Chaptar 2 : model validation

1 .Introduction [article]

2. Experimental errors

3. The Test of Significance

4. confidence intervals

5. analysis of variance , model validation

5.1. ANOVA table

5.2. Correlation coefficient [4]

6. Application : example

Design of experiments (DOI)

Master 2 GCH GEV GPH

Chaptar 3 : Fractional factorial design

1. Introduction
2. Conception of Fractional factorial design
3. Analysis of Fractional factorial design
4. Application : example
5. Other designs : placket; Burman and Tagguchi

Chaptar 4 : Response surface design

1. Introduction
2. Response surface meaning
3. Design study of second degree

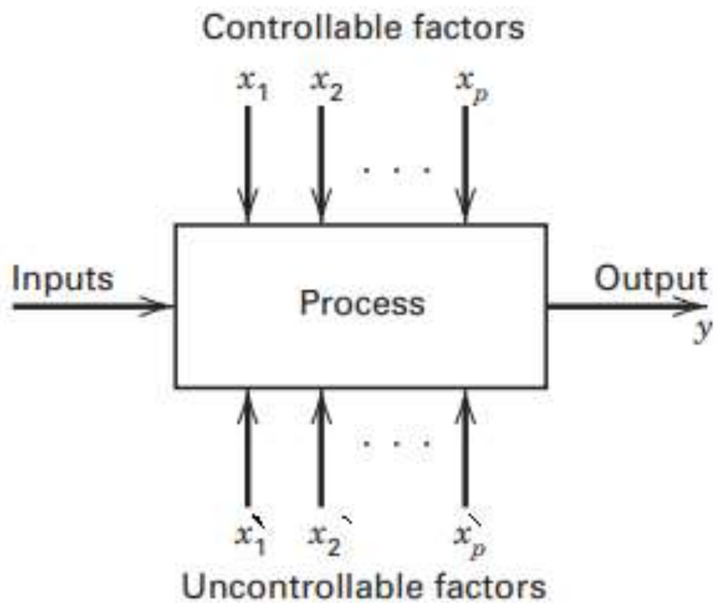
3.1. Box Behnken design

3.2. Composites design

4. Optimization with DOE
5. Application : example

INTRODUCTION
[1]

any process or system can be represented by the model shown in the Figure bellow :



Results = Out put = **response (Y)**

Chaptar1
Introduction

to understand what happens to a process when you change certain input factors, you have to change the factors. In other words, you need to conduct experiments on the system. Each **experimental run** is a test. More formally, we can define an experiment as a test or series of runs in which changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes that may be observed in the output **response**.

DOE

Experimental Design is a collection of experiments or runs that is planned in advance of the actual execution. The particular runs selected in an experimental design will depend upon the purpose of the design.

What is design of experience?

We may want to determine which input variables are responsible for the observed changes in the response, develop a model relating the response to the important input variables and to use this model for process or system improvement or other decision-making.

Response surface [Book 5]

Response surface methodology, or **RSM**, is a collection of mathematical and statistical techniques useful for the modeling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response**. For example**, suppose that a chemical engineer wishes to find **the levels of temperature (x1) and pressure (x2) that maximize the yield (y) of a process.** The process yield is a function of the levels of temperature and pressure, say

$$y = f(x_1, x_2) + \epsilon$$

where :

Ɛ : represents the noise or error observed in the response y.

If we denote the expected response by E(y) = f(x1 , x2) = η , then the surface represented by η = f(x1 , x2) is called a **response surface**.

