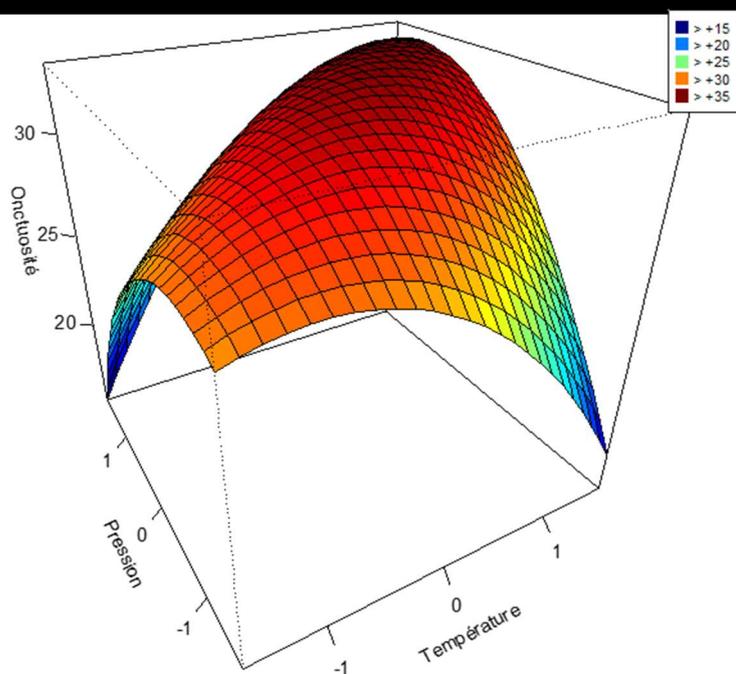


Recueil d'exercices sur les plans d'expériences



Les exercices présents dans ce recueil ont été construits par les étudiants de l'Institut Agro (Rennes) dans le cadre d'un module de planification expérimentale.

Ils sont le résultat de leur apprentissage sur ce cours, mais aussi de leur expérience acquise lors de stage. Certains exercices sont issus d'un contexte réel, avec des données réelles, d'autres exercices sont le fruit de leur imagination.

Exercices proposés par les élèves ingénieurs suivant le module de planification expérimentale

Enseignant responsable : François Husson
Institut Agro - Rennes

Juin 2023

Table des matières

Exercices sur les plans fractionnaires	3
Exercice 1. Préparation de cookies (exemple fictif, par Nassira Chupeau)	3
Exercice 2. Evaluation de l'impact du pâturage porcin sur la qualité des sols (EXERCICE FICTIF).....	5
Exercices sur les plans continus et surface de réponses	6
Exercice 1. Optimisation d'un élevage de Cardina CF cantonesis (exemple fictif, par Bastien DELEUZE)	6
Exercice 2. Un distributeur automatique « flash » ou pas « flash » ? (plans continus, exemple très fictif, Eric KAPISIZ) 11	
Exercice 3. Comment se débarrasser des limaces ? (Exemple fictif, par Morgan LEBOURDAIS)	14
Exercice 4. Impact de la température et de l'apport d'eau sur la croissance d'un plant de tomate (EXERCICE FICTIF) 18	
Exercice 5. Développement de burgers végétariens à base de légumineuses fermentées (exemple issu d'une situation réelle, par Adèle Husson)	20
Exercice 6. Élaboration d'une recette de falafels à base d'haricots mungo (exemple fictif par Kim Lenotte)	24
Exercices sur les plans optimaux	29
Exercice 1. Améliorer le croquant du chocolat (exemple fictif, par Nassira Chupeau)	29
Exercice 2. Stratégies de marketing et distributeurs automatiques (plans optimaux, exemple fictif, Eric KAPISIZ) ...	32
Exercice 3. La meilleure recette de gnocchis (Exemple fictif, par Morgan LEBOURDAIS).....	33
Exercice 4. Développement de burgers végétariens à base de légumineuses (exemple issu d'une situation réelle, par Adèle Husson) 35	
Exercice 5. Amélioration de la texture d'un caviar d'aubergines (exemple fictif par Kim Lenotte).....	38
Exercice 6. Boissons rafraîchissantes (exemple réel, par Alban Krauffel)	40
Exercices divers	54
Exercice 1. Caractérisation d'une boisson par les sensations trigéminales perçues relatives à l'éthanol (exemple fictif, par Bastien DELEUZE).....	54
Exercice 2. Cauchemar en cuisine (thématique réelle - DAVID Solène).....	57

Exercices sur les plans fractionnaires

Exercice 1. Préparation de cookies (exemple fictif, par Nassira Chupeau)

Ce week-end, Anne décide de réaliser une recette de cookies écrite dans un vieux carnet de notes appartenant à sa grand-mère. Malheureusement, les pages du livret se sont abîmées avec le temps et l'encre s'est estompée, rendant difficile la lecture des recettes. Avec l'aide de sa mère, Anne est parvenue à décrypter la liste des ingrédients ainsi que les étapes de préparation à suivre. Seul problème, elle ne parvient pas à se mettre d'accord avec sa mère sur ce qui est écrit concernant les informations suivantes : la température de cuisson, la quantité de farine, de cassonade, de beurre ainsi que de pépites de chocolat au lait.

Voici le tableau énumérant ce qu'Anne et sa mère parviennent à lire :

Informations	Code	Anne	Sa mère
Température de cuisson (°C)	A	160	180
Quantité de farine (g)	B	235	265
Quantité de cassonade (g)	C	170	110
Quantité de beurre (g)	D	100	160
Quantité de pépites de chocolat (g)	E	180	150

- Si Anne et sa mère souhaitent réaliser l'ensemble des recettes possibles en prenant en compte leurs avis respectifs, combien de pâtes à cookies différentes devraient-elles préparer ?

Cette situation peut se résumer à la réalisation d'un plan d'expériences à 5 facteurs de 2 modalités chacun. Il faudrait donc $2^5 = 32$ essais au total pour construire le plan complet.

Malheureusement, Anne et sa mère n'ont ni le temps ni les ingrédients en quantité suffisante pour réaliser l'ensemble de ces essais. Elles décident de passer l'après-midi en cuisine et se limitent à la préparation de 8 pâtes à cookies.

- Quel est le nom de ce nouveau plan d'expérience à construire ?

Le nouveau plan à construire est un plan 2^{5-2} .

- Quels sont les effets pouvant être estimés grâce à ce plan ?

Ce plan 2^3 permet d'estimer les effets :

I	A	B	C	AB	AC	BC	ABC
---	---	---	---	----	----	----	-----

- En déduire les confusions résultantes pour les facteurs manquants.

Nous devons ajouter 2 facteurs au plan de base pour avoir les 5 effets. Il faut donc confondre les facteurs D et E avec des interactions d'ordre le plus élevé -1. D'après la question précédente, cela revient à confondre le facteur D avec AB et le facteur E avec l'interaction AC.

- Quelle est la résolution du plan et qu'induit le choix de se limiter à seulement 8 essais ?

Les effets principaux sont confondus avec des interactions d'ordre 2 ou plus, le plan est donc de résolution III.

Pour estimer ces effets principaux, il faut donc émettre l'hypothèse que les interactions d'ordre 2 et plus sont négligeables.

- En vous aidant du logiciel R et de la fonction FrF2, détaillez les 8 essais à suivre pour Anne et sa mère.

```
library(FrF2)
```

```
plan <-FrF2 (nruns=8,nfactors=5, factor.names=list(temperature=c("160","180"),  
farine=c("235","265"),cassonade=c("170","110"),beurre=c("100","160"),  
chocolat=c("180","150"))  
summary(plan)
```

Call:

```
FrF2(nruns = 8, nfactors = 5, factor.names = list(temperature = c("160","180"),farine =  
c("235","265"),cassonade = c("170","110"),  
beurre =c("100","160"),chocolat = c("180","150"))
```

Experimental design of type FrF2

8 runs

Factor settings (scale ends):

	temperature	farine	cassonade	beurre	chocolat
1	160	235	170	100	180
2	180	265	110	160	150

Design generating information:

\$legend

[1] A=temperature B=farine C=cassonade D=beurre E=chocolat

\$generators

[1] D=AB E=AC

Alias structure:

\$main

[1] A=BD=CE B=AD C=AE D=AB E=AC

\$fi2

[1] BC=DE BE=CD

The design itself:

	temperature	farine	cassonade	beurre	chocolat
1	180	235	110	100	150
2	180	265	170	160	180
3	160	235	110	160	180
4	160	265	170	100	150
5	180	235	170	100	180
6	180	265	110	160	150
7	160	235	170	160	150
8	160	265	110	100	180

class=design, type= FrF2

Le soir même, Anne et sa mère décident de goûter les cookies fraîchement préparés afin de déterminer leur recette préférée. Pour cela, chacune goûte un cookie par essai et lui accorde une note d'appréciation globale.

- Si nous voulions analyser ces résultats, quel modèle devrions-nous utiliser ?

Pour analyser ces résultats, il faudrait effectuer une analyse de variance avec la variable Y la note d'appréciation et 6 variables explicatives comprenant les 5 facteurs ainsi que l'effet juge, soit :

$$Y_{ijklmn} = \mu + A_i + B_j + C_k + D_l + E_m + F_n + \epsilon_{ijklmn}$$

Avec A_i l'effet de la température, B_j l'effet de la quantité de farine, C_k l'effet de la quantité de cassonade, D_l l'effet de la quantité de beurre, E_m l'effet de la quantité chocolat, F_n l'effet juge et ϵ_{ijklmn} la résiduelle.

Exercice 2. Evaluation de l'impact du pâturage porcin sur la qualité des sols (EXERCICE FICTIF)

Le pâturage porcin est peu développé, notamment à cause de l'attrait des porcs pour gratter le sol. Cette perturbation crée des reliefs difficiles à réajuster par la mécanisation. Certaines pratiques permettent de limiter le risque de dégradation des parcelles mais ces pratiques sont encore peu étudiées. On cherche à évaluer l'impact sur le sol de la destruction d'un couvert par pâturage porcin.

On étudie 6 variables à 2 modalités : le stade du couvert (tallage ou épiaison), le sexe, la nature du couvert (radis fourrager ou topinambour), la nature du sol (limoneux ou argileux), la saison (pluvieuse ou sèche), la race des cochons (Duroc et Large White).

1) Combien d'essais faut-il pour construire le plan complet ?

Pour étudier le plan complet à 6 facteurs à 2 modalités on doit faire : $2^6=62$ essais

2) On a une surface limitée donc on ne pourra faire que 16 essais. Quel plan est adapté ?

C'est un plan $L_{16}^{2^{6-2}}$

Avec ce type de plan il est possible de faire seulement 16 essais car $2^4 = 16$

3) Donner les générateurs d'alias et la résolution du plan :

On doit ajouter les facteurs supplémentaires 5 et 6 au plan de base, ils seront confondus avec les interactions d'ordre le plus élevé-1 (d'ordre 3). On choisit de confondre le facteur 5 avec 123 et le facteur 6 avec 124

On obtient les générateurs d'alias suivants :

$$I_a = 1235$$

$$I_b = 1246$$

$$I_{ab} = 3546$$

On a donc un plan de résolution 4 car on choisit l'interaction d'ordre la plus basse.

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

Un plan de résolution 4 confond les interactions d'ordre 2 entre elles. En négligeant les interactions d'ordre 3 on pourra estimer tous les effets principaux car ils sont confondus.

5) Construire le plan sous R et vérifier la qualité du plan

On commence par construire le plan puis nous créerons une matrice pour vérifier la qualité du plan.

```
library(FrF2)
plan <- FrF2(nruns=16, nfactors=6, factor.names=list(avancement_couvert=c("tallage", "epiaison"),
                                                    sexe=c("H", "F"), nature_couvert=c("radis
fourrager", "topinambour"), nature_du_sol=c("limoneux", "argileux")
, saison=c("pluvieuse", "seche"), race_cochons=c("Duroc", "Large white")))
summary(plan)
```

Call:

```
FrF2(nruns = 16, nfactores = 6, factor.names = list(avancement_couvert = c("tallage",
"epiaison"), sexe = c("H", "F"), nature_couvert = c("radis fourrager",
"topinambour"), nature_du_sol = c("limoneux",
"argileux"), saison = c("pluvieuse", "seche"),
race_cochons = c("Duroc", "Large white")))
```

```
Experimental design of type FrF2
16 runs
```

Factor settings (scale ends):

```
avancement_couvert sexe nature_couvert nature_du_sol saison race_cochons
1 tallage H radis fourrager limoneux pluvieuse Duroc
2 epiaison F topinambour argileux seche Large white
```

Design generating information:

\$legend

```
[1] A=avancement_couvert B=sexe C=nature_couvert D=nature_du_sol E=saison
F=race_cochons
```

\$generators

```
[1] E=ABC F=ABD
```

Alias structure:

\$fi2

```
[1] AB=CE=DF AC=BE AD=BF AE=BC AF=BD CD=EF CF=DE
```

The design itself:

```
avancement_couvert sexe nature_couvert nature_du_sol saison race_cochons
1 epiaison H topinambour argileux pluvieuse Duroc
2 tallage H radis fourrager limoneux pluvieuse Duroc
3 tallage F radis fourrager limoneux seche Large white
4 tallage H topinambour limoneux seche Duroc
5 tallage F topinambour limoneux pluvieuse Large white
6 epiaison F topinambour argileux seche Large white
7 epiaison F radis fourrager argileux pluvieuse Large white
8 tallage H topinambour argileux seche Large white
9 epiaison H radis fourrager limoneux seche Large white
10 epiaison F topinambour limoneux seche Duroc
11 tallage F topinambour argileux pluvieuse Duroc
12 epiaison H topinambour limoneux pluvieuse Large white
13 tallage F radis fourrager argileux seche Duroc
14 epiaison H radis fourrager argileux seche Duroc
15 tallage H radis fourrager argileux pluvieuse Large white
16 epiaison F radis fourrager limoneux pluvieuse Duroc
```

class=design, type= FrF2

On voit sur ce tableau les essais réalisés avec les modalités correspondantes pour chaque facteur.

```
X <- model.matrix(~.,data=plan) # Construire la matrice X
round(solve(t(X)%*%X),2)
```

```
(Intercept) (Intercept) avancement_couvert1 sexe1 nature_couvert1 nature_du_sol1 saison1 race_cochons1
avancement_couvert1 0.06 0.00 0.00 0.00 0.00 0.00 0.00
sexe1 0.00 0.06 0.00 0.00 0.00 0.00 0.00
nature_couvert1 0.00 0.00 0.06 0.00 0.00 0.00 0.00
nature_du_sol1 0.00 0.00 0.00 0.06 0.00 0.00 0.00
saison1 0.00 0.00 0.00 0.00 0.06 0.00 0.00
race_cochons1 0.00 0.00 0.00 0.00 0.00 0.00 0.06
```

Comme on le voit sur les éléments surlignés, la matrice est bien en diagonale donc le plan est orthogonal et de bonne qualité. En effet, on peut dire qu'un plan est orthogonal si les valeurs de la diagonale sont égales et les valeurs en dehors de la diagonale sont nulles.

