

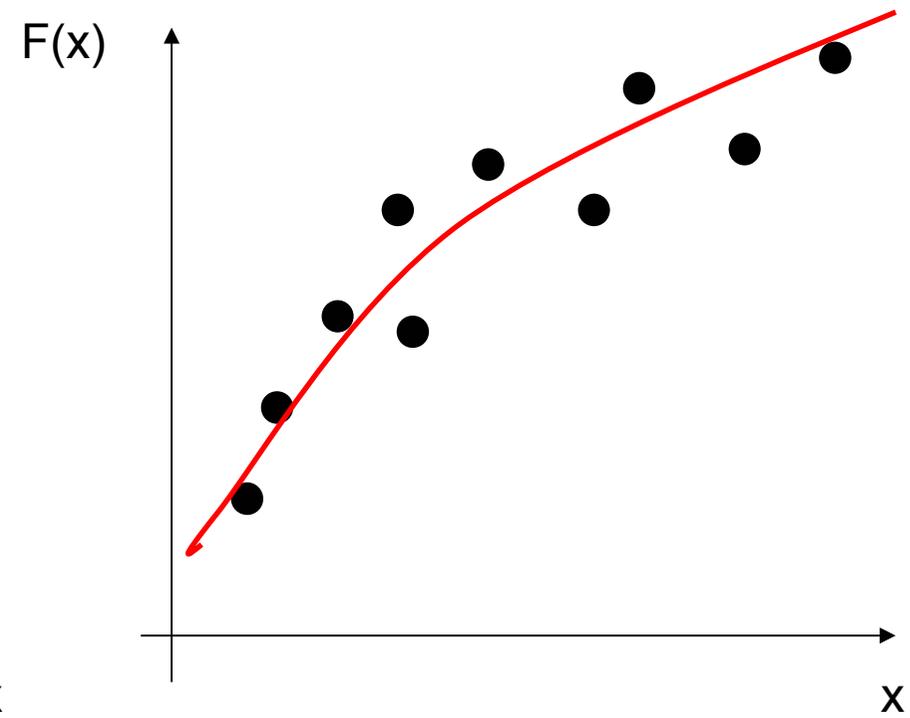
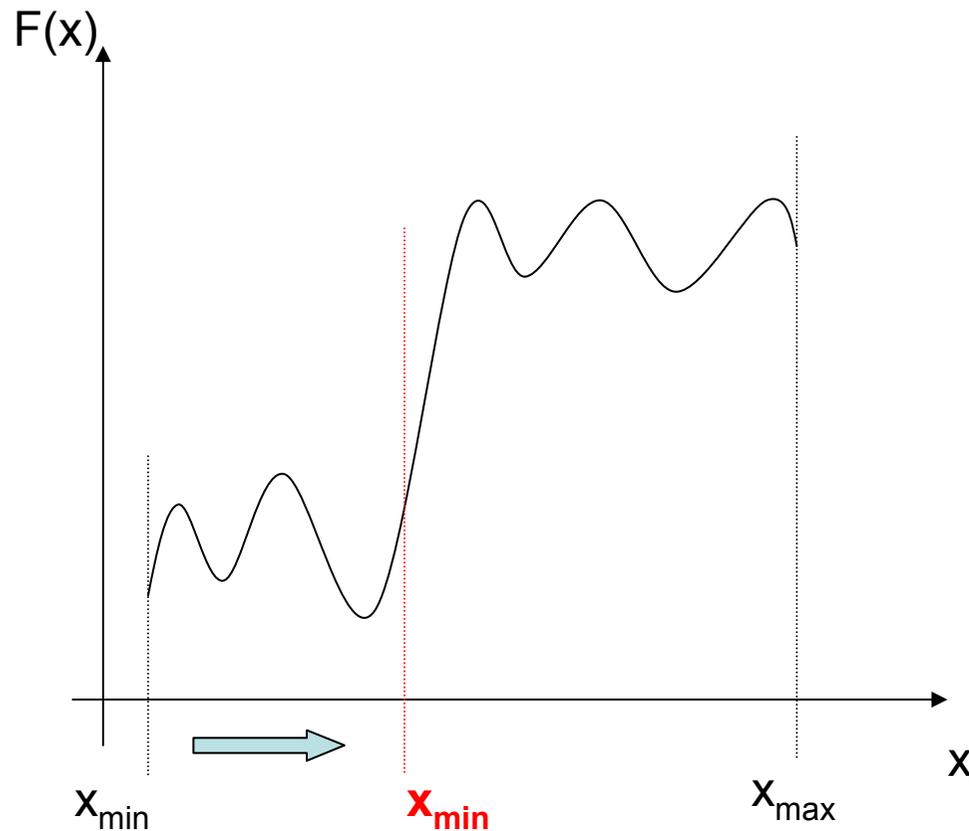
Design of experiments (DOE)

e

Analisi statistica

L'utilizzo fondamentale della metodologia Design of Experiments è approfondire la conoscenza del sistema in esame

- Determinare le variabili più significative;
- Determinare il campo di variazione delle variabili;
- Determinare le relazioni tra variabili-obiettivi e obiettivi-obiettivi



- **Pro:**

- Ridurre in numero di configurazioni calcolate;
- Eliminare le configurazioni ridondanti;
- trovare le maggiori informazioni possibili sulle relazioni tra variabili ed obiettivi.

