

# Plackett Burman

*Giorgio Marrubini e Camillo Melzi*

*06 giugno 2020*

I disegni frazionari  $2_{III}^{k-p}$  di risoluzione *III* sono usati prevalentemente per condurre studi di screening e di robustezza. Il numero di esperimenti di un qualsiasi piano fattoriale frazionario è comunque uguale ad un numero di esperimenti esponenziale  $2^n$ : 8, 16, 32, ... . Plackett e Burman nel 1946 hanno introdotto una classe di disegni fattoriali a 2 livelli ortogonali che permettono di condurre lo screening di  $n - 1$  fattori prevedendo lo svolgimento di un numero  $n = 4m$  di esperimenti (4, 8, 12, 16, 20, 24, ...), in modo tale da offrire le opzioni per colmare i vuoti esistenti ad esempio tra piani di 8 e 16, oppure di 16 e 32 esperimenti. Nel caso in cui  $n$  sia anche una potenza di 2 (4, 8, 16, ...), il disegno di PB corrispondente è un particolare disegno frazionario di risoluzione *III*.

La matrice sperimentale di questi disegni si costruisce seguendo un procedimento iterativo in cui, data la prima riga della matrice (vedi [Tabella 1](#)) di lunghezza  $n - 1$ , si generano le righe successive per permutazione circolare di questa riga (ogni riga ha al primo posto l'elemento che era all'ultimo posto della riga superiore e agli altri  $i = 2, \dots, n - 1$  l'elemento di posto  $i - 1$  della riga superiore; in totale in questo modo si scrivono  $n - 1$  righe, compresa la prima riga data, e a queste si aggiunge una ultima riga composta da solo  $-1$ .

Tabella 1:

$n$																			
4	1	1	-1																
8	1	1	1	-1	1	-1	-1												
12	1	1	-1	1	1	1	-1	-1	-1	1	-1								
16	1	1	1	1	-1	1	-1	1	1	-1	-1	1	-1	-1	-1				
20	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1
...																			

Vediamo ad esempio la costruzione della matrice per  $n = 4$ . La prima riga della matrice, vedi [Tabella 1](#), è:

$$1 \quad 1 \quad -1.$$

Generiamo le altre righe 2 righe per permutazione circolare e aggiungiamo la quarta riga costituita da  $-1$ .

Tabella 2:

	X1	X2	X3
1	1	1	-1
2	-1	1	1
3	1	-1	1
4	-1	-1	-1

Il modello associato a questi disegni tiene conto di tutti i fattori lineari

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{n-1} x_{i,n-1} + \epsilon_i, \quad i = 1, \dots, n,$$

la matrice del modello é uguale alla matrice sperimentale a cui va aggiunta una prima colonna di tutti 1. Abbiamo per lo stimatore  $b = X^{-1}Y$

$$Cov(b) = \frac{\sigma^2}{n} I_n,$$

abbiamo cioè

$$Var(b) = \frac{\sigma^2}{n}, \quad j = 1, \dots, n$$

e

$$Corr(b_i, b_j) = 0, \quad i \neq j.$$

Nell'applicativo, selezionando il numero di fattori da studiare, viene proposto il disegno di Plackett-Burman con un numero di esperimenti pari al minimo multiplo di 4 maggiore del numero dei fattori, **Figura 1**.



Figura 1: Disegno di Plackett-Burman per 5 fattori

Si noti che le prime 5 colonne sono “occupate” dai 5 fattori in analisi indicate con  $x_1, \dots, x_5$  mentre vengono indicate con  $e_1, e_2$  le 2 colonne rimanenti. Queste sono 2 variabili cosiddette *dummy* (variabili fittizie) che possono essere utili per stabilire la significatività dei fattori in studio. Si vedano al riguardo gli esempi nel seguito.

I disegni di Plackett-Burman sono disegni di risoluzione *III*, nel senso che le interazioni sono confuse con i termini lineari. Nel caso in cui il numero di esperimenti  $n$  sia una potenza di 2, come abbiamo già detto, il disegno di Plackett-Burman è un particolare disegno frazionario di risoluzione *III* e quindi i termini lineari sono “completamente confusi” con le interazioni (si veda la sezione dedicata ai disegni frazionari). Nei casi in cui  $n$  non è una potenza di 2, per determinare le relazioni “confuse” si considerano la matrice  $X_1$  del modello

