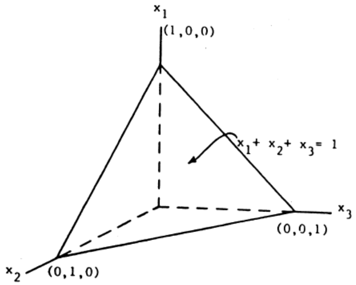
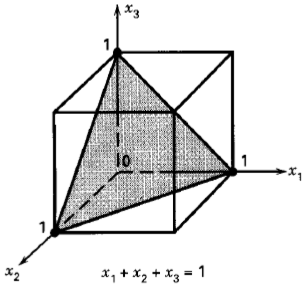
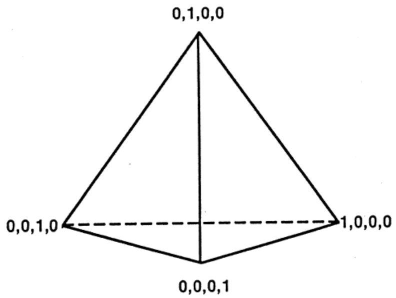
DISEGNO DEGLI ESPERIMENTI PER MISCELE

Molte volte il prodotto d’interesse in uno studio DoE è una miscela di vari componenti.



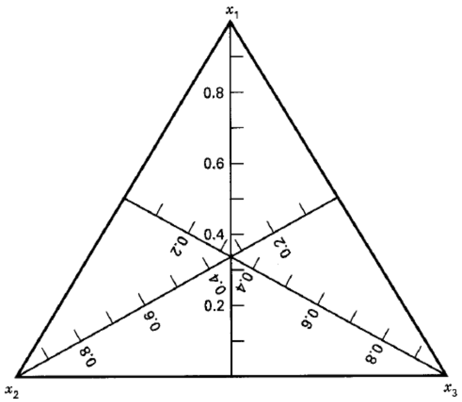


Con 4 componenti lo spazio dei fattori è un tetraedro, con i vertici corrispondenti ai componenti puri.



Quando vi sono 3 componenti nella miscela, la regione sperimentale vincolata può essere rappresentata convenientemente con un grafico a coordinate trilineari.

Ciascuno dei 3 lati rappresenta una miscela in cui è del tutto assente uno dei 3 componenti, che è quello corrispondente al vertice opposto.



I 9 segmenti della griglia in ciascuna direzione indicano incrementi del 10% nei relativi componenti.

In una descrizione più generale, per studiare gli effetti dei componenti delle miscele sulla variabile risposta sono utilizzati **i piani simplesso**.

In pratica, lo spazio dei fattori si riduce a un simplesso regolare di p - 1 dimensioni.

In matematica, il simplesso n-dimensionale è il politopo n- dimensionale col minor numero di vertici.

Il simplesso di dimensione zero è un singolo punto, il simplesso bidimensionale un triangolo e quello tridimensionale un tetraedro.

Il simplesso n - dimensionale ha n + 1 vertici.

Come tutti i politopi, il simplesso ha facce di ogni dimensione: queste sono tutte a loro volta simplessi.

Per la sua semplicità, il simplesso è generalmente ritenuto il "blocco base" con cui costruire spazi n-dimensionali più complicati tramite un processo detto triangolazione.

Un piano simplesso a griglia [**p,m**] per **p** componenti e per **m** numero di divisioni della percentuale di ogni costituente, il quale descrive l'ordine del modello che è supportato dal design, consiste di punti definiti da precise posizioni sulle coordinate.

Le proporzioni assunte da ciascun componente assumono gli m+1 valori x_i equi spaziati tra 0 e 1:

$$\boldsymbol{x_i = \frac{0}{m}, \frac{1}{m}, \frac{2}{m}, \ldots.., \frac{m}{m} \qquad i = 1, 2, \ldots.., p}$$

e si usano tutte le possibili combinazioni, cioè miscele, delle proporzioni della formula sopra.

Se ad esempio, si avessero **p = 3 componenti** e **m = 2**, si avrebbe:

$$x_i = 0, \tfrac{1}{2}, 1 \qquad i = 1, 2, 3$$

La griglia del simplesso avrebbe le seguenti 6 prove:

$$(x_1, x_2, x_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1), (\tfrac{1}{2}, \tfrac{1}{2}, 0), (\tfrac{1}{2}, 0, \tfrac{1}{2}), (0, \tfrac{1}{2}, \tfrac{1}{2})$$

Questo piano prove è riportato nella figura che segue, piano simplesso a griglia [3,2].

I 3 vertici: (1,0,0), (0,1,0) e (0,0,1) corrispondono alle miscele pure.

i punti (1/2, 1/2, 0), (1/2, 0, 1/2) e (0, 1/2, 1/2) corrispondono alle miscele binarie o a 2 componenti, posizionate nei punti centrali dei 3 lati del triangolo.

Inoltre la figura riporta anche i piani simplesso a griglia [3,3], [4,2] e [4,3].

In generale, il numero di punti in un piano simplesso a griglia [p,m] è:

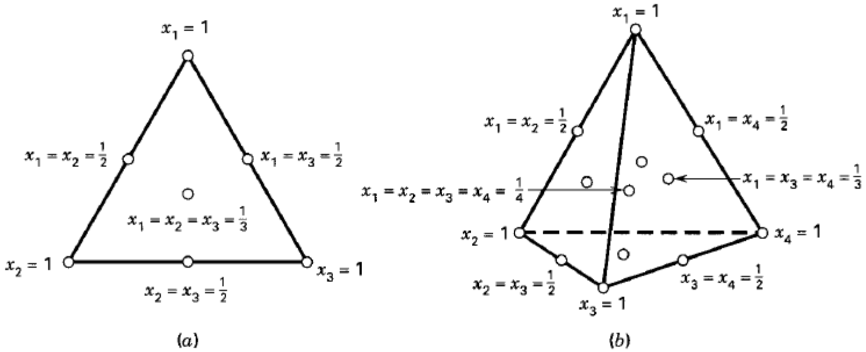
$$N = \frac{(p + m - 1)!}{m!(p - 1)!}$$

Una alternativa al piano simplesso a griglia è il **piano simplesso con il baricentro**.

In un piano simplesso con il baricentro, a p componenti, ci sono $(2^p - 1)$ punti corrispondenti alle p permutazioni di (1,0,0,…,0), le $\binom{p}{2}$ permutazioni di (1/2, 1/2, 0,…..,0), le $\binom{p}{3}$ permutazioni di (1/3, 1/3, 1/3, 0,….,0), … ed il baricentro generale (1/p, 1/p,……,1/p).

La figura di seguito riporta 2 piani simplesso con baricentro:

- piano (a) con p = 3
- piano (b) con p = 4



Nei piani a simplesso visti la maggior parte delle prove sperimentali si trovano sui confini della regione e pertanto includono solo p -1 delle p componenti.

Se si vogliono fare previsioni sulle proprietà delle miscele in modo completo, sarebbe opportuno condurre ulteriori prove all’interno del simplesso.

Viene raccomandato di aumentare i comuni piani simplesso con prove dette “assiali” e con il baricentro, generale, nel caso non fosse già un punto del piano sperimentale.

Modello Canonico Polinomiale

Oltre allo spazio del design, gli esperimenti con le miscele differiscono anche nei modelli matematici da quelli polinomiali comunemente impiegati per le superfici di risposta, a causa del vincolo $\sum x_i = 1$.

Lo spazio sperimentale ammissibile è limitato ad un solo asse.

L’interpretazione dei coefficienti lineari in un modello per miscele può essere ingannevole.