- **Contro:**

- Normalmente si presuppone una relazione tra variabili ed obiettivi (lineare, quadratica, ecc.)

Il DOE **non** viene sempre utilizzato in una fase di ottimizzazione.

Quando il DOE viene utilizzato durante una fase di ottimizzazione, normalmente la **precede** (diminuzione di variabili, nuovi campi di variazione).

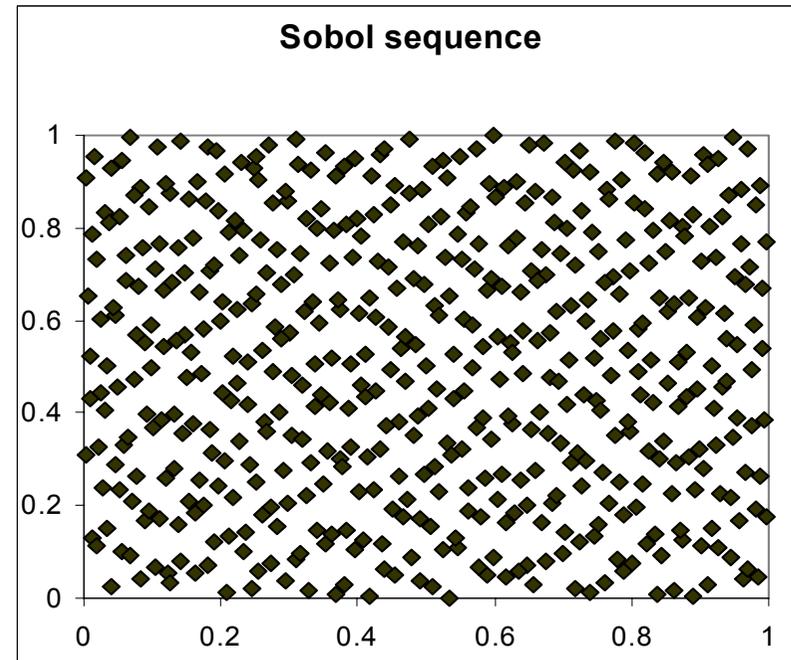
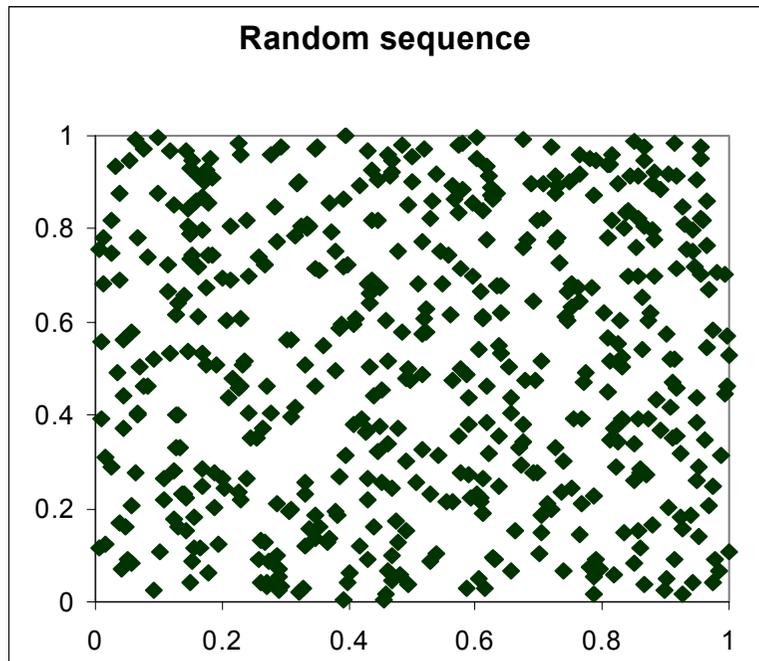
Il DOE può servire a creare in database per la successiva superficie di risposta.

Le metodologie DOE **Random e Sobol** sono basate sulla teoria numerica della generazione casuale di numeri random (normalmente compresi tra 0 e 1)