**Tabella 3**

Tabella 3: Matrice del modello

Int.	x1	x2	x3	x4	x5	e1	e2
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
1	1	-1	1	-1	-1	1	1
1	1	1	-1	1	-1	-1	1
1	-1	-1	-1	-1	-1	-1	-1

e la matrice  $X_2$  dei termini (interazioni) confusi [Tabella 4](#)

Tabella 4: Matrice delle interazioni confuse

x1x2	x1x3	x1x4	x1x5	x1e1	x1e2	...	x4x5	x4e1	x4e2	x5e1	x5e2	e1e2
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	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	1	1	-1	-1
1	1	1	1	1	1		1	1	1	1	1	1

La matrice  $A$ , spesso indicata anche con il nome di “*alias matrix*”, o matrice di confusione, è definita dalla relazione algebrica

$$A = (X_1^t X_1)^{-1} (X_1^t X_2)$$

La matrice  $A$  ha tante righe quanti sono i termini del modello (in questo caso 8: intercetta, 5 fattori e due variabili fittizie) e tante colonne quante sono le interazioni possibili (in questo caso 21 interazioni a due termini), e si chiama matrice di confusione.

In [Figura 2](#) è illustrata la matrice di confusione relativa al modello di Plackett-Burman per 5 fattori di cui è data la matrice in [Figura 1](#). Nella matrice di confusione, in ogni riga si può leggere la “confusione” di ogni termine lineare con le interazioni di due termini.

Il caso in esame è un frazionario  $2_{III}^{5-2}$  e le confusioni, come ci aspettiamo, sono “totali” o “complete”: ciò significa che ogni interazione è “confusa” con un solo termine lineare. Si noti che nella matrice di confusione di questo modello appaiono solo segni 0 o -1.

Scegliendo 12 esperimenti, la matrice di confusione è quella data in [Figura 3](#)

In questo caso le “confusioni” sono “parziali”, cioè le interazioni si confondono “parzialmente” con più termini lineari. Questo ci permette di stimare le interazioni con l’analisi di regressione. Possiamo sostituire gli effetti lineari meno significativi con le interazioni più importanti. Per determinare le interazioni più rilevanti si moltiplicano per un’interazione fissata le confusioni con i termini lineari più grandi per il segno del coefficiente del termine lineare e se ne fa la somma. Nell’applicativo è proposto il grafico delle confusioni (Plot alias) in cui è rappresentata visivamente tale somma [Figura 11](#)

Per meglio chiarire quanto finora detto facciamo 2 esempi numerici.

### Esempio: confusioni

## Confusioni

	x1:x2	x1:x3	x1:x4	x1:x5	x1:e1	x1:e2	x2:x3	x2:x4	x2:x5	x2:e1	x2:e2
(Intercept)	0	0	0	0	0	0	0	0	0	0	0
x1	0	0	0	0	0	0	0	0	0	-1	0
x2	0	0	0	0	-1	0	0	0	0	0	0
x3	0	0	-1	0	0	0	0	0	0	0	-1
x4	0	-1	0	0	0	0	0	0	-1	0	0
x5	0	0	0	0	0	-1	0	-1	0	0	0
e1	-1	0	0	0	0	0	0	0	0	0	0
e2	0	0	0	-1	0	0	-1	0	0	0	0

	x3:x4	x3:x5	x3:e1	x3:e2	x4:x5	x4:e1	x4:e2	x5:e1	x5:e2	e1:e2
(Intercept)	0	0	0	0	0	0	0	0	0	0
x1	-1	0	0	0	0	0	0	0	-1	0
x2	0	0	0	-1	-1	0	0	0	0	0
x3	0	0	0	0	0	0	0	-1	0	0
x4	0	0	0	0	0	0	0	0	0	-1
x5	0	0	-1	0	0	0	0	0	0	0
e1	0	-1	0	0	0	0	-1	0	0	0
e2	0	0	0	0	0	-1	0	0	0	0

Figura 2: Matrice di confusione

Come fatto nel caso del disegno frazionario, per capire le “confusioni” in un disegno di Plackett-Burman, consideriamo un modello teorico di un ipotetico fenomeno in esame

$$y = x_1 + 5x_2 - 3x_3 + 15x_4 - 15x_1x_3 + \epsilon$$

e supponiamo che si vogliano analizzare gli effetti di 5 fattori

$$x_1, x_2, x_3, x_4, x_5$$

perché dall’analisi preliminare del problema è emerso che potrebbero essere importanti nell’influencare la risposta.

