Σχεδιασμοί και Ανάλυση Πειραμάτων με Εφαρμογές σε περιβάλλον *R*

Αναστάσιος Κατσιλέρος

Γεωπονικό Πανεπιστήμιο Αθηνών
Εργαστήριο Βελτίωσης Φυτών και Γεωργικού Πειραματισμού

katsileros@aua.gr

Αθήνα 2022

Μεθοδολογία αποκριτικής επιφανείας (Response Surface Methodology)

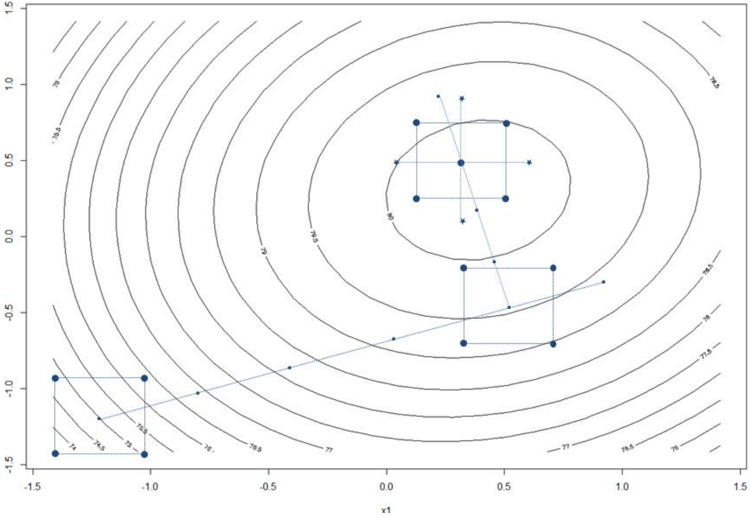
Η μεθοδολογία των αποκριτικών επιφανειών αναπτύχθηκε από τους Box και Wilson (1951) και είναι μία σειρά από μαθηματικές και στατιστικές τεχνικές που χρησιμοποιούνται στην ανάλυση και μοντελοποίηση των πειραματικών δεδομένων με σκοπό την βελτιστοποίηση μιας μεταβλητής απόκρισης.

Χρησιμοποιούνται διαδοχικά πειραματικοί σχεδιασμοί (πρώτης και δεύτερης τάξης) με σκοπό την εύρεση του κατάλληλου μοντέλου, το οποίο προσεγγίζει καλύτερα την επιφάνεια απόκρισης.

Για να εντοπίσει ο ερευνητής την βέλτιστη περιοχή θα πρέπει να ακολουθήσει τα παρακάτω βήματα.

- Επιλογή των κατάλληλων παραγόντων, επιπέδων (προκαταρκτικά πειράματα, screening).
- Επιλογή της περιοχής σχεδιασμού και προσαρμογή μοντέλου πρώτης τάξης με έλεγχο για καμπυλότητα (παραγοντικός σχεδιασμός με κεντρικά σημεία).
- Υπολογισμός της κατεύθυνσης - **μονοπάτι της πιο απότομης μεταβολής** (path of steepest ascent) (μεγιστοποίηση ή ελαχιστοποίηση) της μεταβλητής απόκρισης.

- Εκτέλεση επαναλαμβανόμενων πειραματικών δοκιμών κατά μήκος του μονοπατιού και παρατήρηση μεταβολών της τιμής απόκρισης.
- Στην θέση του μονοπατιού όπου η τιμή της απόκρισης είναι η επιθυμητή, προσαρμόζεται ένα άλλο μοντέλο πρώτης τάξης όπου και γίνεται έλεγχος για καμπυλότητα.
  Αν η έλλειψη προσαρμογής δεν είναι σημαντική, υπολογίζεται ένα δεύτερο μονοπάτι βασισμένο στο νέο μοντέλο και διεξάγονται πειράματα κατά μήκος του δεύτερου μονοπατιού. Αν η έλλειψη προσαρμογής είναι σημαντική, προσαρμόζεται ένα άλλο μοντέλο δεύτερης τάξης (π.χ. κεντρικός σύνθετός σχεδιασμός) και υπολογίζεται το στάσιμο σημείο.



x1

Κεντρικός Σύνθετος Σχεδιασμός (Central Composite Design)

Ο κεντρικός σύνθετος σχεδιασμός αποτελείται από τρία τμήματα:

Έναν πλήρη παραγοντικό σχεδιασμό 2^k ή ένα κλάσμα του, στον οποίο τα επίπεδα είναι κωδικοποιημένα ως -1 και 1 και οι συνδυασμοί των επεμβάσεων ονομάζονται **παραγοντικά σημεία** (factorial points) (n_f).