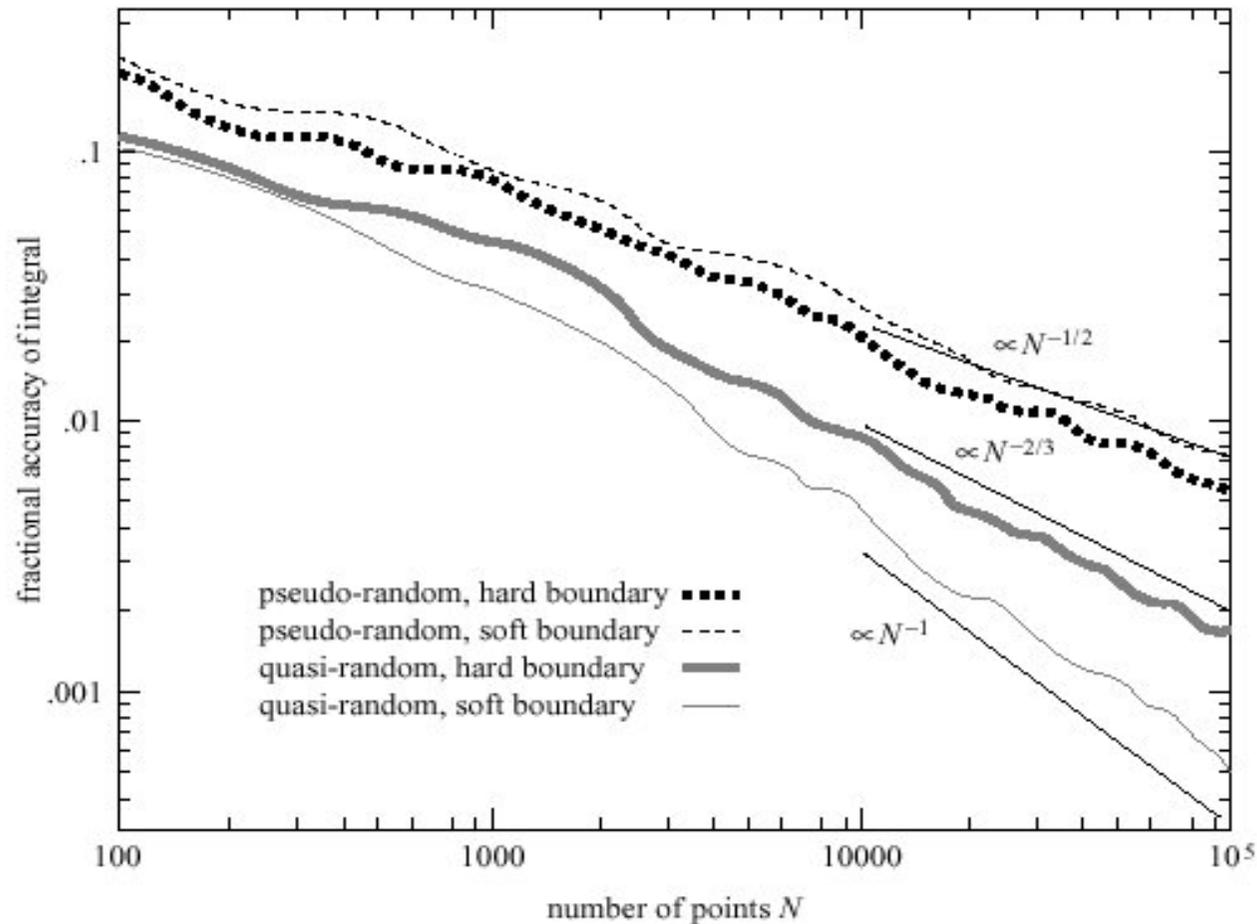
Sono metodologie semplice che coprono in modo adeguato il dominio di definizione delle funzioni obiettivo

- Sequenza Random (migliore per funzioni definite a “molte” variabili)
- Sequenza Sobol (migliore per “basso” numero di variabili)

Essendo metodologie basate su generazione casuale di numeri si evita il problema di creare configurazioni correlate tra loro.



Si nota come l'algorithmo **SOBOL** copre in modo migliore uno spazio a 2 variabili



Esempio di integrazione con tecnica MonteCarlo:

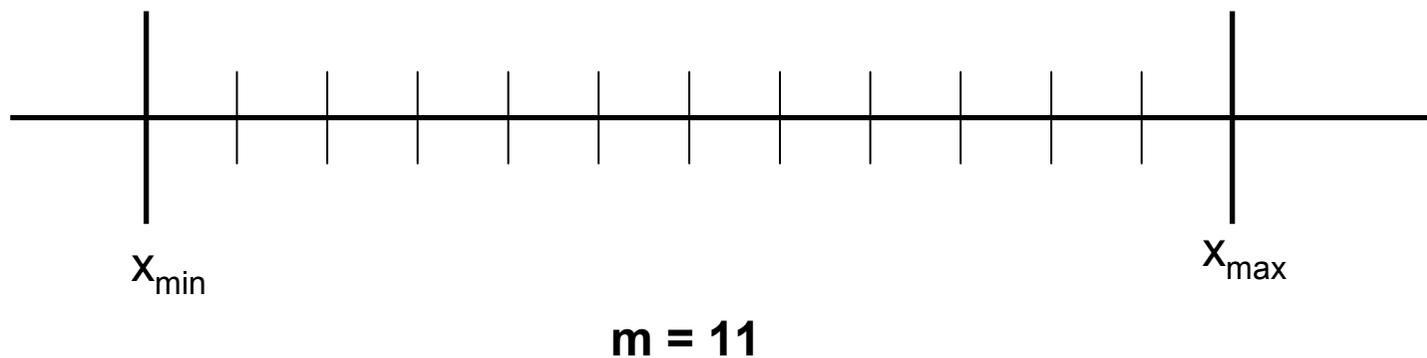
Si nota come l'accuratezza è migliore per una generazione di numeri casuale con algoritmo Sobol: ciò significa una migliore copertura dello spazio di definizione della funzione da integrare

Metodologie fattoriali:

- Full Factorial
- Reduced Factorial

n = numero variabili

m = numero di divisioni interne della variabile (**base**)

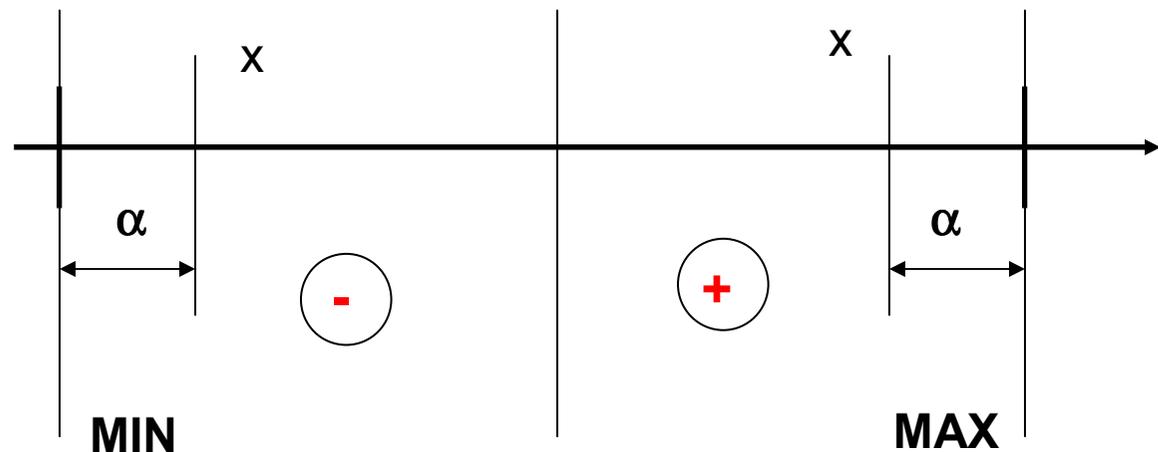


Con l'algoritmo DOE Full Factorial si creano **tutte** le combinazioni possibili delle variabili (date la base)

Numero di configurazioni = m^n

Poiché il numero di configurazioni da calcolare è molto elevato, Full Factorial si applica normalmente quando il numero di variabili e/o base è limitato, oppure quando il calcolo delle funzioni obiettivo è rapido.

**Concetto di intrusione
nel dominio (parametro α)**



ESEMPIO

N = 3 (variabili)

M = 2 (base + e -)

Full Factorial

$2^3=8$ configurazioni totali da calcolare

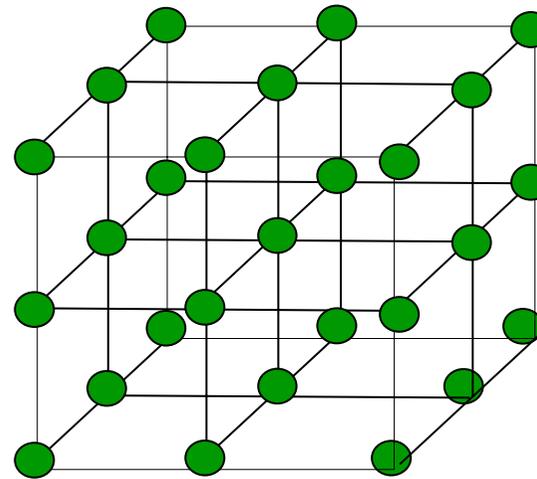
Come si può osservare dall'esempio un vantaggio della metodologia Full Factorial è la presenza dello stesso numero di configurazioni con le variabili nel campo + o -

Svantaggio: per non aumentare troppo le configurazioni base = 2 allora solo interazioni lineari.

	X ₁	X ₂	X ₃
1	+	+	+
2	+	+	-
3	+	-	+
4	+	-	-
5	-	+	+
6	-	+	-
7	-	-	+
8	-	-	-

Full Factorial 3 livelli

Con base 3 è possibile
calcolare le interazioni di
secondo ordine



3 variabili

27 esperimenti

Reduced Factorial

Metodologia fattoriale sviluppata per diminuire le configurazioni necessarie

n = numero di variabili

m = base

Numero configurazioni m^p $p < n$

n. design	x1	x2	x3	x4 (=x1*x2)
1	+	+	+	+
2	+	+	-	+
3	+	-	+	-
4	+	-	-	-
5	-	+	+	-
6	-	+	-	-
7	-	-	+	+
8	-	-	-	+

Da un punto di vista operativo si effettua un Full Factorial su p variabili (normalmente vengono scelte che abbiano un peso maggiore); le rimanenti saranno combinazione delle precedenti.

Pro:

- Il numero di configurazioni da calcolare è minore rispetto al Full Factorial

Contro:

- Impossibilità di studiare tutte le interazioni tra le variabili;
- Limite al numero delle variabili (saturated factorial con $m=3$, max 6 variables ($x_4=x_1*x_2$, $x_5=x_1*x_3$, $x_6=x_2*x_3$))

Latin square

Quando la base aumenta troppo ($m > 2,3$) con le metodologie Factorial il numero di configurazioni richieste può diventare insostenibile; per questo motivo è stata sviluppata la metodologia **Latin Square**.

