Construction de plans sous R

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

1 Plans fractionnaires à 2 modalités

Construction du plan fractionnaire

```
library(FrF2)
plan1 <- FrF2(nruns=8, nfactors=4, factor.names=list(temp=c("min","max"),
     press=c("low","normal"),material=c("M1","M2"),state=c("new","aged")))
plan2 <- FrF2(nfactors=5, resolution=5)
summary(plan2)
```

Vérification de la qualité d'un plan selon le modèle voulu

Ici pour le modèle avec les effets prncipaux et les interactions A:B et A:C.

```
options(contrasts=c("contr.sum","contr.sum"))
X <- model.matrix(~A+B+C+D+E+A:B+A:C, data=plan2)
solve(t(X)%*%X)
```

2 Plans continus

Construction de plan composite centré ou de Box Benhken

```
library(rsm)
plan <- ccd(2)   # donne le plan composite centré standard
plan<-ccd(2, coding=list (x1~(Temp-130)/10, x2~(Duree-50)/10))

Benhken <- bbd(3)  # plan Box Benhken
```

Vérification de la qualité d'un plan selon le modèle voulu

Ici pour le modèle

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i$$

```
X <- model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=plan)
solve(t(X)%*%X)
```

Dépouillement d'un plan selon le modèle voulu

```
CR.rsm <- rsm(Y~SO(x1,x2),data=plan) ## SO pour 2nd order
summary(CR.rsm)                      ##  FO(x1,x2)+TWI(x1,x2)+PQ(x1,x2)
contour(CR.rsm,~x1+x2,image=TRUE)
persp(CR.rsm,~x1+x2,col=rainbow(50), contours="colors")
```

3 Plan pour variables qualitatives à plus de 2 modalités

```
# Construction de plans orthogonaux
library(DoE.base)
fac.design(nlevels=c(4,3,3,2))  # nb d'essais calculé pour plan fractionnaire
fac.design(nlevels=c(2,2,3,3,6), blocks=6, seed=12345)
oa.design(nlevels=c(2,2,2,3,3,3), nruns=36) # plan orthogonal
```

```
# Comparer 6 variétés (A,B,C,D,E,F) en utilisant 4 blocs de 6 parcelles
plan <- oa.design(nlevels=c(6,4),factor.names=list(variete=LETTERS[1:6], bloc=1:4))

# Comparer 3 variétés (A,B,C), 2 doses d'engrais (1,2) en 6 lignes * 6 col
plan <- oa.design(nlevels=c(6,6,3,2))

library(planor)   # Pour aller vraiment bceaucoup plus loin !
# Analyse de sensibilité d'un modèle d'épidémiologie animale
# Plan d'expérience numérique : 12 facteurs à 4 niveaux, 7 facteurs à 2 niv
# Modèle : effets principaux + interactions entre 2 facteurs
# 4^12*2^7=2^31 combinaisons possibles, moins de 2^12 = 4096 expé
frac.key <- planor.designkey(factors = LETTERS[1:19], nlevels = c(rep(4,12),rep(2,7)),
     model = ~(A+B+C+D+E+F+G+H+I+J+K+L+M+N+O+P+Q+R+S)^2, nunits = 4096)
frac.plan <- planor.design(frac.key)
```

4 Plans optimaux

```
# EXEMPLE 1: modèle quadratique avec 3 variables
library(AlgDesign)
dat <- gen.factorial(levels=3,nVars=3,varNames=c("A","B","C"))
desD <- optFederov(~quad(A,B,C),dat,nTrials=14,eval=TRUE)

levels<-seq(-1,1,by=.1)
dat <- expand.grid(list(A=levels,B=levels,C=levels))
desL <- optFederov(~quad(.),dat,nTrials=14,eval=TRUE)

# EXEMPLES 2 : plan orthogonal de Plackett-Burman
dat <- gen.factorial(levels=2,nVars=11,varNames=LETTERS[1:11])
desPB <- optFederov(~.,dat,12,nRepeats=20)
X <- model.matrix(~.,data=desPB$design) ## pour vérifier l'orthogonalité
t(X)%*%X

# EXEMPLE 3: essais imposés OU ajout d'essais
dat<-gen.factorial(levels=3,nVars=3,varNames=c("A","B","C"))
desD <- optFederov(~quad(.),dat,nTrials=14,eval=TRUE)

desA <- optFederov(~quad(.),dat,nTrials=20,augment=TRUE,rows=desD$rows) ## ajout d'essais
```