Τα **αξονικά σημεία** (axial points) τα οποία είναι τοποθετημένα ανά δύο στους άξονες των παραγόντων του σχεδιασμού και απέχουν απόσταση $\pm$ a από το κέντρο του $(n_{ax} = 2 * k)$.

Τα **κεντρικά σημεία** (center points) (n_c).

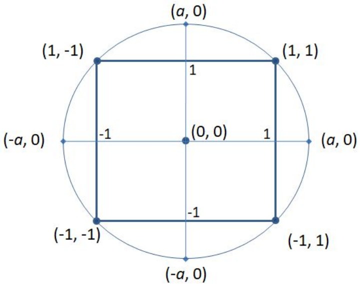
Ο συνολικός αριθμός των σημείων σε έναν κεντρικό σύνθετο σχεδιασμό είναι $N = n_f + n_{ax} + n_c$

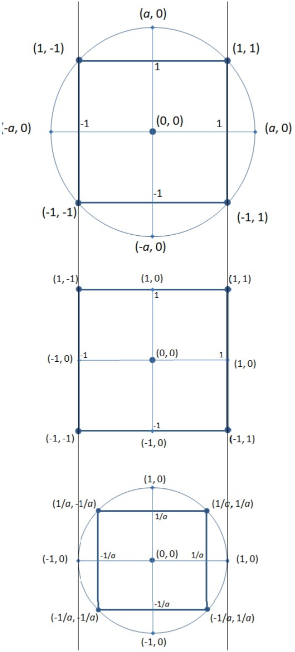
Επιλογή αξονικού σημείου a

Οι τιμές του a επιλέγονται κατά τρόπο ώστε ο σχεδιασμός να έχει επιθυμητές ιδιότητες (rotatability and orthogonality).

Περιστρέψιμο σχεδιασμό (Rotatable): $a = n_f^{1/4}$
Ορθογώνιο σχεδιασμό (Orthogonal):
$a = (\frac{n_f}{4} \cdot (\sqrt{n_f + n_a + n_c} - \sqrt{n_f})^2)^{1/4}$
Σφαιρικό σχεδιασμό (Spherical): $a = k^{1/2}$
Κυβοειδή σχεδιασμό (Face-centered): $a = 1$
Περιστρέψιμο και ορθογώνιο:$a = n_f^{1/4}$ και $n_c = 4n_f^{1/2} + 4 - 2k$

Κεντρικός Σύνθετος Σχεδιασμός ($k = 2$)



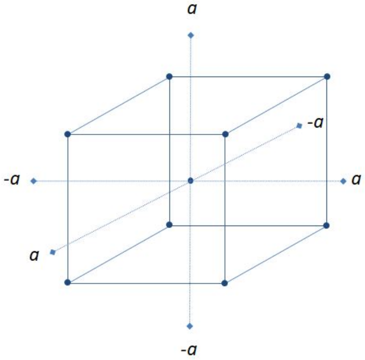


Κεντρικός Σύνθετος Σχεδιασμός

Κυβοειδής
(a=1)

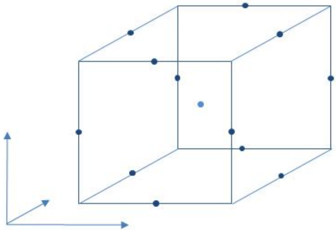
Εγγεγραμμένος
(a=1, levels=1/a)

Κεντρικός Σύνθετος Σχεδιασμός ($k = 3$)



