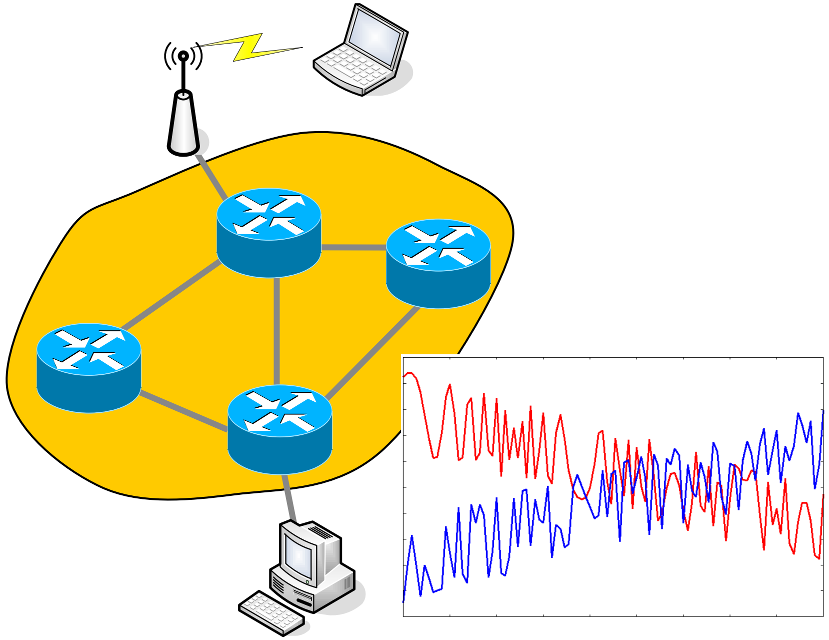
Chapter 13

Design of Experiments (DoE)



Contents

- Introduction to DoE
- Types of experimental designs
- 2^k Factorial design
- $2^k r$ Factorial design with replications
- 2^{k-p} Fractional factorial design

Introduction to DoE

Design of Experiments

- Example: Study the performance of a system in respect to particular parameters
  - System: routing algorithm for a MANET
  - Parameters:
    - Number of nodes: N = {10, 20, 50, 100, 1000, 10000}
    - Mobility: M = {1 m/s, 3 m/s, 5 m/s, 10 m/s}
    - Packet size: P = {64 byte, 256 byte, 512 byte, 1024 byte}
    - Number of parallel flows: F = {1, 3, 5, 7, 10}
  - Parameter space: N x M x P x F = 6 x 4 x 4 x 5 = 480

- Question: how to perform the experiments to understand the effects of the parameters?

Design of Experiments

- Answer: Design of Experiments (DoE)
  - The goal is to obtain

maximum information

with the

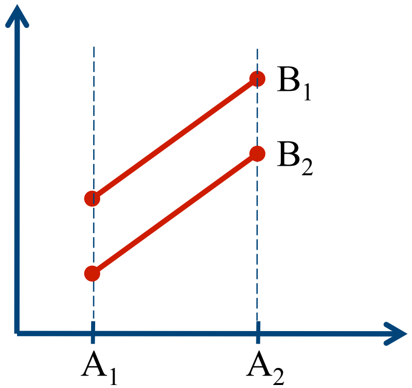
minimum number of experiments

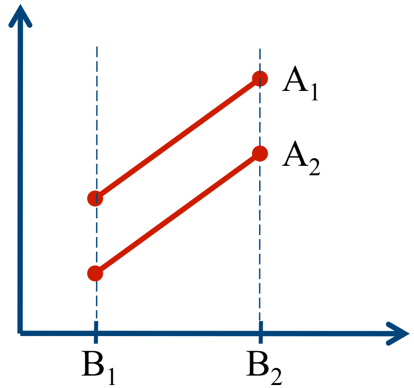
Terminology

Response variable:	The outcome of an experiment
Factor:	Each variable that affects the response variable and has several alternatives
Level:	The values that a factor can assume
Primary Factor:	The factors whose effects need to be quantified
Secondary Factor:	Factors that impact the performance but whose impact we are not interested in quantifying
Replication:	Repetition of all or some experiments
Experimental Unit:	Any entity that is used for the experiment
Interaction:	Two factors A and B interact if the effect of one depends upon the level of the other

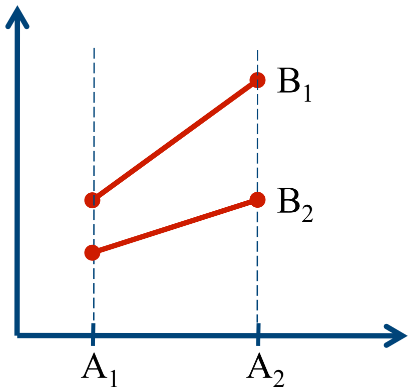
Interaction of factors

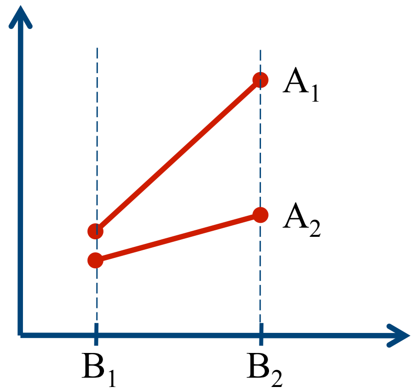
No Interaction





Interaction





Design

- Design: An experimental design consists of specifying the number of experiments, the factor level combinations for each experiment, and the number of replications.