Travaux dirigés sur les plans fractionnaires

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

Exercice 1 : Confusions dans un plan d'expériences 2^{3-1}

On étudie l'influence de trois facteurs qualitatifs A, B, C, présentant chacun 2 modalités (notées 1 et 2) sur une réponse Y à partir de 4 essais. Les conditions de l'expérience sont présentées dans le tableau 1.

Essai	A	B	C
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

Table 1 – Présentation du plan pour 3 facteurs

1. Écrire la matrice des essais.

2. Écrire la matrice des effets associée au modèle contenant tous les effets principaux et toutes les interactions (d'ordre 2 et 3). En déduire la liste des confusions de ce plan puis la résolution du plan.

3. Quel modèle peut-on étudier avec ce plan ? Quelles hypothèses doit-on poser ?

Exercice 2 : Plan 2^{6-2}

L'objectif de cet exercice est de construire un plan de résolution IV permettant d'étudier 6 facteurs à 2 niveaux en 16 essais.

1. Combien d'essais sont nécessaires pour construire un plan complet à 6 facteurs ?

2. Quel plan de base choisissez-vous pour étudier les 6 facteurs en 16 essais ?

On se propose d'étudier le modèle saturé associé à ce plan. La matrice des effets contient les colonnes suivantes :

I	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234

3. Avec quels effets choisissez-vous de confondre les facteurs 5 et 6 ? Déduire toutes les autres confusions du plan. Quelle est la résolution de ce plan ?

4. Peut-on estimer sans ambiguïté tous les effets principaux et toutes les interactions d'ordre 2 ?

5. Combien doit-on réaliser d'essais (au minimum) pour pouvoir estimer sans ambiguïté tous les facteurs principaux et toutes les interactions d'ordre 2 ? Et quelles hypothèses doit-on faire ?

Exercice 3 : Dépouillement d'un plan d'expériences pour 4 facteurs

On étudie l'influence de 4 facteurs $F1$, $F2$, $F3$ et $F4$, présentant chacun 2 modalités (notées 1 et 2) sur une réponse Y. Les conditions et les résultats de l'expérience sont regroupés dans le tableau 2. Ce tableau présente également la matrice des essais recodée : la modalité « 2 » est simplement remplacée par la modalité « -1 ».

1. Commenter brièvement ce plan : les effets principaux sont-ils orthogonaux ? quelle est la résolution de ce plan ? ce plan est-il fractionnaire ?

2. Quelle méthode permet de dépouiller ces résultats ?

3. On s'intéresse uniquement aux effets principaux. Écrire le modèle correspondant sous forme indicée.

4. Saisir les données puis faire l'analyse avec R. Utiliser la fonction `AovSum` de `FactoMineR`.

Essai	$F1$	$F2$	$F3$	$F4$	Y
1	1	1	1	1	29
2	1	1	2	2	1
3	1	2	1	2	-3
4	1	2	2	1	5
5	2	1	1	2	-9
6	2	1	2	1	3
7	2	2	1	1	-5
8	2	2	2	2	-21

$F1$	$F2$	$F3$	$F4$
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

Table 2 – Plan proposé pour 4 facteurs et matrice des essais recodée

Exercice 4 : Construction d'un plan 2^{8-4}

On veut construire un plan à 8 facteurs à 2 niveaux tels que les effets principaux soient confondus avec des interactions d'ordre 3 ou plus.