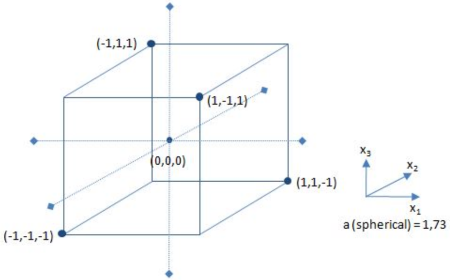
Box-Behnken Σχεδιασμός

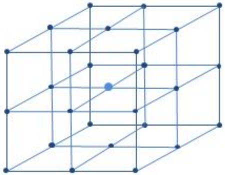
Είναι μία ειδική κατηγορία 3^k παραγοντικού σχεδιασμού όπου οι συνδυασμοί των επεμβάσεων δεν είναι στις κορυφές, αλλά στα μεσαία σημεία και στο κέντρο. Είναι χρήσιμα όταν επιθυμούμε αποφυγή ακραίων συνδυασμών επεμβάσεων. Είναι περιστρέψιμα ή σχεδόν περιστρέψιμα.



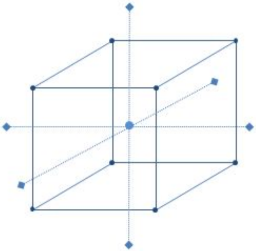
Μικρός Κεντρικός Σύνθετος Σχεδιασμός

Ο μικρός κεντρικός σύνθετος σχεδιασμός είναι κλασματικός παραγοντικός σχεδιασμός τριών παραγόντων, διαχωριστικής ικανότητας III, με προσθήκη αξονικών και κεντρικών σημείων.

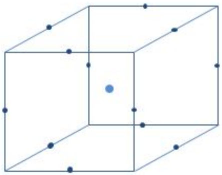




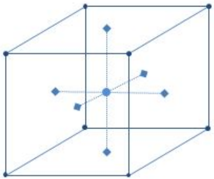
Παραγοντικός Σχεδιασμός 3^k



Κεντρικός Σύνθετος Σχεδιασμός (k = 3)



Box-Behnken Σχεδιασμός



Κεντρικός Σύνθετος Σχεδιασμός (α = 1)

Κεντρικός Σύνθετος Σχεδιασμός	k = 2	k = 3	k = 4	k = 5	k = 6
Παραγοντικά σημεία 2^k	4	8	16	32	64
Κεντρικά σημεία nc	5	5	6	6	6
Αξονικά σημεία	4	6	8	10	16
Κυβοειδή	1	1	1	1	1
Ορθογώνια	1,078	1,287	1,483	1,662	1,824
Σφαιρικά	1,414	1,732	2	2,236	2,449
Περιστρέψιμα	1,414	1,682	2	2,378	2,828
Σύνολο εκτελέσεων CCD	13	19	30	48	82
Box-Behnken	-	15	27	46	54
Παραγοντικός Σχεδιασμός 3^k	9	27	81	243	729

Συνδυασμοί επεμβάσεων		X1	X2	X1	X2	X1	X2	X1	X2
nf	- -	-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
nf Inscribe (a = 1)		± 0,707		± 0,707		± 0,927			
Αξονικά σημεία		Περιστρέψιμος		Σφαιρικός		Ορθογώνιος		Κυβοειδής	
	0a	0	-1,414	0	-1,414	0	-1,078	0	-1
	0A	0	1,414	0	1,414	0	1,078	0	1
	a0	-1,414	0	-1,414	0	-1,078	0	-1	0
	A0	1,414	0	1,414	0	1,078	0	1	0
nc	00	0	0	0	0	0	0	0	0
	00	0	0	0	0	0	0	0	0

Παράδειγμα 1.

Ένας χημικός μηχανικός ενδιαφέρεται να προσδιορίσει τις συνθήκες λειτουργίας που μεγιστοποιούν την απόδοση μιας διαδικασίας. Ο μηχανικός αυτή τη στιγμή λειτουργεί τη διαδικασία με χρόνο αντίδρασης 35 λεπτών και θερμοκρασία αντίδρασης 155°F, που έχουν ως αποτέλεσμα αποδόσεις περίπου 40 τοις εκατό.