- In planning an experiment, you have to decide
  1. what measurement to make (the response)
  2. what conditions to study
  3. what experimental material to use (the units)

- Example
  1. Measure goodput and overhead of a routing protocol
  2. Network with n nodes in chain
  3. Routing protocol, type of nodes, type of links, traffic

Types of experimental designs

Types of experimental designs: Simple design

- Simple design
  - Start with a configuration and **vary one factor** at a time
  - Given k factors and the i-th factor having n_i levels
  - The required number of experiments

$$n = 1 + \sum_{i=1}^{k} (n_i - 1)$$

  - Example:
    - $k=3$, $\{n_1=3, n_2=4, n_3=2\}$
    - $n = 1 + (2 + 3 + 1) = 7$

Types of experimental designs: Full factorial design

- Full factorial design
  - Use all possible combinations at all levels of all factors
  - Given k factors and the i-th factor having n_i levels
  - The required number of experiments

$$n = \prod_{i=1}^{k} n_i$$

  - Example:
    - $k=3$, $\{n_1=3, n_2=4, n_3=2\}$
    - $n = 3 \times 4 \times 2 = 24$

Types of experimental designs
Fractional factorial design

- Fractional factorial design
  - When full factorial design results in a huge number of experiments, it may be not possible to run all
  - Use subsets of levels of factors and the possible combinations of these
  - Given k factors and the i-th factor having n_i levels, and selected subsets of levels $m_i \leq n_i$.
  - The required number of experiments

$$n = \prod_{i=1}^{k} m_i$$

  - Example:
    - k=3, {n_1=3, n_2=4, n_3=2}, but use {m_1=2, m_2=2, m_3=1}
    - $n = 2\times2\times1 = 4$

Types of experimental designs

- Comparison of the design types

Design Type	Factors	Number of experiments
Simple design	$k=3$, $\{n_1=3, n_2=4, n_3=2\}$	7
Full factorial design		24
Fractional factorial design	Use subset $\{m_1=2, m_2=2, m_3=1\}$	4

2^k Factorial Designs

2^k Factorial Designs

- A 2^k factorial design is used to determine the effect of k factors
  - Each factor has two levels
- Advantages
  - It is easy to analyze
  - Helps to identify important factors
    ➡reduce the number of factors
  - Often effect of a factor is unidirectional, i.e., performance increase or decrease
  - Begin by experimenting at the minimum and maximum level of a factor ➡ two levels

2^k Factorial Designs
Example for k=2

- Study impact of memory and cache on performance of a workstation
- Memory size, two levels
- Cache size, two levels

- Performance of workstation as regression model

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

Factor 1

Factor 2

		Memory Size	
		4 MB	16 MB
Cache Size	1	-1,-1	1,-1
	2	-1,1	1,1

$$x_A = \begin{cases} -1 & \text{if 4MB memory} \\ 1 & \text{if 16MB memory} \end{cases}$$

$$x_B = \begin{cases} -1 & \text{if 1kb cache} \\ 1 & \text{if 2kb cache} \end{cases}$$

2^k Factorial Designs
Example for $k=2$

- Regression model

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

- Substitute the results into the model

$$y_1 = q_0 - q_A - q_B + q_{AB}$$
$$y_2 = q_0 + q_A - q_B - q_{AB}$$
$$y_3 = q_0 - q_A + q_B - q_{AB}$$
$$y_4 = q_0 + q_A + q_B + q_{AB}$$

Experiment	A	B	y	AB
1	-1	-1	y_1	1
2	1	-1	y_2	-1
3	-1	1	y_3	-1
4	1	1	y_4	1

- Solve equantions for q_i

$$q_0 = \tfrac{1}{4}(y_1 + y_2 + y_3 + y_4)$$
$$q_A = \tfrac{1}{4}(-y_1 + y_2 - y_3 + y_4)$$
$$q_B = \tfrac{1}{4}(-y_1 - y_2 + y_3 + y_4)$$
$$q_{AB} = \tfrac{1}{4}(y_1 - y_2 - y_3 + y_4)$$

$$y = 40 + 20x_A + 10x_B + 5x_A x_B$$

2^k Factorial Designs
Example for k=2: Sign table method

- Sign table contains the effect of factors

I	A	B	AB	y
1	-1	-1	1	15
1	1	-1	-1	45
1	-1	1	-1	25
1	1	1	1	75
160	**80**	**40**	**20**	**Total**
40	**20**	**10**	**5**	**Total/4**

Result (last two rows)

