

Disegni Frazionari

Giorgio Marrubini e Camillo Melzi

06 giugno 2020

All'aumentare del numero k dei fattori, il numero degli esperimenti da eseguire in un disegno fattoriale completo (a 2 livelli) aumenta esponenzialmente come 2^k .

In [Tabella 1](#) è riportato il numero di esperimenti richiesti in funzione del numero di fattori k nei piani sperimentali fattoriali completi.

Tabella 1: Esperimenti richiesti per disegni fattoriali completi 2^k

Fattori (k)	Esperimenti (2^k)
4	16
5	32
6	64
7	128
8	256
9	512

E' possibile ridurre il numero di esperimenti, riducendolo di $\frac{1}{2}, \frac{1}{4}, \dots$, costruendo a partire da un disegno fattoriale completo 2^k un disegno fattoriale frazionario 2^{k-p} pur di accettare di "confondere" tra loro alcuni termini del modello. In generale, la strategia per fare questo consiste nel cercare di "confondere" termini di ordine maggiore che possono essere considerati trascurabili a priori, secondo il principio empirico della economia degli effetti (v. Glossario).

Vediamo come possiamo costruire un disegno frazionario 2^{k-p} con un esempio specifico. Supponiamo di voler costruire il disegno 2^{5-2} ossia $1/4$ del disegno fattoriale completo 2^5 . Partiamo dal disegno 2^3 , vedi [Tabella 2](#)

Tabella 2: Disegno fattorialo completo 2^3

x1	x2	x3
-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
1	-1	1
-1	1	1
1	1	1

a cui dobbiamo aggiungere i fattori mancanti x_4 e x_5 "confondendoli" con le interazioni $x_4 = x_1x_2$ e $x_5 = x_1x_3$. Otteniamo così il disegno [Tabella 3](#) in cui la quarta colonna risulta il prodotto della prima con seconda e la quinta il prodotto tra la prima e la terza

Tabella 3: Disegno frazionario 2^{5-2}

x1	x2	x3	x4	x5
-1	-1	-1	1	1
1	-1	-1	-1	-1
-1	1	-1	-1	1
1	1	-1	1	-1
-1	-1	1	1	-1
1	-1	1	-1	1
-1	1	1	-1	-1
1	1	1	1	1

Diremo che $x_4 = x_1x_2$ e $x_5 = x_1x_3$ sono i generatori del disegno frazionario 2^{5-2} .

Nell'applicativo, nel menù *Frazionario*, vengono costruiti i disegni frazionari 2^{k-p} indicando il numero di fattori k e il numero di generatori p .

In **Figura 1** abbiamo il risultato che si ottiene per $k = 5$ e $p = 2$, ossia il disegno 2^{5-2} .



Figura 1: Disegno frazionario 2^{5-2}

Si noti che i generatori sono indicati con le lettere maiuscole, e che queste corrispondono alle colonne della matrice disposte in ordine alfabetico (A è la prima colonna, B la seconda, C la terza, D la quarta ed E la quinta colonna)

Poiché ogni colonna della matrice del disegno elevata al quadrato è la colonna cosiddetta *identità*, *Int.*, costituita solo da valori uguali a 1, dalle relazioni $x_4 = x_1x_2$ e $x_5 = x_1x_3$ si ottiene facilmente che

$$Int. = x_1x_2x_4 \quad \text{e} \quad Int. = x_1x_3x_5$$

dette *Relazioni di identità*. La lunghezza (ordine delle interazioni) minima delle relazioni d'identità è chiamata *risoluzione* del disegno e, di solito, è indicata con un numero romano. Nel nostro esempio la risoluzione è *III* e il disegno viene indicato con 2^{5-2}_{III} .

In linea di principio conviene scegliere relazioni di risoluzione massima (V e superiore) in quanto confondono termini di ordine maggiore. Questa di solito è l'indicazione data dalla maggior parte dei software professionali commerciali (v. Figura 2). Ma affinando l'esperienza e la pratica nell'uso dei fattoriali frazionari si constateranno le notevoli potenzialità offerte anche da disegni di ordine III e IV.

Nella **Figura 2**, è riportata la copia della prima schermata da Design Expert®: sono indicati i disegni frazionari che si possono costruire in funzione del numero di fattori (per colonna) e numero di esperimenti (per riga) con la relativa risoluzione.