Exercices sur les plans continus et surface de réponses

Exercice 1. Optimisation d'un élevage de Cardina CF cantonesis (exemple fictif, par Bastien DELEUZE)

L'exercice se base sur une problématique et des données fictives. Bob est un passionné d'aquariophilie et plus particulièrement Cardina CF cantonesis, une variété de crevettes très recherchée pour sa diversité de couleurs et de paternes de couleurs par les amateurs d'aquariophilie. L'élevage amateur et professionnel de cette espèce est une course effrénée dont le seul objectif est d'être le premier à obtenir une nouvelle mutation de souche, c'est-à-dire d'obtenir par croisement et reproduction des paternes de coloration d'exosquelette jusqu'alors inconnus (ci-dessous les principales souches connues à ce jour et une des souches les plus rares actuellement, dont le prix d'un individu peut être estimé à plusieurs centaines d'euro).

Par soucis de moyen, Bob ne peut dédier que 10 bacs à cette étude.

4. En utilisant R, présenter le plan des essais permettant d'établir les conditions d'élevage optimales.

```
library(rsm)
plan <- ccd(2, n0=1, randomize = F, coding = list(x1~(Temperature-22)/2, x2~(pH-7.25)/0.75))
plan
```

	run.order	std.order	Temperature	pH	Block
1	1	1	20.00000	6.50000	1
2	2	2	24.00000	6.50000	1
3	3	3	20.00000	8.00000	1
4	4	4	24.00000	8.00000	1
5	5	5	22.00000	7.25000	1
6	1	1	19.17157	7.25000	2
7	2	2	24.82843	7.25000	2
8	3	3	22.00000	6.18934	2
9	4	4	22.00000	8.31066	2
10	5	5	22.00000	7.25000	2

5. Commenter la qualité de ce plan.

Pour étudier la qualité de ce plan, on calcule la matrice $(X'X)^{-1}$. Pour cela, on crée la matrice des effets comprenant les effets linéaires, les effets d'interaction et les effets quadratiques.

```
X <- model.matrix(~quad(.), data=plan[,3:4])
round(solve(t(X)%*%X), 2)
```

	(Intercept)	x1	x2	I(x1^2)	I(x2^2)	I(x1 * x2)
(Intercept)	0.50	0.00	0.00	-0.25	-0.25	0.00
x1	0.00	0.12	0.00	0.00	0.00	0.00
x2	0.00	0.00	0.12	0.00	0.00	0.00
I(x1^2)	-0.25	0.00	0.00	0.22	0.09	0.00
I(x2^2)	-0.25	0.00	0.00	0.09	0.22	0.00
I(x1 * x2)	0.00	0.00	0.00	0.00	0.00	0.25

En analysant les résultats de la matrice $(X'X)^{-1}$, on peut observer que :

- L'interception est indépendante des effets principaux et des effets d'interactions et est faiblement liée aux effets quadratiques.
- Les effets principaux sont indépendants de tous les autres effets.
- Les effets d'interactions sont faiblement confondus entre eux
- Les effets quadratiques sont indépendants des autres effets.
- Au coefficient σ^2 près (non pris en compte car indépendant du plan), les variances des estimateurs des effets se lisent sur la diagonale de la matrice $(X'X)^{-1}$: elles sont hétérogènes. Si la variance des effets linéaires (X1 et X2) est la plus faible, la variance des estimations des coefficients reste relativement faible.

Après 1 an, Bob obtient les résultats suivants :

```
Y<- c(5.36, 11.78, 3.15, 5.26,12.95, 4.52, 12.63, 8.52, 4.23, 13.11)
```

6. Commenter les résultats de l'analyse des résultats.

On teste en premier l'ajustement du modèle complet aux résultats. Pour cela, on utilise ces lignes de code du R :

```
Y<- c(5.36, 11.78, 3.15, 5.26,12.95, 4.52, 12.63, 8.52, 4.23, 13.11)
CR.rsm <- rsm(Y~SO(x1,x2), data=plan)
summary(CR.rsm)
```

```
Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)

Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 13.03000    0.73693 17.6815 6.01e-05 ***
x1           2.49991    0.36846  6.7847 0.002464 **
x2          -1.84962    0.36846 -5.0198 0.007386 **
x1:x2       -1.07750    0.52109 -2.0678 0.107498
x1^2        -2.49937    0.48743 -5.1276 0.006850 **
x2^2        -3.59937    0.48743 -7.3844 0.001793 **
---
```

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

Multiple R-squared:  0.9712, Adjusted R-squared:  0.9351
F-statistic: 26.95 on 5 and 4 DF,  p-value: 0.003532
```

Analysis of Variance Table

```
Response: Y
      Df Sum Sq Mean Sq F value Pr(>F)
FO(x1, x2)  2  77.365   38.683   35.6153 0.002827
TWI(x1, x2)  1   4.644    4.644    4.2758 0.107498
PQ(x1, x2)   2  64.352   32.176   29.6245 0.004000
Residuals    4   4.344    1.086
Lack of fit   3   4.332    1.444  112.8044 0.069077
Pure error   1   0.013    0.013
```

Stationary point of response surface:

```
      x1      x2
0.5740103 -0.3428537
```

Stationary point in original units:

```
Temperature      pH
 23.14802      6.99286
```

Eigenanalysis:

```
eigen() decomposition
$values
[1] -2.279471 -3.819279
```

\$vectors

```
      [,1]      [,2]
x1 -0.9258442  0.3779053
x2  0.3779053  0.9258442
```

En étudiant les résultats, on peut affirmer que l'ajustement du modèle est bon car on obtient une valeur de R^2 de 0.9712, qui est une valeur très proche de 1 (ajustement parfait du modèle).

On peut aussi voir que la probabilité critique associée à l'ajustement du modèle est significative (p -value= 0.003532).

Si on s'intéresse maintenant à l'erreur pure, c'est dire à la variance de mesure estimée grâce aux répétitions de l'expérience dans de mêmes conditions, conditions expérimentales correspondant aux points au centre du Plan Composite Centré (T_{mp} = 22°C & pH = 7.250). Pour ces essais, l'erreur pure, c'est-à-dire la variance des résultats obtenus, est égale à 0.013.

L'erreur d'ajustement mesure elle l'écart entre les points moyens par condition expérimentale et les valeurs prédites correspondantes. Dans le cas de notre modèle, l'erreur d'ajustement vaut 4.332 et a une p -value associée de 0.069077. On en conclut donc que le déficit d'ajustement n'est pas significatif au seuil 5%, on a donc aucune raison de remettre en cause le modèle.

En étudiant maintenant le tableau d'analyse de la variance, on peut constater que tous les effets sont significatifs au seuil 5% à l'exception des effets d'interactions (p -value TWI(x₁,x₂) = 0.107498).

7. Après l'analyse des résultats, exprimer le modèle que vous conservez et donner sa surface de réponse

Tous les coefficients sont significatifs excepté celui de l'interaction. On retient donc le modèle sans interaction. Il n'est pas utile de réestimer les coefficients car avec un plan composite centré, l'interaction est orthogonale à tous les autres effets. On obtient donc :

$$Y = 13,03 + 2,50X_1 - 1,85X_2 - 2,50(X_1)^2 - 3,60(X_2)^2$$

```
CR.rsm <- rsm(Y~FO(x1,x2) + PQ(x1,x2), data=plan)
summary(CR.rsm)
contour(CR.rsm, ~FO(x1,x2) + PQ(x1,x2), image=T)
persp(CR.rsm, ~FO(x1,x2) + PQ(x1,x2), col=rainbow(50), contours="colors")
```

```
Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)

            Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.03000    0.73693  17.6815 6.01e-05 ***
x1           2.49991    0.36846   6.7847 0.002464 **
x2          -1.84962    0.36846  -5.0198 0.007386 **
x1:x2       -1.07750    0.52109  -2.0678 0.107498
x1^2        -2.49937    0.48743  -5.1276 0.006850 **
x2^2        -3.59937    0.48743  -7.3844 0.001793 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9712, Adjusted R-squared:  0.9351
F-statistic: 26.95 on 5 and 4 DF,  p-value: 0.003532
```

Analysis of Variance Table

```
Response: Y
            Df Sum Sq Mean Sq F value Pr(>F)
FO(x1, x2)  2  77.365   38.683   35.6153 0.002827
TWI(x1, x2) 1   4.644    4.644    4.2758 0.107498
PQ(x1, x2)  2  64.352   32.176   29.6245 0.004000
Residuals   4   4.344    1.086
Lack of fit  3   4.332    1.444  112.8044 0.069077
Pure error   1   0.013    0.013
```

Stationary point of response surface:

```
      x1      x2
0.5740103 -0.3428537
```

Stationary point in original units:

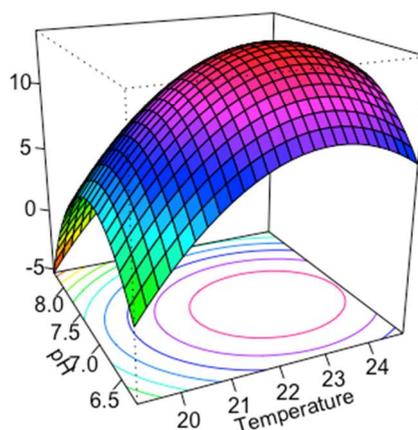
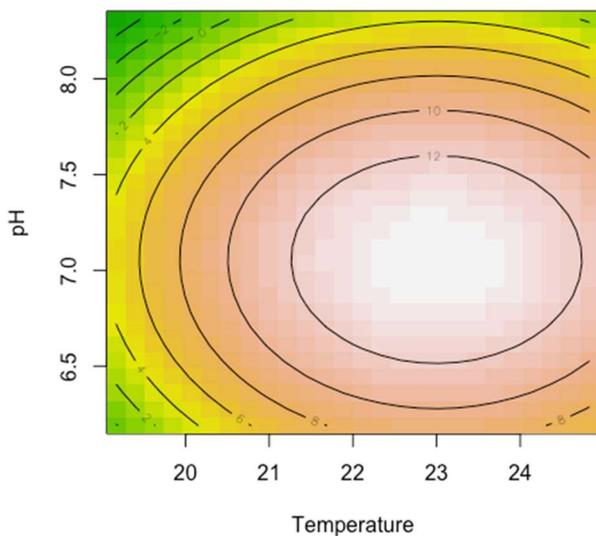
```
Temperature      pH
    23.14802     6.99286
```

Eigenanalysis:

```
eigen() decomposition
$values
[1] -2.279471 -3.819279
```

\$vectors

```
      [,1]      [,2]
x1 -0.9258442 0.3779053
x2  0.3779053 0.9258442
```



En étudiant des valeurs propres (eigen values), on constate que les valeurs propres de x_1 et x_2 sont des valeurs négatives, on peut donc en déduire que le point stationnaire est un maximum. On confirme cela en observant l'allure de la surface de réponse.

On peut donc chercher maintenant à déterminer le couple de valeur des facteurs Température/pH pour lequel Y est optimal. Pour cela, on annule les dérivées partielles de y soit :

$$dy/dx_1 = 2.49991 - 4.99982x_1 = 0 \quad dy/dx_2 = -1,85 - 7,2 x_2 = 0$$

On trouve donc $x_1 \approx 0,57$ et $x_2 \approx -0,34$

Ces valeurs de x correspondent à des valeurs de Température = 23,15 et de pH = 6,99.

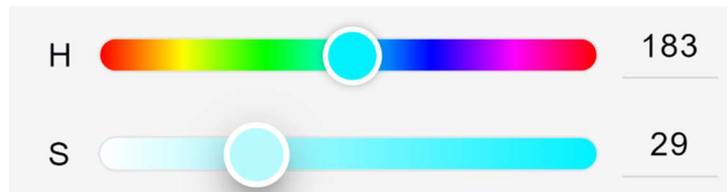
8. Formulez un avis à Bob pour lui permettre d'optimiser son élevage.

Les paramètres optimaux pour favoriser la reproduction de ses individus et donc d'obtenir le plus rapidement un grand nombre d'individus, en se positionnement d'un point de vue de réalité pratique de paramètres obtenables pour un éleveur « amateur » comme Bob, sont de régler ses chauffages ou bien la pièce d'élevage à une température de 23°C (on préférera une température plus froide que l'optimal permettant ainsi de tamponner une variabilité de la température sans engendrer un stress aux individus, stress souvent favorisé par une hausse de la température. Quant au pH, Bob devrait paramétrer l'eau de ses bacs pour obtenir un pH de 7.

Exercice 2. Un distributeur automatique « flash » ou pas « flash » ? (plans continus, exemple très fictif, Eric KAPISIZ)

CONTEXTE :

La même entreprise a cette fois-ci conduit une étude sur la couleur que devrait être un distributeur automatique afin d'être plus attirant. Un distributeur automatique en noir mate (témoin) illuminé sous lumière dans une salle obscure et un



distributeur peint en une teinte (variable *hue*, « H » sur la figure, valeurs allant de 0 à 360) et une saturation (variable *sat*, « S » sur la figure, valeurs allant de 0 à 100) spécifique. La figure ci-dessous illustre comment ces paramètres fonctionnent.

Voici le plan qui a été utilisé pour savoir comment peindre les distributeurs automatiques :

```
>plan <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(hue-180)/120, x2~(saturation-50)/35))
```

On a répété les mêmes expériences 10 fois. La réponse Y correspond au nombre de fois (pourcentage arrondi à 2 chiffres) que le distributeur « témoin » a été choisi à la place du distributeur teint.

```
> Y<-c(0.763,0.745,0.147,0.137,0.813,0.820,0.290,0.287,0.988,0.384,0.764,0.724)
```

QUESTIONS :

1. Analysez le plan d'expérience composite centré. Commentez. Emettez une hypothèse sur les effets et interactions qu'on pourrait observer.

-----REPONSE

Selon le composite distributeurs peints ont les suivantes :

	> plan	run.order	std.order	hue	saturation	Block
1	1	1	60.00000	15.0000000	1	
2	2	2	300.00000	15.0000000	1	
3	3	3	60.00000	85.0000000	1	
4	4	4	300.00000	85.0000000	1	
5	5	5	180.00000	50.0000000	1	
6	6	6	180.00000	50.0000000	1	
7	1	1	10.29437	50.0000000	2	
8	2	2	349.70563	50.0000000	2	
9	3	3	180.00000	0.5025253	2	
10	4	4	180.00000	99.4974747	2	
11	5	5	180.00000	50.0000000	2	
12	6	6	180.00000	50.0000000	2	

plan d'expérience centré, les distributeurs automatiques ont les couleurs

couleurs différentes	sat	hue
	15	60
	15	300
	85	60
	85	300
	50	180
	50	11
	50	350
	1	180
	100	180

Hypothèses : les couleurs saturées pourraient ressortir beaucoup plus car elles frappent à l'œil. Utiliser des couleurs saturées pour attirer l'attention est une technique marketing assez commune. Cependant, les couleurs pastel moins saturées pourraient également être favorisées car c'est des couleurs à la mode.

2. Analysez les résultats du plan construit.

-----REPONSE

```
>CR.rsm <- rsm(Y~SO(x1,x2),data=plan)
>summary(CR.rsm)
```

```
Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)

              Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  0.7802500  0.0331209  23.5576 3.839e-07 ***
x1           -0.0040303  0.0234200  -0.1721  0.86903
x2           -0.2597731  0.0234200 -11.0919 3.198e-05 ***
x1:x2         0.0020000  0.0331209   0.0604  0.95381
x1^2         -0.2556875  0.0261844  -9.7649 6.632e-05 ***
x2^2         -0.0569375  0.0261844  -2.1745  0.07261 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9733,    Adjusted R-squared:  0.951
F-statistic: 43.69 on 5 and 6 DF,  p-value: 0.0001216
```

Les lignes surlignées nous indiquent qu'il n'y a pas d'effet significatif de la variable x1 sur la réponse Y. Il n'y a également pas d'interaction entre x1 et x2. La réponse Y ne dépend donc uniquement que de la variable x2 qui correspond à la variable *sat*. On a un modèle très bien ajusté car nous observons un R^2 de $0.951 \approx 1$.

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)  2  0.53999  0.269993  61.5303 0.0001005
TWI(x1, x2)  1  0.00002  0.000016   0.0036 0.9538102
PQ(x1, x2)  2  0.41863  0.209316  47.7022 0.0002071
Residuals   6  0.02633  0.004388
Lack of fit  3  0.02025  0.006749   3.3297 0.1747853
Pure error   3  0.00608  0.002027
```

L'analyse de la table de variance ci-dessus nous confirme qu'il n'y a pas d'interaction entre x1 et x2.