Se ad esempio tutti e due i coefficienti lineari fossero abbastanza grandi, ma approssimativamente simili, si concluderebbe che non si evidenzia un significativo effetto dovuto alla variazione della composizione della miscela.

Nel caso di 2 variabili, **il modello quadratico a superficie di risposta**, sarebbe:

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2.$$

poichè $x_1 + x_2 = 1$ si può ottenere $x_1 = 1 - x_2$ e $x_2 = 1 - x_1$.

perciò si possono sostituire x_1^2 e x_2^2 con:

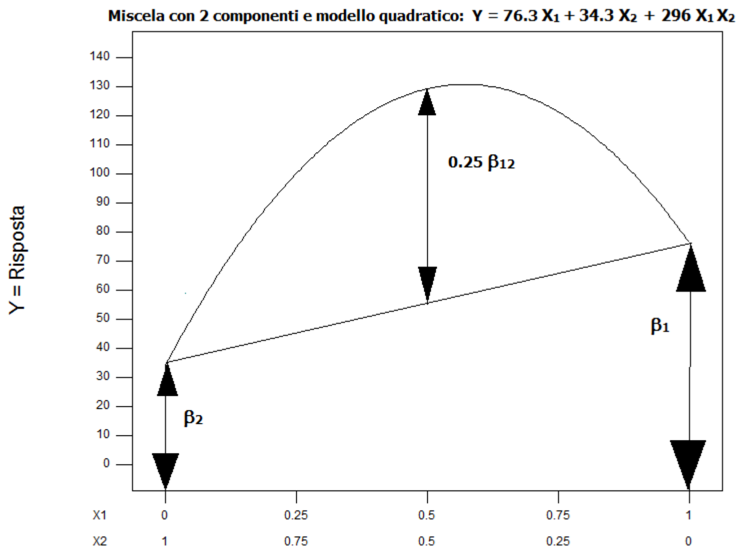
$$x_1^2 = x_1 x_1 = x_1(1 - x_2) = x_1 - x_1 x_2$$

$$x_2^2 = x_2 x_2 = x_2(1 - x_1) = x_2 - x_1 x_2$$

Per cui si ottiene:

$$\begin{aligned} \widehat{y} &= b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 \\ &= b_0(x_1 + x_2) + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1(1 - x_2) + b_{22} x_2(1 - x_1) \\ &= (b_0 + b_1 + b_{11}) x_1 + (b_0 + b_2 + b_{22}) x_2 + (b_{12} - b_{11} - b_{22}) x_1 x_2 \\ &= b_1^* x_1 + b_2^* x_2 + b_{12}^* x_1 x_2 \end{aligned}$$

dove $b_1^* = (b_0 + b_1 + b_{11})$ $b_2^* = (b_0 + b_2 + b_{22}),$ and $b_{12}^* = (b_{12} - b_{11} - b_{22}).$



Ai punti estremi, sinistro e destro, corrispondenti a $(x_1, x_2) = (0, 1)$ e $(1, 0)$, i valori di $\hat{y}$ sono rispettivamente β_1 e β_1.

Così, i coefficienti dei termini lineari rappresentano la risposta prevista per le miscele pure.

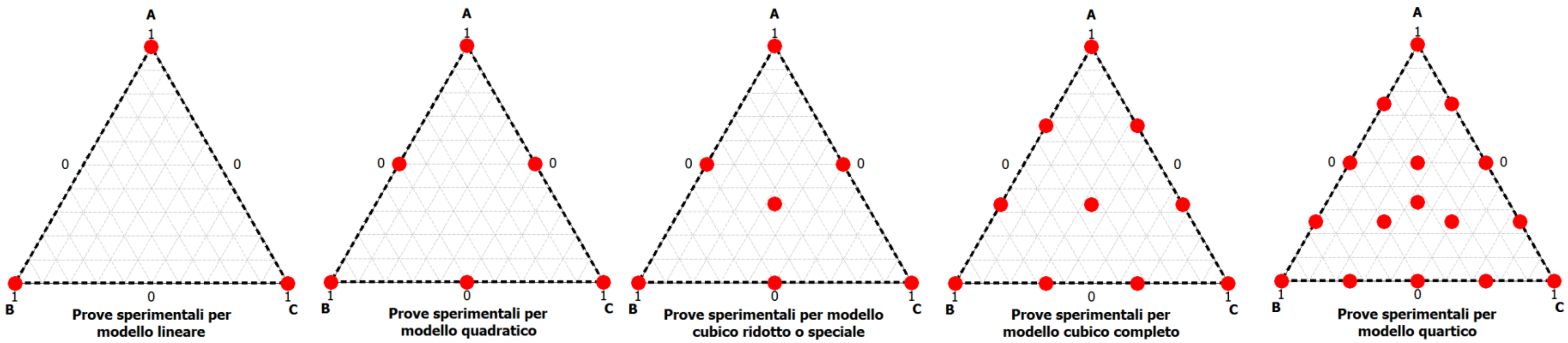
Le forme canoniche o di Scheffè dei modelli polinomiali per le miscele così derivate e maggiormente utilizzate, sono le seguenti:

Numero dei termini di un Polinomio Canonico:

Numero dei Componenti, p	Lineare	Quadratico	Cubico Speciale	Cubico Completo
2	2	3	—	—
3	3	6	7	10
4	4	10	14	20
5	5	15	25	35
6	6	21	41	56
7	7	28	63	84
8	8	36	92	120
⋮	⋮	⋮	⋮	⋮
p	p	$p(p+1)/2$	$p(p^2+5)/6$	$p(p+1)(p+2)/6$

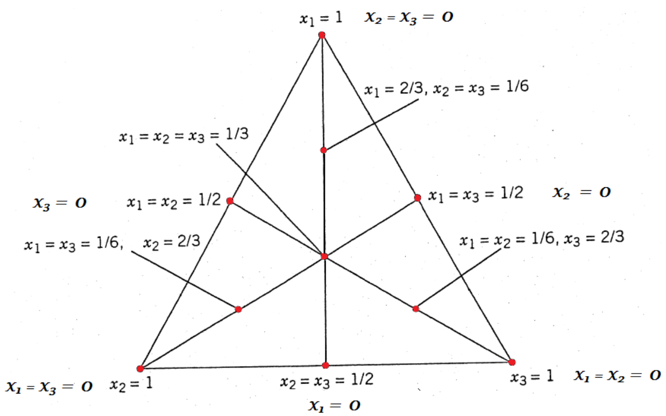
Di solito, nei modelli matematici per le miscele sono necessari termini di ordine superiore, poiché i fenomeni studiati possono essere complessi, e la regione sperimentale di solito è l'intera regione di esercizio, che essendo grande richiede pertanto un modello complesso.

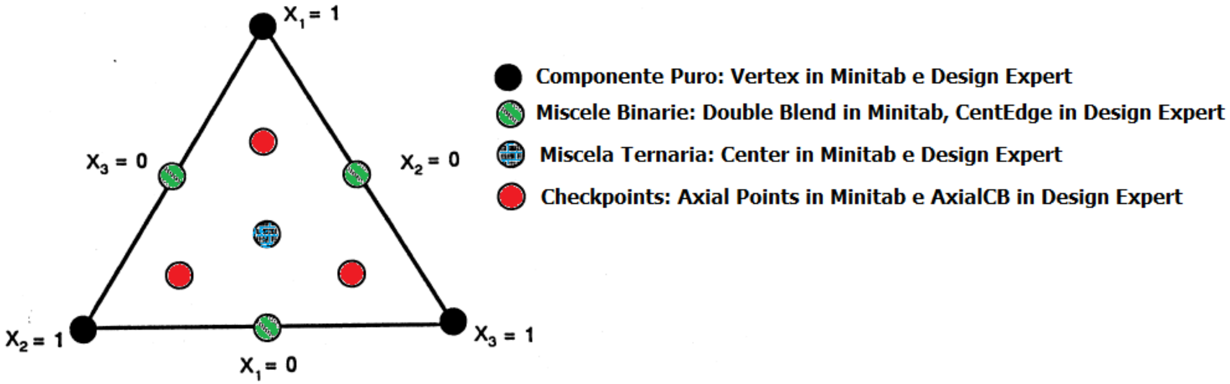
Nel caso di 3 componenti, gli esperimenti candidati in funzione del modello matematico scelto, sono:



Usualmente si desidera incrementare il simplesso, a griglia o con il baricentro, con punti aggiuntivi all'interno della regione sperimentale, in cui le miscele sono costituite da tutte le p componenti della miscela.