		Number of Factors																		
Runs		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	4	2 ²	2 ³⁻¹ _{III}																	
8			2 ³	2 ⁴⁻¹ _{IV}	2 ⁵⁻² _{III}	2 ⁶⁻³ _{III}	2 ⁷⁻⁴ _{III}													
16				2 ⁴	2 ⁵⁻¹ _V	2 ⁶⁻² _{IV}	2 ⁷⁻³ _{IV}	2 ⁸⁻⁴ _{IV}	2 ⁹⁻⁵ _{III}	2 ¹⁰⁻⁶ _{III}	2 ¹¹⁻⁷ _{III}	2 ¹²⁻⁸ _{III}	2 ¹³⁻⁹ _{III}	2 ¹⁴⁻¹⁰ _{III}	2 ¹⁵⁻¹¹ _{III}					
32					2 ⁵	2 ⁶⁻¹ _{VI}	2 ⁷⁻² _{IV}	2 ⁸⁻³ _{IV}	2 ⁹⁻⁴ _{IV}	2 ¹⁰⁻⁵ _{IV}	2 ¹¹⁻⁶ _{IV}	2 ¹²⁻⁷ _{IV}	2 ¹³⁻⁸ _{IV}	2 ¹⁴⁻⁹ _{IV}	2 ¹⁵⁻¹⁰ _{IV}	2 ¹⁶⁻¹¹ _{IV}	2 ¹⁷⁻¹² _{III}	2 ¹⁸⁻¹³ _{III}	2 ¹⁹⁻¹⁴ _{III}	2 ²⁰⁻¹⁵ _{III}
64						2 ⁶	2 ⁷⁻¹ _{VII}	2 ⁸⁻² _V	2 ⁹⁻³ _{IV}	2 ¹⁰⁻⁴ _{IV}	2 ¹¹⁻⁵ _{IV}	2 ¹²⁻⁶ _{IV}	2 ¹³⁻⁷ _{IV}	2 ¹⁴⁻⁸ _{IV}	2 ¹⁵⁻⁹ _{IV}	2 ¹⁶⁻¹⁰ _{IV}	2 ¹⁷⁻¹¹ _{IV}	2 ¹⁸⁻¹² _{IV}	2 ¹⁹⁻¹³ _{IV}	2 ²⁰⁻¹⁴ _{IV}
128							2 ⁷	2 ⁸⁻¹ _{VIII}	2 ⁹⁻² _{IV}	2 ¹⁰⁻³ _{IV}	2 ¹¹⁻⁴ _{IV}	2 ¹²⁻⁵ _{IV}	2 ¹³⁻⁶ _{IV}	2 ¹⁴⁻⁷ _{IV}	2 ¹⁵⁻⁸ _{IV}	2 ¹⁶⁻⁹ _{IV}	2 ¹⁷⁻¹⁰ _{IV}	2 ¹⁸⁻¹¹ _{IV}	2 ¹⁹⁻¹² _{IV}	2 ²⁰⁻¹³ _{IV}
256								2 ⁸	2 ⁹⁻¹ _{IX}	2 ¹⁰⁻² _{VI}	2 ¹¹⁻³ _{VI}	2 ¹²⁻⁴ _{VI}	2 ¹³⁻⁵ _V	2 ¹⁴⁻⁶ _V	2 ¹⁵⁻⁷ _V	2 ¹⁶⁻⁸ _V	2 ¹⁷⁻⁹ _V	2 ¹⁸⁻¹⁰ _{IV}	2 ¹⁹⁻¹¹ _{IV}	2 ²⁰⁻¹² _{IV}
512									2 ⁹	2 ¹⁰⁻¹ _X	2 ¹¹⁻² _{VII}	2 ¹²⁻³ _{VI}	2 ¹³⁻⁴ _{VI}	2 ¹⁴⁻⁵ _{VI}	2 ¹⁵⁻⁶ _{VI}	2 ¹⁶⁻⁷ _{VI}	2 ¹⁷⁻⁸ _{VI}	2 ¹⁸⁻⁹ _{VI}	2 ¹⁹⁻¹⁰ _V	2 ²⁰⁻¹¹ _V