3. Présentez le modèle conservé après l'analyse.

Nous réécrivons le modèle en enlevant l'interaction x1:x2.

```
>CR.rsm <- rsm(Y~FO(x1,x2)+PQ(x1,x2),data=plan)
>summary(CR.rsm)
```

```
Call:
rsm(formula = Y ~ FO(x1, x2) + PQ(x1, x2), data = plan)

            Estimate Std. Error  t value Pr(>|t|)
(Intercept)  0.7802500  0.0306733  25.4374 3.706e-08 ***
x1           -0.0040303  0.0216893  -0.1858  0.85785
x2           -0.2597731  0.0216893 -11.9770 6.441e-06 ***
x1^2         -0.2556875  0.0242494 -10.5441 1.507e-05 ***
x2^2         -0.0569375  0.0242494  -2.3480  0.05124 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9733,    Adjusted R-squared:  0.958
F-statistic: 63.68 on 4 and 7 DF,  p-value: 1.379e-05
```

On obtient un modèle extrêmement bien ajusté avec un R² de 0.958 (p-value : 1.379x10⁻⁰⁵).

```
Analysis of Variance Table

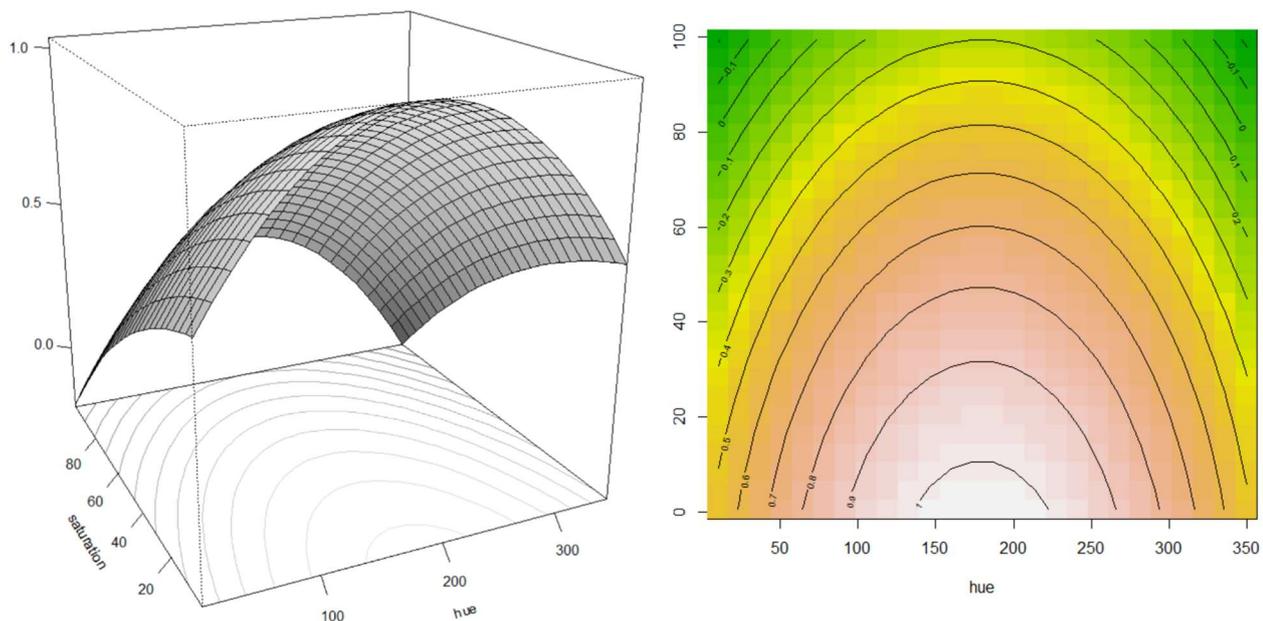
Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)  2  0.53999  0.269993  71.7418 2.171e-05
PQ(x1, x2)  2  0.41863  0.209316  55.6187 5.049e-05
Residuals   7  0.02634  0.003763
Lack of fit  4  0.02026  0.005066   2.4992  0.2388
Pure error   3  0.00608  0.002027
```

L'erreur du modèle n'est pas significative (p-value = 0.2388 > 0.05), on a donc un modèle bien ajusté.

4. Donnez la surface de réponse.

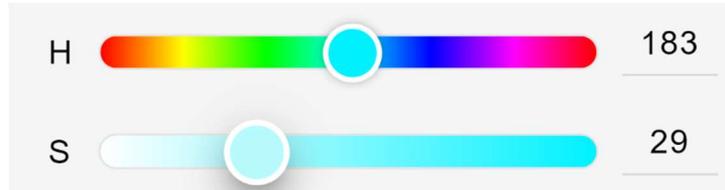
-----REPONSE

```
>contour(CR.rsm,~x1+x2,image=TRUE)
>persp(CR.rsm,~x1+x2,col = gray.colors(100), contours = "colors")
```



Sachant que la réponse Y correspond au nombre de fois (en pourcentage) que le distributeur automatique « témoin » est choisi au lieu du distributeur automatique teint. Les courbes de niveau avec une valeur basse correspondant aux paramètres couleurs qui attirent le plus l'attention. Ces zones sont en vert sur le premier graphe. Les couleurs les plus saturées, « flash », sont donc largement préférées par rapport aux autres. De plus, le graphe en 3D semble nous montrer qu'à quelconque niveau de saturation, les teintes les plus proches du rouge semblent être préférées (support figure ci-dessous) même si l'analyse nous indiquait que l'influence de la variable teinte « hue » sur la préférence des consommateurs (variable réponse Y) n'était pas significative.

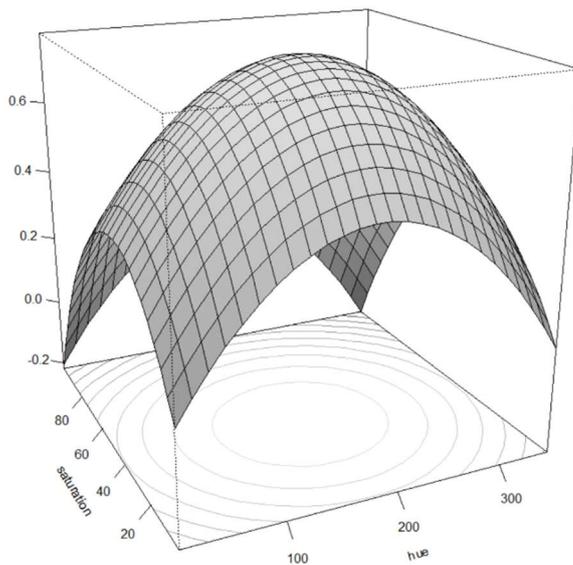
5. En réalité, les données étaient inexploitables. Les valeurs n'ont été changé qu'après qu'un stagiaire ait



remarqué une observation qui revenait excessivement souvent durant les expériences qui se résume à « la machine blanche est vraiment beaucoup plus belle que toutes les autres machines que je n'ai jamais vues ». A votre avis, expliquez quelle valeur de Y le stagiaire a truqué et pourquoi. Comment aurait-on pu planifier l'expérience dès le départ pour éviter cette erreur ? Veuillez retrouver le graphe inexploitable ci-dessous.

-----REPONSE

D'après les analyses nous avons affirmé notre hypothèse



```
> plan
run.order std.order hue saturation Block
1 1 1 60.00000 15.0000000 1
2 2 2 300.00000 15.0000000 1
3 3 3 60.00000 85.0000000 1
4 4 4 300.00000 85.0000000 1
5 5 5 180.00000 50.0000000 1
6 6 6 180.00000 50.0000000 1
7 1 1 10.29437 50.0000000 2
8 2 2 349.70563 50.0000000 2
9 3 3 180.00000 0.5025253 2
10 4 4 180.00000 99.4974747 2
11 5 5 180.00000 50.0000000 2
12 6 6 180.00000 50.0000000 2
```

de départ selon laquelle les couleurs « flash », c'est-à-dire les couleurs les plus saturées, étaient préférables aux couleurs moins brillantes et pastel. Or, la valeur de saturation *sat* de la couleur blanche (surligné sur le graphe ci-contre) est une exception parmi les couleurs proposées car elle est tout aussi « flash » que les couleurs très saturées mais a une valeur saturation *sat* de 0.5. La valeur modifiée par le stagiaire était donc la neuvième valeur de la réponse Y qu'il a remplacé par une valeur très faible pour être parfaitement en accord avec le modèle.

Il y a plusieurs aspects de cette expérience qui aurait pu être changé. Tout d'abord le choix des deux variables quantitatives et le choix de laisser le plan d'expérience composite centré choisir les teintes sachant qu'il n'y avait pas tant de diversité dans les couleurs. Il aurait au moins été préférable d'avoir du vert, du rouge absolu, plus de teintes de bleu et de jaune dans l'expérience. Et encore, vu que le modèle avec un seul et unique point stationnaire n'est clairement pas adapté à ce type d'expérience comme nous l'avons vu avec le problème de la couleur blanche tout aussi « flash » mais sans le même critère de *sat* élevé. Dans ce modèle d'analyse ce point paraît aberrant tandis qu'il devrait avoir tout autant d'importance.

Le dernier aspect le plus important qui aurait pu être modifié est l'aspect fictif du sujet qui suppose que les gens ont envie d'acheter à manger devant une machine qui fait mal aux yeux plutôt qu'un distributeur aux couleurs pastel.

Exercice 3. Comment se débarrasser des limaces ? (Exemple fictif, par Morgan LEBOURDAIS)

Un maraîcher exaspéré de voir ses laitues dévorées, veut déterminer comment réduire au maximum le nombre de limaces dans ses parcelles. Pour cela il peut agir sur deux leviers :
 - L'ajout de molluscicide (entre 0 et 20g par parcelle, 20g étant la dose maximale recommandée par le producteur), variable notée x_1 .

- La hauteur du mulch (entre 0 et 10cm), variable notée x2.

Le maraîcher dispose de 20 parcelles pour réaliser ses essais.

1. Quel modèle sera utilisé pour analyser les résultats de ces expériences ?

Nous essayons d'expliquer Y : le nombre de limaces par parcelle grâce à deux variables quantitatives, nous utiliserons donc un modèle de régression linéaire multiple.

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

2. Construire sur R le plan composite centré correspondant à l'aide du package rsm. Faites en sorte que le nombre total d'essai de ce plan corresponde au nombre de parcelles disponibles. Quel problème remarquez-vous ?

On peut utiliser la fonction `ccd` pour générer un plan composite centré, on précise l'argument `n0=6` pour avoir 12 points au centre et obtenir un plan composite centré en 20 essais.

```
PCC <- ccd(2, n0=6, coding=list(x1~(molluscicide-10)/10, x2~(mulch-5)/5), randomize=FALSE)
PCC
```

	run.order	std.order	molluscicide	mulch	Block
1	1	1	0.000000	0.000000	1
2	2	2	20.000000	0.000000	1
3	3	3	0.000000	10.000000	1
4	4	4	20.000000	10.000000	1
5	5	5	10.000000	5.000000	1
6	6	6	10.000000	5.000000	1
7	7	7	10.000000	5.000000	1
8	8	8	10.000000	5.000000	1
9	9	9	10.000000	5.000000	1
10	10	10	10.000000	5.000000	1
11	1	1	-4.142136	5.000000	2
12	2	2	24.142136	5.000000	2
13	3	3	10.000000	-2.071068	2
14	4	4	10.000000	12.071068	2
15	5	5	10.000000	5.000000	2
16	6	6	10.000000	5.000000	2
17	7	7	10.000000	5.000000	2
18	8	8	10.000000	5.000000	2
19	9	9	10.000000	5.000000	2
20	10	10	10.000000	5.000000	2

Data are stored in coded form using these coding formulas ...

```
x1 ~ (molluscicide - 10)/10
x2 ~ (mulch - 5)/5
```

Le problème de ce plan composite centré est que certaines expériences, correspondant aux points en étoiles, sont irréalisables. En effet, même s'il est envisageable de dépasser les 10cm de mulch ou de dépasser les doses de molluscicide recommandées, il est impossible de mettre des doses négatives de molluscicide ou des épaisseurs négatives de mulch ...

3. Malheureusement, le maraîcher s'est rendu compte de ce problème après avoir déjà mis en place 4 expérimentations correspondant à des points au centre. Par ailleurs, il ne veut pas dépasser les doses recommandées de molluscicide par peur de polluer. Générez un plan expérimental en 20 essais prenant en compte ces contraintes.

Premièrement, il faut générer un plan contenant tout les essais possibles compte tenu des contraintes. Pour rappel, on code -1 la valeur minimale des variables (0 pour x1 et x2) et 1 valeur maximale de ces dernières (20 pour x1 et 10 pour x2).

```
niveauA<-seq(-1,1,by=.1)
niveauB<-seq(-1,1.4,by=.1) #on peut dépasser la valeur max de mulch donc 1.4
essai_possible<- expand.grid(list(molluscicide=niveauA,mulch=niveauB))
```

La liste des essais possibles comprend 525 essais, par manque de place elle ne sera pas affichée.