Προσαρμογή πρώτης τάξης μοντέλου

A	B	x1	x2	Y
30	150	-1	-1	39,3
30	160	-1	1	40
40	150	1	-1	40,9
40	160	1	1	41,5
35	155	0	0	40,3
35	155	0	0	40,5
35	155	0	0	40,7
35	155	0	0	40,2
35	155	0	0	40,6

```
> design=FrF2(4,2,ncenter = 5,randomize = F)
> design
   A  B
1 -1 -1
2  1 -1
3 -1  1
4  1  1
5  0  0
6  0  0
7  0  0
8  0  0
9  0  0
class=design, type= full factorial.center
```

```
> Y=c(39.3, 40.9, 40, 41.5, 40.3, 40.5, 40.7, 40.2, 40.6)
> data=cbind(design, Y)
> data
   A  B    Y
1 -1 -1 39.3
2  1 -1 40.9
3 -1  1   40
4  1  1 41.5
5  0  0 40.3
6  0  0 40.5
7  0  0 40.7
8  0  0 40.2
9  0  0 40.6
```

```
> fit=lm(Y~ A*B, data)
> library(alr3)
> pureErrorAnova(fit)
Analysis of Variance Table
Response: Y
             Df  Sum Sq   Mean Sq  F value   Pr(>F)
A             1  2.40250  2.40250  55.8721   0.001713 **
B             1  0.42250  0.42250   9.8256   0.035030 *
A:B           1  0.00250  0.00250   0.0581   0.821316
Residuals     5  0.17472  0.03494
Lack of fit   1  0.00272  0.00272   0.0633   0.813741
Pure Error    4  0.17200  0.04300
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> fit=lm(Y~ A*B, data)
> summary(fit)
Call:
lm.default(formula = Y ~ A * B, data = data)
Residuals:
1 2 3 4 5 6 7 8 9
-0.01944 -0.01944 -0.01944 -0.01944 -0.14444 0.05556 0.25556
-0.24444 0.15556
Coefficients:
             Estimate   Std. Error   t value   Pr(>|t|)
 (Intercept) 40.44444   0.06231      649.069   1.65e-13 ***
 A           0.77500    0.09347      8.292     0.000417 ***
 B           0.32500    0.09347      3.477     0.017713 *
 A:B         -0.02500   0.09347      -0.267    0.799787
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1869 on 5 degrees of freedom
Multiple R-squared: 0.9418, Adjusted R-squared: 0.9069
F-statistic: 26.97 on 3 and 5 DF, p-value: 0.00163
```

Υπολογισμός της κατεύθυνσης

Το μοντέλο που επιλέγεται είναι το εξής:

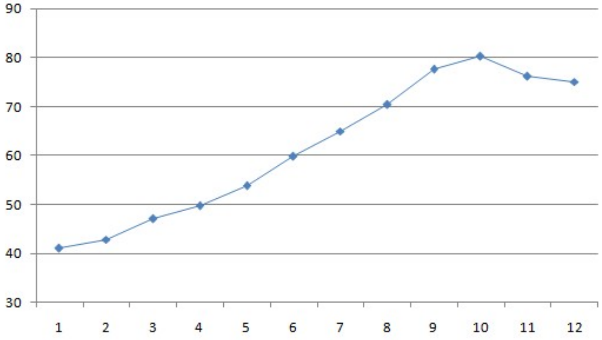
$$Y = 40,44 + 0,775x_1 + 0,325x_2$$

Για την εύρεση της βέλτιστης περιοχής, θα πρέπει να μετακινηθούμε στην κατεύθυνση - μονοπάτι, 0,775 μονάδες στην x_1 και 0,325 μονάδες στην x_2. Το μονοπάτι θα ξεκινάει από τα σημεία $x_1 = 0$ και $x_2 = 0$ και θα έχει κλίση 0,325/0,775 = 0,42.

Εκτέλεση πειραματικών δοκιμών κατά μήκος του μονοπατιού

Βήματα	x1	x2	A	B	Y
Origin	0	0	35	155	
Δ	1	0,42	5	2	
Origin + 1Δ	1	0,42	40	157	41
Origin + 2Δ	2	0,84	45	159	42,9
Origin + 3Δ	3	1,26	50	161	47,1
Origin + 4Δ	4	1,68	55	163	49,7
Origin + 5Δ	5	2,1	60	165	53,8
Origin + 6Δ	6	2,52	65	167	59,9
Origin + 7Δ	7	2,94	70	169	65
Origin + 8Δ	8	3,36	75	171	70,4
Origin + 9Δ	9	3,78	80	173	77,6
Origin + 10Δ	10	4,2	85	175	**80,3**
Origin + 11Δ	11	4,62	90	179	76,2
Origin + 12Δ	12	5,04	95	181	75,1