Figura 2: Disegni frazionari con risoluzione

Per la ragione detta sopra, i programmatori di Design Expert® hanno assegnato dei codici-colore “semaforici” ai diversi disegni frazionari. Il colore del riquadro corrisponde al rischio di ottenere risultati inconcludenti. Perciò disegni con risoluzione *III*, in cui si “confondono” i termini lineari con le interazioni di ordine 2 sono colorati in rosso (rischio/attenzione alti); i disegni di risoluzione *IV* in cui si “confondono” i termini lineari con le interazioni di ordine 3, e le interazioni di ordine 2 sono confuse a coppie tra loro, sono in giallo (rischio/attenzione medi). In verde, invece, sono indicati i disegni di risoluzione superiore a V (via libera, nessuna attenzione?). Questo tipo di classificazione è discutibile, come detto, ed è da considerare al pari di un consiglio di prudenza, peraltro scontato, perché l'uso dei fattoriali frazionari è da pensare per risolvere il problema chimico in esame, e non viceversa (i.e. come adattare il problema ad un disegno frazionario di risoluzione sufficientemente alta). La **Figura 2** è la prima immagine del menù *Frazionari* dell'applicativo.

Con semplice algebra, sempre osservando che $x_i^2 = Int.$, i.e. ogni colonna al quadrato è la colonna identità *Int.* formata da tutti 1, si ottengono tutte le altre “confusioni”.

Nell'applicativo compaiono automaticamente sia il modello relativo al disegno sia tutte le “confusioni”, vedi **Figura 3**

Confusioni	
Modello	
$y \sim 1 + x1 + x2 + x3 + x4 + x5 + x2:x3 + x3:x4$	
	$x1 = x3:x5 = x1:x2:x3:x4:x5 = x2:x4$ $x2 = x3:x4:x5 = x1:x2:x3:x5 = x1:x4$ $x3 = x1:x5 = x2:x4:x5 = x1:x2:x3:x4$ $x4 = x2:x3:x5 = x1:x3:x4:x5 = x1:x2$ $x5 = x2:x3:x4 = x1:x2:x4:x5 = x1:x3$ $x2:x3 = x4:x5 = x1:x3:x4 = x1:x2:x5$ $x3:x4 = x2:x5 = x1:x2:x3 = x1:x4:x5$

Figura 3: Modello e confusioni del disegno frazionario 2^{5-2}_{III}

Per quanto riguarda la rimanente parte di output, tutto è presentato nella stessa logica già vista per i disegni fattoriali completi.

Esempio: confusioni

Per comprendere meglio le “confusioni” in un disegno frazionario consideriamo il seguente esempio didattico. Supponiamo che il modello “vero” (quello che a priori non conosciamo) del fenomeno di studio sia il seguente

$$y = x_1 + 5x_2 - 3x_3 + 15x_1x_3 + \epsilon$$

Nella progettazione degli esperimenti per studiare il fenomeno in osservazione abbiamo ipotizzato che la risposta dipenda da 5 fattori x_1, x_2, x_3, x_4, x_5 .

Per studiare tutte le possibili interazioni consideriamo un disegno fattoriale completo [Tabella 4](#)

Tabella 4: Disegno fattoriale completo 2^5 (32 esperimenti) con risposte costruito dal modello “vero” ipotizzato $y = x_1 + 5x_2 - 3x_3 + 15x_1x_3 + \epsilon$

x1	x2	x3	x4	x5	y
-1	-1	-1	-1	-1	11.96
1	-1	-1	-1	-1	-17.15
-1	1	-1	-1	-1	23.23
1	1	-1	-1	-1	-5.04
-1	-1	1	-1	-1	-22.91
1	-1	1	-1	-1	5.90
-1	1	1	-1	-1	-12.81
1	1	1	-1	-1	18.18
-1	-1	-1	1	-1	11.56
1	-1	-1	1	-1	-17.82
-1	1	-1	1	-1	20.86
1	1	-1	1	-1	-6.44
-1	-1	1	1	-1	-24.14
1	-1	1	1	-1	9.41
-1	1	1	1	-1	-13.74
1	1	1	1	-1	16.17
-1	-1	-1	-1	1	11.64
1	-1	-1	-1	1	-14.71
-1	1	-1	-1	1	20.62
1	1	-1	-1	1	-5.95
-1	-1	1	-1	1	-23.92
1	-1	1	-1	1	7.07
-1	1	1	-1	1	-13.31
1	1	1	-1	1	17.66
-1	-1	-1	1	1	11.69
1	-1	-1	1	1	-15.81
-1	1	-1	1	1	21.73
1	1	-1	1	1	-4.59
-1	-1	1	1	1	-24.53
1	-1	1	1	1	6.81
-1	1	1	1	1	-13.02
1	1	1	1	1	18.17

Il cui grafico dei coefficienti è dato in [Figura 4](#).