Un esempio di sistema delle coordinate in un piano simplesso a griglia aumentato, con 3 componenti, è il seguente:



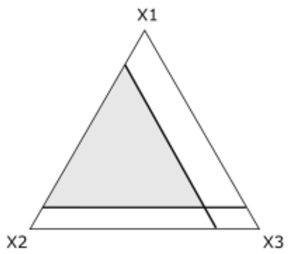


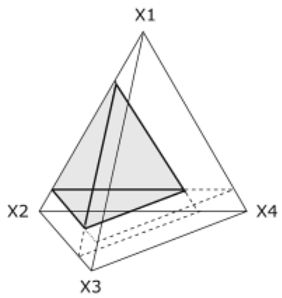
Disegno degli Esperimenti su Miscele per Simplesso con ulteriori vincoli

Frequentemente oltre al vincolo costituito dal fatto che la somma delle percentuali dei costituenti deve essere eguale a 1, nello studio di miscele occorre considerare altri vincoli.

Si consideri l'aggiunta di un secondo vincolo sul limite inferiore, diverso da zero, e le relative figure redatte con 3 e con 4 componenti:

$$\mathbf{0.1 \leq x_1}$$
$$\mathbf{0.2 \leq x_2}$$

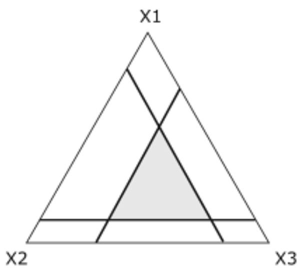




In entrambi i casi, le regioni vincolate grigie hanno ancora la sagoma di un simplesso.

In ultimo, si consideri l'aggiunta di un terzo vincolo sul limite inferiore, diverso da zero, e la relativa figura redatte con 3 componenti:

$$\mathbf{0.1 \leq x_1}$$
$$\mathbf{0.2 \leq x_2}$$
$$\mathbf{0.3 \leq x_3}$$



Di conseguenza, i disegni sperimentali su miscele, che presentano **vincoli solo sui limiti inferiori**, possono essere trattati come simplessi.

Introduzione dei vincoli sottintesi

Se, in una miscela a 3 componenti, il limite inferiore del componente x_1, $L_1 = 0.1$:

$$\mathbf{0.1 \leq x_1}$$

allora le proporzioni massime degli altri 2 componenti della miscela, U_2 e U_3, non possono essere maggiori di 0.9.
Qualunque proporzione di x_2 e x_3, maggiore di 0.9, è inaccessibile.
Se si impostassero limiti superiori su x_2 e x_3 uguali a 1, sarebbero dichiarati incongruenti.
Di consequenza, l'insieme completo dei vincoli dovrebbe essere così scritto:

Qualunque proporzione di x_1, maggiore di 0.8, è inaccessibile.
Qualunque proporzione di x_2, maggiore di 0.9, è inaccessibile.
Qualunque proporzione di x_3, maggiore di 0.7, è inaccessibile.

Se si impostassero limiti superiori su x_1, x_2 e x_3 uguali a 1, sarebbero dichiarati incongruenti.

Di consequenza, l'insieme completo dei vincoli dovrebbe essere così scritto:

$$\mathbf{0.1 \leq x_1 \leq 0.8}$$
$$\mathbf{0.2 \leq x_2 \leq 0.9}$$
$$\mathbf{0.0 \leq x_3 \leq 0.7}$$

Codifica dei Componenti nelle Miscele

Ci sono **3 metriche o scale** che possono essere utilizzate per la regolazione dei componenti delle miscele:

- **Actual** (Design Expert), **Amount** (Minitab): quantità fisica di un particolare ingrediente, espressa come percentuale o frazione della miscela totale. Corrisponde alla scala originale usata per immettere i livelli dei componenti durante la pianificazione del disegno sperimentale.

- **Real** (Design Expert), **Proportion** (Minitab): frazione dei componenti in relazione alla miscela totale, riportando la somma dei componenti uguale a 1.

$$\mathrm{Real_i} = \frac{\mathrm{Actual_i}}{\sum \mathrm{Actual_i}} \qquad \mathrm{Proportion_i} = \frac{\mathrm{Amount_i}}{\sum \mathrm{Amount_i}}$$

Se $\sum Actual_i = 1\ (100\%)$, $\sum Proportion_i = 1\ (100\%)$, **allora la scala Actual è identica alla scala Real, così come la scala Amount è identica alla scala Proportion.**

- **Pseudocomponent** (Design Expert e Minitab): ridefinisce le coordinate della regione sperimentale in termini similari alle variabili codificate (-1 e +1), utilizzate negli esperimenti fattoriali.

  La formula per ottenere la scala Pseudo è derivata dai componenti in scala Real (Design Expert) oppure in scala Proportion (Minitab).

  La scala Real (Design Expert) o Proportion (Minitab) risulta identica alla scala Pseudo, quando tutti i limiti inferiori dei componenti sono uguali a zero.

Nell'esempio riportato:

Miscela di 3 componenti, Piano Sperimentale Simplesso con Centroide, Modello Quadratico Aumentato, **somma totale = 0.8, unità actual o amount**.

$$\mathbf{0.2 \le X_1 \le 0.3}$$
$$\mathbf{0.4 \le X_2 \le 0.5}$$
$$\mathbf{0.1 \le X_3 \le 0.2}$$

Pseudocomponenti

Gli pseudocomponenti sono variabili codificate, utilizzate per semplificare la costruzione dei disegni sperimentali e per l'adattamento di modelli matematici.

Inoltre, riduce la correlazione tra i limiti dei componenti nei disegni vincolati.

I disegni vincolati (quelli in cui si specificano i limiti superiori e/o inferiori) producono coefficienti che sono altamente correlati.

I limiti inferiori sono necessari quando uno qualsiasi dei componenti deve essere presente nella miscela.

Ad esempio, l'aranciata deve contenere succo di arancia.

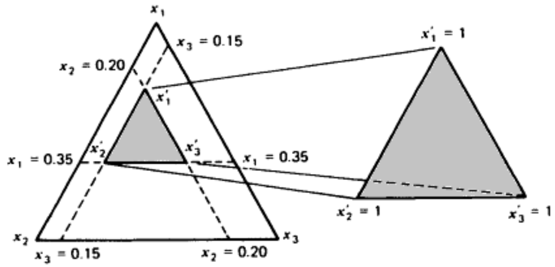
I limiti superiori sono necessari quando la miscela non può contenere più di una determinata percentuale di un ingrediente.
Ad esempio, una miscela di una torta non può contenere più del 5% lievito.

Di solito, è possibile ridurre le correlazioni tra i coefficienti trasformando i componenti in pseudo-componenti.