1. Donner les générateurs d'alias d'un tel plan.

2. Les effets principaux et les interactions d'ordre 2 sont-elles estimables sans ambiguïté ?

3. Est-il possible de tester quelques interactions d'ordre 2 avec le plan que vous avez construit ? Si oui, combien ? Si toutes les interactions d'ordre 2 ne peuvent pas être testées, comment faire pour pouvoir toutes les tester ?

4. Construire ce plan sous R en utilisant la fonction `FrF2` du package `FrF2`. Calculer la matrice $(X'X)^{-1}$ quand vous utilisez ce plan et que vous construisez un modèle avec seulement les effets linéaires (utiliser la fonction `model.matrix` pour calculer X). Refaire la même chose en ajoutant au modèle les interactions entre les facteurs A-B, A-C. Interpréter.

Exercice 5 : Comment peser avec précision ?

On s'intéresse à la précision des mesures de pesées effectuées par une balance de Roberval (voir figure 1).

Figure 1 – Balance de Roberval

On désire peser 4 objets A, B, C et D avec le maximum de précision. On sait que pour une pesée, la précision avec cette balance est de l'ordre de 1 gramme (i.e. l'écart-type est de 1 gramme, indication fournie par le constructeur). On va utiliser la méthodologie des plans d'expériences en considérant qu'un objet correspond à 1 facteur. On utilisera la convention suivante : si l'objet est déposé dans le plateau de gauche, le poids est compté positivement ; s'il est déposé dans le plateau de droite, le poids est compté négativement ; et si l'objet n'est sur aucun plateau, son poids n'est pas compté.
La matrice des essais correspondant à la pesée « classique » dans lequel on pèse les objets séparément est donnée au tableau 3.

Essai	A	B	C	D
pesée 1	1	0	0	0
pesée 2	0	1	0	0
pesée 3	0	0	1	0
pesée 4	0	0	0	1

Table 3 – Présentation du plan pour la pesée classique

On décide de faire les pesées du tableau 4.

Essai	Objets sur le plateau gauche	Objets sur le plateau droit	Poids ajoutés sur le plateau droit
pesée 1	A, B, C, D	aucun	150
pesée 2	A, B	C, D	50
pesée 3	A, C	B, D	10
pesée 4	A, D	B, C	30

Table 4 – Pesées pour 4 essais

Les résultats des pesées (en gramme) sont les suivantes : $Y_1 = 150$, $Y_2 = 50$, $Y_3 = 10$, $Y_4 = 30$.

1. Donner le plan d'expériences qui a été utilisé lors de ces 4 pesées sous forme de matrice.

2. Après avoir précisé à quel paramètre correspondait le poids de chaque objet, donner une estimation du poids de chaque objet.

3. Comment peut-on calculer la précision des estimations de poids ? Quelle est la précision de chacune des estimations de poids (sachant que la précision de la balance est $\sigma = 1$ gramme) ?

Exercice 6 : Analyse conjointe (ou Trade off) en marketing

L'analyse conjointe (ou Trade off) est souvent utilisée en marketing pour analyser le marché avant de lancer un nouveau produit. L'idée est de choisir parmi plusieurs produits celui qui sera le mieux perçu par les consommateurs. Le principe de l'analyse conjointe est le suivant : on demande à des consommateurs de classer plusieurs produits concurrents ou d'accorder une note à ces produits.
On veut choisir un emballage pour une nouvelle recette de crème caramel. Plusieurs critères (7) sont plus particulièrement étudiés :

- Le parfum : le parfum (caramel) est écrit en gros ou la marque est écrite en gros
- la DLC : la date limite de consommation est indiquée sur le couvercle ou sur le côté
- mention magnésium : la mention « le magnésium est bon pour la santé » est indiquée ou non
- Produit français : la mention « fabriqué en France » est indiquée ou non
- la couleur : une seule couleur (couleur caramel) ou plusieurs couleurs
- la taille des caractères : petite ou grande taille
- le fond : une photo en fond ou rien

Afin de réaliser l'analyse conjointe, l'entreprise peut faire des emballages correspondant à n'importe quelle combinaison des 7 critères. Cependant, la construction d'un emballage coûte cher et il n'est pas raisonnable de faire classer plus de 8 produits par un consommateur.