Per ridurre della metà il numero di esperimenti si considera un disegno frazionario 2_{IV}^{5-1} , [Tabella 5](#)

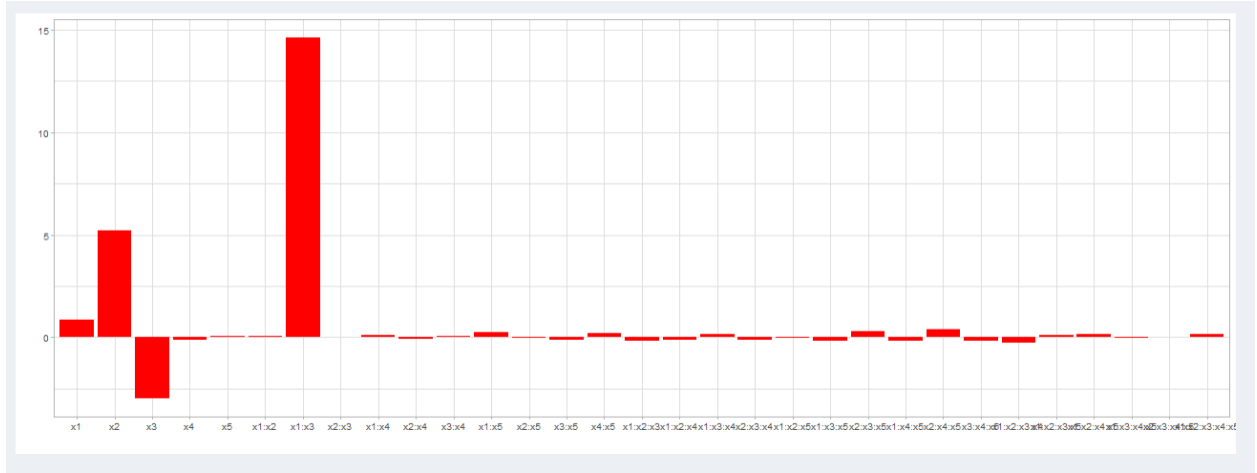


Figura 4: Grafico dei coefficienti del modello 2^5

Tabella 5: Disegno frazionario 2^{5-1}_V (16 esperimenti) con risposte costruito dal modello supposto $y = x_1 + 5x_2 - 3x_3 + 15x_1x_3 + \epsilon$

x1	x2	x3	x4	x5	y
-1	-1	-1	-1	1	11.64
1	-1	-1	-1	-1	-17.15
-1	1	-1	-1	-1	23.23
1	1	-1	-1	1	-5.95
-1	-1	1	-1	-1	-22.91
1	-1	1	-1	1	7.07
-1	1	1	-1	1	-13.31
1	1	1	-1	-1	18.18
-1	-1	-1	1	-1	11.56
1	-1	-1	1	1	-15.81
-1	1	-1	1	1	21.73
1	1	-1	1	-1	-6.44
-1	-1	1	1	1	-24.53
1	-1	1	1	-1	9.41
-1	1	1	1	-1	-13.74
1	1	1	1	1	18.17

In questo caso il grafico dei parametri è dato da [Figura 5](#)