Con l'introduzione degli pseudo-componenti, in effetti, l'area dei dati vincolata viene ricodificata, quindi la quantità minima consentita (limite inferiore) di ciascun componente diventa zero.

Se ad esempio ci fossero i seguenti vincoli sui limiti inferiori dei componenti:

$$\mathbf{0.35 \le x_1 \le 0.65}$$
$$\mathbf{0.20 \le x_2 \le 0.50}$$
$$\mathbf{0.15 \le x_3 \le 0.45}$$



Se ad esempio ci fossero i seguenti vincoli sulle proporzioni dei componenti:

$$\mathbf{0.1 \le x_1 \le 0.4}$$
$$\mathbf{0.3 \le x_2 \le 0.6}$$
$$\mathbf{0.0 \le x_3 \le 0.3}$$

Vincoli inconsistenti

I vincoli vengono definiti inconsistenti quando:

$$U_i + \sum_{j \neq i}^{p} L_j > 1$$

$$L_i + \sum_{j \neq i}^{p} U_j < 1$$

Qui U_i, U_j, L_i e L_j sono espressi in **unità Real** (Design Expert) o in **unità Proportion** (Minitab).

$$U_i + \sum_{j \neq i}^{p} L_j > Total$$

$$L_i + \sum_{j \neq i}^{p} U_j < Total$$

Qui U_i, U_j, L_i e L_j sono espressi in **unità Actual** (Design Expert) o in **unità Amount** (Minitab).

Limiti inferiori inconsistenti

I limiti inferiori vengono definiti inconsistenti quando:

$$\sum_{i}^{p} L_i \geq 1, in\ unità\ Real\ (Design\ Expert)\ oppure\ Proportion\ (Minitab)$$

$$\sum_{i}^{p} L_i \geq Total, in\ unità\ Actual\ (Design\ Expert)\ oppure\ Amount\ (Minitab)$$

Limiti superiori inconsistenti

I limiti superiori vengono definiti inconsistenti quando:

Esempio L - Pseudo

Componente A: 0.4 – 0.5
Componente B: 0.2 – 0.3
Componente C: 0.3 – 0.4

Lati del Triangolo **Vertici del Triangolo**

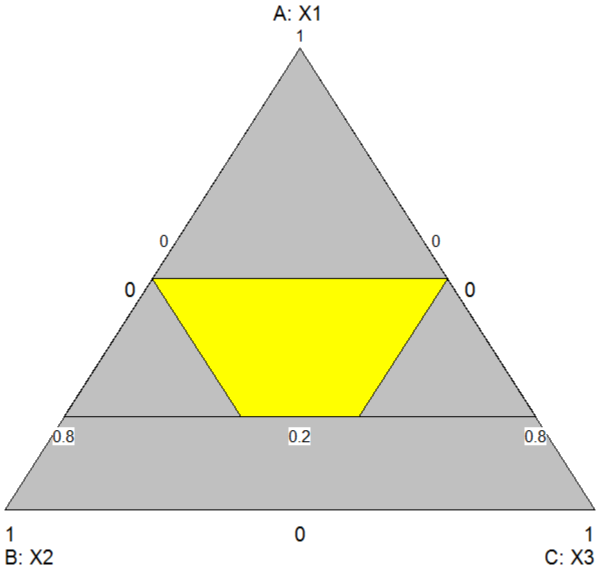
Se così non fosse, il dominio sperimentale è riconducibile ad un politopo irregolare.

Se i componenti hanno vincoli sia superiori, sia inferiori, la regione di ammissibilità non è più un simplesso, ma sarà un politopo irregolare.

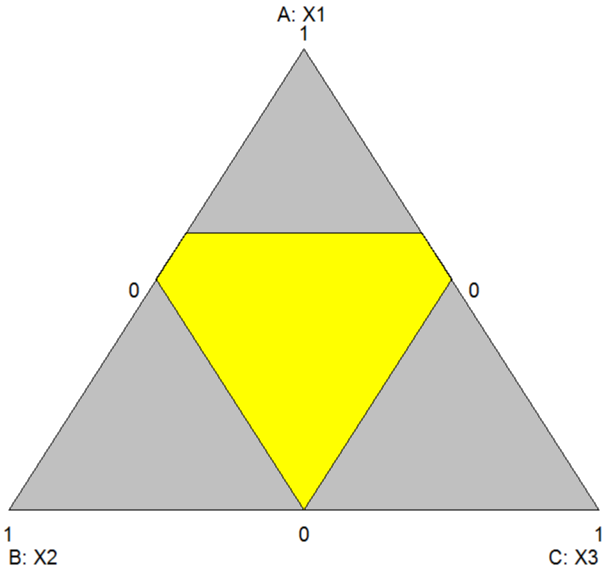
Poiché la regione sperimentale non ha una forma standard, per questi tipi di problematiche sulle miscele si utilizzano i **piani generati dal calcolatore.**

Esempi aggiuntivi

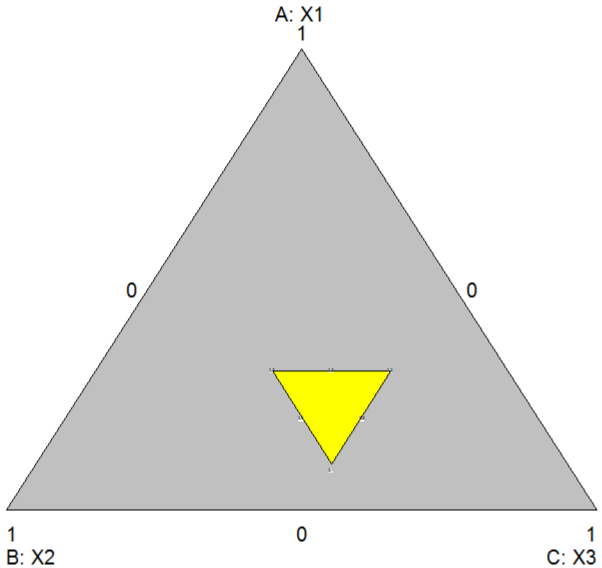
Si considerino questi tre esempi, con 3 componenti e relativi vincoli sperimentali:



Esempio 1



Esempio 2



Esempio 3

Come regolare i vincoli inconsistenti e renderli coerenti.

Primo esempio:

$R_1 = 0.1 = R_U = 0.1$, perciò $L(X_1)$ è accessibile;

$R_2 = 0.1 = R_U = 0.1$, perciò $L(X_2)$ è accessibile;

$R_3 = 0.2 > R_U = 0.1$, allora $L(X_3)$ è inaccessibile;

$R_1 = 0.1 < R_L = 0.3$, perciò $U(X_1)$ è accessibile;

$R_2 = 0.1 < R_L = 0.3$, perciò $U(X_2)$ è accessibile;

$R_3 = 0.2 < R_L = 0.3$, perciò $U(X_3)$ è accessibile.

Nuovo limite inferiore implicito: $L(X_3)^* = U(X_3) - R_U = 0.8 - 0.1 = 0.7$

Secondo esempio:

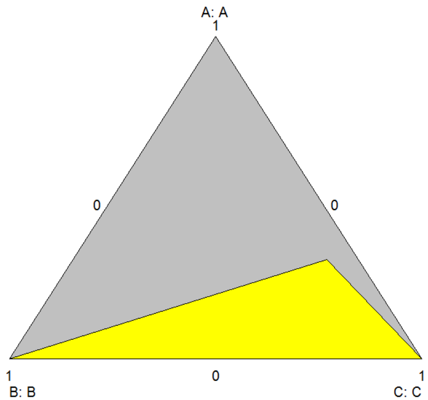
$0.3 \le X_1 \le 0.8$
$0.1 \le X_1 \le 0.5$ $\qquad \sum X_i = 1$
$0.2 \le X_1 \le 0.6$

Nuovo limite inferiore implicito: $U(X_1)^* = L(X_1) + R_L = 0.3 - 0.4 = 0.7$

Vincoli multipli lineari.