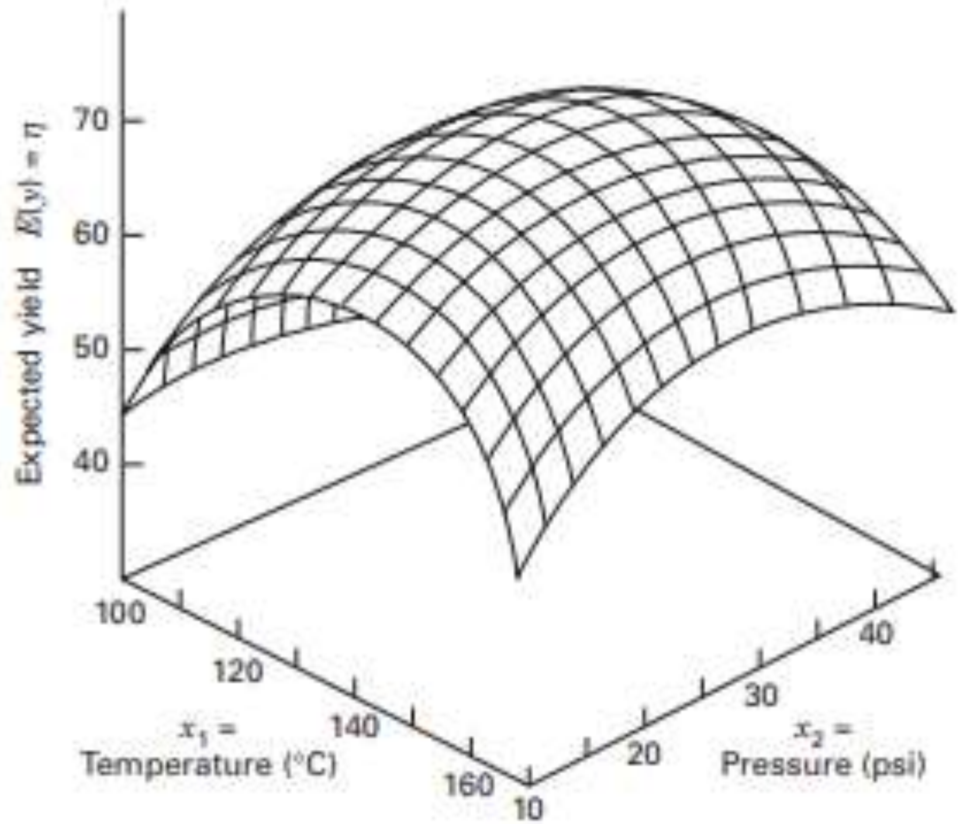


Figure : Representation of the response surface graphically (response surface plot)

Usually, a low-order **polynomial** is employed :

1. first-order model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

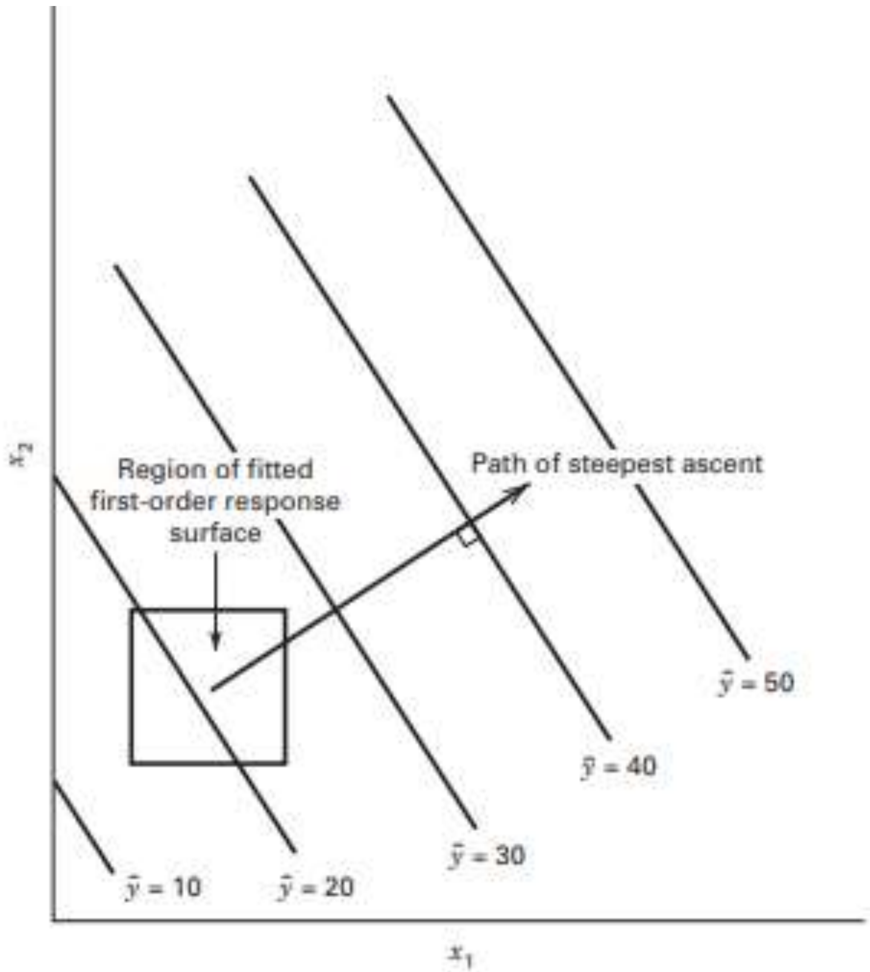
2. second-order model

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum\sum_{i<j} \beta_{ij} x_i x_j + \epsilon$$

Designs for fitting response surfaces are called **response surface designs**

The Method of Steepest Ascent : is a procedure for moving in the direction of the maximum increase in the response. Of course, **if minimization is desired**, then we call this technique **the method of steepest descent**. The fitted first-order model is :

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i$$



x_2
Region of fitted
first-order response
surface
Path of steepest ascent
$\hat{y} = 50$
$\hat{y} = 40$
$\hat{y} = 30$
$\hat{y} = 10$
$\hat{y} = 20$
x_1

Factors[7]

Independent Variable (Factor or Treatment Factor) is one of the variables under study that is being controlled at or near some target value, or **level**, during any given experiment. The level is being changed in some systematic way from run to run in order to determine what effect it has on the **response**(s). We have two kind of factors:

A.real factors

B.coded factors

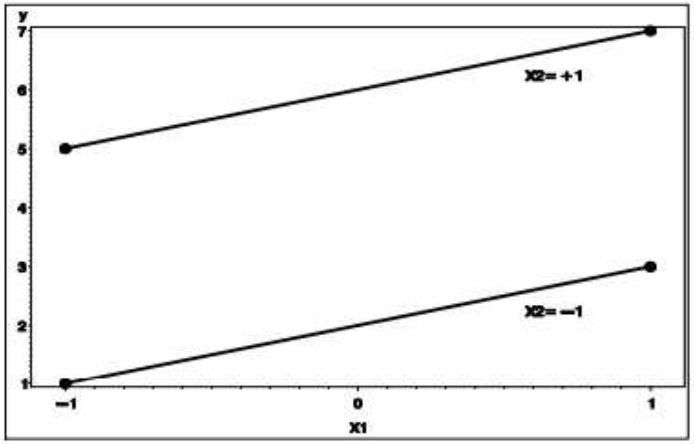
Coding the Data

$$Z1 = \frac{X1 - (X\ low + X\ high)/2}{(X\ high - X\ low)/2}$$

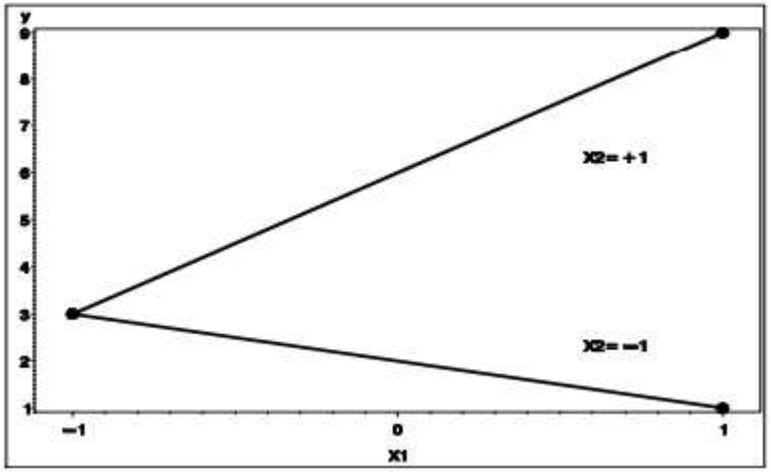
Example: the factor of temperature takes two levels 50 °C and 100°C. The coded factors of this two levels are:

$$\frac{50-75}{50/2} = -1 \quad \text{et} \quad \frac{100-75}{50/2} = +1$$

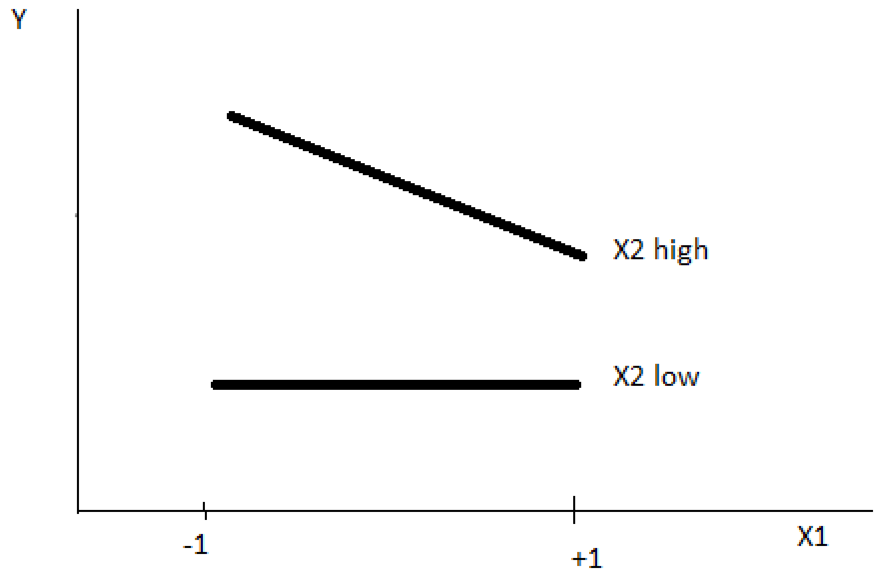
Interaction : it is joint effect between two factors, the effect of one factor upon the response will differ depending on the level of the other factor.