Con 5 fattori il primo disegno di Plackett-Burman “utile” è quello che prevede di eseguire 8 esperimenti  
**Tabella 9**

Tabella 5: Disegno di Plackett-Burman con 8 esperimenti con risposte costruite dal modello supposto  $y = x_1 + 5x_2 - 3x_3 + 15x_4 - 15x_1x_3 + \epsilon$

x1	x2	x3	x4	x5	e1	e2	y
1	1	1	-1	1	-1	-1	2.66
-1	1	1	1	-1	1	-1	1.26
-1	-1	1	1	1	-1	1	-9.53
1	-1	-1	1	1	1	-1	-0.81
-1	1	-1	-1	1	1	1	5.62
1	-1	1	-1	-1	1	1	-9.10

x1	x2	x3	x4	x5	e1	e2	y
1	1	-1	1	-1	-1	1	8.56
-1	-1	-1	-1	-1	-1	-1	-3.04

E' un disegno fattoriale frazionario a due livelli con risoluzione *III*. I termini lineari sono confusi con i termini di interazione di ordine 2. Si noti il valore  $-1$  nella matrice delle confusioni [Figura 4](#)

Il grafico dei coefficienti è dato da [Figura 5](#)

### Confusioni

	x1:x2	x1:x3	x1:x4	x1:x5	x1:e1	x1:e2	x1:e3	x1:e4	x1:e5	x1:e6	x2:x3
(Intercept)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
x1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.33
x2	0.00	-0.33	-0.33	-0.33	0.33	-0.33	-0.33	0.33	0.33	-0.33	0.00
x3	-0.33	0.00	0.33	-0.33	-0.33	0.33	-0.33	0.33	-0.33	-0.33	0.00
x4	-0.33	0.33	0.00	0.33	0.33	-0.33	-0.33	-0.33	-0.33	-0.33	-0.33
x5	-0.33	-0.33	0.33	0.00	-0.33	-0.33	-0.33	-0.33	0.33	0.33	-0.33
e1	0.33	-0.33	0.33	-0.33	0.00	-0.33	0.33	-0.33	-0.33	-0.33	-0.33
e2	-0.33	0.33	-0.33	-0.33	-0.33	0.00	0.33	-0.33	0.33	-0.33	0.33
e3	-0.33	-0.33	-0.33	-0.33	0.33	0.33	0.00	-0.33	-0.33	0.33	-0.33
e4	0.33	0.33	-0.33	-0.33	-0.33	-0.33	-0.33	0.00	-0.33	0.33	-0.33
e5	0.33	-0.33	-0.33	0.33	-0.33	0.33	-0.33	-0.33	0.00	-0.33	0.33
e6	-0.33	-0.33	-0.33	0.33	-0.33	-0.33	0.33	0.33	-0.33	0.00	0.33
	x2:x4	x2:x5	x2:e1	x2:e2	x2:e3	x2:e4	x2:e5	x2:e6	x3:x4	x3:x5	x3:e1
(Intercept)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
x1	-0.33	-0.33	0.33	-0.33	-0.33	0.33	0.33	-0.33	0.33	-0.33	-0.33
x2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.33	-0.33	-0.33
x3	-0.33	-0.33	-0.33	0.33	-0.33	-0.33	0.33	0.33	0.00	0.00	0.00
x4	0.00	0.33	-0.33	-0.33	0.33	-0.33	0.33	-0.33	0.00	-0.33	-0.33
x5	0.33	0.00	0.33	0.33	-0.33	-0.33	-0.33	-0.33	-0.33	0.00	0.33
e1	-0.33	0.33	0.00	-0.33	-0.33	-0.33	-0.33	0.33	-0.33	0.33	0.00
e2	-0.33	0.33	-0.33	0.00	-0.33	0.33	-0.33	-0.33	-0.33	-0.33	0.33
e3	0.33	-0.33	-0.33	-0.33	0.00	0.33	-0.33	0.33	0.33	-0.33	0.33
e4	-0.33	-0.33	-0.33	0.33	0.33	0.00	-0.33	-0.33	-0.33	0.33	-0.33
e5	0.33	-0.33	-0.33	-0.33	-0.33	-0.33	0.00	-0.33	-0.33	-0.33	-0.33
e6	-0.33	-0.33	0.33	-0.33	0.33	-0.33	-0.33	0.00	0.33	0.33	-0.33

Figura 3: Parte della matrice di confusione di un disegno di Plackett-Burman con 12 esperimenti

Se a priori non possiamo escludere nessuna interazione di ordine 2, potrebbe accadere che la non significatività dei termini  $x_4$  e  $x_5$  sia dovuta al fatto che si annullino con le interazioni di ordine 2 con cui sono confusi. Ad esempio,  $x_4$  potrebbe annullarsi con  $x_1x_3$  ( $x_4$  si confonde con  $-x_1x_3$  e, se i coefficienti di questi termini hanno valore numerico simile, come nel nostro esempio, si annullano).