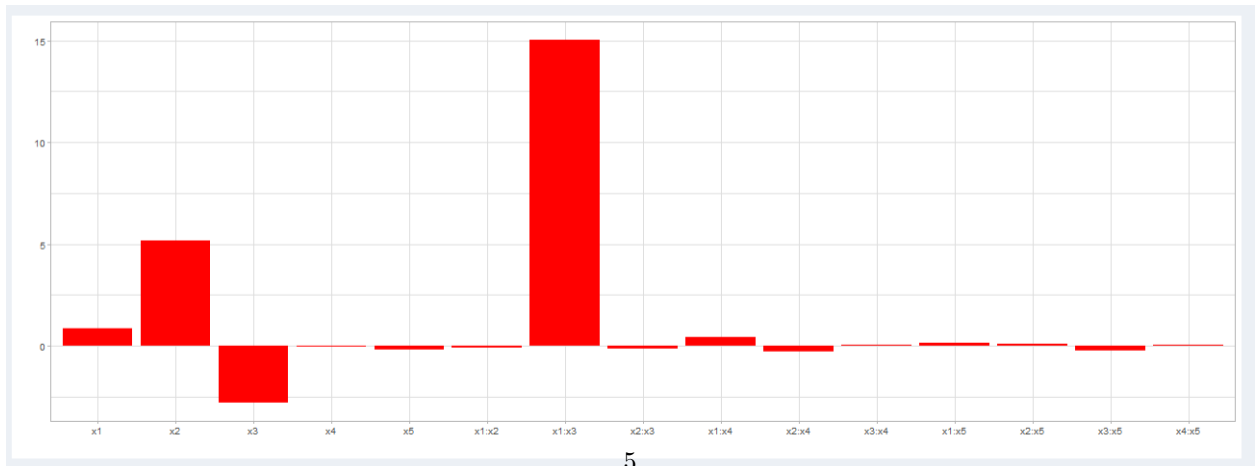


Figura 5: Grafico dei coefficienti del modello frazionario 2^{5-1}_V

Il disegno scelto ha risoluzione V , ciò significa che i termini lineari sono confusi con le interazioni di ordine 4 e quindi possiamo supporre che il valore di ciascun parametro si riferisca al termine lineare corrispondente (consideriamo trascurabili tutte le interazioni di ordine 4). E' possibile fare un ragionamento analogo per le interazioni di ordine 2 che si confondono con le interazioni di ordine 3. Supponendo che queste ultime siano trascurabili, possiamo dunque concludere che il valore di ciascun parametro di interazione si riferisca esclusivamente alle interazioni di ordine 2.

Se si volesse diminuire ulteriormente il numero di esperimenti, è possibile ricorrere ad un disegno frazionario 2_{III}^{5-2} , **Tabella 6**

Tabella 6: Disegno frazionario 2_{III}^{5-2} (8 esperimenti) con risposte costruito dal modello supposto $y = x_1 + 5x_2 - 3x_3 + 15x_1x_3 + \epsilon$

x1	x2	x3	x4	x5	y
-1	-1	-1	1	1	11.69
1	-1	-1	-1	-1	-17.15
-1	1	-1	-1	1	20.62
1	1	-1	1	-1	-6.44
-1	-1	1	1	-1	-24.14
1	-1	1	-1	1	7.07
-1	1	1	-1	-1	-12.81
1	1	1	1	1	18.17

Questo è un disegno di risoluzione III , bisogna quindi fare molta attenzione alle confusioni **Figura 6**

Confusioni	
x_1	$x_2:x_4 = x_3:x_5 = x_1:x_2:x_3:x_4:x_5$
x_2	$x_1:x_4 = x_3:x_4:x_5 = x_1:x_2:x_3:x_5$
x_3	$x_1:x_5 = x_2:x_4:x_5 = x_1:x_2:x_3:x_4$
x_4	$x_1:x_2 = x_2:x_3:x_5 = x_1:x_3:x_4:x_5$
x_5	$x_1:x_3 = x_2:x_3:x_4 = x_1:x_2:x_4:x_5$
$x_2:x_3$	$x_4:x_5 = x_1:x_3:x_4 = x_1:x_2:x_5$
$x_3:x_4$	$x_2:x_5 = x_1:x_2:x_3 = x_1:x_4:x_5$

Figura 6: Confusioni del disegno 2_{III}^{5-2}

I termini lineari si confondono con le interazione di ordine 2. Come si vede dal grafico dei coefficienti **Figura 7** ad esempio il valore di x_5 è circa 15, ma non siamo in grado di stabilire se è dovuto dal termine lineare x_5 , dal termine “confuso” x_1x_3 (come in questo caso, ricordo la forma del modello $y = x_1 + 5x_2 - 3x_3 + 15x_1x_3 + \epsilon$) o dalla combinazione di entrambi.