No interaction between X1 and X2



Fort interaction between X1 and X2



Weak interaction between X1 and X2

Simple linear regression [8]

DOE application in the future can be predicted with linear regression model. A general multiple linear regression model with k repressor variables is express as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

where, βj , βj , j = 0. 1, 2, …, k, are regression coefficients

Factorial design 2 power K [1]

is that of k factors, each at only two levels. These levels may be quantitative, such as two values of temperature, pressure, or time; or they may be qualitative, such as two machines, two operators, the “high” and “low” levels of a factor, or perhaps the presence and absence of a factor

Determination of effects

is the change in the response that is caused by a change in a factor or independent variable. the effect can be estimated by calculating it from the observed response data.

example

Suppose that we are interested in improving the yield of a chemical process. We know from the results of a characterization experiment that the two most important process variables that influence the yield are operating temperature and reaction time. The process currently runs at 145°F and 2.1 hours of reaction time, producing yields of around 80 percent. the contour lines for yields of 60, 70, 80, 90, and 95 percent.

reaction time [1,5 -2,5]

Temperature [140 -150]

Matrix design with real factors

time	temperature	yield
1,5	140	60
2,5	140	70
1,5	150	90
2,5	150	95

Matrix design with coded factors

time	temperature	yield
-1	-1	60
+1	-1	70
-1	+1	90
+1	+1	95

E = average Effect

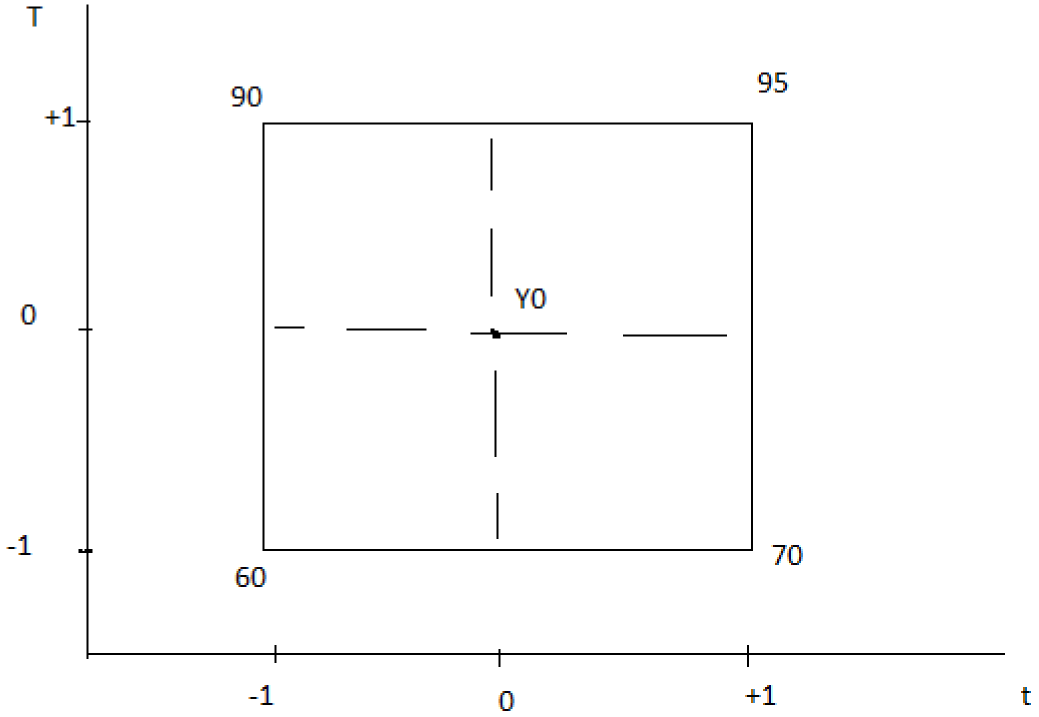
E T = -60-70+90+95/4 = 13,75

Et = -60+70-90+95/4 = 3,75

global effect

Gobal effect = average effect * 2

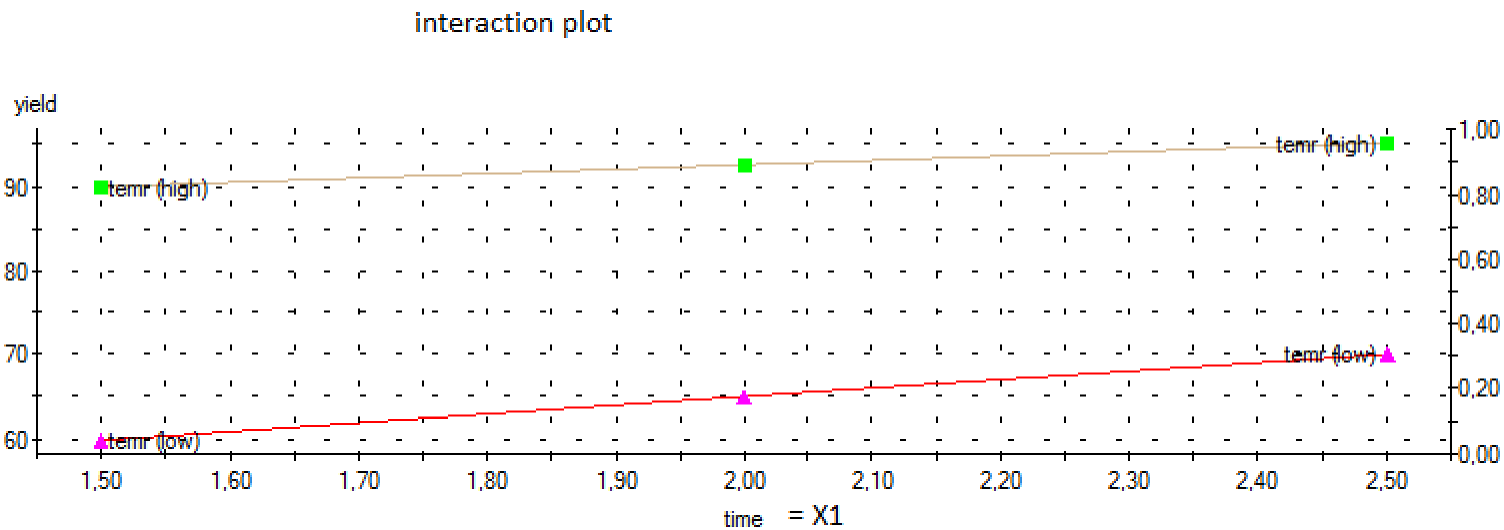
graphical presentation of effects



Determination of interaction type

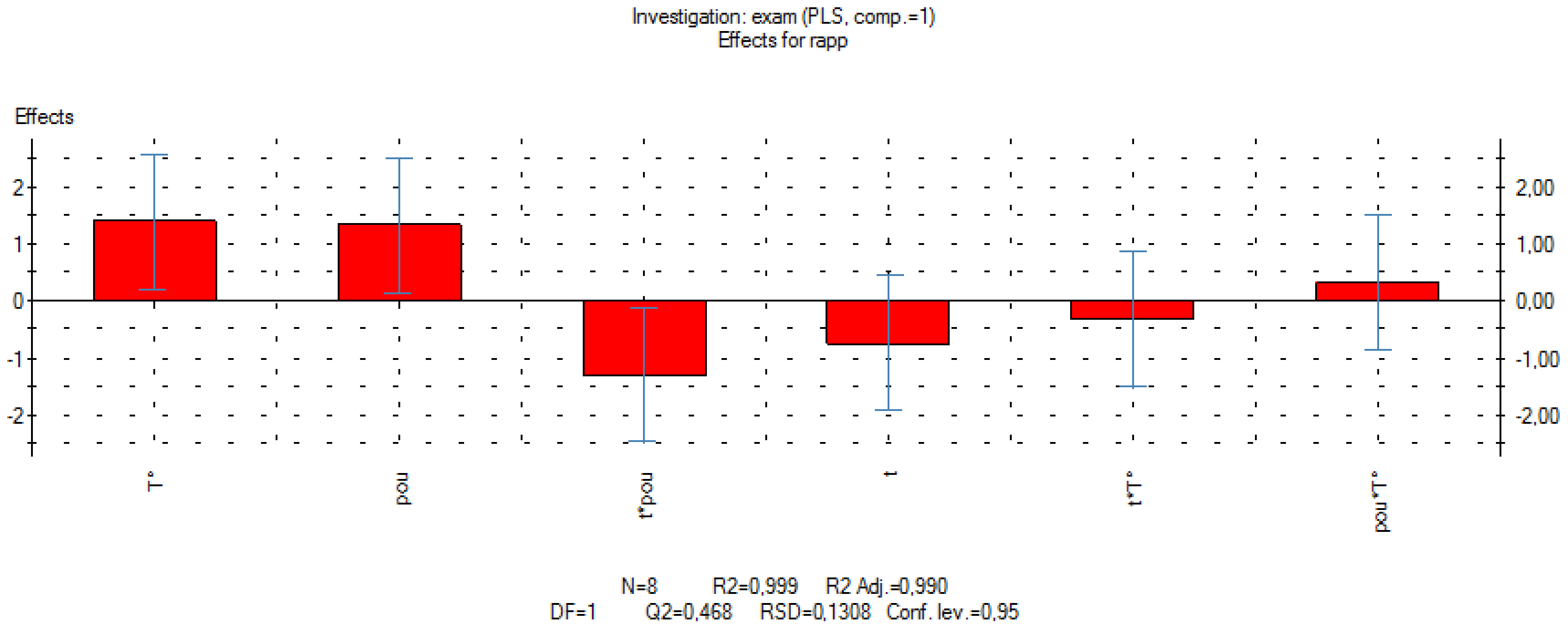
time	temperature	yield
-1	-1	60
+1	-1	70
-1	+1	90
+1	+1	95

Determination of interactions and effects



Weak interaction between time and temperature

Plot interpretation



Each factor has a negative sign, that means this factor decrease the response and vice versa.