Εκτέλεση πειραματικών δοκιμών κατά μήκος του μονοπατιού



Προσαρμογή πρώτης τάξης μοντέλου στη νέα περιοχή

A	B	x1	x2	Y
80	170	-1	-1	76,5
80	180	-1	1	77
90	170	1	-1	78
90	180	1	1	79,5
85	175	0	0	79,9
85	175	0	0	80,3
85	175	0	0	80
85	175	0	0	79,7
85	175	0	0	79,8

```
> design1=FrF2(4,2,ncenter = 5,randomize = F)
> design1
   A  B
1 -1 -1
2  1 -1
3 -1  1
4  1  1
5  0  0
6  0  0
7  0  0
8  0  0
9  0  0
class=design, type= full factorial.center
```

```
> Y1=c(76.5, 78, 77, 79.5, 79.9, 80.3, 80, 79.7, 79.8)
> data1=cbind(design1, Y1)
> data1
   A  B    Y
1 -1 -1 76.5
2  1 -1 78.0
3 -1  1 77.0
4  1  1 79.5
5  0  0 79.9
6  0  0 80.3
7  0  0 80.0
8  0  0 79.7
9  0  0 79.8
```

```
> fit1=lm(Y1~A*B,data1)
> pureErrorAnova(fit1)
Analysis of Variance Table
Response: Y1
              Df  Sum Sq  Mean Sq  F value    Pr(>F)
 A             1   4.000    4.000   75.472  0.000966 ***
 B             1   1.000    1.000   18.868  0.012217 *
 A:B           1   0.250    0.250    4.717  0.095610 .
 Residuals     5  10.870    2.174
 Lack of fit   1  10.658   10.658  201.094  0.000143 ***
 Pure Error    4   0.212    0.053
—
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Προσαρμογή ενός δεύτερης τάξης μοντέλου

A	B	x1	x2	Y
80	170	-1	-1	76,5
80	180	-1	1	77
90	170	1	-1	78
90	180	1	1	79,5
85	175	0	0	79,9
85	175	0	0	80,3
85	175	0	0	80
85	175	0	0	79,7
85	175	0	0	79,8
92,07	175	1,414	0	78,4
77,93	175	-1,414	0	75,6
85	182,07	0	1,414	78,5
85	167,93	0	-1,414	77

```
> library(rsm)
> design2=ccd(2, n0 = c(5,0), alpha = "rotatable", randomize
= F, oneblock = T)
> design2
   run.order std.order  x1.as.is  x2.as.is
1          1         1 -1.000000 -1.000000
2          2         2  1.000000 -1.000000
3          3         3 -1.000000  1.000000
4          4         4  1.000000  1.000000
5          5         5  0.000000  0.000000
6          6         6  0.000000  0.000000
7          7         7  0.000000  0.000000
8          8         8  0.000000  0.000000
9          9         9  0.000000  0.000000
10         1         1 -1.414214  0.000000
11         2         2  1.414214  0.000000
12         3         3  0.000000 -1.414214
13         4         4  0.000000  1.414214
```

```
> Y2=c(76.5, 78, 77, 79.5, 79.9, 80.3, 80, 79.7, 79.8, 75.6,
78.4, 77, 78.5 )
> data2=cbind(design2, Y2)
> data2
```

	run.order	std.order	x1.as.is	x2.as.is	Y
1	1	1	-1.000000	-1.000000	76.5
2	2	2	1.000000	-1.000000	78
3	3	3	-1.000000	1.000000	77
4	4	4	1.000000	1.000000	79.5
5	5	5	0.000000	0.000000	79.9
6	6	6	0.000000	0.000000	80.3
7	7	7	0.000000	0.000000	80
8	8	8	0.000000	0.000000	79.7
9	9	9	0.000000	0.000000	79.8
10	1	1	-1.414214	0.000000	75.6
11	2	2	1.414214	0.000000	78.4,
12	3	3	0.000000	-1.414214	77
13	4	4	0.000000	1.414214	78.5