Par la suite, on utilise un algorithme de Federov pour choisir une liste de 16 essais parmi tous les essais possibles (4 sont déjà mis en place).

```
set.seed(1234)
```

```
planopt16 <- optFederov(~quad(.),data=essai_possible,nTrials = 16, criterion="D")
```

Enfin, on ajoute les 4 essais déjà mis en place au plan contenant 16 essais pour obtenir notre plan expérimental final en 20 essais, comprenant 4 points au centre (déjà mis en place).

```
niveau0 <- c(0,0)
```

```
point_au_centre <- expand.grid(list(molluscicide=niveau0,mulch=niveau0))
```

```
#On créer un plan avec uniquement 4 points au centre
```

```
planopt <- rbind(planopt16$design, point_au_centre)
```

```
#On ajoute ces 4 essais au 16 essais choisis par Federov
```

```
print(planopt)
```

```
> print(planopt)
  molluscicide mulch
1          -1.0 -1.0
11           0.0 -1.0
20           0.9 -1.0
21           1.0 -1.0
22          -1.0 -0.9
252           1.0  0.1
253          -1.0  0.2
262          -0.1  0.2
263           0.0  0.2
273           1.0  0.2
484          -1.0  1.3
504           1.0  1.3
505          -1.0  1.4
515           0.0  1.4
516           0.1  1.4
525           1.0  1.4
12           0.0  0.0
2            0.0  0.0
3            0.0  0.0
4            0.0  0.0
```

4. Un algorithme de Federov fournira toujours une solution. Vérifiez la qualité de ce plan.

```
X<-model.matrix(~quad(.),data=planopt)
round(solve(t(X)%*%X),2)
```

```
> round(solve(t(X)%*%X),2)
```

	(Intercept)	molluscicide	mulch
(Intercept)	0.14	0.00	0.01
molluscicide	0.00	0.10	0.00
mulch	0.01	0.00	0.09
I(molluscicide^2)	-0.09	-0.01	0.03
I(mulch^2)	-0.06	0.01	-0.06
molluscicide:mulch	0.00	-0.02	0.00

	I(molluscicide^2)	I(mulch^2)
(Intercept)	-0.09	-0.06
molluscicide	-0.01	0.01
mulch	0.03	-0.06
I(molluscicide^2)	0.24	-0.06
I(mulch^2)	-0.06	0.12
molluscicide:mulch	0.00	0.00

	molluscicide:mulch
(Intercept)	0.00
molluscicide	-0.02
mulch	0.00
I(molluscicide^2)	0.00
I(mulch^2)	0.00
molluscicide:mulch	0.09

Le plan n'est pas parfaitement orthogonal, il y a de légères confusions pour les effets quadratiques mais cela reste

globalement un plan convenable.

5. Le maraîcher met en place et réalise les essais, les résultats (en nombre de limaces par parcelles) sont enregistrés dans le vecteur Y (voir ci dessous). Analyser les résultats sous R. Quelle quantité de molluscicide et quelle épaisseur de mulch doit-il utiliser à l'avenir pour avoir un nombre minimal de limaces dans ses parcelles.

```
Y<-c(47,45,46,42,48,45,48,45,45,44,52,47,53,49,50,48,44,44,44,47)
reg<-rsm(formula=Y~SO(molluscicide,mulch), data=planopt)
summary(reg)
```

```
Call:
rsm(formula = Y ~ SO(molluscicide, mulch), data = planopt)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	44.83432	0.42401	105.7394	< 2.2e-16	***
molluscicide	-1.95611	0.35328	-5.5369	7.324e-05	***
mulch	1.34738	0.34711	3.8817	0.001661	**
molluscicide:mulch	-0.29370	0.34888	-0.8418	0.414029	
molluscicide^2	0.93637	0.55311	1.6929	0.112596	
mulch^2	1.37894	0.40005	3.4469	0.003929	**

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

```
Multiple R-squared: 0.8819, Adjusted R-squared: 0.8397
F-statistic: 20.91 on 5 and 14 DF, p-value: 4.843e-06
```

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(molluscicide, mulch)	2	106.069	53.034	41.2144	1.36e-06
TWI(molluscicide, mulch)	1	0.935	0.935	0.7268	0.408263
PQ(molluscicide, mulch)	2	27.531	13.766	10.6976	0.001515
Residuals	14	18.015	1.287		
Lack of fit	11	11.265	1.024	0.4552	0.854109
Pure error	3	6.750	2.250		

```
Stationary point of response surface:
molluscicide  mulch
 0.9843368   -0.3837290
```

```
Eigenanalysis:
eigen() decomposition
$values
[1] 1.4232320 0.8920755
```

```
$vectors
           [,1]      [,2]
molluscicide -0.2887783 -0.9573960
mulch         0.9573960 -0.2887783
```

On constate tout d'abord que l'erreur d'ajustement n'est pas significative au seuil de 5%, on ne rejette pas l'hypothèse H0, le modèle est donc bien ajusté. On remarque également que l'effet quadratique de molluscicide et l'interaction entre molluscicide et mulch ne sont pas significatifs. Si l'on s'intéresse aux valeurs propres, on voit qu'elle sont toutes les deux > 0, ce qui signifie que le point stationnaire fournit par R correspond à un minimum. Si on convertit avec le codage évoqué en question 3, ce minimal est atteint avec 19,69g de molluscicide (pour faire simple il arrondira à la dose maximal : 20g) et un mulch d'une épaisseur d'environ 3cm.

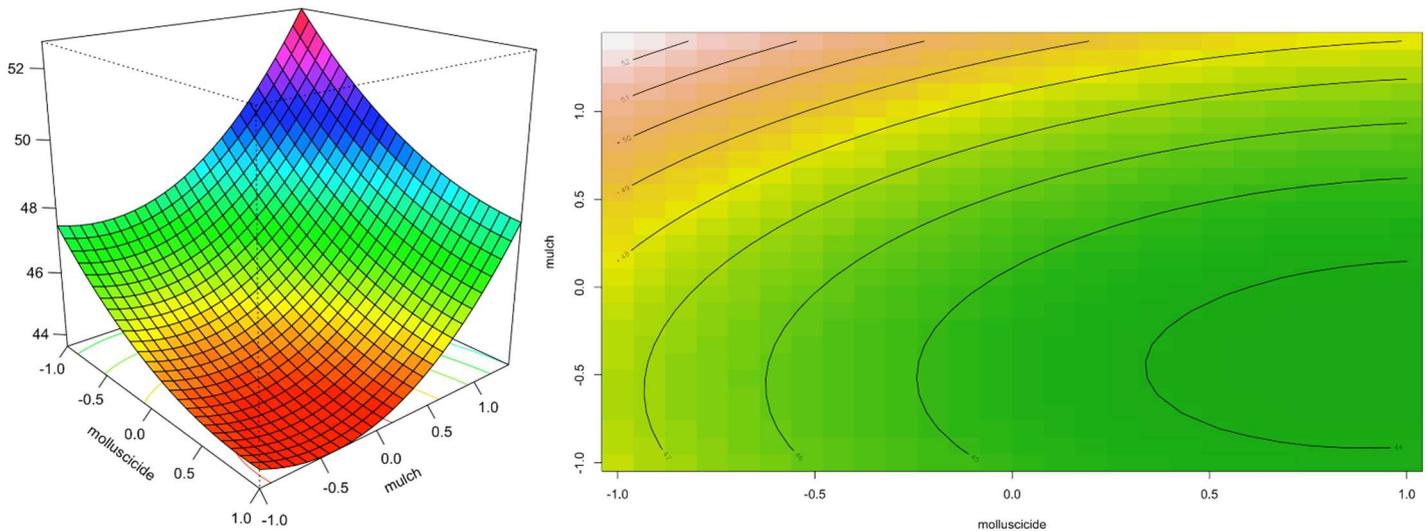
6. Représentez sous R la surface réponse de Y.

```
contour(reg,~molluscicide+mulch,image=TRUE)
persp(reg,~molluscicide+mulch,col=rainbow(50),contours="colors", theta="50")
```

Exercice 4. Impact de la température et de l'apport d'eau sur la croissance d'un plant de tomate (EXERCICE FICTIF)

Avec le dérèglement climatique, les questions d'apport d'eau et de température se posent à nouveau car les conditions auxquelles les maraîcher.e.s étaient habitué.e.s évoluent plus rapidement qu'auparavant. Dans cette étude, on cherche à maximiser le gain de biomasse en gramme de la plante par jour (en juillet-août) selon les apports d'eau et la température.

- 1) Pour estimer ces besoins, on cherche à proposer un plan d'expériences permettant d'estimer au mieux la surface de réponse. On souhaite considérer dans cette étude une température entre 15 et 45 degrés et un apport d'eau quotidien entre 1,5 et 8,5 décilitres d'eau par



pied.

On fait un plan composite centré. On choisit de mettre 4 points au centre pour tester les effets quadratiques. On obtient le code suivant :

```
library(rsm)
plan_continu<-ccd(2, n0=2, randomize=FALSE, coding=list(x1~(arrosage-5)/3.5,x2~(temperature-30)/15))
plan_continu
```

run.order	std.order	arrosage	temperature	Block
1	1	1.50000000	15.000000	1
2	2	8.50000000	15.000000	1
3	3	1.50000000	45.000000	1
4	4	8.50000000	45.000000	1
5	5	5.00000000	30.000000	1
6	6	5.00000000	30.000000	1
7	1	0.05025253	30.000000	2
8	2	9.94974747	30.000000	2
9	3	5.00000000	8.786797	2
10	4	5.00000000	51.213203	2
11	5	5.00000000	30.000000	2
12	6	5.00000000	30.000000	2

Data are stored in coded form using these coding formulas ...

$x_1 \sim (\text{arrosage} - 5)/3.5$

$x_2 \sim (\text{temperature} - 30)/15$

On a surligné en jaune les points au centre qui permettent de tester les effets quadratiques. Sur le graphique, on observe pour chacun des douze essais, les valeurs prises pour les deux facteurs.

2) On veut vérifier la qualité d'ajustement du modèle sachant qu'on obtient les réponses Y suivantes (52,63,52,66,62,63,45,68,58,59,62,63) qui correspondent à la biomasse prise en gramme par jours.

```
Y<-c(52,63,52,66,62,63,45,68,58,59,62,63)
CR.rsm<-rsm(Y~SO(x1,x2),data = plan_continu)
summary(CR.rsm)
```

```
Call:
rsm(formula = Y ~ SO(x1, x2), data = plan_continu)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	62.50000	0.62985	99.2296	7.059e-11	***
x1	7.19086	0.44537	16.1457	3.589e-06	***
x2	0.55178	0.44537	1.2389	0.261645	
x1:x2	0.75000	0.62985	1.1908	0.278718	
x1^2	-2.81250	0.49794	-5.6482	0.001321	**
x2^2	-1.81250	0.49794	-3.6400	0.010835	*

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

Multiple R-squared: 0.9805, Adjusted R-squared: 0.9643
F-statistic: 60.42 on 5 and 6 DF, p-value: 4.741e-05

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	416.10	208.052	131.1095	1.119e-05
TWI(x1, x2)	1	2.25	2.250	1.4179	0.278718
PQ(x1, x2)	2	61.04	30.521	19.2335	0.002457
Residuals	6	9.52	1.587		
Lack of fit	3	8.52	2.840	8.5211	0.055929
Pure error	3	1.00	0.333		

Stationary point of response surface:

x1	x2
1.3355128	0.4285273

Stationary point in original units:

arrosage	temperature
9.674295	36.427909

Eigenanalysis:

eigen() decomposition

\$values

[1] -1.6875 -2.9375

\$vectors

	[,1]	[,2]
x1	-0.3162278	-0.9486833
x2	-0.9486833	0.3162278

Le R2 est très bon 0,9643, cela veut dire que l'ajustement global du modèle est excellent. Ce R2 est très significatif (p value = 4.741e-05).

Pour les 4 points au centre, la variance, correspondant à l'erreur pure est de 0,333, à l'échelle de nos valeurs c'est une valeur minimale donc l'erreur pure est très faible.

On regarde le "lack of fit" ce dernier part de l'hypothèse H0 que le modèle n'est pas ajusté or on rejette cette hypothèse car p > 0,05. Donc le modèle est bien ajusté, nous n'avons pas besoin d'utiliser d'autres variables. Ensuite, on regarde la significativité des effets linéaires, d'interactions et quadratiques. Pour les effets d'interaction, la p-value est supérieure à 0,05 on rejette l'hypothèse qu'il y ait des effets d'interaction et on relance le modèle sans les interactions.

Call:

```
CR.rsm<-rsm(Y~FO(x1,x2)+PQ(x1,x2),data = plan_continu)
```

Call:

```
rsm(formula = Y ~ FO(x1, x2) + PQ(x1, x2), data = plan_continu)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	62.50000	0.64838	96.3939	3.407e-12	***
x1	7.19086	0.45847	15.6843	1.037e-06	***
x2	0.55178	0.45847	1.2035	0.2678973	
x1^2	-2.81250	0.51259	-5.4868	0.0009193	***
x2^2	-1.81250	0.51259	-3.5360	0.0095213	**

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

Multiple R-squared: 0.9759, Adjusted R-squared: 0.9622
F-statistic: 70.94 on 4 and 7 DF, p-value: 9.562e-06

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	416.10	208.052	123.7232	3.453e-06
PQ(x1, x2)	2	61.04	30.521	18.1500	0.001699
Residuals	7	11.77	1.682		
Lack of fit	4	10.77	2.693	8.0784	0.058748
Pure error	3	1.00	0.333		

Stationary point of response surface:

x1 x2
1.2783758 0.1522143

Stationary point in original units:

arrosage temperature
9.474315 32.283214

Eigenanalysis:

eigen() decomposition

\$values

[1] -1.8125 -2.8125

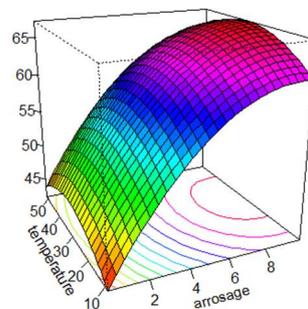
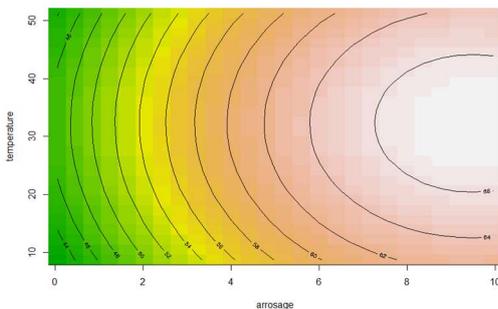
\$vectors

[,1] [,2]
x1 0 -1
x2 -1 0

3) Quelle est la combinaison de lumière et de chaleur optimale pour avoir le plus grand gain de biomasse sur un plant de tomate par jour ?

A partir de la manipulation précédente, on obtient la valeur exacte du gain de biomasse maximum d'après les expérimentations. Ceci correspond à 0,95 L par jour et 32 °C. Pour s'assurer que ce résultat correspond bien à un maximum, on va représenter la surface de réponse.