Anche per le dummy bisogna fare attenzione e essere sicuri che non appaiano importanti perchè “rappresentano” in realtà una interazione confusa con esse.

Non possiamo sapere quindi se la significatività che osserviamo per alcuni termini è autentica, ossia dovuta realmente al fattore stesso, oppure derivi dalle interazioni confuse con ciascun termine. In questi casi, è solo la conoscenza tecnica del problema chimico (o fisico o ingegneristico) che può aiutare a definire la reale importanza dei diversi termini e a giungere alla conclusione di escludere alcune interazioni (perché ad esempio prive di senso fisico o impossibili tecnicamente).

Se a priori invece non si può escludere nessuna interazione, si può prendere in considerazione il piano sperimentale di Plackett-Burman successivo, ossia quello che prevede 12 esperimenti. Questo numero di prove è comunque inferiore a quello del fattoriale frazionario corrispondente che è il  $2_V^{5-1}$ , e conta quindi 16 esperimenti, [Tabella 6](#)

## Confusioni

	x1:x2	x1:x3	x1:x4	x1:x5	x1:e1	x1:e2	x2:x3	x2:x4	x2:x5	x2:e1	x2:e2
(Intercept)	0	0	0	0	0	0	0	0	0	0	0
x1	0	0	0	0	0	0	0	0	0	-1	0
x2	0	0	0	0	-1	0	0	0	0	0	0
x3	0	0	-1	0	0	0	0	0	0	0	-1
x4	0	-1	0	0	0	0	0	0	-1	0	0
x5	0	0	0	0	0	-1	0	-1	0	0	0
e1	-1	0	0	0	0	0	0	0	0	0	0
e2	0	0	0	-1	0	0	-1	0	0	0	0

	x3:x4	x3:x5	x3:e1	x3:e2	x4:x5	x4:e1	x4:e2	x5:e1	x5:e2	e1:e2
(Intercept)	0	0	0	0	0	0	0	0	0	0
x1	-1	0	0	0	0	0	0	0	-1	0
x2	0	0	0	-1	-1	0	0	0	0	0
x3	0	0	0	0	0	0	0	-1	0	0
x4	0	0	0	0	0	0	0	0	0	-1
x5	0	0	-1	0	0	0	0	0	0	0
e1	0	-1	0	0	0	0	-1	0	0	0
e2	0	0	0	0	0	-1	0	0	0	0

Figura 4: Matrice delle confusioni di un pb con 8 esperimenti

Tabella 6: Disegno di Plackett-Burman con 12 esperimenti con risposte costruite dal modello ipotizzato  $y = x_1 + 5x_2 - 3x_3 + 15x_4 - 15x_1x_3 + \epsilon$

x1	x2	x3	x4	x5	e1	e2	e3	e4	e5	e6	y
1	1	-1	1	1	1	-1	-1	-1	1	-1	10.41
-1	1	1	-1	1	1	1	-1	-1	-1	1	-28.31
1	-1	1	1	-1	1	1	1	-1	-1	-1	24.41
-1	1	-1	1	1	-1	1	1	1	-1	-1	36.73
-1	-1	1	-1	1	1	-1	1	1	1	-1	-38.92
-1	-1	-1	1	-1	1	1	-1	1	1	1	26.56
1	-1	-1	-1	1	-1	1	1	-1	1	1	-29.71
1	1	-1	-1	-1	1	-1	1	1	-1	1	-20.04
1	1	1	-1	-1	-1	1	-1	1	1	-1	3.18
-1	1	1	1	-1	-1	-1	1	-1	1	1	1.26
1	-1	1	1	1	-1	-1	-1	1	-1	1	21.81
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-3.04

Il grafico dei coefficienti è dato da [Figura 6](#)

Quando si applicano piani di Plackett-Burman quindi è molto importante prestare la massima attenzione alla significatività dei fattori e alle possibili interazioni confuse con essi. Questo vale a maggior ragione per le variabili dummy che, come visto, possono essere confuse con le interazioni coppie al pari di qualsiasi altro fattore compreso nel piano degli esperimenti.

Nel caso di un piano sperimentale di 12 esperimenti, le confusioni sono parziali. Si noti il valore  $-0.33$  nella matrice delle confusioni [Figura 3](#). Tale valore permette di stimare le interazioni con l'analisi di regressione. Possiamo sostituire gli effetti lineari meno importanti con le interazioni più significative.