Application TP 01 : full factorial 2power 2

Suppose that we are interested in improving the yield of a chemical process. We know from the results of a characterization experiment that the two most important process variables that influence the yield are operating temperature and reaction time. The process currently runs at 145°F and 2.1 hours of reaction time, producing yields of around 80 percent. the contour lines for yields of 60, 70, 80, 90, and 95 percent.

reaction time [1,5 -2,5]

Temperature [140 -150]

1- determine the effects of each factor

2- give the interaction between the factors

3- write the mathematical model

A- for first order

B- for second order

Determination of effects in case of three factors

determination of interactions in case of three factors

Tp 2 : full factorial 2 power 3

Chaptar 2
model validation

Introduction

Model validation is defined as the process of determining the degree to which a model is an accurate representation of the real world system. The model is valid means that it behaves similar to the real world system for any effective input. The increasing of validation experiments will enhance the confidence in using the computational model as well as reduce the risks of using an invalid model to some extent.

Experimental errors [1]

the variance among experimental units treated alike, often symbolized as σ2 or σe 2

suppose that y1, y2, . . . , yn represents a sample. Then the sample mean

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

and the sample variance S2

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

the sample variance S2 is a point estimator of the population variance σ2 .

Example :

time	temperature	yield
-1	-1	11,1
+1	-1	12,6
-1	+1	10,4
+1	+1	11,9

sample mean

$$\bar{y} = \frac{1}{4}(11{,}1 + 12{,}6 + 10{,}4 + 11{,}9) = 11{,}5$$

$$11{,}1 - 11{,}5 = -0{,}4$$

$$12{,}6 - 11{,}5 = +1{,}1$$

$$10{,}4 - 11{,}5 = -1{,}1$$

$$11{,}9 - 11{,}5 = 0{,}4$$

$$(-0{,}4)^2 + (1{,}1)^2 + (-1{,}1)^2 + (0{,}4)^2 = +0{,}16 + 1{,}21 + 1{,}21 + 0{,}16 = +2{,}74$$

sample variance

$$S2 = \frac{2{,}74}{4-1} = 0{,}9133$$

sample standard deviation

$$\text{écart-type} = \sqrt{0{,}9133} = 0{,}9557 \cong 0{,}96$$

sample standard deviation, is used as a measure of dispersion

S= $$\sigma_{population} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{i=n}(y_i - \bar{y})^2}$$

The Test of Significance

- ✓State the Research Hypothesis.
- ✓State the Null Hypothesis (no effect).
- ✓Type I and Type II Errors. Select a probability of error level (alpha level)
- ✓Chi Square Test. Calculate Chi Square. Degrees of freedom. Distribution Tables. Interpret the results.
- ✓T-Test.

The Test of Significance : application

time	temperature	yield
-1	-1	11,1
+1	-1	12,6
-1	+1	10,4
+1	+1	11,9

A company produce a mean yield of product A of 11,5% . The decision maker in this company believes that the mean yield could be 15%.

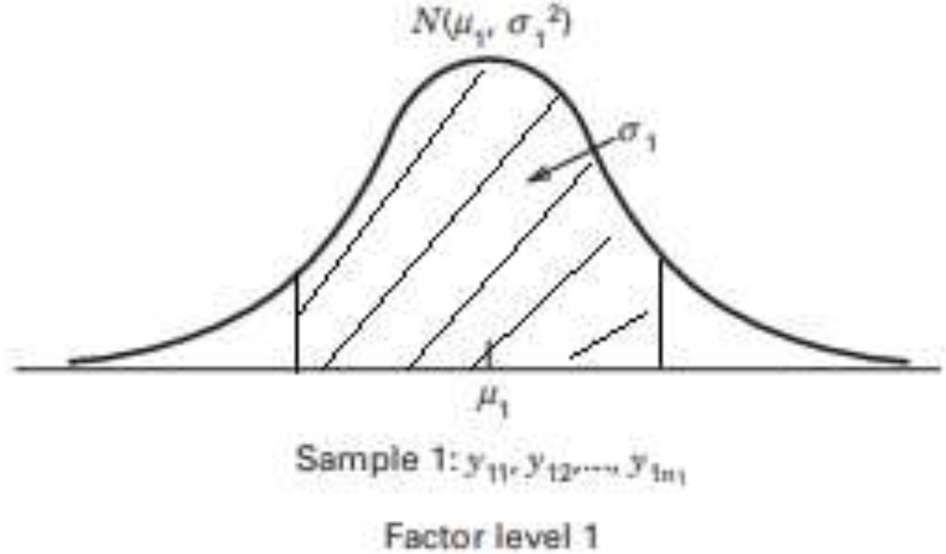
To do a test of significance:

-State the Research Hypothesis : it is possible to get the mean yield equal to 15% (Ha : alternative hypothesis).

-State the Null Hypothesis H0 (no effect) : $\eta = 11{,}5\%$.

-Select probability of error (calculate P- Value):

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$



$Z = (\bar{Y} - \eta^\circ)/(S/\sqrt{n})$

Y : the mean sample (in this case the mean yield)

η° : 15%

S : sample standard deviation

n : sample size

P- value : is the probability of obtaining a sample (more extreme) that the ones observed in your data assuming H0 is true.

