

Travaux dirigés sur les plans continus

Module de plan d'expériences - F. Husson - Agrocampus

Exercice 1 : DÉTERMINATION D'UN OPTIMUM : LA RÉACTION DE MAILLARD

On étudie la réaction de Maillard obtenue lors de la cuisson d'une protéine P avec un sucre S. Dans cette réaction, deux facteurs (notés X_1 et X_2) ont été identifiés comme importants : le ratio protéine/sucre et le pH de début de réaction. La variable réponse Y est une mesure de la capacité anti-oxydante du composé obtenu.

On recherche pour quelles valeurs de X_1 et X_2 la réponse Y est minimum. Un plan d'expérience a été mis en place grâce aux lignes de code suivantes :

```
library(rsm)
plan <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(ratio-30)/10, x2~(pH-7)/2))
Y <- c(20, 8, 28, 16, 14, 10, 25.4853, 8.5147, 9.3431, 20.6569, 14, 10)
```

1. *Expliciter le modèle utilisé classiquement pour analyser les résultats d'un tel plan.*
2. *Recopier les lignes de code permettant de générer le plan et de saisir les réponses dans R puis, à l'aide de la fonction `rsm` du package `rsm`, analyser les résultats. Quelle est l'erreur pure ? Que pouvez-vous dire sur l'ajustement du modèle ? Quel modèle conservez-vous ?*
3. *Ecrire le modèle finalement conservé pour la recherche de l'optimum et donner la valeur du couple (X_1, X_2) pour lequel la variable Y est minimum ? Donner une estimation de ce minimum.*
4. *A l'aide des fonctions `contour` ou `persp`, représenter les surfaces de réponses et dire si l'optimum est un minimum ou un maximum.*

Exercice 2 : OPTIMISATION DE CONCENTRATION CELLULAIRE

Un laboratoire biologique cherche à élaborer une solution ayant la plus grande concentration cellulaire possible. Les biologistes ont établi que la réponse obtenue (concentration cellulaire mesurée en $\mu\text{g/l}$) semble dépendre principalement de 5 facteurs : la température, le pH de la solution, la vitesse d'agitation, le taux d'oxygénation ainsi que la durée de la culture. Les diverses plages d'utilisation possibles sont

- Température en degré : 30 - 40
- pH : 6 - 8
- Vitesse agitation en tour/min : 100 - 200
- Taux d'oxygénation en % : 10-30
- Durée culture en h : 2 - 4

Pour étudier le phénomène on va mettre en œuvre un PCC mais on souhaite réaliser le moins d'essais possibles.

1. *Proposer un tel plan. Pour réduire le nombre d'essais, vous remplacez le bloc factoriel complet avec un bloc fractionnaire de résolution 5. Un tel plan peut être construit avec la ligne de commande :*

```
plan <- ccd(basis=~A+B+C+D, generators = E~A*B*C*D, randomize=FALSE)
```

2. *En revenant aux unités initiales, décrire à l'expérimentateur 3 expériences qu'il va réaliser (une par partie du PCC : factoriel, étoiles, centre)*
3. *Les expériences sont réalisées et la concentration cellulaire est obtenue. Par contre, l'expérimentateur n'a réalisé que 3 expériences des conditions (35, 7, 150, 20, 3). On totalise 29 essais. Les réponses sont collectées dans un vecteurs $Y=c(23.2, 27.9, 24.2, \dots)$. On concatène ce vecteur à la matrice des essais et on lance l'analyse suivante. Commenter les lignes de codes. Quel modèle est ajusté ?*

```
library(rsm)
reg <- rsm(Y~SO(Temp,pH,Vit,Oxy,Dur),data=dondon)
summary(reg)
```