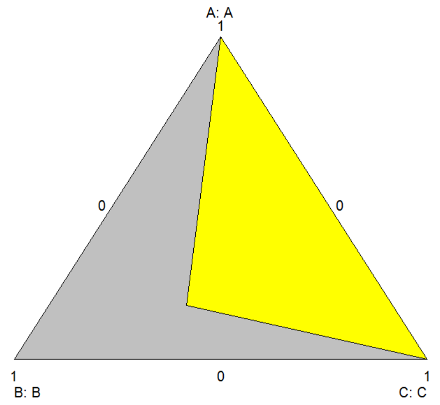
Primo esempio

Unendo i 2 vincoli lineari, si ottiene:



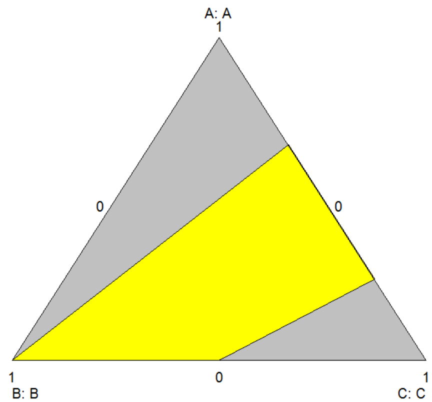
Secondo esempio

Unendo i 2 vincoli lineari, si ottiene:



Terzo esempio

Unendo i 2 vincoli lineari, si ottiene:



Nel disegno degli esperimenti per miscele, quando sui componenti si introducono:

- limiti, sia superiori, sia inferiori;
- vincoli lineari singoli e/o multipli;

che danno origine ad una **regione sperimentale non riconducibile ad un simplesso**, si deve creare il piano delle prove sperimentali attaverso il **Disegno Ottimale** (Design Expert) o **Extreme Vertices** (Minitab).

Progettazione degli esperimenti che combinano i componenti della miscela con i fattori di processo.

La strategia tipica per il disegno degli esperimenti nell'industria chimica è:

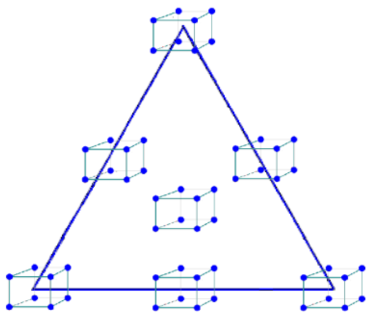
- la buona messa a punto della formulazione attraverso il disegno degli esperimenti con miscele;
- l' ottimizzazione del processo con il disegno degli esperimenti fattoriale e con la superficie di risposta.

Generalmente questi due step sono trattati separatamente dagli sperimentatori.

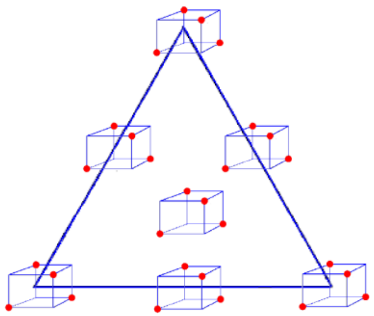
Tuttavia, le interazioni tra le variabili relative alla composizione ed i fattori di processo non possono essere scoperte tramite questo approccio così semplificato.

Si può condurre un esperimento completo, che combini i componenti della miscela con i fattori di processo, in un disegno degli esperimenti "incrociato" miscela-processo.

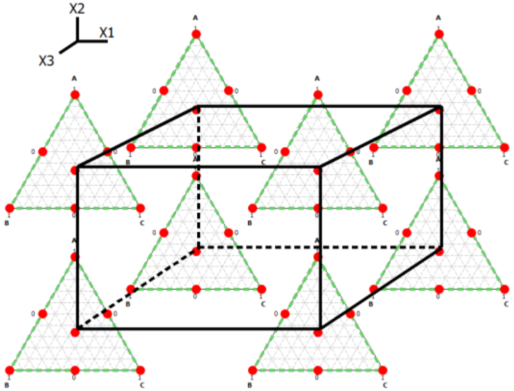
Rappresentazioni di domini sperimentali di piani prove congiunti, costituiti da miscele con **3 componenti e 3 fattori** di processo:



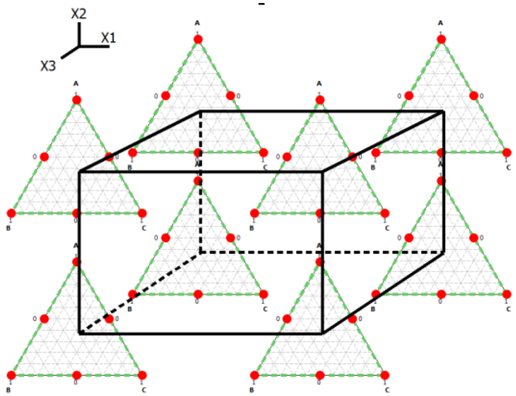
Disegno combinato di un simplesso con centroide e di un piano fattoriale completo, con **56** prove sperimentali.



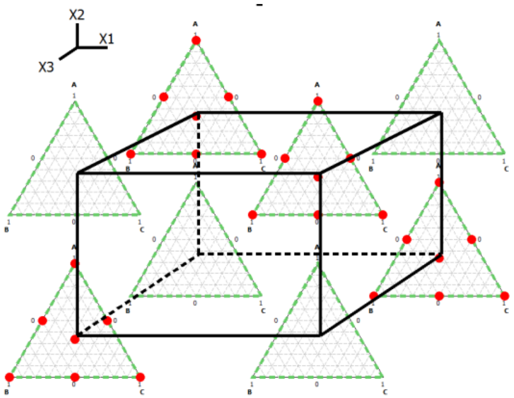
Disegno combinato di un simplesso con centroide e di un piano fattoriale frazionato, con **28** prove sperimentali



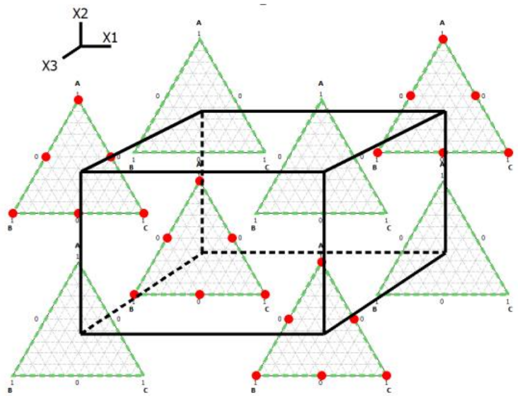
Disegno combinato di un piano fattoriale completo
con un simplesso con centroide, con **56** prove sperimentali.



Disegno combinato di un piano fattoriale completo con
un simplesso senza centroide, con **48** prove sperimentali



Disegno combinato di piano fattoriale frazionato
(frazione principale) con un simplesso
con centroide, con **28** prove sperimentali.



Disegno combinato di un piano fattoriale frazionato
(frazione secondaria) con un simplesso
senza centroide, con **24** prove sperimentali.

Il modello matematico per un disegno degli esperimenti combinato miscela-processo può essere molto strutturato, anche per uno studio relativamente semplice.

La **seconda linea** dell'equazione **rivela l'effetto lineare del primo fattore** di processo (Z_1), il quale **sposta la risposta media** ad ogni combinazione data, **dei componenti della miscela**.