Esempio: studio dei fattori dell'estrazione liquido-liquido

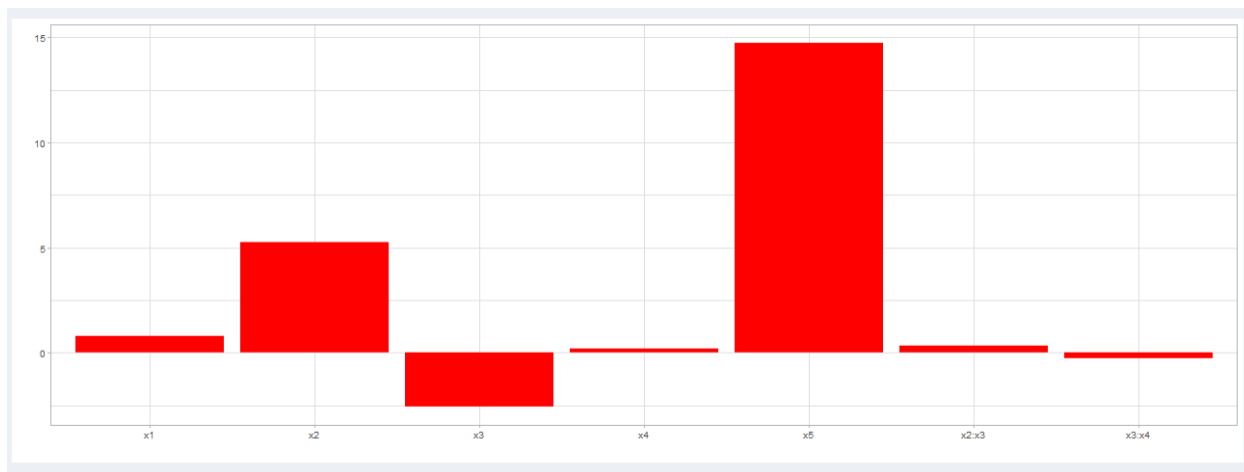


Figura 7: Grafico dei coefficienti del disegno 2^{5-2}_{III}

Per meglio comprendere i disegni frazionari e l'utilizzo dell'applicativo in questi disegni consideriamo il seguente esempio.

Si vogliono studiare i seguenti 4 fattori:

- volume miscela solventi per estrazione
- tempo di centrifuga dell'estratto
- forza ionica del campione (quantità di NaCl da aggiungere)
- tempo di estrazione

Definiamo innanzitutto il dominio sperimentale. Per ogni fattore determiniamo l'intervallo di valori compreso tra un massimo e un minimo entro i quali studiare il fenomeno, [Tabella 7](#)

Tabella 7: Definizione dei livelli

Fattori	-1	+1
vol. solvente	10	40
t. centrifuga	5	20
forza ionica	1	5
t. estrazione	1	5

Il piano fattoriale completo prevede 16 esperimenti. L'impegno del laboratorio è considerato troppo oneroso. Si decide quindi di utilizzare un disegno frazionario 2^{4-1}_{IV} per eseguire la metà degli esperimenti. Vedi [Figura 8](#)

n° fattori

n° generatori

Generatori

generators
"D=ABC"

Disegno

Exp#	x1	x2	x3	x4
1	-1	-1	-1	-1
2	1	-1	-1	1
3	-1	1	-1	1
4	1	1	-1	-1
5	-1	-1	1	1
6	1	-1	1	-1

Tabella 8: Piano sperimentale

vol. solvente	t. centrifuga	forza ionica	t. estrazione
10	5	1	1
40	5	1	5
10	20	1	5
40	20	1	1
10	5	5	5
40	5	5	1
10	20	5	1
40	20	5	5

Il modello e le “confusioni” sono dati in [Figura 9](#)

Modello	Confusioni
$y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_1x_2 + x_1x_3 + x_2x_3$	$x_1 = x_2x_3x_4$ $x_2 = x_1x_3x_4$ $x_3 = x_1x_2x_4$ $x_4 = x_1x_2x_3$ $x_1x_2 = x_3x_4$ $x_1x_3 = x_2x_4$ $x_2x_3 = x_1x_4$

Figura 9: Modello e condusioni del frazionario di 2_{IV}^{4-1}

Il disegno è di risoluzione IV quindi, come si vede in [Figura 9](#), abbiamo “confusione” tra i termini lineari e le interazioni di ordine 3 e tra le coppie di interazioni di ordine 2.