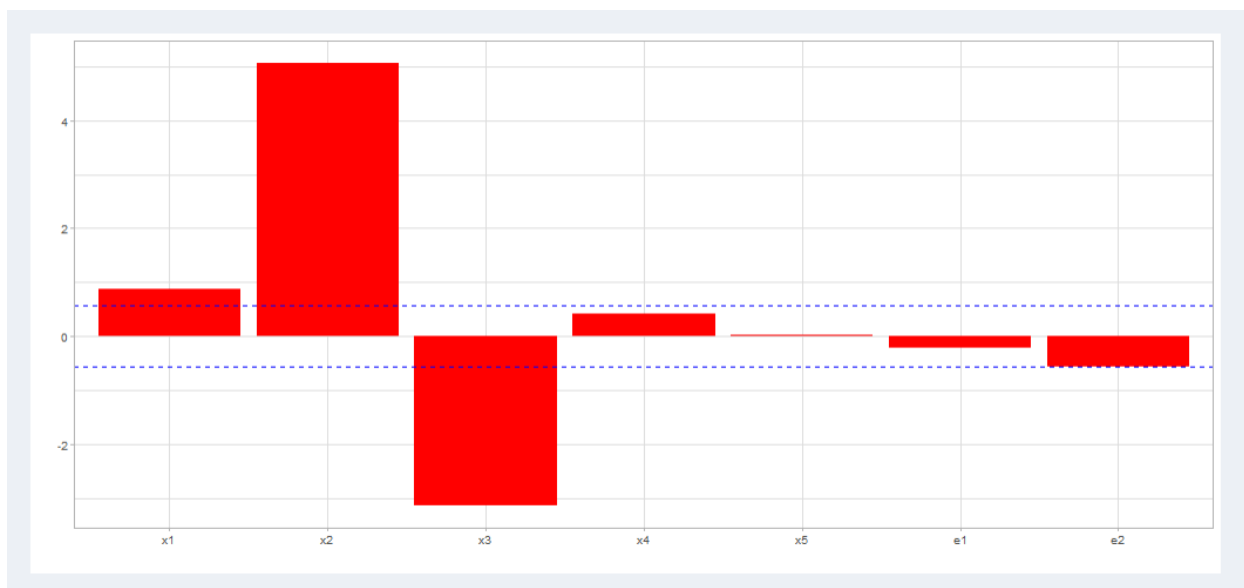


Figura 5: Grafico dei coefficienti del pb con 8 esperimenti

Per determinare le interazioni più significative, si moltiplicano queste per una interazione fissata, si moltiplicano le confusioni con i termini lineari più importanti per il segno del coefficiente del termine lineare e se ne fa la somma [Figura 7](#).

Con l'applicativo possiamo innanzitutto scaricare un disegno di Plackett-Burman con 12 esperimenti e quindi ricaricare il file salvato nel menù *Piano Personalizzato* (se si preferisce si può anche importare il disegno con copia/incolla). Vedi [Figura 8](#)

Una volta importato il dataset selezioniamo i termini del modello che vogliamo studiare. Nel nostro caso ad esempio possiamo considerare i 5 termini lineari e le 4 interazioni  $x_1x_3, x_2x_5, x_2x_4, x_4x_5$  che hanno maggiore probabilità di essere significative, vedi [Figura 7](#).

Inserendo le 12 risposte nell'opportuno riquadro otteniamo il seguente grafico dei coefficienti [Figura 9](#).

Tale grafico evidenzia la significatività sia di  $x_4$  che dell'interazione  $x_1x_3$  che non riuscivamo a distinguere poiché questi termini, a causa delle confusioni, si annullavano a vicenda.

### Esempio Elvitegravir

Il problema è lo sviluppo di un metodo semplice per l'isolamento di un principio attivo da plasma umano preliminare all'analisi HPLC-ESI/MS/MS. La risposta è la sensibilità S/N, che si vuole massimalizzare.