Call:

```
rsm(formula = Y ~ SO(Temp, pH, Vit, Oxy, Dur), data = dondon)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	67.152174	2.702443	24.8487	7.356e-09 ***
Temp	0.091667	0.999921	0.0917	0.9292111
pH	-2.308333	0.999921	-2.3085	0.0498044 *
Vit	-1.358333	0.999921	-1.3584	0.2113880
Oxy	-0.258333	0.999921	-0.2584	0.8026564
Dur	1.575000	0.999921	1.5751	0.1538761
Temp:pH	0.212500	1.224648	0.1735	0.8665535
Temp:Vit	2.125000	1.224648	1.7352	0.1209246
Temp:Oxy	0.612500	1.224648	0.5001	0.6304394
Temp:Dur	-1.587500	1.224648	-1.2963	0.2310174
pH:Vit	-0.850000	1.224648	-0.6941	0.5072845
pH:Oxy	-0.312500	1.224648	-0.2552	0.8050254
pH:Dur	-0.137500	1.224648	-0.1123	0.9133698
Vit:Oxy	2.150000	1.224648	1.7556	0.1172284
Vit:Dur	-0.050000	1.224648	-0.0408	0.9684335
Oxy:Dur	-1.512500	1.224648	-1.2350	0.2518527
Temp^2	-6.882609	0.988993	-6.9592	0.0001173 ***
pH^2	-8.007609	0.988993	-8.0967	4.004e-05 ***
Vit^2	-8.770109	0.988993	-8.8677	2.066e-05 ***
Oxy^2	-8.582609	0.988993	-8.6781	2.420e-05 ***
Dur^2	-9.982609	0.988993	-10.0937	7.917e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared: 0.9642, Adjusted R-squared: 0.8746

F-statistic: 10.76 on 20 and 8 DF, p-value: 0.0009082

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(Temp, pH, Vit, Oxy, Dur)	5	233.5	46.70	1.9462	0.1920
TWI(Temp, pH, Vit, Oxy, Dur)	10	243.3	24.33	1.0140	0.5020
PQ(Temp, pH, Vit, Oxy, Dur)	5	4689.0	937.80	39.0812	2.058e-05
Residuals	8	192.0	24.00		
Lack of fit	6	180.4	30.07	5.2085	0.1698
Pure error	2	11.5	5.77		

Stationary point of response surface:

Temp	pH	Vit	Oxy	Dur
-0.01836893	-0.14043266	-0.07679570	-0.03015076	0.08379137

4. Commenter le tableau d'analyse de la variance et l'erreur pure.

5. Au vu des résultats, que conseillez-vous de faire ?

6. On obtient les résultats suivants. Commenter :

```
reg <- rsm(Y~ FO(Temp, pH, Vit, Oxy, Dur) + PQ(Temp,pH,Vit,Oxy,Dur),data=dondon)
summary(reg)
```

Call:

```
rsm(formula = Y ~ FO(Temp, pH, Vit, Oxy, Dur) + PQ(Temp, pH,
  Vit, Oxy, Dur), data = dondon)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	67.152174	2.712942	24.7525	2.361e-15 ***
Temp	0.091667	1.003806	0.0913	0.92825
pH	-2.308333	1.003806	-2.2996	0.03366 *

```

Vit      -1.358333    1.003806   -1.3532    0.19275
Oxy      -0.258333    1.003806   -0.2574    0.79982
Dur      1.575000    1.003806    1.5690    0.13405
Temp^2   -6.882609    0.992835   -6.9323   1.769e-06 ***
pH^2     -8.007609    0.992835   -8.0654   2.182e-07 ***
Vit^2    -8.770109    0.992835   -8.8334   5.812e-08 ***
Oxy^2    -8.582609    0.992835   -8.6445   7.991e-08 ***
Dur^2    -9.982609    0.992835  -10.0546   8.212e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9188,      Adjusted R-squared:  0.8736
F-statistic: 20.36 on 10 and 18 DF,  p-value: 8.142e-08

Analysis of Variance Table
Response: Y

              Df Sum Sq Mean Sq F value    Pr(>F)
FO(Temp, pH, Vit, Oxy, Dur)  5  233.5    46.70   1.9311    0.1389
PQ(Temp, pH, Vit, Oxy, Dur)  5 4689.0   937.80  38.7793 5.043e-09
Residuals                18   435.3    24.18
Lack of fit               16   423.7    26.48   4.5873    0.1935
Pure error                 2    11.5     5.77

Stationary point of response surface:

      Temp      pH      Vit      Oxy      Dur
0.006659297 -0.144133750 -0.077441077 -0.015049814  0.078887195

Eigenanalysis:
$values
[1] -6.882609 -8.007609 -8.582609 -8.770109 -9.982609

```

Exercice 3 : PLAN DE DOEHLERT POUR DEUX FACTEURS : UNE BONNE GALETTE BRETONNE

On cherche à mettre au point une nouvelle recette de galette bretonne (de blé noir) en incorporant du lait dans la fabrication de la pâte (dans la recette traditionnelle la farine n'est mélangée qu'avec de l'eau). Un premier facteur (noté X_1) est donc le pourcentage de lait dans le liquide que l'on mélangera avec la farine : *a priori*, il est raisonnable de fixer les limites de variation de ce facteur entre 0 et 20%.