Application : $Z = (11.5 - 15)/(0.96/\sqrt{4}) = -7{,}29$

Anova table :

The ANOVA table has rows for every term in the model and columns for source squares, degrees of freedom (DF), mean squares, and F-statistics.

SS: sum of square; MS : sample variance ; SD : sample standard deviation ; application for the same example using MODDE

	1	2	3	4	5	6	7
1	yield	DF	SS	MS	F	p	SD
2				(variance)			
3	Total	4	531,74	132,935			
4	Constant	1	529	529			
5							
6	Total Corrected	3	2,73999	0,91333			0,955683
7	Regression	3	2,73999	0,91333	--	--	0,955683
8	Residual	0	0	0			--
9							
10	Lack of Fit	0	--	--	--	--	--
11	(Model Error)						
12	Pure Error	0	--	--			--
13	(Replicate Error)						
14							
15	N = 4	Q2 =	--	Cond. no. =	1,0000		
16	DF = 0	R2 =	--	Y-miss =	0		
17		R2 Adj. =	--	RSD =	--		

Correlation coefficient

R2 is the portion of the total variation in Y that is explained away by using the x information in a regression. R2 is always between 0 and 1. An R2 of 0 means that x provides no information about y. An R2 of 1 means that use of x information allows perfect prediction of y.

$$R^2 = \frac{\text{explained variability}}{\text{total variability}} = \frac{SS_{tot} - SS_{res}}{SS_{tot}}.$$

R2 = sum of square calculated from the model/sum of square calculated from experimental data

Example :

Essai n°	Réponses mesurées	Réponses calculées	Résidus
1	26,1	25,825	0,275
2	22,2	23	–0,8
3	10,1	11,025	–0,925
4	12,2	10,75	1,45
5	14,2	13,95	0,25
6	12,7	12,425	0,275
7	5,9	5,5	0,4
8	5,6	6,525	–0,925
9	23	22	1
10	20,1	20,575	–0,475
11	2,4	2,75	–0,35
12	3,7	3,875	–0,175
13	11	12,525	–1,525
14	13,4	12,4	1
15	0,5	–0,375	0,875
16	1,7	2,05	–0,35

$$\hat{y} = 11{,}55 - 0{,}1x_1 - 6{,}2875x_2 - 3{,}425x_3 - 2{,}075x_4$$
$$+ 0{,}6375x_1x_2 + 0{,}325x_1x_3 + 0{,}35x_1x_4 + 1{,}5875x_2x_3 - 1{,}1125x_2x_4 + 0{,}6x_3x_4$$

sum of square calculated from experimental data

$$26{,}1 - 11{,}55 = +14{,}55$$
$$22{,}2 - 11{,}55 = +10{,}65$$
$$10{,}1 - 11{,}55 = -1{,}45$$
$$\ldots$$
$$0{,}5 - 11{,}55 = -11{,}05$$
$$1{,}7 - 11{,}55 = -9{,}85$$

$$(14{,}55)^2 + (10{,}65)^2 + \ldots + (-11{,}05)^2 + (-9{,}85)^2$$
$$= +\,211{,}702\,5 + 113{,}422\,5 + \ldots + 122{,}102\,5 + 97{,}022\,5 = +\,975{,}72$$

sum of square calculated from the model

$$25{,}825 - 11{,}55 = +14{,}275$$
$$23{,}0 - 11{,}55 = +11{,}45$$
$$11{,}025 - 11{,}55 = -0{,}525$$
$$\ldots$$
$$-0{,}375 - 11{,}55 = -11{,}925$$
$$2{,}05 - 11{,}55 = -9{,}5$$

$$(14{,}275)^2 + (11{,}45)^2 + \ldots + (-11{,}925)^2 + (-9{,}5)^2$$
$$= +\,203{,}775\,625 + 131{,}102\,5 + \ldots + 142{,}205\,625 + 90{,}25 = +\,965{,}30$$

R2 = 965.30/975.72 = 0.989

Degree of freedom

The number of degrees of freedom of a sum of squares is equal to the number of independent elements in that sum of squares.
For example, diapo 27

$$SS = \Sigma_{i=1}^{n}(y_i - \bar{y})^2$$

single degree of freedom = n -1

Confidence interval

The confidence interval (CI) expresses the degree of uncertainty around a certain effect, it give an idea about the power of a study.