Tra tutti i fattori considerati si è deciso di considerare i 5 fattori:

- $x_1$ : Solvente (agente precipitante)
- $x_2$ : Volume di plasma ( $\mu\text{L}$ )
- $x_3$ : Volume di solvente (rapp v.campione)
- $x_4$ : Tempo miscelazione (sec)
- $x_5$ : Temperatura centrifuga ( $^{\circ}\text{C}$ )

Il dominio sperimentale è definito in [Tabella 7](#)



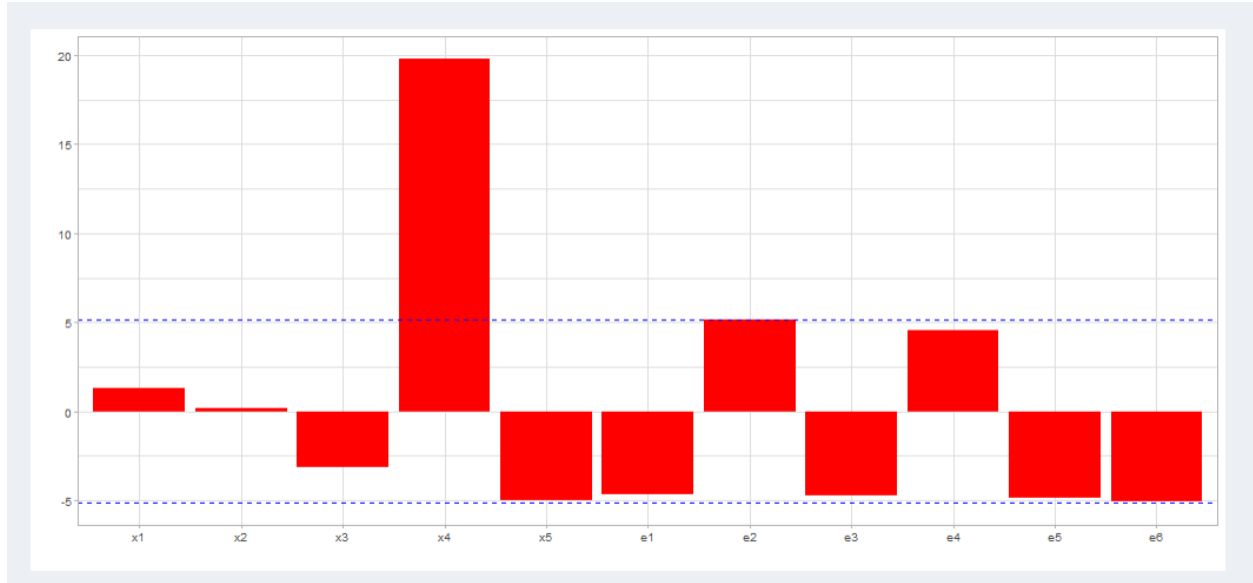


Figura 6: Grafico dei coefficienti del piano di Plackett-Burman con 12 esperimenti

Tabella 7: Definizione dei livelli dei 5 fattori

Fattori	-1	+1
Solvente	ACN	MeOH
Plasma	50	200
Solvente (vol)	1:3	1:7
Miscelazione	20	60
Centrifuga	4	25

Si sceglie un disegno di Plackett-Burman con 8 esperimenti. Sono quindi aggiunte al piano sperimentale 2 variabili dummy

- e1: orologio al polso
- e2: canto mentre lavoro

I livelli delle 2 variabili fittizie sono dati nella [Tabella 8](#). Sono 2 variabili che sicuramente non hanno alcuna influenza sulla risposta e che saranno utili quindi per avere un benchmark di significatività (o se si preferisce, una stima della importanza del rumore di fondo casuale nei risultati degli esperimenti).

Tabella 8: Definizione dei livelli dei 2 fattori dummy

Fattori	-1	+1
Orologio	sin	dx
Canto	si	no

Il disegno con le relative risposte è dato in [Tabella 9](#)