Vengono quindi eseguiti, in ordine casuale, gli 8 esperimenti ottenendo per la risposta *Resa* i seguenti valori, [Tabella 9](#)

Tabella 9: Piano sperimentale 2_{IV}^{4-1} con risposte

vol. solvente	t. centrifuga	forza ionica	t. estrazione	Resa
10	5	1	1	17
40	5	1	5	37.9
10	20	1	5	17
40	20	1	1	24.6
10	5	5	5	28.4
40	5	5	1	22.7
10	20	5	1	30.3
40	20	5	5	36.3

Inserendo nell’applicativo gli 8 valori della *Resa* ottenuti (in ordine come il [Tabella 9](#)), otteniamo la stima puntuale dei parametri [Figura 10](#) e il relativo grafico [Figura 11](#).

Ricordando le “confusioni” [Figura 9](#) e che il disegno è di risoluzione IV abbiamo che il termini lineari sono “confusi” con le interazioni di ordine 3, possiamo quindi supporre che i valori dei parametri x_1, x_2, x_3 e x_4 si riferiscano ai termini lineari mentre rimangono le “confusioni” a coppie per le interazioni di ordine 2.

Il modello risulta quindi

$$y = 26.77 + 3.60x_1 + 0.28x_2 + 2.65x_3 + 3.12x_4 - 0.2(x_1x_2 + x_3x_4) - 3.53(x_1x_3 + x_2x_4) + 3.60(x_1x_4 + x_2x_3)$$

Parametri regressione

Stima puntuale

(Intercept)	x1	x2	x3	x4	x1:x2
26.77	3.60	0.28	2.65	3.12	-0.20
x1:x3	x2:x3				
-3.53	3.60				

Figura 10: Stima puntuale dei parametri

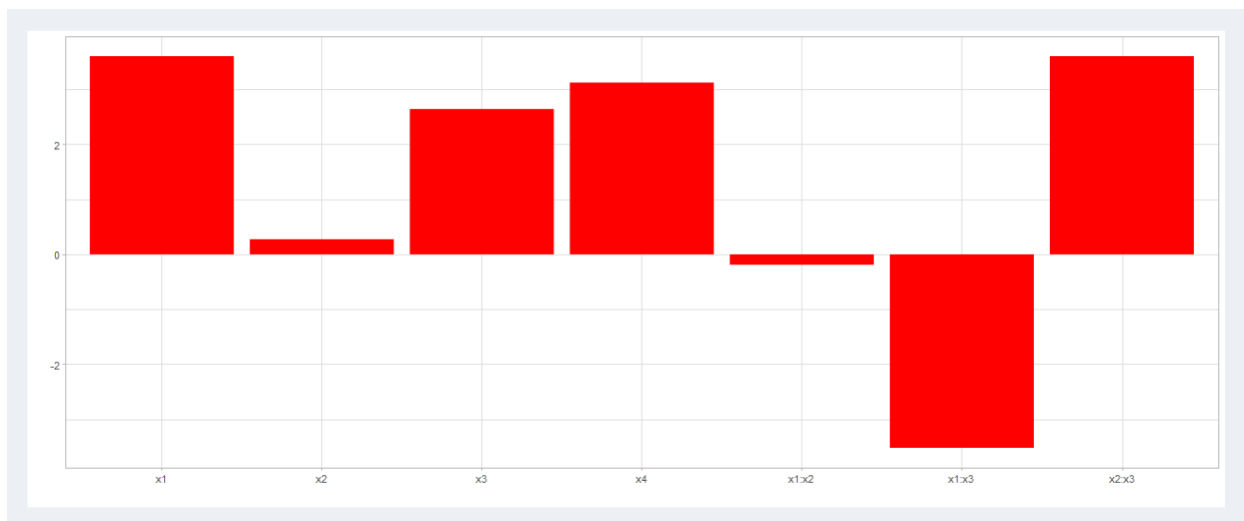


Figura 11: Grafico dell'estima puntuale dei parametri

Il fattore x_2 ha coefficiente piccolo e non sembra importante, mentre gli altri tre termini lineari lo sono sicuramente.

