of 3 treatments

Partially balanced incomplete block design

b	t		
1	1	2	3
2	1	2	4
3	1	5	6
4	2	5	6
5	3	4	5
6	3	4	6

6 treatments, 6 blocks of size 3

Latin square design

row	column 1	2	3	4	5	6	7
1	A	B	C	D	E	F	G
2	B	C	D	E	F	G	A
3	C	D	E	F	G	A	B
4	D	E	F	G	A	B	C
5	E	F	G	A	B	C	D
6	F	G	A	B	C	D	E
7	G	A	B	C	D	E	F

Sudoku design

row	column 1	2	3	4	5	6	7	8	9
1	3	1	8	7	2	4	9	6	5
2	4	2	7	6	9	5	8	1	3
3	9	6	5	1	8	3	4	7	2
4	1	8	4	9	3	6	2	5	7
5	2	9	3	5	7	1	6	8	4
6	5	7	6	2	4	8	3	9	1
7	7	5	9	4	6	2	1	3	8
8	6	3	2	8	1	7	5	4	9
9	8	4	1	3	5	9	7	2	6

Limitations of standard designs

- what if n is not a power of 2 or a multiple of 4?
- what if some of the level combinations of a design are forbidden?
- what if some of the blocks are smaller than others?
- what if you don't find an orthogonal array for your problem?