Il numero di configurazioni richieste è m^2 dove m è la base delle variabili

Example:

1	2	3
3	1	2
2	3	1

a	b	c
c	a	b
b	c	a

A	B	C
B	C	A
C	A	B

Latin Square for 3 variables (X1,X2,X3)
with 3 levels :

X1(1,2,3),
X2(A,B,C),
X3(a,b,c)

a1A	b2B	c3C
c3B	a1C	b2A
b2C	c3A	a1B

Vantaggi Latin Square:

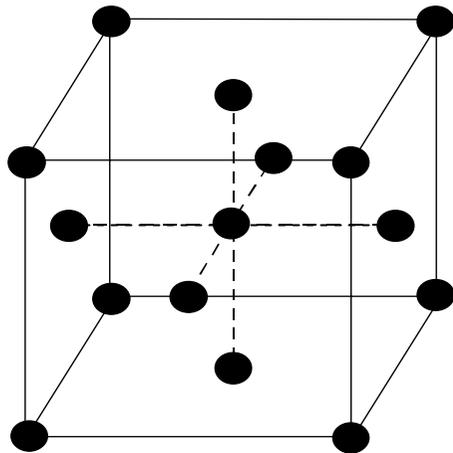
- **Il numero di design calcolati non dipende dal numero di variabili;**
- **Possibilità di avere una base elevata senza aumentare il numero di configurazioni richieste;**

Svantaggi Latin Square:

- **Il DOE non rappresenta l'intero spazio delle variabili (perdita di informazioni).**

Cubic Face Centered

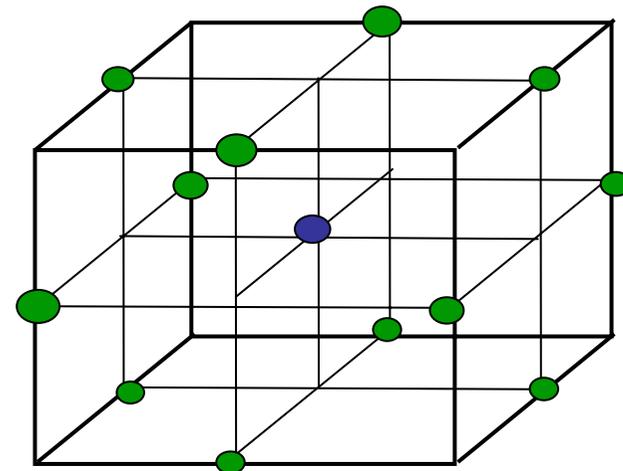
- $2n + 2*n + 1$ configurazioni
 - permette lo studio delle interazioni del secondo ordine
 - Meno costoso del 3 livelli Full Factorial



3 variabili
15 configurazioni

Box-Behnker

Simile alla metodologia Cubic Face Centered ma non vengono utilizzati i punti estremi degli edge



Dopo che la tabella DOE delle configurazioni è stata valutata, è necessario processare i dati in modo da ottenere le informazioni del sistema in esame:

- quali sono le variabili **più influenti**?
- si può **ridurre** lo spazio delle variabili ?
- qual è la zona dello spazio di definizione delle variabili **più efficiente** per gli obiettivi?
- il numero di obiettivi e vincoli è **adeguato** per il problema in esame?

Analisi statistica dei dati

Analisi statistica semplificata

$f: \mathcal{R}^4 \Rightarrow \mathcal{R}$ $n = 4$ $m = 2$ Reduced Factorial $p = 3$

	A	B	C	D	OBJ
1	-	-	-	-	65,6
2	-	-	+	+	79,3
3	-	+	-	+	51,3
4	-	+	+	-	69,6
5	+	-	-	+	59,8
6	+	-	+	-	77,7
7	+	+	-	-	74,2
8	+	+	+	+	87,9

Analisi statistica semplificata in quanto si osserva solamente la differenza media del valore della funzione all'interno del range + o -.

	A	B	C	D	AB	AC	AD
Tot +	300	283	315	278	307	231,2	282,9
Tot -	266	282	251	287	258,4	282,9	282,5
Diff	34	0,6	64	-9	48,6	-51,7	0,4
Effect	8,5	0,2	16	-2	12,5	-12,925	0,1