2^k Factorial Designs
Example for $k=2$: Allocation of variation

- Determine the importance of a factor
  - Calculate the variance

$$s_y^2 = \frac{\sum_{i=1}^{2^2} (y_i - \bar{y})^2}{2^2 - 1}$$

  - Sum of squares total (SST): Total variation of y

$$y = SST = \sum_{i=1}^{2^2} (y_i - \bar{y})^2$$

  - For 2^2 design, the variation is given by

$$SST = \underbrace{2^2 q_A^2}_{SSA} + \underbrace{2^2 q_B^2}_{SSB} + \underbrace{2^2 q_{AB}^2}_{SSAB}$$

  - SSA: part explained by factor A
  - Fraction of variation explained by A: SSA/SST

2^k Factorial Designs
The General Case

- In the general case there are k factors, each factor has two levels
- A total of 2^k experiments are required
- Analysis produces 2^k effects (results)
  - k main effects
  - $\binom{k}{2}$ two-factor interactions
  - $\binom{k}{3}$ three-factor interactions
  - …

- Sign table method is used!

2^k Factorial Designs
The General Case

- Sign table, example for k=3

I	A_1	A_2	A_3	A_1A_2	A_1A_3	A_2A_3	$A_1A_2A_3$	y
+	-	-	-	+	+	+	-	y_1
+	+	-	-	-	-	+	+	y_2
+	-	+	-	-	+	-	+	y_3
+	+	+	-	+	+	-	-	y_4
+	-	-	+	+	+	-	+	y_5
+	+	-	+	-	-	-	-	y_6
+	-	+	+	-	-	+	-	y_7
+	+	+	+	+	+	+	+	y_8

2^k Factorial Designs
The General Case

- Sign table

I	A_1	A_2	A_3	…	A_1A_2	A_1A_3	…	$A_1A_2A_3$	…	y
1	-1									y_1
1	1									y_2
1	-1									y_3
…	…									…
SumI SumI/2^k										Total Total/2^k

$2^k r$ Factorial Design with Replications

$2^k r$ Factorial Design with Replications

- Problem with 2^k factorial design is that it does not provide the estimation of experimental errors, since no repetitions
- Solution: Repeat an experiment r times ➡ replication
  - If each of the 2^k experiments is repeated r times ➡ $2^k r$ factorial design with replications
- Extended model

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$$

Experimental error

$2^k r$ Factorial Design with Replications

- For analysis, the same method is used, except for y, the mean of the replications is used.

I	A	B	AB	y	$\bar{y}$
1	-1	-1	1	(15,18,12)	15
1	1	-1	-1	(45,48,51)	48
1	-1	1	-1	(25,28,19)	24
1	1	1	1	(75,75,81)	77
164 **41**	**86** **21.5**	**38** **9.5**	**20** **5**		**Total** **Total/4**

- Experimental error is given: $e_{ij} = y_{ij} - \bar{y}$
- Sum of squared errors (SSE) and the standard deviation of errors:

$$SSE = \sum_{i=1}^{2^2} \sum_{j=1}^{r} e_{ij}^2 \qquad s_e = \sqrt{\frac{SSE}{2^2(r-1)}}$$

2^{k-p} Fractional Factorial Design

2^{k-p} Fractional Factorial Design

- When the number of factors is large, a full factorial design requires a large number of experiments
- In that case fractional factorial design can be used
  - Requires fewer experiments, e.g., 2^{k-1} requires half of the experiments as a full factorial design

2^{k-p} Fractional Factorial Design

- Preparing the sign table
  - Choose k-p factors and prepare a complete sign table.
    ➨Sign table with 2^{k-p} rows and 2^{k-p} columns
  - The first column will be marked I and consists of all 1s
  - The next k-p columns will be marked with the k-p factors that were chosen
  - The remaining columns are simply products of these factors

2^{k-p} Fractional Factorial Design

- Sign table, example for $k=7, p=4 \Rightarrow 2^{7-4}=2^3$

k-p chosen factors (F_1, F_2, F_3); *products of chosen factors* (F_1F_2, F_1F_3, F_2F_3, $F_1F_2F_3$)

I	F_1	F_2	F_3	F_1F_2	F_1F_3	F_2F_3	$F_1F_2F_3$
+	-	-	-	+	+	+	-
+	+	-	-	-	-	+	+
+	-	+	-	-	+	-	+
+	+	+	-	+	+	-	-
+	-	-	+	+	+	-	+
+	+	-	+	-	-	-	-
+	-	+	+	-	-	+	-
+	+	+	+	+	+	+	+

2^{k-p} rows

2^{k-p} columns

2^{k-p} Fractional Factorial Design

- Confounding
  - with fractional factorial design some of the effects can not be determined
  - only combined effects of several factors can be computed
- A fractional factorial design is not unique

- Design resolution
  - The resolution of a design is measured by the **order of effects** that are confounded
  - The **order of effect** is the number of factors included in it
    I = ABC order of 3 ➡Resolution R_{III}
    I = ABCD order of 4 ➡Resolution R_{IV}
  - A design of higher resolution is considered a better design.

Summary

- Design of experiments provides a method for planned experiments
- Goal: Obtain maximum information with minimum experiments
- Basic techniques
  - Factorial design
  - Factorial design with replications
  - Fractional factorial design