## Plot alias

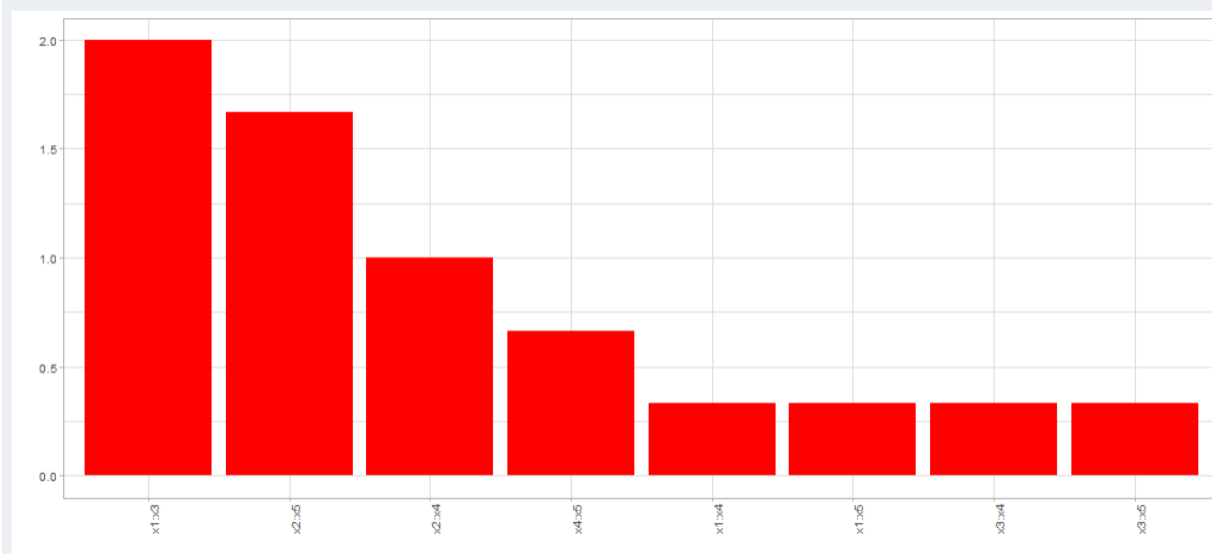


Figura 7: Grafico dei coefficienti del pb con 12 esperimenti

**Termini modello (x)**

x1 x2 x3 x4 x5 x1:x3 x2:x5 x2:x4 x4:x5

Figura 8: Selezione delle variabili

Tabella 9: Disegno di Plackett-Burman per lo sviluppo del metodo di estrazione di Elvitegravir con risposte

Exp.	x1	x2	x3	x4	x5	e1	e2	S.N
1	1	1	1	-1	1	-1	-1	31795
2	-1	1	1	1	-1	1	-1	33313
3	-1	-1	1	1	1	-1	1	32264
4	1	-1	-1	1	1	1	-1	31559
5	-1	1	-1	-1	1	1	1	35150
6	1	-1	1	-1	-1	1	1	21201
7	1	1	-1	1	-1	-1	1	32344
8	-1	-1	-1	-1	-1	-1	-1	21087

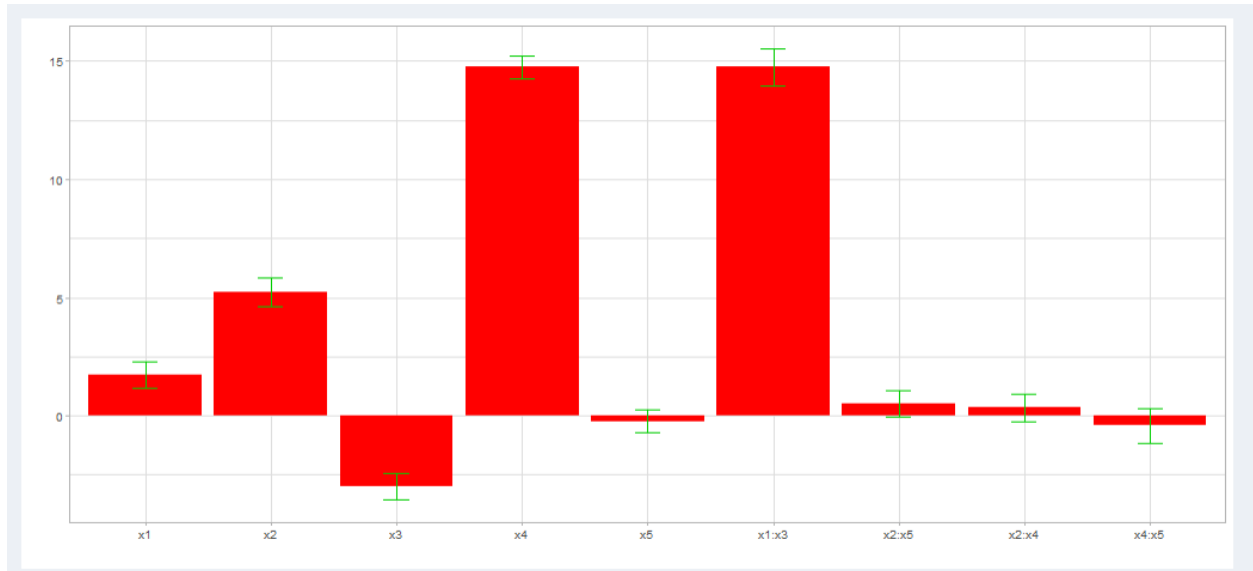


Figura 9: Selezione delle variabili

Inserendo le risposte nell'applicativo otteniamo la stima dei parametri in [Figura 10](#)

Come detto, le 2 variabili dummy possono essere usate come benchmark di significatività. Quindi tutti i parametri che in valore assoluto sono minori del valore assoluto dei coefficienti delle 2 variabili dummy possono essere considerati non significativi. In [Figura 10](#) sono tracciate 2 linee orizzontali indicanti la fascia di non significatività. La variabile  $x_4$  (Tempo miscelazione, misurata in sec) è non significativa.