Diff = Tot+ - Tot-
Effect = Diff / 4

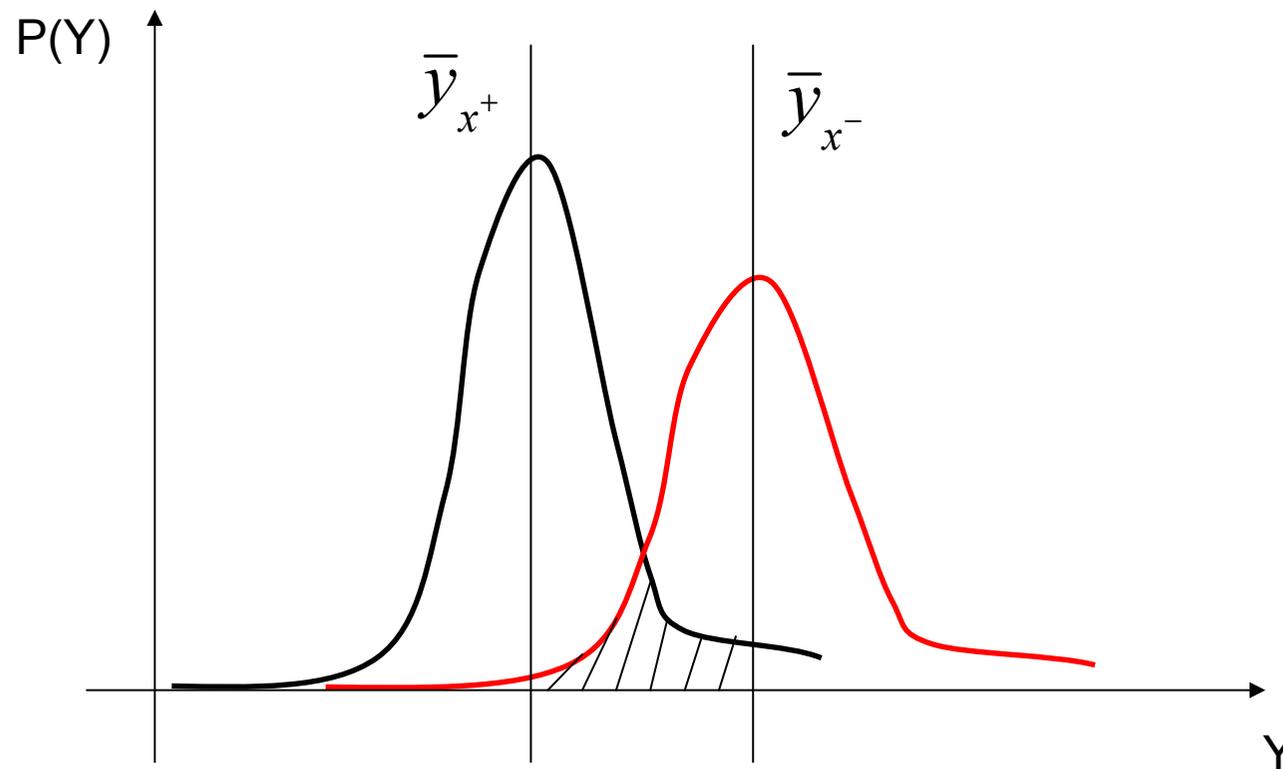
Per migliorare l'accuratezza dei risultati si utilizza la metodologia **t-student**

$$\bar{y}_{x_{1+}} = \frac{\sum_{i=1}^{n_{x_{1+}}} y_{i,n_{x_{1+}}}}{n_{x_{1+}}}$$

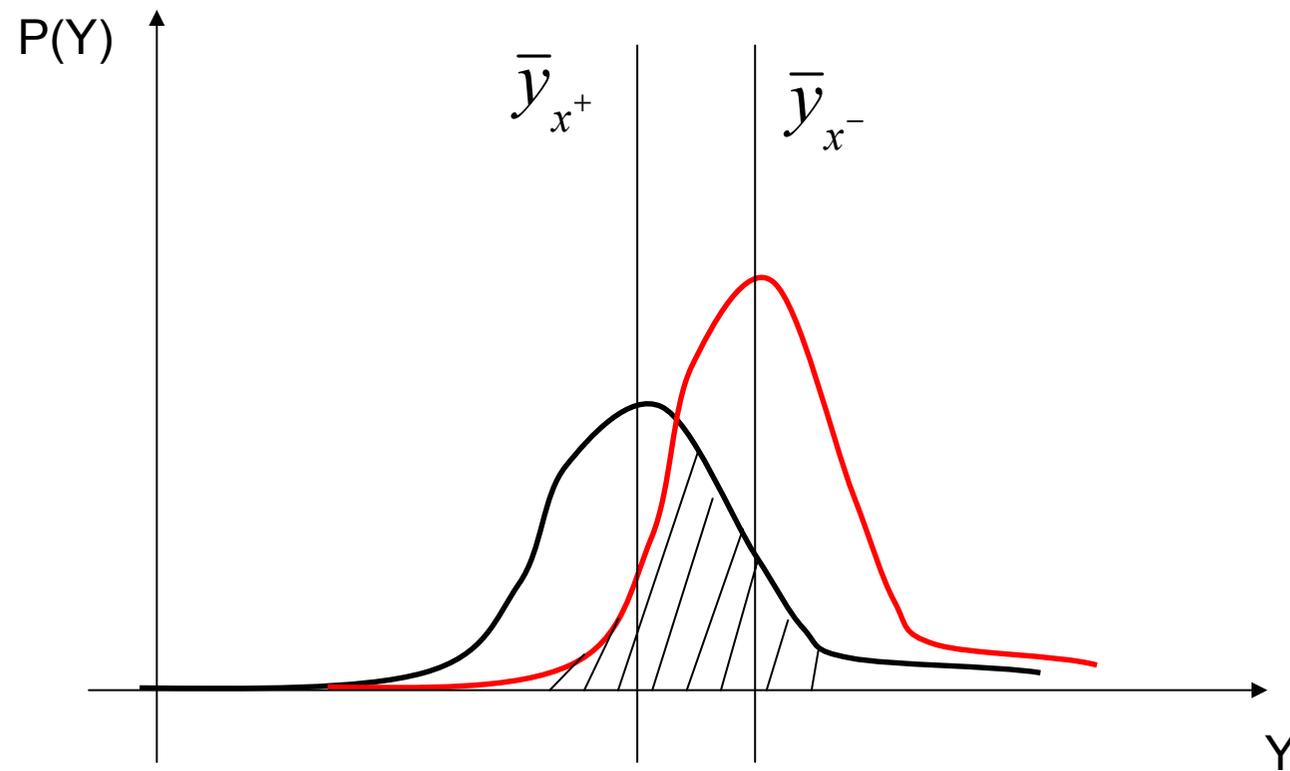
$$\bar{y}_{x_{1-}} = \frac{\sum_{i=1}^{n_{x_{1-}}} y_{i,n_{x_{1-}}}}{n_{x_{1-}}}$$