```
contour(CR.rsm,~x1+x2,image=TRUE)  
persp(CR.rsm,~x1+x2,col=rainbow(50),contours="colors")
```



En effet, au point de coordonnées, (9.5, 32), on est au point avec la valeur la plus grande de Y.

Exercice 5. Développement de burgers végétariens à base de légumineuses fermentées (exemple issu d'une situation réelle, par Adèle Husson)

Une entreprise spécialisée dans le développement de burgers végétariens à base de légumineuses fermentées décide d'étendre sa gamme avec un nouveau produit à base de lentilles corail. Dans les burgers existants, la teneur en légumineuse est comprise entre 40 et 60% selon la légumineuse en question. De plus, les temps de fermentation varient entre 28 et 32h selon le produit considéré. Le nouveau salarié R&D en charge du projet ne sait donc pas quel dosage ni quel temps de fermentation choisir afin d'optimiser sa recette.

Il dispose de 12 essais au maximum car la taille de son incubateur pour fermenter les produits est limitée. De plus, il doit présenter les échantillons en analyse sensorielle d'ici peu.

1) Quel plan d'expérience faut-il construire ?

Nous disposons de deux variables explicatives quantitatives ainsi qu'une variable réponse quantitative. Nous allons donc construire un plan composite centré.

2) Construire ce plan à l'aide du logiciel R

```
library(rsm) #chargement du package
plan <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(dosage-50)/10, x2~(temps-30)/2))# construction du plan composite centré à 2 facteurs avec 4 points au centre (n0=2) pour respecter les 12 essais maximum
> plan
```

	run.order	std.order	dosage	temps	Block
1	1	1	40.00000	28.00000	1
2	2	2	60.00000	28.00000	1
3	3	3	40.00000	32.00000	1
4	4	4	60.00000	32.00000	1
5	5	5	50.00000	30.00000	1
6	6	6	50.00000	30.00000	1
7	1	1	35.85786	30.00000	2
8	2	2	64.14214	30.00000	2
9	3	3	50.00000	27.17157	2
10	4	4	50.00000	32.82843	2
11	5	5	50.00000	30.00000	2
12	6	6	50.00000	30.00000	2

Les essais 1 à 4 correspondent aux points du plan factoriel complet 2^2 ; les essais 5,6, 11 et 12 correspondent aux points au centre et les essais 7 à 10 aux points en étoile.

3) Commentez la qualité de ce plan

```
X <- model.matrix(~X1+X2+I(X1^2)+I(X2^2)+I(X1*X2), data=plan)
solve(t(X)%*%X)
```

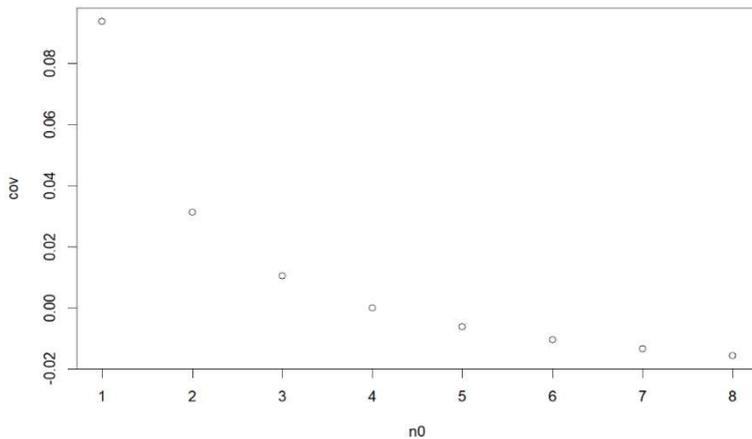
```
> solve(t(X)%*%X)
```

	(Intercept)	x1	x2	I(X1^2)	I(X2^2)	I(X1 * X2)
(Intercept)	0.250	0.000	0.000	-0.12500	-0.12500	0.00
x1	0.000	0.125	0.000	0.00000	0.00000	0.00
x2	0.000	0.000	0.125	0.00000	0.00000	0.00
I(X1^2)	-0.125	0.000	0.000	0.15625	0.03125	0.00
I(X2^2)	-0.125	0.000	0.000	0.03125	0.15625	0.00
I(X1 * X2)	0.000	0.000	0.000	0.00000	0.00000	0.25

Nous pouvons voir que la constante est complètement indépendante des effets principaux et de l'interaction. Par contre, celle-ci est un peu liée aux effets quadratiques. De plus, il n'y a aucune confusion des effets principaux. Enfin, les effets quadratiques sont confondus entre eux.

4) Si nous avons voulu optimiser le modèle (dans le cas où réaliser plus d'essais aurait été possible), quel aurait été le nombre d'essais optimal ?

```
n0=1:8
cov <- NULL
for (i in 1:length(n0)){
  plan_optimal <- ccd(2,n0=n0[i],randomize=FALSE,coding=list (x1~(dosage-50)/10,x2~(temps-30)/2))
  X <- model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=plan_optimal)
  cov <- c(cov,solve(t(X)%*%X) [5,4])
}
plot(cov~n0)
```



Le graphique permet de voir que la covariance aurait été nulle pour $n_0=4$ et donc pour 8 essais au centre. Il aurait donc fallu faire 16 essais en tout afin d'optimiser le modèle.

La matrice du plan en 16 essais est la suivante :

```
> solve(t(X)%*%X)
      (Intercept)      x1      x2 I(x1^2) I(x2^2) I(x1 * x2)
(Intercept)    0.1250  0.000  0.000 -0.0625 -0.0625    0.00
x1              0.0000  0.125  0.000  0.0000  0.0000    0.00
x2              0.0000  0.000  0.125  0.0000  0.0000    0.00
I(x1^2)         -0.0625  0.000  0.000  0.1250  0.0000    0.00
I(x2^2)         -0.0625  0.000  0.000  0.0000  0.1250    0.00
I(x1 * x2)      0.0000  0.000  0.000  0.0000  0.0000    0.25
```

Nous pouvons voir qu'il n'y a plus de confusion des effets quadratiques entre eux. La variance des effets principaux et de l'interaction reste inchangée mais celle de la constante et des effets quadratiques a diminuée.

Une analyse sensorielle est ensuite réalisée avec les différents produits. Tous les juges donnent une note d'appréciation globale de chaque produit allant de 1 à 10.

La moyenne des notes par produit est donnée ici : $Y \leftarrow c(0.2 \ 4.9 \ 3.1 \ 7.7 \ 7.0 \ 6.9 \ 1.1 \ 7.8 \ 1.8 \ 5.8 \ 7.2 \ 6.9)$

5) Explicitez le modèle utilisé afin d'analyser les résultats du plan ?

Il faut utiliser le modèle de régression linéaire multiple avec les effets linéaires, les effets quadratiques et les interactions. Celui-ci s'écrit de la façon suivante :

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

Hypothèses sur la résiduelle : $\forall i=1, \dots, n$ ϵ_i i.i.d., $E(\epsilon_i)=0$, $V(\epsilon_i)=\sigma^2$ et, $\forall i \neq k$ $cov(\epsilon_i, \epsilon_k) = 0$

6) Analyser les résultats obtenus en commentant le tableau d'analyse de variance

```
res <- rsm(Y~SO(x1,x2), data=plan)
summary(res)
```

```
Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.000000    0.067414 103.8363 5.377e-11 ***
x1           2.346904    0.047669  49.2336 4.709e-09 ***
x2           1.419607    0.047669  29.7807 9.506e-08 ***
x1:x2        -0.025000    0.067414  -0.3708  0.7235
x1^2         -1.312500    0.053295 -24.6269 2.949e-07 ***
x2^2         -1.637500    0.053295 -30.7250 7.891e-08 ***
```

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

```
Multiple R-squared:  0.9987, Adjusted R-squared:  0.9976
F-statistic: 922.2 on 5 and 6 DF, p-value: 1.438e-08
```

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	60.186	30.0930	1655.4163	5.919e-09
TWI(x1, x2)	1	0.002	0.0025	0.1375	0.7235
PQ(x1, x2)	2	23.629	11.8146	649.9211	9.700e-08
Residuals	6	0.109	0.0182		
Lack of fit	3	0.049	0.0164	0.8178	0.5637
Pure error	3	0.060	0.0200		

Premièrement, nous pouvons voir que le modèle est bien ajusté car l'hypothèse H_0 : le modèle est bien ajusté n'est pas rejetée (p-value = 0.56). L'erreur pure et l'erreur d'ajustement sont du même ordre de grandeur.

De plus, le R^2 est très élevé (0.9987) avec une p-value très inférieure à 0.05.

Ensuite, nous pouvons voir que globalement, il y a des effets linéaires et quadratiques. Par contre, il n'y a pas d'effet d'interaction (P-Value = 0.72 > 0.05).

7) Que conseillez-vous de faire en considérant les résultats obtenus ?

Il faut reconstruire le modèle sans interaction :

```
res2 <- rsm(Y~FO(x1,x2)+PQ(x1,x2), data=plan)
```

Call:

```
rsm(formula = Y ~ FO(x1, x2) + PQ(x1, x2), data = plan)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	7.000000	0.063124	110.892	1.278e-12	***
x1	2.346904	0.044636	52.579	2.359e-10	***
x2	1.419607	0.044636	31.804	7.853e-09	***
x1^2	-1.312500	0.049904	-26.300	2.941e-08	***
x2^2	-1.637500	0.049904	-32.813	6.320e-09	***

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

Multiple R-squared: 0.9987, Adjusted R-squared: 0.9979
F-statistic: 1315 on 4 and 7 DF, p-value: 3.851e-10

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	60.186	30.0930	1888.0434	2.725e-10
PQ(x1, x2)	2	23.629	11.8146	741.2512	7.115e-09
Residuals	7	0.112	0.0159		
Lack of fit	4	0.052	0.0129	0.6446	0.6678
Pure error	3	0.060	0.0200		

Stationary point of response surface:

x1	x2
0.8940586	0.4334677

Stationary point in original units:

dosage	temps
58.94059	30.86694

Eigenanalysis:

eigen() decomposition

\$values

[1] -1.3125 -1.6375

Nous pouvons observer que le modèle est encore mieux ajusté que le précédent (0.6678 > 0.5637).

Le modèle peut s'écrire : $Y = 7 + 2.3x_1 + 1.4x_2 - 1.3x_1^2 - 1.64x_2^2 + \epsilon$

L'effet linéaire du dosage (x1) est plus important que l'effet temps de fermentation (x2).

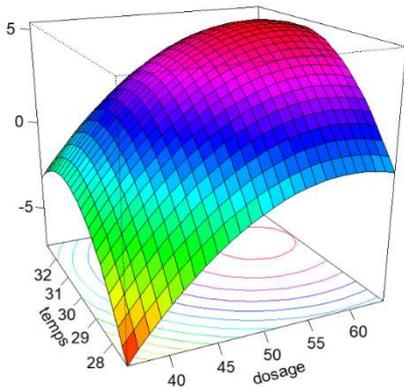
8) Pour quelles valeurs de x1 et x2 la note attribuée par les juges est-elle optimale ? Vérifiez que cette valeur correspond bien à un maximum.

L'optimum de Y est atteint pour des valeurs de $x_1 = 58.9\%$ et $x_2 = 30.9h$.

Les deux valeurs propres sont négatives (-1.31 et -1.64) ce qui signifie que l'optimum est bien un maximum.

Il est possible d'observer cela graphiquement :

```
persp(res, ~x1+x2, col = rainbow(50), contours = "colors")
```



Exercice 6. Élaboration d'une recette de falafels à base d'haricots mungo (exemple fictif par Kim Lenotte)

L'équipe R&D d'une entreprise connue pour les alternatives végé cherche à mettre au point une nouvelle recette de falafels à base d'haricots mungo. Les autres recettes de falafels à base de différentes légumineuses de l'entreprise étant un succès auprès des consommateurs, la meilleure combinaison possible est donc recherchée pour maximiser l'appréciation de ce nouveau produit. Étant limité en temps et en argent (nombre d'essais limité) l'équipe R&D fait appel à vous pour la conseiller.

Pour l'élaboration de cette recette, l'équipe R&D hésite sur deux facteurs : Le premier (noté x_1) correspond au pourcentage de Haricot mungo à utiliser (entre 40% et 60%) et le deuxième (noté x_2) à la température de cuisson (entre 170 et 180°C).

De plus, on vous informe que pour évaluer ces recettes, les falafels seront soumises à un jury. Chaque juge donnera une note allant de 0 à 10 d'appréciation globale à chaque produit.

La note d'appréciation comprise entre 0 et 10 est la variable à expliquer (Y) et votre objectif est de la maximiser. Autrement dit, vous cherchez la combinaison $\{x_1, x_2\}$ des facteurs permettant d'obtenir une valeur maximum de la réponse Y .

Ainsi, on vous demande de faire le moins d'essais possibles (pas plus de 20) tout en permettant de garantir que les effets des différents facteurs sont indépendants les uns des autres.

Q1) Que sous-entend l'équipe quand elle dit qu'elle souhaite garantir que les effets des différents facteurs soient indépendants les uns des autres ?

→ Elle souhaite l'orthogonalité.

Q2) Quel type de plan choisiriez-vous à la place de l'équipe R&D de cette entreprise ? Vous préciserez le nombre d'essais.

→ Nous nous intéressons à l'effet de deux variables explicatives quantitatives (x_1 et x_2) sur une variable réponse (Y) quantitative, ainsi le plan d'expérience à construire est un plan composite centré à 2 facteurs.

Afin de calculer le nombre d'essais à réaliser, nous allons utiliser cette formule :

$$\text{Nb d'expériences} : 2^{k-p} + 2k + n_0$$

Avec : $2^{k-p} = 2^2 = 4$ essais (plan factoriel complet)

$2k = 2*2 = 4$ essais (points en étoiles)

$n_0 = 8$ essais (au centre, en gardant en tête qu'on veut l'orthogonalité des facteurs)

Ainsi, au total il y aura 16 essais. La demande de l'entreprise de ne pas dépasser 20 essais est respectée.

Q3) Quel modèle allez-vous utiliser afin d'analyser les résultats d'un tel plan ?

→ Un modèle de régression linéaire avec lequel la variable réponse Y est expliquée par les effets principaux (les deux var explicatives x1 et x2), leurs effets quadratiques et leur interaction :

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

Q4) À l'aide des lignes de code données ci-dessous, générez le plan sur R.

```
#Chargement du package rsm permettant de construire des plans composites centrés
library(rsm)
#Construction du PCC à 2 facteurs avec 8 points au centre (n0*2, les noms des facteurs
(avec leur valeur au centre et leur étendue de variation)
plan<-ccd(2,n0=4, randomize=FALSE, coding=list(x1~(Haricot- 50)/10,x2~(Temp-175)/5))
print(plan[,3:5])
```

	Haricot	Temp	Block
1	40.00000	170.0000	1
2	60.00000	170.0000	1
3	40.00000	180.0000	1
4	60.00000	180.0000	1
5	50.00000	175.0000	1
6	50.00000	175.0000	1
7	50.00000	175.0000	1
8	50.00000	175.0000	1
9	35.85786	175.0000	2
10	64.14214	175.0000	2
11	50.00000	167.9289	2
12	50.00000	182.0711	2
13	50.00000	175.0000	2
14	50.00000	175.0000	2
15	50.00000	175.0000	2
16	50.00000	175.0000	2

Q5) Analysez la qualité du plan.

```
#Création de la matrice des effets (linéaires, quadratiques et d'interaction)
X<-model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2), data=plan)
#Création de la matrice (X'X)-1
solve(t(X)%*%X)
```

	(Intercept)	x1	x2	I(x1^2)	I(x2^2)	I(x1 * x2)
(Intercept)	0.1250	0.000	0.000	-0.0625	-0.0625	0.00
x1	0.0000	0.125	0.000	0.0000	0.0000	0.00
x2	0.0000	0.000	0.125	0.0000	0.0000	0.00

I (x1^2)	-0.0625	0.000	0.000	0.1250	0.0000	0.00
I (x2^2)	-0.0625	0.000	0.000	0.0000	0.1250	0.00
I (x1 * x2)	0.0000	0.000	0.000	0.0000	0.0000	0.25

→ Nous pouvons voir ici que les effets principaux (x1 et x2) sont indépendants des autres effets. Il y a juste une légère **confusion** entre les effets quadratiques et la constante. Nous pouvons donc dire que globalement ce plan est bon car les essais choisis vont permettre d'estimer tous les coefficients sans qu'il y ait de corrélation entre les coefficients. De plus, il répond aux attentes de l'entreprise car l'orthogonalité entre les effets est de qualité.