Dans l'étude de ce facteur, il apparaît nécessaire de faire varier simultanément la fluidité de la pâte, en faisant varier la quantité de farine mélangée à un volume fixe de liquide. On obtient ainsi un second facteur (noté X_2) dont la plage de variation s'étend de 460g à 540g de farine par litre de liquide.

Pour évaluer ces recettes, on fabriquera des galettes que l'on soumettra à un jury. Chaque juge donnera une note d'appréciation globale à chaque galette : la réponse Y sera la moyenne des notes données par le jury.

On recherche donc la combinaison $\{X_1, X_2\}$ des facteurs conduisant à une valeur maximum de la réponse Y . Enfin, on dispose de peu de temps pour cette expérience et l'on accordera une grande importance à limiter le nombre d'essais.

Afin de déterminer l'optimum cherché, un collègue vous propose de réaliser un plan de Doehlert qu'il a déjà utilisé (cf. tableau 1).

Essai	X_1	X_2
1	0	0
2	1	0
3	-1	0
4	0.5	0.866
5	-0.5	-0.866
6	0.5	-0.866
7	-0.5	0.866

TABLE 1 – Plan de Doehlert pour deux facteurs

N'ayant pas d'indications sur les propriétés de ce plan, et n'en trouvant pas sur internet, vous décidez de les étudier vous-même directement. Pour cela, vous considérez le modèle incluant les effets linéaires, quadratiques et l'interaction (entre effets linéaires) :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$$

Etudiez les propriétés de ce plan en abordant les points suivants

1. *Orthogonalités ou confusions entre effets.*
 2. *Variance (des estimateurs) des effets. Pour la constante β_0 , dire à partir de quelle combinaison d'essais la variance de l'estimateur est calculée.*
 3. *Comparer avec un plan composite centré (utiliser la fonction `ccd` du package `rsm` pour le construire)*
- Il est possible d'ajouter des points au centre. Si l'on ajoute deux points au centre, on obtient un plan en 9 essais.
4. *Quelles modifications apportent ces deux points supplémentaires ?*
 5. *Conclure sur l'intérêt du plan de Doehlert pour deux facteurs.*

Exercice 4 : ETUDE DES PROPRIÉTÉS DES PLANS DE BOX-BENHKEN

On s'intéresse à la propriété des plans de Box-Benhken.

1. *A l'aide de la fonction `bdd` du package `rsm`, construire un plan de Box-Benhken mettant en jeu 3 variables explicatives. On considérera 2 points au centre pour ce premier plan.*
2. *A l'aide de la fonction `model.matrix`, évaluer la qualité du plan construit en étudiant notamment les orthogonalité et la précision des estimateurs des paramètres d'un modèle complet mettant en jeu tous les effets linéaires, les effets quadratiques et les interactions entre les variables prises 2 à 2 (on pourra charger le package `AlgDesign` et utiliser `~quad(.)` pour définir un tel modèle.*
3. *Faites varier le nombre de points au centre (1, 4, 10) et interpréter l'impact du nombre de points au centre sur la précision des estimateurs du modèle ainsi que l'orthogonalité entre tous les effets.*

Exercice 5 : CONSTRUCTION DE SURFACE DE RÉPONSE

On va simuler des données selon un modèle et étudier ce modèle ensuite. On donne les lignes de commandes suivantes :

```
set.seed(1234)
library(rsm)
PCC <- ccd(3)
X <- model.matrix(~quad(.), data=PCC[,3:5])
beta = c(1, -0.8, -2.5, -1, -0.4, -2, -1, 0.8, 3, 1)
Y = X%*%beta + rnorm(nrow(PCC))
```

1. *Commenter les lignes de code du programme ci-dessus.*
2. *A l'aide de la fonction `rsm` du package `rsm`, étudier les résultats de votre modèle. Quel modèle conservez-vous ?*
3. *Construire le modèle conservé à la question précédente. Tracer les surfaces de réponse à l'aide de la fonction `contour`. Par défaut, les surfaces sont représentées pour 2 variables, la troisième étant au centre du domaine. Tracer également les surfaces avec la troisième variable au point stationnaire (il suffit de préciser avec l'argument `at` où construire la surface de réponse). Comparer alors les surfaces de réponse au centre du domaine et au point stationnaire.*

Vous pourrez utiliser la ligne de code suivante pour avoir les 6 figures sur le même graphe :

```
par(mfrow=c(2,3))
```

Faire de même avec la fonction `persp`.