$$t_{x_1} = \frac{|\bar{y}_{x_{1+}} - \bar{y}_{x_{1-}}|}{\sigma_{x_1}}$$

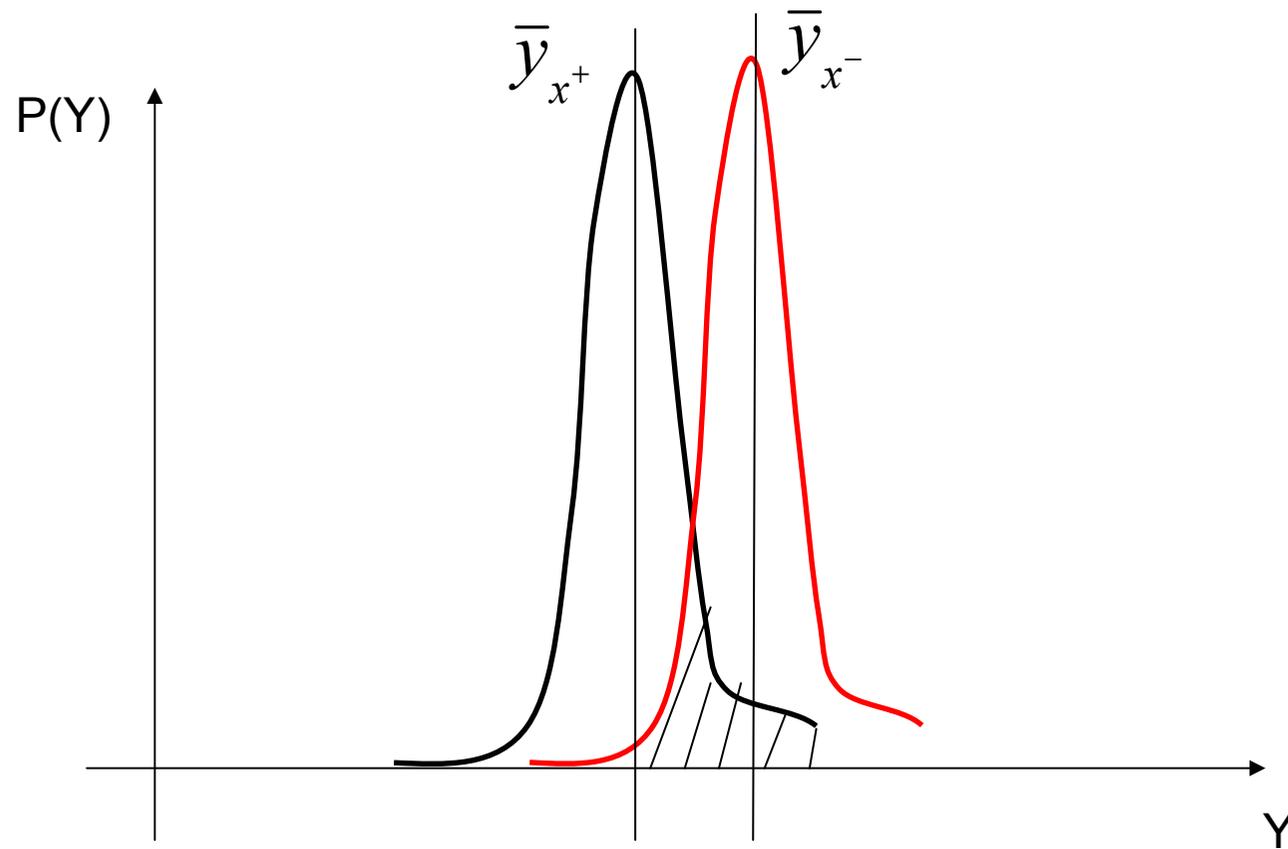
$$\sigma_{x_1} = \sqrt{\frac{\sum_{i=1}^{n_{1,x_{1+}}} (\bar{y}_{x_{1+}} - y_{i,x_{1+}})^2 + \sum_{i=1}^{n_{1,x_{1-}}} (\bar{y}_{x_{1-}} - y_{i,x_{1-}})^2}{(n_{1,x_{1+}} + n_{1,x_{1-}} - 2)(n_{1,x_{1+}} \cdot n_{1,x_{1-}})}} (n_{1,x_{1+}} + n_{1,x_{1-}})}$$