Partie résultats

En attendant la réalisation des essais et leur dégustation, les membres de l'équipe R&D vous demande de leur expliquer comment analyser les résultats. Pour cela, vous allez simuler des données à l'aide des lignes de code ci-dessous :

```
#Renseigne des valeurs pour beta (le 1 correspond à l'intercept et 0,9 à l'effet de x1
etc.)
#Pour cet exemple nous ne voulons pas d'effet d'interaction alors il faut mettre 0 pour
la dernière valeur de beta

beta<-c(1,0.9,-2.5,-1,-1.4,0)

#Initialise le générateur de nombres hasard pour avoir les mêmes résultats
set.seed(123)
#Simulation des données pour notre plan
Y<-X%*%beta + rnorm(nrow(plan))
```

Q6) À l'aide des lignes de code données ci-dessous, analysez les résultats. Vous commenterez l'ajustement du modèle, les différents effets et vous préciserez si vous conservez le modèle de la question Q3. Si non, quel modèle conserverez-vous ?

```
#Création du modèle comprenant les effets linéaires, les effets quadratiques et l'effet
d'interaction grâce au terme SO(=FO+TWI+PQ)
reg<-rsm(Y~ SO(x1,x2), data=plan)
summary (reg)
```

Call:

```
rsm(formula = Y ~ SO(x1, x2), data = plan)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.34784	0.34343	3.9246	0.002845	**
x1	0.79790	0.34343	2.3233	0.042532	*
x2	-2.35030	0.34343	-6.8436	4.494e-05	***
x1:x2	-0.45462	0.48569	-0.9360	0.371304	
x1^2	-1.43285	0.34343	-4.1721	0.001911	**
x2^2	-1.15375	0.34343	-3.3595	0.007249	**

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

Multiple R-squared: 0.8911, Adjusted R-squared: 0.8366

F-statistic: 16.36 on 5 and 10 DF, p-value: 0.000156

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	49.284	24.6422	26.1160	0.0001071
TWI(x1, x2)	1	0.827	0.8267	0.8762	0.3713043
PQ(x1, x2)	2	27.074	13.5368	14.3464	0.0011531
Residuals	10	9.436	0.9436		

```
Lack of fit 3 1.958 0.6526 0.6109 0.6290946
Pure error 7 7.478 1.0683
```

```
Stationary point of response surface:
      x1      x2
0.4542137 -1.1080389
```

```
Stationary point in original units:
haricot Temp
54.54214 169.45981
```

```
Eigenanalysis:
eigen() decomposition
$values
[1] -1.026568 -1.56003
```

→ Tout d'abord, nous pouvons voir que l'erreur d'ajustement n'est pas significative (lack of fit 0.629 > 0.05) ce qui signifie que le modèle est bien ajusté. Ensuite, si on s'intéresse aux effets, nous pouvons voir qu'il y a des effets linéaires (p-value de FO(x1, x2) = 0.0001071 < 0.05). Ceci signifie qu'il y a au moins un des deux facteurs x1 ou x2 qui a un effet linéaire sur Y. Dans notre cas, les deux facteurs semblent avoir un effet linéaire significatif sur Y (p-value x1 = 0.042532 < 0.05 et p-value x2 = 4.494 * 10⁻⁵). Il y a également des effets quadratiques (p-value de PQ(x1, x2) = 0.0011531) et comme pour les effets linéaires, les deux facteurs x1² et x2² ont un effet significatif sur Y. Enfin, il n'y a pas d'effet d'interaction (TWI) car la probabilité critique est supérieure à 0.05. Le modèle conservé est donc celui sans effet d'interaction.

Q7) Écrivez le modèle conservé et analysez les résultats obtenus sur R comme si c'était ceux obtenus à la suite de la dégustation des falafels.

→ Modèle sans effet d'interaction :

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

```
#Modèle sans effet d'interaction
reg2<-rsm(Y~FO(x1,x2)+PQ(x1,x2), data=plan)
summary (reg2)
```

```
Call:
rsm(formula = Y ~ FO(x1, x2) + PQ(x1, x2), data = plan)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.34784	0.34149	3.9469	0.002285	**
x1	0.79790	0.34149	2.3365	0.039415	*
x2	-2.35030	0.34149	-6.8824	2.648e-05	***
x1^2	-1.43285	0.34149	-4.1958	0.001496	**
x2^2	-1.15375	0.34149	-3.3785	0.006159	**

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

```
Multiple R-squared: 0.8815, Adjusted R-squared: 0.8384
F-statistic: 20.46 on 4 and 11 DF, p-value: 4.699e-05
```

Analysis of Variance Table

```
Response: Y
      Df Sum Sq Mean Sq F value Pr(>F)
FO(x1, x2) 2 49.284 24.6422 26.4133 6.312e-05
PQ(x1, x2) 2 27.074 13.5368 14.5097 0.0008226
Residuals 11 10.262 0.9329
Lack of fit 4 2.784 0.6961 0.6516 0.6438989
```

```
Pure error    7    7.478    1.0683
```

```
Stationary point of response surface:
```

```
      x1      x2  
0.2784307 -1.0185494
```

```
Stationary point in original units:
```

```
haricot    Temp  
52.78431 169.90725
```

```
Eigenanalysis:
```

```
eigen() decomposition
```

```
$values
```

```
[1] -1.153748 -1.432851
```

→ L'effet de la température de cuisson (x_2) semble avoir une influence plus forte sur l'appréciation du falafel que la quantité d'haricot mungo : ($|\text{coef}(\text{Temp})| > |\text{coef}(\text{haricot})|$) → $2.35 > 0.79$).

Le coefficient négatif du facteur x_2 (température de cuisson) signifie un effet négatif sur l'appréciation des falafels. Les falafels cuits à une température plus faible auront tendance à être plus appréciés.

À l'inverse, le coefficient positif du facteur x_1 (quantité d'haricot mungo) correspond à un effet positif et significatif du facteur sur l'appréciation des falafels. Une augmentation de la quantité d'haricot mungo sera associée à une meilleure appréciation du produit. En d'autres termes, plus il y aura d'haricot mungo dans la recette, plus les falafels seront susceptibles d'être appréciés. L'effet est cependant moins fort que celui de la température de cuisson.

Q8) Donnez les valeurs de x_1 et x_2 qui permettent d'optimiser Y .

→ Stationary point in original units:

```
haricot    Temp  
52.78431 169.90725
```

Q9) Précisez si l'optimum est un point selle, un maximum ou un minimum et commentez.

→ Sans la figure nous pouvons voir qu'il s'agit d'un maximum car les valeurs propres sont toutes les deux négatives :

```
Eigenanalysis:
```

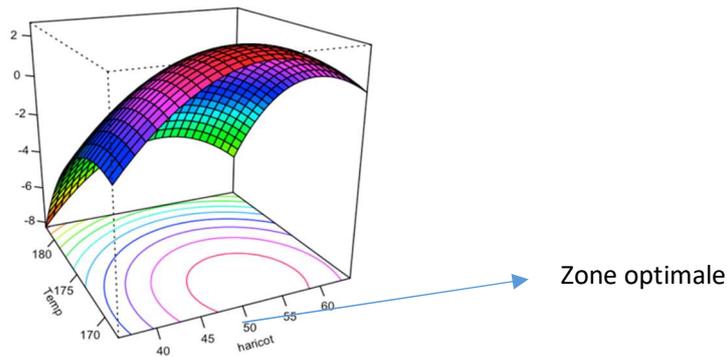
```
eigen() decomposition
```

```
$values
```

```
[1] -1.153748 -1.432851
```

Q10) Représentez sur R les surfaces de réponses.

```
persp(reg2, ~x1+x2, contours="colors")
```



Exercices sur les plans optimaux

Exercice 1. Améliorer le croquant du chocolat (exemple fictif, par Nassira Chupeau)

Un groupe d'étudiants à l'Institut Agro travaillent sur le développement d'une tablette de chocolat. Malheureusement, ces derniers ne parviennent pas à optimiser la formulation de la recette pour améliorer la force de résistance de la tablette exprimée en newton et mesurée à l'aide d'un texturomètre. Leur objectif est de se rapprocher au plus près de celles relevées pour des tablettes similaires achetées dans le commerce.

Pour pallier ce problème, ce groupe d'étudiants listent 5 facteurs qui pourraient avoir un effet sur la force de résistance de la tablette de chocolat :

- F1 = Le pourcentage de cacao : 55%-65%-75%
 - F2 = Les conditions de refroidissement de la tablette : au réfrigérateur (4°C) ou à température ambiante (20°C)
 - F3 = Le type du moule : modèle 1, modèle 2 ou modèle 3
 - F4 = La méthode de tempérage : traditionnelle (avec les trois paliers de températures) ou utilisation de Mycryo (beurre de cacao en poudre)
 - F5 = Le temps écoulé entre la formation des tablettes et le démoulage : 8h, 12h, 24h ou 48h
- **Quel serait le nombre d'essais à réaliser par ce groupe d'étudiants dans le cadre de la mise en place d'un plan complet ?**

Nous avons au total 5 facteurs à 2, 2, 3, 3 et 4 modalités. Pour calculer le plan complet, il suffit de multiplier le nombre de modalités entre eux, soit : $2 \times 2 \times 3 \times 3 \times 4 = 144$.

Il faudrait donc réaliser 144 essais pour réaliser le plan complet.

- **Quelles sont les lignes de code à utiliser dans R pour afficher ce plan complet ?**

```
library(DoE.base)
```

```
plan <- fac.design(nlevels=c(3,2,3,2,4),factor.names=(c("cacao", "refroidissement", "moule", "tempérage", "démoulage")))
```

```
plan
```

- Déterminez le nombre minimal d'essais à prévoir pour estimer les effets principaux et comparez ce résultat avec le nombre d'essais à réaliser pour assurer l'orthogonalité entre les facteurs 2 à 2. Commentez ces résultats.

Le nombre minimal d'essais se calcule en déterminant le nombre minimal de degré de liberté à avoir pour estimer tous les effets : $1 + 2 \times (2-1) + 2 \times (3-1) + (4-1) = 10$ essais.

Pour l'orthogonalité des facteurs, il faut :

- $F2 \perp F4$ donc minimum $2 \times 2 = 4$ essais
- $F1 \perp F2$, $F1 \perp F4$, $F2 \perp F3$ et $F3 \perp F4$ donc minimum $3 \times 2 = 6$ essais
- $F2 \perp F5$ et $F4 \perp F5$ donc minimum $2 \times 4 = 8$ essais
- $F1 \perp F3$ donc minimum $3 \times 3 = 9$ essais
- $F1 \perp F5$ et $F3 \perp F5$ donc minimum 12 essais

Le plus petit commun multiple (PPCM) de ces nombres est :

$$\text{PPCM}(4, 6, 8, 9, 12) = 72$$

Il faut au minimum 72 essais pour avoir l'orthogonalité des facteurs 2 à 2 (plan $L_{72} 2^2 3^2 4^1$) mais 10 essais suffisent pour estimer les paramètres principaux.

La réalisation de 72 essais semble difficilement réalisable dans cette situation (temps imparti, manque de moyens, quantité d'ingrédients). Néanmoins, s'ils décident d'effectuer 10 essais, l'analyse des données sera limitée. Ils pourront utiliser un modèle d'analyse de variance à 5 facteurs, mais il sera complètement saturé, il ne restera en effet plus de degré de liberté pour la résiduelle.

Le groupe d'étudiants souhaite trouver une piste d'amélioration concernant leur problématique avant les vacances scolaires. Il leur reste donc trois jours pour préparer l'ensemble des tablettes et effectuer les mesures au texturomètre. En faisant l'inventaire des quantités d'ingrédients restant ainsi que du temps imparti, les étudiants se fixent un objectif de 18 essais à réaliser.

- Construire sur R le plan optimal en 18 essais en utilisant la fonction `optFederov` du package `AlgDesign`.

```
library(AlgDesign)
```

```
opt<-optFederov(~.,data=plan,nTrials=18,eval=TRUE)
```

```
opt
```

```
$D #la racine k de la variance généralisée
```

```
[1] 0.4145571
```

```
$A #la moyenne de la variance des coefficients
```

```
[1] 2.879489
```

```
$I #la variance de prédiction Moyenne sur X
```

```
[1] 10.53636
```

```
$Ge #la variance normalisée minimax sur X
```

```
[1] 0.801
```

```
$Dea #la limite inférieure de l'efficacité D pour les conceptions théoriques approximatives
```

```
[1] 0.781
```

```
$design
```

```
  cacao refroidissement moule tempérage démoulage
```

12	1	1	2	1	3
13	1	2	3	2	4
20	1	2	3	1	1
21	2	2	2	2	3
26	1	1	1	2	4
34	3	1	3	1	3
38	1	1	1	1	1
46	3	2	1	2	3
76	2	2	1	1	4
90	2	1	1	1	2
93	2	2	3	1	2
96	3	1	2	1	4
108	3	2	2	1	4
109	3	1	3	2	2
110	2	1	3	2	4
119	3	2	1	2	1
127	2	1	2	2	1
133	1	2	2	2	2

\$rows

```
[1] 12 13 20 21 26 34 38 46 76 90 93 96 108 109 110 119 127 133
```

- Vérifiez la qualité de ce plan et commentez.

Pour vérifier la qualité du plan à 18 essais, il faut construire la matrice $(X'X^{-1})$ c'est-à-dire l'inverse du produit entre la matrice des essais et sa transposée.

```
options(contrasts = c("contr.sum", "contr.sum"))
X <- model.matrix(~.,data=opt$design)
round(t(X)%*%X,2) #on a arrondi le plan à un chiffre après virgule
```

	(Intercept)	cacao.L	cacao.Q	refroidissement1	moule.L	moule.Q	tempéragel
(Intercept)	18.00	0.00	0.00	0	0.00	0.00	0
cacao.L	0.00	6.00	0.00	0	0.00	0.00	0
cacao.Q	0.00	0.00	6.00	0	0.00	0.00	0
refroidissement1	0.00	0.00	0.00	18	0.00	0.00	2
moule.L	0.00	0.00	0.00	0	6.00	0.00	0
moule.Q	0.00	0.00	0.00	0	0.00	6.00	0
tempéragel	0.00	0.00	0.00	2	0.00	0.00	1
démoulage.L	1.34	0.63	0.00	0	0.32	-0.55	0
démoulage.Q	1.00	-0.71	0.41	0	-0.71	0.41	0
démoulage.C	0.45	-0.32	-0.91	0	0.63	0.73	0

	démoulage.L	démoulage.Q	démoulage.C
(Intercept)	1.34	1.00	0.45
cacao.L	0.63	-0.71	-0.32
cacao.Q	0.00	0.41	-0.91
refroidissement1	0.00	0.00	0.00
moule.L	0.32	-0.71	0.63
moule.Q	-0.55	0.41	0.73
tempéragel	0.00	0.00	0.00
démoulage.L	4.90	0.67	0.30
démoulage.Q	0.67	4.50	0.22
démoulage.C	0.30	0.22	4.10

Nous remarquons que le plan à 18 essais ne permet pas d'atteindre une orthogonalité parfaite entre les facteurs, ce qui était prévisible car l'orthogonalité est atteinte pour un plan à 72 essais ou plus. Il y a des valeurs non nulles en-dehors de la diagonale, ce qui signifie qu'il n'y a pas indépendance entre tous les facteurs. Il y a des confusions entre les effets même si ces valeurs de covariances restent faibles. Le plan peut être considéré comme de bonne qualité mais ces observations devront être prises en compte lors de l'analyse et l'interprétation des résultats. Le mieux serait d'augmenter le nombre d'essais à réaliser pour se rapprocher d'une pseudo-orthogonalité.