Lo studio può continuare affinando l'indagine con un nuovo piano di esperimenti che prenda in esame solo questi 4 fattori significativi.

## Parametri regressione

Stima puntuale

(Intercept)	x1	x2	x3	x4	x5
29839.12	-614.38	3311.37	-195.88	2530.88	2852.87
e1	e2				
466.63	400.63				

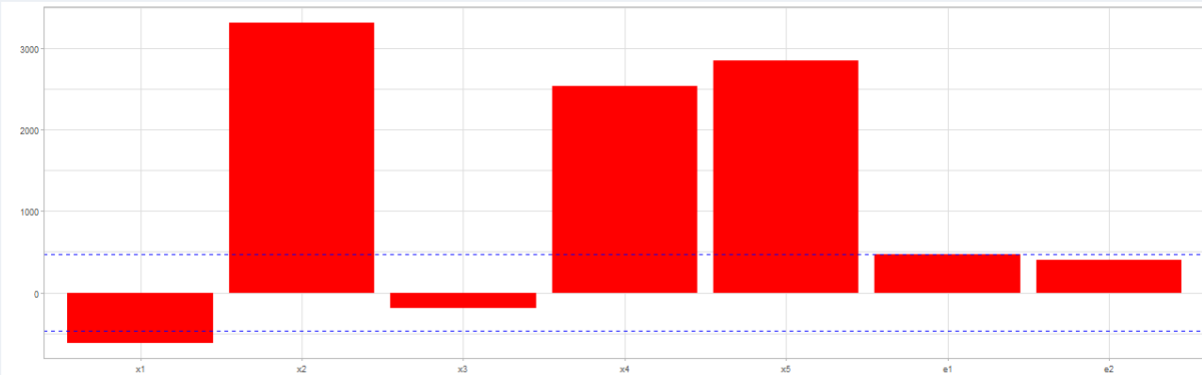


Figura 10: Stima dei parametri

## Plot alias

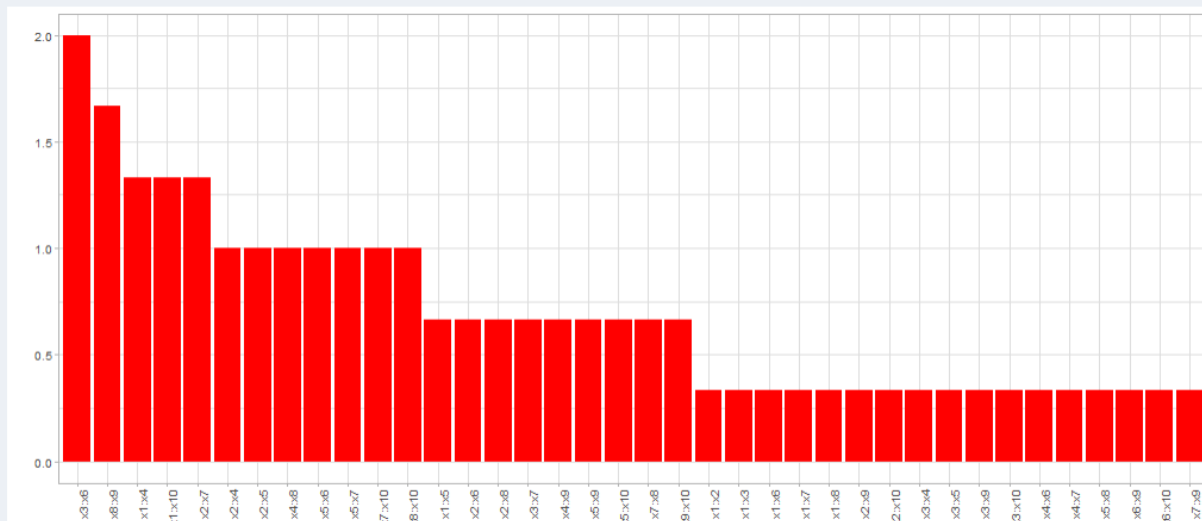


Figura 11: Plot alias