$$CI = \overline{Y} \pm z \cdot \frac{s}{\sqrt{n}}$$

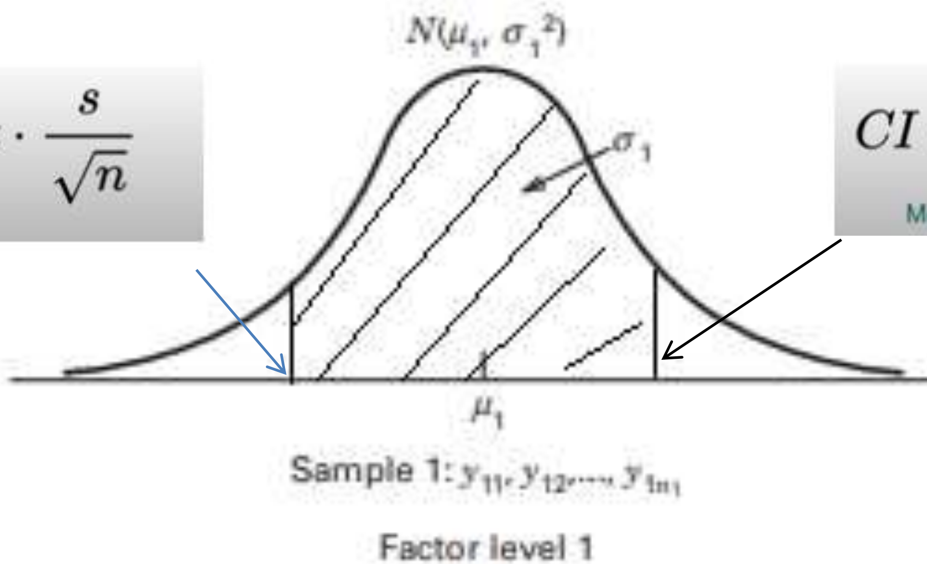
Mean value

Where :
Y : is the mean value
Z : is Z value for confidence level
S : standard deviation
n : sample size

Confidence level	Z-value
80%	1.28
90%	1.645
95%	1.96
98%	2.33
99%	2.58

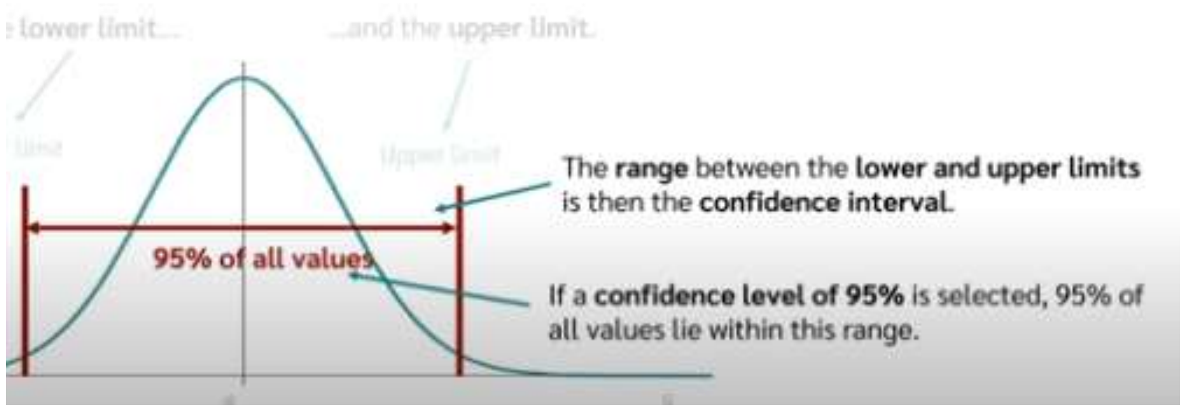
$$CI = \overline{Y} - z \cdot \frac{s}{\sqrt{n}}$$

Mean value



$$CI = \overline{Y} + z \cdot \frac{s}{\sqrt{n}}$$

Mean value



Chaptar 3 : Fractional factorial design 2 power (k−p)

INTRODUCTION

Suppose there are k factors (A,B,...,J,K) in an experiment. All possible factorial effects include

effects of order 1: A, B, ..., K (main effects)

effects of order 2: AB, AC, ...,JK (2-factor interactions)

– Lower order effects are more likely to be important than higher order effects.

– Effects of the same order are equally likely to be important.

Number of runs required for full factorial grows quickly!!!

Consider 2 power k design

– If k = 7 → 128 runs required

Construction and Analysis of the One-Half Fraction

A one-half fraction of the 2power k design of the highest resolution may be constructed by writing down a basic design consisting of the runs for a full 2k1 factorial

EXAMPLE

Chemists are trying to determine the tellurium (Te) content in seawater. But the nature and concentration of other metals distort the measurements. The content in tellurium can be either too high or too low depending on the other metals and their concentration. The disruptive metals could be sodium (Na), potassium (K) and calcium (Ca).

Basic matrix of a plan 2 power3.

factors	Level -1 (µg/ml)	LEVEL 0 (µg/ml)	Level +1 (µg/ml)
Concentration Na (1)	2,5	25	250
Concentration K (2)	2,5	25	250
Concentration Ca (3)	2,5	25	250

MATRIX DESIGN

2 power 3

	A (-1, +1)	B (-2, +2)	C (-4, +4)
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

2 power 4

	A (-1, +1)	B (-2, +2)	C (-4, +4)	D (-8,+8)
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4	+	+	-	-
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16	+	+	+	+