Il parametro di t-student dà l'indicazione dell'area di ricoprimento tra le due gaussiane; se t-student tende ad **uno** l'area è piccola è la variabile è **significativa**



Se t-student tende ad **zero** l'area è grande è la variabile **non è significativa**



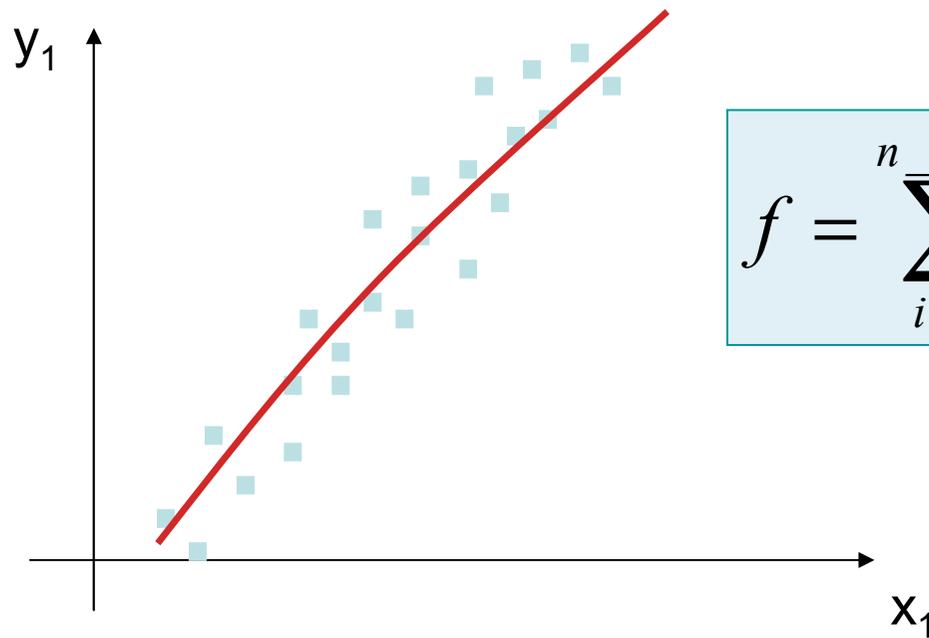
A differenza dell'analisi statistica semplificata non è importante solo la differenza tra le medie, ma anche la distribuzione attorno al valore medio

- Nel caso di t-Student **elevato** sarà possibile:
 - affermare che nel dominio di definizione di una variabile, la differenza tra intervallo + e - è elevata;
 - restringere le ricerche successive nell'intervallo (+,-) di interesse maggiore.
- Nel caso di t-Student **basso** sarà possibile:
 - affermare che la funzione obiettivo non varia quando si passa dal dominio - a quello +;
 - possibilità di eliminare la variabile nelle analisi successive.

Uno degli aspetti più importanti dell'analisi statistica è trovare le interazioni tra le variabili di progetto e le funzioni obiettivo e, nel caso multi obiettivo, tra gli obiettivi stessi

Per calcolare queste interazioni si può utilizzare il **criterio di Chauvenet**

Il criterio di Chauvenet aiuta a trovare la funzione di regressione $y=a_0+a_1x+a_2x^2$ che minimizza la funzione:



$$f = \sum_{i=1}^{n \text{ dati}} \left[y_i - (a_0 + a_1 x_i + a_2 x_i^2) \right]^2$$

- Il problema numerico (trovare i parametri a_0, a_1, a_2) può essere risolto con il metodo di Kramer (risoluzione sistemi lineari) usando le relazioni:

$$\begin{cases} \frac{\partial f}{\partial a_0} = 0 \\ \frac{\partial f}{\partial a_1} = 0 \\ \frac{\partial f}{\partial a_2} = 0 \end{cases}$$

- Dopo avere effettuato la correlazione, sarà possibile determinare l'insieme dei punti (x_1, y_1) troppo distanti dalla curva di correlazione stessa.
- L'utilità di eliminare questi punti sta nel fatto che essi possono essere **NON** rappresentativi per una curva di correlazione (errori numerici, zone del dominio di non interessanti, ecc.)

Criterio di Chauvenet:

$$\Delta_i = y_i - (a_0 + a_1 x_i + a_2 x_i^2)$$

$$\sigma_{\Delta} = \sqrt{\frac{\sum_{i=1}^{n_dati} (\bar{\Delta} - \Delta_i)^2}{n_dati - 1}}$$

Verranno eliminati i punti i-esimi che: $\Delta_i > \varepsilon_{\max} \sigma_{\Delta}$

ε_{\max} È funzione del numero di punti n_dati (la tolleranza cresce con il numero di dati)

Il criterio di Chauvenet è molto utile per determinare le relazioni esistenti tra varie entità (variabili-obiettivi).

Molto interessante è l'applicazione del Latin Square assieme al criterio di Chauvenet (multi livello).

Con la possibilità di eliminare alcuni design si riesce ad ottenere una funzione di correlazione più precisa e meno influenzata da punti troppo diversi.