```
> fit2=rsm(Y2 ~ SO(x1, x2), data2)
> summary(fit2)
Call:
rsm(formula = Y ~ SO(x1, x2), data = data)
              Estimate   Std. Error   t value     Pr(>|t|)
 (Intercept)  79.94000   0.118959     671.9974    < 2.2e-16 ***
 X1           0.99497    0.094045     10.5798     1.474e-05 ***
 X2           0.51516    0.094045     5.4778      0.0009281 ***
 x1:x2        0.25000    0.133000     1.8797      0.1022110
 x1^2         -1.37625   0.100852     -13.6462    2.672e-06 ***
 x2^2         -1.00125   0.100852     -9.9279     2.244e-05 ***
—
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Multiple R-squared: 0.9828, Adjusted R-squared: 0.9705
F-statistic: 79.85 on 5 and 7 DF, p-value: 5.108e-06
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	10.0430	5.0215	70.9690	2.251e-05
TWI(x1, x2)	1	0.2500	0.2500	3.5333	0.1022
PQ(x1, x2)	2	17.9548	8.9774	126.8785	3.170e-06
Residuals	7	0.4953	0.0708		
Lack of fit	3	0.2833	0.0944	1.7817	0.2897
Pure error	4	0.2120	0.0530		

Stationary point of response surface:

x1	x2
0.3892604	0.3058577

Eigenanalysis:

eigen() decomposition

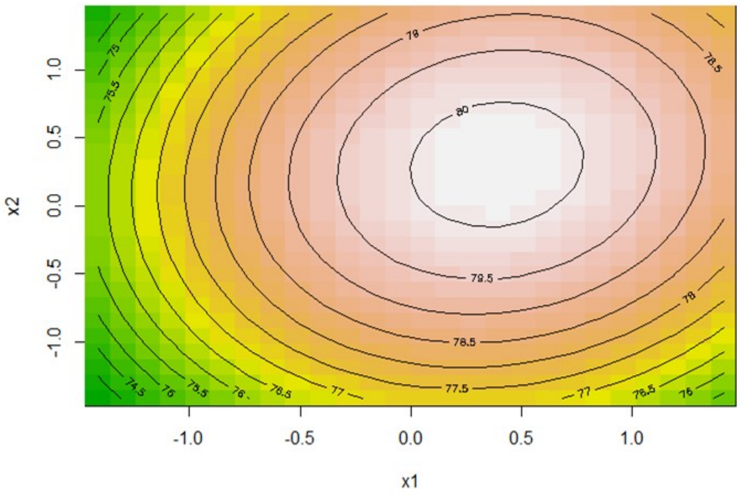
$values

[1] -0.963403 -1.414097

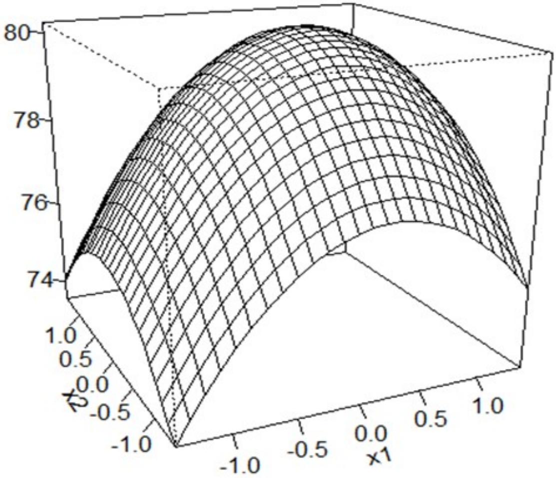
$vectors

	[,1]	[,2]
x1	-0.2897841	-0.9570920
x2	-0.9570920	0.2897841

```
> contour(fit, ~ x1 + x2, image=TRUE)
```



```
> persp(fit, x2 ~ x1)
```



```
> design=bbd(3,n0=4,randomize=F) # Box-Behnken Design
   run.order std.order x1.as.is x2.as.is x3.as.is
1          1         1       -1       -1        0
2          2         2        1       -1        0
3          3         3       -1        1        0
4          4         4        1        1        0
5          5         5       -1        0       -1
6          6         6        1        0       -1
7          7         7       -1        0        1
8          8         8        1        0        1
9          9         9        0       -1       -1
10        10        10        0        1       -1
11        11        11        0       -1        1
12        12        12        0        1        1
13        13        13        0        0        0
14        14        14        0        0        0
15        15        15        0        0        0
16        16        16        0        0        0
```

```
> Y=c(6,14,13,16,10,10,5,13,11,9,14,15, 17, 16, 16, 17)
> data=cbind(design, Y)
> fit=rsm(Y ~ SO(x1, x2, x3), data)
> summary(fit)
Call:
rsm(formula = Y ~ SO(x1, x2, x3), data = data)
              Estimate  Std. Error  t value   Pr(>|t|)
 (Intercept)  16.50000  1.11337     14.8199   5.936e-06 ***
 x1           2.37500   0.78727     3.0168    0.02349 *
 x2           1.00000   0.78727     1.2702    0.25104
 x3           0.87500   0.78727     1.1114    0.30892
 x1:x2        -1.25000  1.11337     -1.1227   0.30446
 x1:x3        2.00000   1.11337     1.7964    0.12257
 x2:x3        0.75000   1.11337     0.6736    0.52563
 x1^2         -3.50000  1.11337     -3.1436   0.01998 *
 x2^2         -0.75000  1.11337     -0.6736   0.52563
 x3^2         -3.50000  1.11337     -3.1436   0.01998 *
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1,x2,x3)	3	59.25	19.750	3.9832	0.07066
TWI(x1,x2,x3)	3	24.50	8.167	1.6471	0.27573
PQ(x1,x2,x3)	3	100.25	33.417	6.7395	0.02385
Residuals	6	29.75	4.958		
Lack of fit	3	28.75	9.583	28.7500	0.01036
Pure error	3	1.00	0.333		

Stationary point of response surface:

x1	x2	x3
0.3220621	0.5354767	0.2743902

Eigenanalysis:

eigen() decomposition

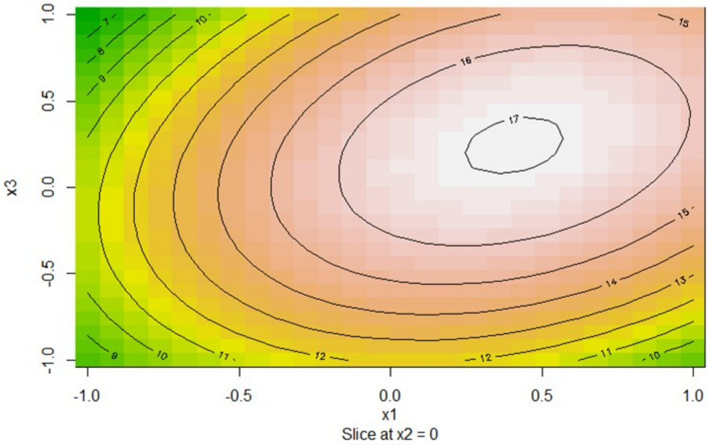
$values

[1] -0.6051339 -2.5154899 -4.6293761

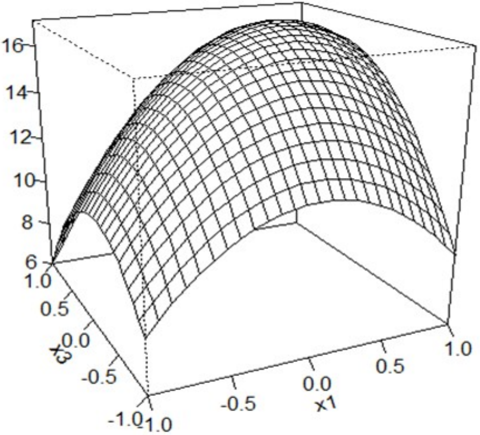
$vectors

	[,1]	[,2]	[,3]
x1	-0.19041565	0.68208513	0.7060466
x2	0.97979748	0.08724755	0.1799577
x3	0.06114563	0.72604941	-0.6849186

```
> contour(fit, ~ x1 + x3, image=TRUE)
```



```
> persp(fit, x3 ~ x1)
```



Slice at x2 = 0

Βιβλιογραφία

- Kuehl, R. (2000). *Design of experiments: statistical principles of research design and analysis*, 2nd ed. Pacific Grove (Calif.): Duxbury press.
- Montgomery, D. C. (2012). *Design and analysis of experiments*, 8th ed. Hoboken (N.J.): Wiley.