Exercice 2. Stratégies de marketing et distributeurs automatiques (plans optimaux, exemple fictif, Eric KAPISIZ)

CONTEXTE :

Après avoir participé à une conférence sur le marketing en grande surface, une entreprise de distributeurs automatiques apprend que l'emplacement d'un produit sur un rayon exerce une influence sur sa visibilité. Par exemple, les produits placés au niveau des yeux sont favorisés par les consommateurs par rapport à ceux qui sont proche du sol. Cette entreprise se demande alors s'il n'y aurait pas lieu à faire entre les rayons en grande surface et les emplacements sur un distributeur.

On s'intéresse donc ici à savoir si certains emplacements sont favorisés par rapport à d'autres.

- On dispose de deux types de distributeurs, une classique où le produit choisi tombe sur la surface en mousse en bas du distributeur (cela évite d'endommager le produit) et une « armature » qui récupère le produit avec un système mécanique.
- Les distributeurs contiennent un de ces trois produits concernés ; une bouteille d'eau en plastique, une cannette de soda en aluminium et un petit paquet M&M's en plastique.
- Une des deux méthodes de paiement sont imposées au consommateur, le paiement plus interactif par pièce ou le paiement sans contact, plus discret, avec une carte.

Voici le procédé de l'expérience : le consommateur entre dans une grande salle blanche vêtue d'un toit véranda couvrable avec un des deux distributeurs au milieu de la salle. On choisit d'illuminer ou non la salle. Si la salle n'est pas illuminée, le consommateur est dans le noir hormis la lumière venant de l'intérieur du distributeur qui lui permet de voir les produits et légèrement ses environs. Si la salle est illuminée, on utilise soit une lumière artificielle comme dans les grandes surfaces, soit on découvre la véranda pour laisser entrer la lumière naturelle. Tous les emplacements du distributeur sont remplis au maximum et cela avec qu'un seul des trois produits concernés (c'est-à-dire que le distributeur ne peut contenir qu'un type de produit à la fois). On impose une des deux méthodes de paiement au consommateur. Il n'y a aucun bruit dans la salle. Les consommateurs font à peu près la même taille. La seule instruction donnée au consommateur est qu'il ne peut sortir de la salle qu'après avoir acheté un produit.

L'étude est menée avec 100 consommateurs. Chaque consommateur refait le test avec une combinaison différente de distributeur automatique 8 fois. Les données seront analysées par la suite afin de savoir s'il y a un emplacement favorisé selon les différents paramètres.

QUESTIONS :

1. Identifier les différentes modalités. Quel est le nombre d'essais associé au plan complet ? Combien faut-il d'essais au minimum si on veut l'orthogonalité entre les facteurs 2 à 2 ? Quel plan mettons-nous en place ?

-----REPONSE

On identifie les facteurs et leurs modalités, on a les facteurs :

- type de distributeur à deux modalités
- produits à trois modalités
- méthode de paiement à deux modalités
- les types de luminosités dans la salle à trois modalités (obscurité, lumière artificielle, lumière naturelle)

Le nombre d'essais associés au plan complet est donc de $2^2 + 3^2 = 36$ (aussi le nombre maximal d'essais).

Le nombre minimal d'essais : $1 + 2 \times (2 - 1) + 2 \times (3 - 1) = 7$ (aussi le nombre de paramètres à estimer).

Il faut $N \geq PPCM(4, 6, 9) = 36$ essais minimums pour avoir l'orthogonalité des facteurs. Seul le plan complet permet d'estimer les confusions d'effets.

2. Construire le plan complet et évaluez sa qualité.

-----REPONSE

```
>library(DoE.base)
>library(AlgDesign)
>#le plan complet
>Design.1 <- fac.design(nlevels=c(2,2,3,3), factor.names= c("type","paiement","produit","luminosite"))
>Xopti=model.matrix(~ . , Design.1)
>VARCOVopti <- solve(t(Xopti)%*%Xopti)
>round(VARCOVopti, 3)
```

```
> round(VARCOVopti, 3)
      (Intercept) type1 paiement1 produit.L produit.Q luminosite.L luminosite.Q
(Intercept)    0.028 0.000    0.000    0.000    0.000    0.000    0.000
type1          0.000 0.028    0.000    0.000    0.000    0.000    0.000
paiement1     0.000 0.000    0.028    0.000    0.000    0.000    0.000
produit.L     0.000 0.000    0.000    0.083    0.000    0.000    0.000
produit.Q     0.000 0.000    0.000    0.000    0.083    0.000    0.000
luminosite.L  0.000 0.000    0.000    0.000    0.000    0.083    0.000
luminosite.Q  0.000 0.000    0.000    0.000    0.000    0.000    0.083
```

Le plan est de très bonne qualité.

3. Construisez un plan optimal à 8 essais. Quel pourrait être l'utilité d'un plan optimal à la place d'un plan complet ? Argumentez.

-----REPONSE

```
>Design.1 <- fac.design(nlevels=c(2,2,3,3), factor.names= c("type","paiement","produit","luminosite"))
>plan.1.Dopt<-optFederov(~.,Design.1,nTrials=8,criterion="D")#plan optimal à 8 essais
>Xopti=model.matrix(~., plan.1.Dopt$design)
>VARCOVopti <- solve(t(Xopti)%*%Xopti)
>round(VARCOVopti, 2)
>plan.1.Dopt$D#efficacité du plan optimal par rapport au plan complet
[1] 0.4962913
```

L'efficacité de ce plan est d'environ 50% par rapport au plan complet.

Supposons avoir 720 essais. La question est donc de savoir si l'on préfère faire faire les 8 mêmes essais du plan optimal 90 fois (8x90=720) ou faire faire les 36 essais du plans complet 30 fois (30x36=720). Vu que la réponse Y de l'emplacement choisi peut potentiellement être dû à beaucoup de hasard, il pourrait être préférable de répéter le même plan optimal 90 fois afin de réduire la variabilité de Y même si cela signifierait qu'il y a plus de confusion des différents facteurs.

Exercice 3. La meilleure recette de gnocchis (Exemple fictif, par Morgan LEBOURDAIS)

S'ennuyant un jeudi après-midi, deux étudiants cherchent à déterminer la meilleure recette possible de gnocchis. Pour cela ils vont faire plusieurs recettes en faisant varier certains facteurs :

<u>Facteur 1</u> : Ratio farine/purée	Quantitative	0,2 < F1 < 0,6
<u>Facteur 2</u> : Méthode d'obtention de la purée	Qualitative : 2 modalités	Moulin à légume ou presse-purée
<u>Facteur 3</u> : Variété de pomme de terre	Qualitative : 4 modalités	Bintje, Charlotte, Désirée ou Monalisa
<u>Facteur 4</u> : Aromates	Qualitative : 3 modalités	Origan, sauge ou laurier

1. Nous sommes ici en présence de variables qualitatives et quantitatives, quelle est la meilleure stratégie à adopter pour planifier les recettes à réaliser ?

Dans ce cas de figure (variables mixtes), deux possibilités sont envisageables :

3. On peut réaliser un plan continu avec seulement la variable quantitative et un plan avec seulement les variables qualitatives et on répète le plan quantitatif pour chaque essai du plan qualitatif. Cette solution n'est pas envisageable car il y a beaucoup de variables qualitatives et cela représenterait un nombre d'expériences trop important.
 3. On transforme la variable quantitative en une variable qualitative avec au moins 3 modalités et on construit un plan pour les variables qualitatives. On aura donc 3 ratios : 0,2 ; 0,4 et 0,6. Cette solution permet également de faciliter les pesées pour les étudiants.
2. Construire ce plan. Combien d'essais comprend le plan complet ? Combien d'essais faut-il au minimum pour avoir l'orthogonalité entre les facteurs 2 à 2 ?

La ligne de code R permettant de construire ce plan asymétrique avec des facteurs à 2, 3 et 4 niveaux est la suivante :

```
plan <- gen.factorial(levels=c(2,3,3,4),nvars=4,factors = "all")
```

Nombre d'essais du plan complet : $2 \times 3 \times 3 \times 4 = 72$ expériences.

Nombre minimum d'essais pour avoir l'orthogonalité entre les facteurs 2 à 2 = PPCM (6, 8, 9, 12) = 72 expériences. Le plan complet correspond donc au nombre minimum d'essais pour avoir l'orthogonalité entre les facteurs 2 à 2.

3. Les deux étudiant étant attendus au bar en début de soirée, ils estiment qu'ils ne pourront réaliser que 10 recettes chacun avant de partir. Par ailleurs, ils considèrent avoir les mêmes talents culinaires, l'effet « cuisinier » sera donc négligeable. Quel est le meilleur plan d'expérimentation en 10 essais par étudiant ?

On peut utiliser l'algorithme de Fedorov sur R pour déterminer le meilleur plan d'expérience en 20 essais, puis les deux étudiants pourront réaliser au hasard 10 essais chacun car l'effet « cuisinier » est ici considéré négligeable. Si ce n'était pas le cas, il aurait fallu le prendre en compte en ajoutant un facteur à 2 modalités : étudiant 1 et étudiant 2.

On peut construire un tel plan avec les lignes de code suivantes :

```
planopt <- optFedorov(~.,data=plan,nTrials = 20, criterion="D")
planopt$design
```

ID	method	ratio	aromate	variete	
1	1	1	1	1	1
6	2	1	2	1	1
16	2	2	1	2	1
35	1	2	3	3	1
39	1	2	1	1	2
48	2	2	3	1	2
54	2	1	2	2	2
57	1	1	3	2	2
61	1	1	1	3	2
68	2	2	2	3	2
79	1	2	2	1	3
81	1	1	3	1	3
96	2	2	3	2	3
98	2	1	1	3	3
112	2	2	1	1	4
118	2	1	3	1	4
123	1	2	1	2	4
125	1	1	2	2	4

139	1	2	2	3	4
142	2	1	3	3	4

4. Quelle est la qualité de ce plan ?

```
X<-model.matrix(~.,data=planopt$design)
round(solve(t(X)%*%X),2)
```

	(Intercept)	method1	ratio1	ratio2	aromate1
(Intercept)	0.05	0.00	0.00	0.01	0.01
method1	0.00	0.05	0.01	0.00	0.00
ratio1	0.00	0.01	0.10	-0.05	0.00
ratio2	0.01	0.00	-0.05	0.11	0.00
aromate1	0.01	0.00	0.00	0.00	0.11
aromate2	0.01	0.00	0.00	0.00	-0.06
variete1	-0.01	0.00	0.00	-0.01	-0.01
variete2	0.01	0.00	-0.02	0.01	0.01
variete3	-0.01	0.00	0.00	-0.01	-0.01
	aromate2	variete1	variete2	variete3	
(Intercept)	0.01	-0.01	0.01	-0.01	
method1	0.00	0.00	0.00	0.00	
ratio1	0.00	0.00	-0.02	0.00	
ratio2	0.00	-0.01	0.01	-0.01	
aromate1	-0.06	-0.01	0.01	-0.01	
aromate2	0.11	-0.01	0.01	-0.01	
variete1	-0.01	0.14	-0.05	-0.03	
variete2	0.01	-0.05	0.18	-0.05	
variete3	-0.01	-0.03	-0.05	0.14	

On remarque que le plan n'est pas vraiment orthogonal car il y a des valeurs différentes de 0 en dehors de la diagonale. Cependant, les confusions restent faibles et la qualité de ce plan est satisfaisante vu le nombre réduit d'essais réalisés par rapport au plan complet.

Exercice 4. Développement de burgers végétariens à base de légumineuses (exemple issu d'une situation réelle, par Adèle Husson)

Une entreprise décide de développer un burger végétarien à base de légumineuses. L'appréciation de ce burger par les consommateurs peut dépendre de plusieurs facteurs :

La légumineuse : pois chiche, lentille verte, lentille jaune, haricot rouge, haricot noir, haricot blanc

L'épice utilisée : curry, paprika, cumin, curcuma, poivre

L'huile végétale utilisée : colza, tournesol, olive, lin

La forme du burger final : carré, rond

L'entreprise a pour objectif de formuler le burger qui sera le plus apprécié par les consommateurs.

1) Quel est le nombre d'essais associé au plan complet ? Combien d'essais faudrait-il au minimum pour respecter l'orthogonalité des facteurs 2 à 2 ?

Le plan contient 4 facteurs à 6, 5, 4 et 2 modalités. Le nombre d'essais associé au plan complet est donc de : $6 \times 5 \times 4 \times 2 = 240$ essais.

Pour respecter l'orthogonalité des facteurs deux à deux et donc estimer sans confusion d'effets, il faut :

A orthogonal à B : $6 \times 5 = 30$ essais

A orthogonal à C : $6 \times 4 = 24$ essais

A orthogonal à D : $6 \times 2 = 12$ essais

B orthogonal à C : $5 \times 4 = 20$ essais

B orthogonal à D : $5 \times 2 = 10$ essais

C orthogonal à D : $4 \times 2 = 8$ essais

PPCM (8,10,12,20,24,30) = 120 essais minimum pour un plan orthogonal.

2) Quel est le nombre d'essai minimum que l'on peut faire ?