La **terza linea** mostra **l'effetto lineare del secondo fattore** di processo (Z_2).

L'**ultima linea** dell'equazione rappresenta le **interazioni** fra i **fattori** di processo e la **miscela**.

Quando sono presenti termini del modello della miscela e del modello fattoriale di ordine superiore, il modello combinato sarà più complesso.

Se ad esempio si volesse incrociare il modello **cubico speciale** della miscela costituita da **3 componenti**:

$$y_{MC} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \varepsilon$$

con il modello polinomiale **quadratico** con **3 variabili** di processo:

$$y_{PV} = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + \alpha_{12} z_1 z_2 + \alpha_{13} z_1 z_3 + \alpha_{23} z_2 z_3 + \alpha_{11} z_1^2 + \alpha_{22} z_2^2 + \alpha_{22} z_3^2 + \varepsilon.$$

si otterrebbe il modello combinato miscela-processo costituita da **70 termini** (7 x 10):

$$\begin{aligned} y_{SC-Q} = {} & g_1^0 x_1 + g_2^0 x_2 + g_3^0 x_3 + g_{12}^0 x_1 x_2 + g_{13}^0 x_1 x_3 + g_{23}^0 x_2 x_3 + g_{123}^0 x_1 x_2 x_3 + (g_1^1 x_1 + g_2^1 x_2 + g_3^1 x_3 + g_{12}^1 x_1 x_2 + g_{13}^1 x_1 x_3 + g_{23}^1 x_2 x_3 + g_{123}^1 x_1 x_2 x_3) z_1 \\ & + (g_1^2 x_1 + g_2^2 x_2 + g_3^2 x_3 + g_{12}^2 x_1 x_2 + g_{13}^2 x_1 x_3 + g_{23}^2 x_2 x_3 + g_{123}^2 x_1 x_2 x_3) z_2 + (g_1^3 x_1 + g_2^3 x_2 + g_3^3 x_3 + g_{12}^3 x_1 x_2 + g_{13}^3 x_1 x_3 + g_{23}^3 x_2 x_3 + g_{123}^3 x_1 x_2 x_3) z_3 \\ & + (g_1^{12} x_1 + g_2^{12} x_2 + g_3^{12} x_3 + g_{12}^{12} x_1 x_2 + g_{13}^{12} x_1 x_3 + g_{23}^{12} x_2 x_3 + g_{123}^{12} x_1 x_2 x_3) z_1 z_2 + (g_1^{13} x_1 + g_2^{13} x_2 + g_3^{13} x_3 + g_{12}^{13} x_1 x_2 + g_{13}^{13} x_1 x_3 \\ & + g_{23}^{13} x_2 x_3 + g_{123}^{13} x_1 x_2 x_3) z_1 z_3 + (g_1^{23} x_1 + g_2^{23} x_2 + g_3^{23} x_3 + g_{12}^{23} x_1 x_2 + g_{13}^{23} x_1 x_3 + g_{23}^{23} x_2 x_3 + g_{123}^{23} x_1 x_2 x_3) z_2 z_3 + (g_1^{11} x_1 + g_2^{11} x_2 + g_3^{11} x_3 \\ & + g_{12}^{11} x_1 x_2 + g_{13}^{11} x_1 x_3 + g_{23}^{11} x_2 x_3 + g_{123}^{11} x_1 x_2 x_3) z_1^2 + (g_1^{22} x_1 + g_2^{22} x_2 + g_3^{22} x_3 + g_{12}^{22} x_1 x_2 + g_{13}^{22} x_1 x_3 + g_{23}^{22} x_2 x_3 \\ & + g_{123}^{22} x_1 x_2 x_3) z_2^2 + (g_1^{33} x_1 + g_2^{33} x_2 + g_3^{33} x_3 + g_{12}^{33} x_1 x_2 + g_{13}^{33} x_1 x_3 + g_{23}^{33} x_2 x_3 + g_{123}^{33} x_1 x_2 x_3) z_3^2 + \varepsilon \end{aligned}$$

Un disegno sperimentale che supporti l'adattamento di questo modello richiederebbe almeno 70 prove, con l'aggiunta di prove extra per valutare il Pure Error (errore sperimentale) e il Lack of Fit (mancanza di adattamento).

Una alternativa può essere quella di generare un disegno ottimale, con un determinato numero di punti sperimentali.

Disegno degli esperimenti per miscele: pianificazione prove in Design Expert.

Simplex Lattice Linear:

Simplex Lattice Quadratic:

Simplex Lattice Special Cubic:

Simplex Lattice Cubic:

Simplex Centroid Linear

Simplex Centroid Quadratic

Simplex Centroid Special Cubic

Piani (Ottimi) generati dal calcolatore

Ci sono 3 circostanze in cui i piani prove standard per superfici di risposta come:

Le 3 situazioni, per le quali un certo tipo disegno degli esperimenti, sviluppato dal calcolatore, può essere adatto, sono:

1) Regione sperimentale di interesse irregolare, talvolta denominata “Lattina Ammaccata”, dove a seguito dell’**applicazione di vincoli** sui livelli dei fattori o dei componenti, vengono rimossi in pratica dei vertici dallo spazio del poligono, poliedro o politopo.
Non esiste un piano standard a superficie di risposta che si adatti esattamente a questa regione.

2) Modello non standard. In genere, si tende a scegliere un modello a superficie di risposta di primo o secondo ordine, poiché questo modello empirico può essere una buona approssimazione del meccanismo vero.
A volte però si ha una particolare conoscenza o intuizione del processo in studio tale da prendere in considerazione un modello non standard, ad esempio con termini cubici.
In questo caso, sia il CCD, sia il BBD, non mettono a disposizione dati sufficienti per poter ottenere il modello di terzo ordine.
Inoltre, quando in aggiunta a fattori di tipo continuo, si lavora con alcune variabili categoriali o attributive, il piano prove generato dal calcolatore minimizza il numero delle prove sperimentali.

3) Richieste di dimensione campionarie insolite. Qualche volta si può avere la necessità di effettuare un numero di prove sperimentali minore di quello previsto da un disegno classico non saturato.
Se si volesse ad esempio accostare un modello matematico del secondo ordine con quattro variabili, il piano composito centrale richiederebbe un numero di prove sperimentali compreso tra 28 e 30, in funzione del numero dei punti centrali deciso.
Il modello matematico ha 15 termini, per cui se ci fossero dei vincoli economici o temporali, si potrebbe optare per un piano prove con un numero inferiore di esperimenti.

Sebbene i piani generati dal calcolatore possono essere utilizzati per questo scopo, di solito si può affrontare la situazione con protocolli migliori.
Ad esempio, si può costruire un piano composito ridotto con quattro fattori conducendo 20 prove sperimentali, compresi 4 punti centrali, oppure un piano ibrido con appena 16 prove.
Queste opzioni sono in genere migliori del piano prove generato dal calcolatore allo scopo di ridurre il numero delle prove sperimentali.

Molti risultati relativi ai piani generati al calcolatore rappresentano sviluppi del lavoro sulla teoria dei piani ottimi, intendendo che un piano è "ottimo" rispetto a qualche criterio.

Ai programmi per calcolatore viene richiesta la costruzione di questi piani.

L'approccio è il seguente:

- specificare un modello;
- determinare la regione di interesse;
- scegliere il numero di prove da condurre;
- specificare il criterio di ottimalità;
- scegliere i punti del piano da un insieme di punti candidati che l'utilizzatore potrebbe considerare.

Tipicamente i punti candidati sono una griglia di punti distribuiti in una regione sperimentale ammissibile.