1. On décide de réaliser un plan en 8 essais. Quel est le nom du plan à construire ?

2. Construire le plan sous R. Donner les générateurs d'aliases et la résolution du plan.

3. On suspecte une interaction possible entre les critères « mention magnésium » et « produit français » (si les deux mentions sont présentes, l'emballage est trop surchargé de mentions). Avec le plan que vous avez construit, pouvez-vous analyser cette interaction ? Si oui, comment faites-vous ? Si non, comment feriez-vous ?

Une fois ces emballages fabriqués, 100 consommateurs sont interrogés et doivent mettre une note comprise entre 0 (l'emballage ne me plaît pas) et 10 (l'emballage me plaît beaucoup) pour chacun des emballages (il y a alors 800 données).

4. Comment analysez-vous ces résultats (préciser le modèle utilisé) ?

Exercice 7 : Plan $L_8 2^3 4$ et $L_8 2^4 4$

On étudie un processus au travers d'une variable réponse Y. A priori, 4 facteurs influent sur Y :
- 3 facteurs ($F1$, $F2$ et $F3$) à deux modalités chacun ;
- 1 facteur ($F4$) à 4 modalités.

On néglige a priori toutes les interactions.
On cherche à construire un plan en 8 essais permettant d'étudier simultanément ces 4 facteurs. Pour que ces 4 facteurs soient orthogonaux, on se fonde sur la matrice des effets (du modèle complet) du plan combinant $F1$, $F2$ et $F3$. On sait que deux colonnes de cette matrice définissent un facteur à 4 modalités (par exemple en notant ; $- - = 1$; $- + = 2$; $+ - = 3$; $+ + = 4$).

1. *À partir de ce principe, construire un plan tel que les effets principaux des 4 facteurs $F1$, $F2$, $F3$ et $F4$ sont orthogonaux. Ce plan est noté $L_8 2^3 4$ car il y a 8 essais, 3 facteurs à 2 niveaux et un facteur à 4 niveaux.*

2. *Avec quoi (quel facteur ou quelle interaction) est confondu le facteur à 4 niveaux ?*

3. *Le plan $L_8 2^4 4$ existe-t-il ? Si oui, comment le construire ?*

Exercice 8 : Répartition des expériences sur 2 équipes de travail

Vous voulez étudier l'influence de 3 facteurs qualitatifs $F1$, $F2$ et $F3$ présentant chacun 2 modalités :
- $F1$ est le facteur température qui peut prendre les valeurs 30 et 45 degrés ;
- $F2$ est le facteur pression qui peut prendre les valeurs 2 et 4 bars ;
- $F3$ est la vitesse de rotation de l'appareil qui peut prendre les valeurs 900 tours/minute et 1200 tours/minute.

On peut effectuer 8 expériences donc réaliser un plan complet. Vous voulez alors répartir ces 8 essais pour 2 équipes de travail (4 essais par équipe).

Les deux équipes pouvant travailler de façon différente, décrire précisément les 4 essais que vous décidez de faire réaliser par chacune des équipes de travail.

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 $\mu g/l$) 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,....). 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 classique (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. Comparer avec un plan composite centré ayant 2 points au centre (utiliser l'argument $n0 = 1$ dans la fonction `ccd`) puis 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 `bbd` 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 :

```
library(rsm)
library(AlgDesign)
PCC <- ccd(3, randomize=FALSE)
X <- model.matrix(~quad(.),data=PCC[,3:5])
beta <- c(1,-0.8,-2.5,-1,-0.4,-2,-1,0.8,3,1)
set.seed(1234)
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`.