Quindi si può sostenere l'ipotesi che le interazioni confuse siano dovute ai termini diversi da x_2 .

Questa osservazione è di fatto risultata coerente con il dato sperimentale osservato secondo cui il tempo di centrifuga non ha alcun effetto sulla resa di estrazione perché il livello più basso scelto è già più che sufficiente per rompere l'emulsione creata dopo la miscelazione delle fasi del solvente di estrazione e del campione. Tali conclusioni sono confermate dal piano 2^4 poi condotto a termine.

Il modello finale semplificato quindi è

$$y = 26.77 + 3.60x_1 + 2.65x_3 + 3.12x_4 - 0.2x_3x_4 - 3.53x_1x_3 + 3.60x_1x_4$$

Per convalidare il modello sono state eseguite misure indipendenti nel punto test $(1, -1, -1, -1)$. Inserendo le rese osservate 17.2, 16.9, 17.0, 16.8 nell'apposita casella dell'applicativo si ottiene [Figura 12](#)

La risposta predetta nel punto test e gli estremi dell'intervallo di confidenza costruito con la stima di σ ottenuta con le 4 misure indipendenti sono dati in [Figura 13](#).

Il modello è convalidato statisticamente e può quindi essere usato per esplorare in modo attendibile il dominio delle risposte, [Figura 14](#)

Misure indipendenti

17.2 16.9 17 16.8

media 2.5% 97.5%
16.975 16.703 17.247

dev.st.
0.171

gdl
3

Figura 12: Misure indipendenti nel punto test

Punto previsione (coord.separate da spazio)

-1 -1 -1 -1

previsione
17

2.5% 97.5% 0.5% 99.5% 0.05% 99.95%
1 16.456 17.544 16.002 17.998 14.793 19.207

Figura 13: Previsione nel punto test e estremi dell'intervallo di confidenza

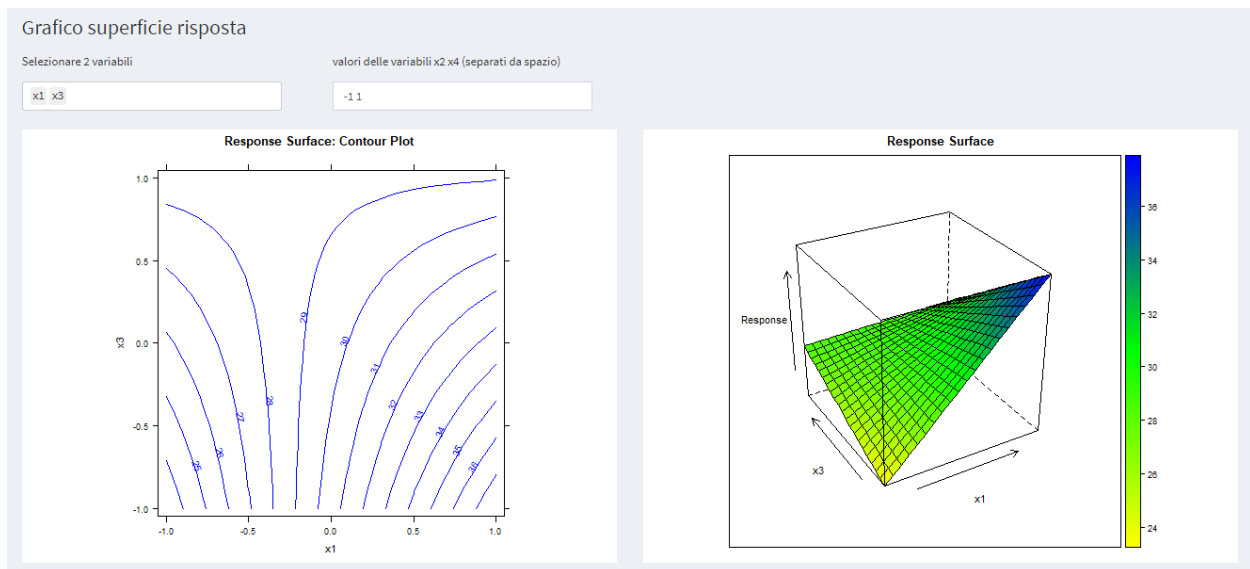


Figura 14: Grafico della superficie di risposta