Il punto di partenza della teoria del disegno sperimentale e dei metodi di regressione è rappresentato da una **matrice X,** che riassume le correlazioni o le covarianze tra la variabile dipendente (misurata per lo meno su una scala ad intervalli equivalenti), le variabili indipendenti (che possono essere quantitative oppure dicotomiche) e tra le variabili indipendenti stesse.

I metodi per generare piani prove ottimali si basano su considerazioni relative a questa matrice X.

X è la matrice (n x p) del modello, che ha tante righe quanti sono gli esperimenti (n) e tante colonne (p) quanti sono i termini del modello matematico di regressione che si vuole stimare.

La matrice **X'X** viene comunemente chiamata **matrice di informazione**.

La sua inversa, $\mathbf{(X'X)^{-1}}$, che è di estrema importanza sia per la teoria del disegno sperimentale che per i metodi di regressione, viene chiamata invece **matrice di dispersione** (o matrice varianza-covarianza).

La matrice di dispersione ha un'importanza fondamentale nella teoria del disegno sperimentale; infatti, gli elementi lungo la sua diagonale principale sono proporzionali, attraverso l'errore sperimentale, all'incertezza che si ha nella stima dei coefficienti del modello di regressione.

In pratica gli elementi diagonali della matrice di dispersione rendono conto della varianza dei coefficienti del modello.

La somma di tali elementi, ovvero la traccia della matrice, divisa per il numero p dei coefficienti, dà la varianza media dei coefficienti.
Invece, i termini fuori dalla diagonale contengono l'informazione riguardante la covarianza dei coefficienti del modello, cioè come si comporta un coefficiente se si commette un certo errore nella stima di un altro.

I termini extradiagonali informano quindi sull'indipendenza relativa delle stime dei coefficienti.

Quindi poiché il volume della regione è proporzionale all'incertezza sulla conoscenza dei coefficienti, per ottenere un'elevata precisione nella stima dei coefficienti del modello di regressione si devono scegliere gli esperimenti in modo che il determinante di **(X' X)**$^{-1}$ sia minimo.

Ci sono diversi criteri di ottimalità che sono stati proposti per scegliere un piano prove: A, D, G e V.

Questi criteri sono relazionati con la minimizzazione di varie funzioni della matrice di dispersione **(X'X)**$^{-1}$.

Quindi il piano G-ottimo minimizza la massima varianza di previsione.

V-ottimalità, chiamata anche -Q, -IV e I-ottimalità: considera il valore medio della varianza di previsione, definita dalla **media** di **$x'(X'X)^{-1}x\sigma^2$**, di uno specificato insieme di punti del piano, nella regione di interesse del piano, cioè si cerca tra le combinazioni di esperimenti possibili, quella con il valore minimo.

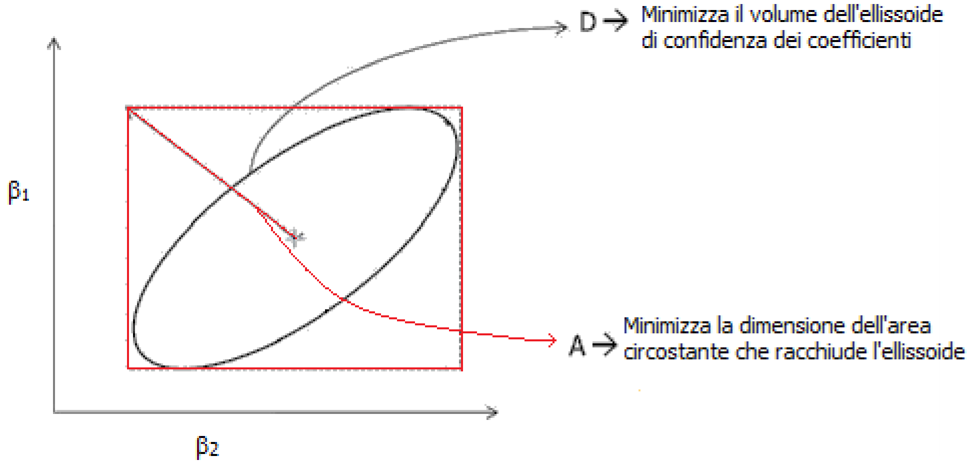
Quindi il piano V-ottimo minimizza la varianza media di previsione di questo insieme di punti.
In altre parole, i **criteri di ottimalità A e D sono indirizzati alle varianze delle stime dei parametri**, mentre i **criteri di ottimalità G e V sono indirizzati alla varianza dell'equazione di predizione.**

Considerando ogni criterio, la desiderabilità di un piano prove sperimentale aumenta al diminuire dei seguenti valori:

- **traccia di $(X'X)^{-1}$;**
- **determinante di $(X'X)^{-1}$;**
- **massimo di $x'(X'X)^{-1}x\sigma^2$;**
- **media di $x'(X'X)^{-1}x\sigma^2$.**

Riassumendo:

Disegno A-ottimale: minimizza la varianza media degli estimatori dei parametri del modello matematico;
Disegno D-ottimale: minimizza la varianza generalizzata degli estimatori dei parametri del modello matematico;
Disegno G-ottimale: minimizza la varianza massima dei valori predetti;
Disegno V-ottimale: minimizza la varianza media dei valori predetti.



Spesso l'insieme dei criteri enunciati e discussi per l'ottimalità viene chiamato con il nome di **"Criteri di Ottimalità Alfabetica"**

Vi sono situazioni in cui il piano di ottimalità alfabetica è noto o può essere costruito analitacamente.

Un buon esempio è il piano fattoriale a 2 livelli, 2^k , il quale è D -, A -, G - e V – ottimo per l'accostamento del modello lineare con k variabili o per l'accostamento del modello di primo ordine con interazioni.
Tuttavia nella maggior parte dei casi, il piano ottimo non è noto e si deve impiegare un algoritmo per trovarlo.

Molti software statistici, che si occupano di disegno degli esperimenti, hanno la capacità di realizzare questi piani sperimentali.

La ricerca della combinazione ottima non può essere effettuata in modo sistematico, dato il numero generalmente enorme di combinazioni.

La maggior parte delle procedure per la costruzione dei piani sono basate sull'**algoritmo di scambio**.

Nel caso si volesse applicare il criterio più ampiamente utilizzato di D-ottimalità:

1) si inizia con una scelta casuale degli esperimenti;
2) si calcola il determinante della matrice di informazione con gli esperimenti selezionati;
3) sempre casualmente si sceglie uno degli esperimenti selezionati in precedenza;
4) si sostituisce questo esperimento con un altro scelto casualmente (evitando duplicazioni);
5) se il determinante è maggiore di quello calcolato al punto 2, si aggiorna il complesso degli esperimenti selezionati e si torna al punto 2.

Tuttavia, i risultati sono generalmente soddisfacenti da un punto di vista pratico, e sono di gran lunga superiori rispetto a qualsiasi design ad hoc.

Piano D-ottimale

Si prenda in considerazione un disegno degli esperimenti per miscele con 2 componenti, per facilitare la comprensione del metodo dal punto di vista grafico.

Il modello matematico di Scheffè è:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Nella tabella vengono riportati 3 possibili piani sperimentali:

Disegno Sperimentale

A		B		C	
0.75	0.25	1.0	0.0	1.0	0.0
0.75	0.25	1.0	0.0	1.0	0.0
0.50	0.50	0.5	0.5	1.0	0.0
0.50	0.50	0.5	0.5	0.0	1.0
0.25	0.75	0.0	1.0	0.0	1.0
0.25	0.75	0.0	1.0	0.0	1.0

Il piano sperimentale A è costituito da 2 repliche dei punti axial check blend e 2 repliche del centroide.

Nel piano sperimentale B, i punti axial check blend sono stati spostati al bordo del simplesso, dove sono diventati vertici.

Nel piano sperimentale C, una replica del centroide è stata spostata verso un bordo, cioè è diventata un vertice, e l'altra replica del centroide è stata spostata all'altro bordo, cioè all'altro vertice.

Nella figura sono mostrate le regioni di confidenza congiunta al 95% dei 2 coefficienti del modello matematico, presupponendo β_1=10 e β_2= 20, con σ=0.25

L'ellisse più grande rappresenta la regione di confidenza congiunta al 95% del piano sperimentale A.

L’ellisse intermedia corrisponde alla regione del piano sperimentale B, mentre la regione più piccola corrisponde al piano sperimentale C.

Nella tabella che segue, viene mostrata la matrice del disegno A:

$$\begin{bmatrix} 0.75 & 0.25 \\ 0.75 & 0.25 \\ 0.50 & 0.50 \\ 0.50 & 0.50 \\ 0.25 & 0.75 \\ 0.25 & 0.75 \end{bmatrix}$$

In questa matrice, ogni riga rappresenta una osservazione, mentre le colonne rappresentano le proporzioni dei componenti.

Dato che, in questo specifico caso, si è interessati al modello matematico lineare, le matrice del disegno sperimentale è anche la **matrice del modello**, comunemente denominata matrice **X**.

Qualora si stesse pianificando un disegno sperimentale per adattare un modello matematico quadratico, allora la matrice del disegno dovrebbe essere aumentata di una colonna aggiuntiva, corrispondente al prodotto X_1X_2, e la matrice del modello **X** sarebbe costituita da 3 colonne.

Per un modello cubico, la matrice del modello **X** sarebbe costituita da 4 colonne.

Il volume della regione di confidenza congiunta per i coefficienti di regressione, cioè la varianza generalizzata, è inversamente proporzionale alla radice quadrata del determinante di X'X.

Si parte calcolando **X'X** e si calcola poi il determinante.

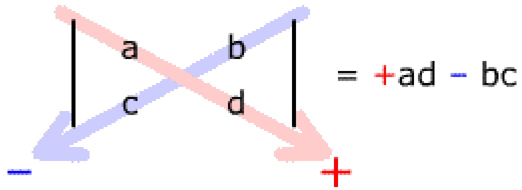
La matrice del modello **X** è di dimensioni n x p, dove n = numero delle osservazioni, e p = numero dei coefficienti del modello matematico.

La matrice trasposta di **X**, cioè **X'**, è perciò di dimensioni p x n.

Come risultato, la matrice informazione **X'X** è sempre una matrice quadrata di dimensioni p x p.

Per il piano sperimentale A:

Una regola mnemonica, nel caso di una matrice 2x2, per ricordare la definizione è illustrata nella figura riportata qui di seguito:

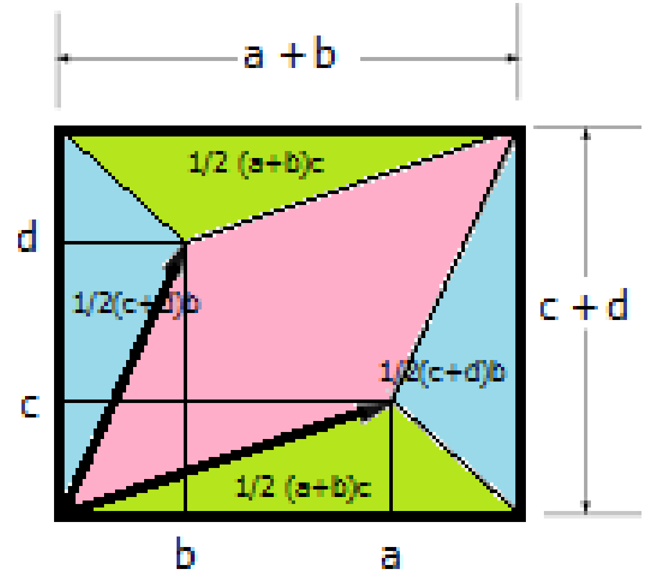


Il determinante della matrice si ottiene prendendo, con il segno +, il prodotto degli elementi sulla diagonale principale (in rosso) e sottraendo ad esso il prodotto degli elementi sulla diagonale opposta (in blu).

La tabella di seguito riassume la matrice informazione **X'X**, il determinante **|X'X|**, l'area dell'ellisse **$(|X'X|)^{-1}$** e l'area relativa dell'ellisse dei piani sperimentali A, B e C, come già visto nel grafico sovrastante che riporta le aree delle ellissi in funzione dei 3 differenti piani prove sperimentali.

Disegno	X'X	|X'X|	$(\|X'X\|)^{-1}$	Aree Relative
A	$\begin{bmatrix} 1.75 & 1.25 \\ 1.25 & 1.75 \end{bmatrix}$	1.5	$1/\sqrt{1.5} = 0.816$	1.0
B	$\begin{bmatrix} 2.5 & 0.5 \\ 0.5 & 2.5 \end{bmatrix}$	6.0	$1/\sqrt{6} = 0.408$	0.5
C	$\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$	9.0	$1/\sqrt{3} = 0.3$	0.41

Poiché il determinante gioca un ruolo importante nell'ottimalità di un disegno sperimentale, vale la pena spendere alcune parole per spiegare l'interpretazione geometrica del determinante.



Il determinante è uguale all'area del parallelogramma.

Può essere ottenuto sottraendo l'area dei 4 triangoli dall'area del rettangolo.

Le aree dei 2 triangoli di colore verde, in alto e in basso, sono uguali e corrispondono a **1/2 (a + b) c.**

Le aree dei 2 triangoli di colore azzurro, a destra e a sinistra, sono uguali e corrispondono a **1/2 (c + d) b**.

L'area del parallelogramma, di colore rosa, risulta, perciò: **(a+b)(c+d) - (a+b)c - (c+d)b = ad - bc**

Il determinante di una matrice 3 x 3 sarebbe il volume di un parallelepipedo.

dove:
p = numero dei termini del modello matematico;
N_D = numero delle osservazioni;
σ^2_{max} = valore massimo di $\mathbf{N\ x'(X'X)^{-1}\ x}$, varianza della predizione del gruppo di punti candidati;
σ^2_{ave} = valore medio di $\mathbf{N\ x'(X'X)^{-1}\ x}$, varianza della predizione del gruppo di punti candidati.

Questi valori di efficienza misurano la bontà di un disegno sperimentale, in relazione all'ipotesi di un ipotetico piano ortogonale che potrebbe non esistere, perciò non sono utili come misure assolute dell'efficienza di un disegno sperimentale, ma sarebbero da usare in modo relativo, per confrontare piani prove sperimentali nelle stesse condizioni.

Ad esempio, quando un valore **D-efficiency è 0%, uno o più parametri non possono essere stimati**.

Quando un valore **D-efficiency è 100%, allora il disegno sperimentale è bilanciato e ortogonale**.

Valori intermedi indicano che tutti i parametri possono essere stimati, ma con una precisione minore dell'ottimale

I disegni degli esperimenti fattoriali completi e frazionari sono sia ortogonali, sia bilanciati.

Un disegno sperimentale è bilanciato quando ogni livello si presenta un ugual numero di volte entro ciascun fattore, il che significa che l'intercetta è ortogonale ad ogni effetto.

Quando ogni coppia di fattori presenta livelli che compaiono lo stesso numero di volte attraverso tutte le coppie di fattori, il disegno è ortogonale.

Più in generale, un disegno è ortogonale quando le frequenze delle coppie dei livelli sono proporzionali o uguali.