On calcule les degrés de liberté : $1 + (6-1) + (5-1) + (4-1) + (2-1) = 14$ essais

3) L'entreprise voulant réaliser le moins d'essais possible, construisez le plan optimal à 14 essais sous R et observez sa qualité

```
library(DoE.base)
#construction du plan complet
plan <- fac.design(nlevels = c(6,5,4,2), factor.names = c("legu", "epice", "huile",
"forme"))
library(AlgDesign)
# construction du plan optimal
plan.opt <- optFederov(~., plan, nTrials = 14)
$design
  legu epice huile forme
1      4      3      1      2
40     6      4      2      2
41     3      3      4      1
54     2      1      4      2
72     1      1      2      1
74     6      1      3      1
85     5      4      1      1
111    5      2      4      2
113    3      4      3      2
139    1      3      1      2
145    5      5      2      2
167    3      1      1      2
190    1      5      4      1
210    2      2      1      1

# Vérification de la qualité du plan
X <- model.matrix(~., plan.opt$design)
round(solve(t(X)%*%X),1)

      (Intercept) legu.L legu.Q legu.C legu^4 legu^5 epice.L
(Intercept)      0.1   0.0  -0.1   0.0   0.1   0.1   0.0
legu.L            0.0   0.6   0.1   0.1   0.2  -0.1   0.0
legu.Q           -0.1   0.1   0.8   0.2  -0.2  -0.4   0.0
legu.C            0.0   0.1   0.2   0.7  -0.2  -0.2   0.2
legu^4            0.1   0.2  -0.2  -0.2   1.1   0.3  -0.1
legu^5            0.1  -0.1  -0.4  -0.2   0.3   0.8   0.0
epice.L           0.0   0.0   0.0   0.2  -0.1   0.0   0.5
epice.Q           0.0   0.1  -0.1   0.0   0.2   0.1   0.2
epice.C           0.1   0.1   0.1   0.1   0.2  -0.1   0.0
epice^4           0.0   0.1   0.1   0.2  -0.6  -0.2   0.1
huile.L           0.0   0.0  -0.1  -0.1   0.0   0.1   0.0
huile.Q          -0.1   0.3   0.2   0.1   0.3   0.0   0.1
huile.C           0.0   0.1  -0.1   0.1   0.2   0.0  -0.1
formel           0.0   0.0   0.1   0.0   0.0   0.0   0.0
      epice.Q epice.C epice^4 huile.L huile.Q huile.C formel
(Intercept)  0.0   0.1   0.0   0.0  -0.1   0.0   0.0
legu.L        0.1   0.1   0.1   0.0   0.3   0.1   0.0
legu.Q       -0.1   0.1   0.1  -0.1   0.2  -0.1   0.1
legu.C        0.0   0.1   0.2  -0.1   0.1   0.1   0.0
legu^4        0.2   0.2  -0.6   0.0   0.3   0.2   0.0
legu^5        0.1  -0.1  -0.2   0.1   0.0   0.0   0.0
epice.L       0.2   0.0   0.1   0.0   0.1  -0.1   0.0
epice.Q       0.5   0.0  -0.1  -0.1   0.2   0.0   0.0
epice.C       0.0   0.7  -0.1  -0.1  -0.1   0.0   0.0
epice^4      -0.1  -0.1   0.8   0.0  -0.2  -0.1   0.0
huile.L      -0.1  -0.1   0.0   0.3   0.0  -0.1   0.0
huile.Q       0.2  -0.1  -0.2   0.0   0.7   0.1   0.0
huile.C       0.0   0.0  -0.1  -0.1   0.1   0.5   0.0
```

```
formel      0.0    0.0    0.0    0.0    0.0    0.0    0.1
```

Nous pouvons observer que le plan n'est pas parfaitement orthogonal. Certains effets principaux sont confondus entre eux. Les covariances sont de l'ordre de 0.1.

4) Construire la matrice du plan orthogonal

```
plan2 <- optFederov(~., plan, nTrials = 120)
X <- model.matrix(~., plan2$design)
round(solve(t(X)%*%X), 2)
Resultat (extrait de la matrice)
```

(Intercept)	legu.L	legu.Q	legu.C	legu^4	legu^5		
(Intercept)	0.01	0.00	0.00	0.00	0.00	0.00	0.00
legu.L	0.00	0.05	0.00	0.00	0.00	0.00	0.00
legu.Q	0.00	0.00	0.05	0.00	0.00	0.00	0.00
legu.C	0.00	0.00	0.00	0.05	0.00	0.00	0.00
legu^4	0.00	0.00	0.00	0.00	0.05	0.00	0.00
legu^5	0.00	0.00	0.00	0.00	0.00	0.05	0.00
epice.L	0.00	0.00	0.00	0.00	0.00	0.00	0.00
epice.Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00
epice.C	0.00	0.00	0.00	0.00	0.00	0.00	0.00
epice^4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
huile.L	0.00	0.00	0.00	0.00	0.00	0.00	0.00
huile.Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00
huile.C	0.00	0.00	0.00	0.00	0.00	0.00	0.00
formel	0.00	0.00	0.00	0.00	0.00	0.00	0.00

	epice.L	epice.Q	epice.C	epice^4	huile.L	huile.Q
(Intercept)	0.00	0.00	0.00	0.00	0.00	0.00
legu.L	0.00	0.00	0.00	0.00	0.00	0.00
legu.Q	0.00	0.00	0.00	0.00	0.00	0.00
legu.C	0.00	0.00	0.00	0.00	0.00	0.00
legu^4	0.00	0.00	0.00	0.00	0.00	0.00
legu^5	0.00	0.00	0.00	0.00	0.00	0.00
epice.L	0.04	0.00	0.00	0.00	0.00	0.00
epice.Q	0.00	0.04	0.00	0.00	0.00	0.00
epice.C	0.00	0.00	0.04	0.00	0.00	0.00
epice^4	0.00	0.00	0.00	0.04	0.00	0.00
huile.L	0.00	0.00	0.00	0.00	0.03	0.00
huile.Q	0.00	0.00	0.00	0.00	0.00	0.03
huile.C	0.00	0.00	0.00	0.00	0.00	0.00
formel	0.00	0.00	0.00	0.00	0.00	0.00

Ici, nous pouvons voir que le plan est de meilleure qualité que précédemment. En effet, les effets principaux sont indépendants (pas de confusion) et la variance de ces derniers est minimisée.

5) L'entreprise se rend compte qu'elle dispose encore d'un peu de budget pour réaliser des essais supplémentaires. Elle décide donc d'en réaliser 6 de plus afin d'optimiser au mieux ses résultats. Construire le plan et comparer sa qualité au plan en 14 essais.

```
plan.opt2 <- optFederov(~., plan, nTrials = 20, augment=TRUE, rows=plan.opt$rows)
X <- model.matrix(~., plan.opt2$design)
round(solve(t(X)%*%X), 1)
```

(Intercept)	legu.L	legu.Q	legu.C	legu^4	legu^5	epice.L	epice.Q	epice.C
(Intercept)	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
legu.L	0.0	0.4	0.0	-0.1	0.0	0.0	-0.1	0.1
legu.Q	0.0	0.0	0.3	0.1	0.0	0.0	-0.1	0.0
legu.C	0.0	-0.1	0.1	0.4	0.0	0.1	0.0	0.0
legu^4	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.1
legu^5	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0
epice.L	0.0	-0.1	0.0	0.1	0.0	0.0	0.3	0.0
epice.Q	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	0.3
epice.C	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.3

epiceA4	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
huile.L	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
huile.Q	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
huile.C	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0
forme1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Nous pouvons voir que la variance des effets principaux a diminué (par exemple, les valeurs surlignées en jaune dans les deux matrices montrent que celle-ci valait 0.6 pour la légumineuse L et elle vaut désormais 0.4). Le plan est donc de meilleure qualité que celui en 14 essais.

Exercice 5. Amélioration de la texture d'un caviar d'aubergines (exemple fictif par Kim Lenotte)

Une entreprise de traiteurs souhaite optimiser la texture d'un caviar d'aubergine, et notamment à limiter la teneur en eau, afin de proposer à ses clients le meilleur produit possible. Les différents caviars seront soumis à un jury avant d'être présentés aux clients. Les facteurs qui a priori devraient influencer la texture sont les suivants :

- Facteur A : la variété des aubergines (noire, fairytale, blanche, italienne, chinoise)
- Facteur B : la maturité des aubergines (immature, mûre, sur mûre)
- Facteur C : le type de cuisson (four, grillade, vapeur)
- Facteur D : méthode de préparation des aubergines (mixeur, blender)
- Facteur E : Prétraitement des aubergines (souple de sel avant cuisson ou non)
- Facteur F : Peau (avec la peau d'aubergine, sans la peau d'aubergine)

On s'intéresse donc à l'effet de 6 facteurs qualitatifs sur la réponse Y (texture du caviar). Le facteur A a 5 modalités, les facteurs B et C ont 3 modalités et les facteurs D, E et F ont 2 modalités.

Q1) Quel est le nombre d'essais associé au plan complet ?

→ Le nombre d'essais associé au plan complet est : $5 \times 3^2 \times 2^3 = 360$ essais

Q2) Donnez le nombre d'essais au minimum si on veut orthogonalité entre les facteurs 2 à 2 ?

→ $A \perp B$ et $A \perp C \rightarrow 5 \times 3 = 15$ essais

$A \perp D$, $A \perp E$ et $A \perp F \rightarrow 5 \times 2 = 10$ essais

$B \perp C \rightarrow 3 \times 3 = 9$ essais

$B \perp D$, $B \perp E$, $B \perp F$, $C \perp D$, $C \perp E$ et $C \perp F \rightarrow 3 \times 2 = 6$ essais

$D \perp E$, $D \perp F$ et $E \perp F \rightarrow 2 \times 2 = 4$ essais

Le nombre d'essais doit être supérieur ou égale au PPCM(4,6,9,10,15), soit à 180 essais.

Q3) Construisez ce plan complet (sans l'afficher) à l'aide des lignes de code ci-dessous :

```
library(DoE.base)
```

```
complet<-fac.design(nlevels=c(2,2,2,3,3,5), factor.names=c("F","E","D","C","B","A"))
```

Q4) Disposant de très peu de temps pour réaliser tous les essais, le cuisinier décide de ne réaliser que 15 essais.

Construisez un plan optimal en 15 essais à l'aide des lignes de code ci-dessous :

```
library(AlgDesign)
```

```
optimal<-optFederov(~.,data=complet, nTrials=15, eval=TRUE)
```

```
print(optimal)
```

```

$design
  A B C D E F
3   2 2 2 2 1 4
17  1 2 1 2 3 3
62  2 2 1 3 2 2
88  1 1 2 2 2 1
99  1 1 2 3 2 4
156 1 1 1 1 1 2
162 2 1 2 3 1 3
192 2 1 1 1 3 4
217 2 1 2 2 3 2
227 1 2 2 1 1 5
247 2 1 1 2 2 5
248 2 2 2 1 2 3
259 2 2 2 1 3 1
288 2 2 1 3 1 1
360 1 2 2 3 3 5

```

Q5) Avec 15 essais, quel modèle d'analyse peut-on considérer ? Précisez, pour chaque terme, le nombre de degrés de liberté associé.

On réalise une analyse de variance à 6 facteurs.

Pour la constante : 1ddl

Pour le facteur à 5 modalités : $1 \times (5-1) = 4$

Pour les 2 facteurs à 3 modalités : $2 \times (3-1) = 4$

Pour les 3 facteurs à 2 modalités : $3 \times (2-1) = 3$

Il y a donc 3 ddl pour estimer la résiduelle. Le plan n'est donc pas saturé

Q6) Sur R, calculez la matrice des effets associée à ce modèle, puis calculez la matrice de dispersion qui va permettre de déterminer la qualité du plan. Commentez.

```

X<-model.matrix(~., data=optimal$design)
round(solve(t(X)%*%X),2)

```

	(Intercept)	F1	E1	D1	C.L	C.Q	B.L	B.Q	A.L	A.Q	A.C	A^4
(Intercept)	0.07	0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	-0.01
F1	0.01	0.08	-0.01	0.00	0.01	0.03	0.03	-0.01	-0.02	0.04	0.03	-0.01
E1	0.00	-0.01	0.08	-0.01	-0.01	-0.03	-0.03	0.01	0.00	0.06	0.00	0.01
D1	-0.01	0.00	-0.01	0.08	0.00	0.00	0.00	0.00	0.00	-0.03	0.00	0.03
C.L	0.00	0.01	-0.01	0.00	0.21	0.00	0.04	0.00	0.00	0.00	0.00	0.00
C.Q	0.00	0.03	-0.03	0.00	0.00	0.22	0.02	0.04	-0.01	-0.01	0.01	-0.01
B.L	0.00	0.03	-0.03	0.00	0.04	0.02	0.22	-0.01	-0.01	-0.01	0.01	-0.01
B.Q	0.00	-0.01	0.01	0.00	0.00	0.04	-0.01	0.21	0.00	0.00	0.00	0.00
A.L	0.00	-0.02	0.00	0.00	0.00	-0.01	-0.01	0.00	0.34	-0.01	-0.01	0.00
A.Q	0.01	0.04	0.06	-0.03	0.00	-0.01	-0.01	0.00	-0.01	0.41	0.02	-0.01
A.C	0.00	0.03	0.00	0.00	0.00	0.01	0.01	0.00	-0.01	0.02	0.35	-0.01
A^4	-0.01	-0.01	0.01	0.03	0.00	-0.01	-0.01	0.00	0.00	-0.01	-0.01	0.35

On observe des confusions. En ayant choisi seulement 15 essais, la qualité du plan a été détériorée (non orthogonal). Le cuisinier devrait réaliser plus d'essais s'il veut avoir des résultats pertinents. Idéalement, il devrait réaliser 180 essais au minimum (question 2).

Q7) En fonction des résultat obtenus refaite l'analyse.

```

library(DoE.base)
library(AlgDesign)
complet<-fac.design(nlevels=c(2,2,2,3,3,5),factor.names=c("F","E","D","C","B","A"))
optimal<-optFederov(~.,data=complet, nTrials=180, eval=TRUE)
X<-model.matrix(~., data=optimal$design)
VARCOV<-round(solve(t(X)%*%X),2)

```

	(Intercept)	F1	E1	D1	C.L	C.Q	B.L	B.Q	A.L	A.Q	A.C	A^4
(Intercept)	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
F1	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E1	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D1	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C.L	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C.Q	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
B.L	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
B.Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
A.L	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
A.Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
A.C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
A^4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03

Avec 180 essais, la matrice obtenue est orthogonale. Cependant, réaliser 180 essais prend du temps et de l'argent, ainsi pour ne faire que 15 essais et avoir des résultats pertinents, le cuisinier devrait enlever une des 15 expériences qui avaient été choisies du plan et la remplacer par une expérience du plan complet qui n'avait pas été choisie. À chaque fois qu'il fait un échange, il devra regarder le déterminant de $(X'X)^{-1}$ à l'aide de la ligne de code ci-dessous en cherchant à le minimiser. Après un échange, si le déterminant diminue alors on garde l'échange. Il faut répéter ces étapes jusqu'à ce qu'il y ait convergence.

Ligne de code pour calculer le déterminant :

```
determinant<-det(VARCOV)
```

Exercice 6. Boissons rafraîchissantes (exemple réel, par Alban Krauffel)

Partie 1 : mélanges de sirops

Quatre colocataires sont amateurs de boissons fraîches au goûter. Ils cherchent à savoir quel mélange à base d'un sirop, d'un diluant et de glaçons, est le meilleur. Voici les facteurs qu'ils souhaitent tester :

-sirop : grenadine et pêche (2 modalités)

-diluant : eau, limonade, jus d'orange, et lait (4 modalités)

-glaçons : avec ou sans glaçons (2 modalités)

La variable réponse sera une note d'appréciation sur 20.

Les 4 colocataires vont se répartir les expériences ; ils sont prêts à boire jusqu'à 6 boissons chacun.e. On risque donc d'avoir un effet juge : ces derniers n'ont pas les mêmes goûts, et n'ont pas une notation aussi exigeante ! On prend donc en compte cet effet juge et ajoutant un 4^{ème} facteur :

-juge : A, S, R, D (4 modalités)

Pour estimer correctement tous les effets, les colocataires ont besoin de concevoir un plan d'expérience.

1) Combien d'essais comporte le plan complet ? Construisez-le sur R.

Il y a 2 facteurs à 2 modalités et 2 facteurs à 4 modalités donc le plan complet comporte : $2^2 * 4^2 = 64$ essais.
On le réalise sur R :

```

library(DoE.base)
plan_complet <- fac.design(nlevels = c(4, 4, 2, 2))
plan_complet

```

	A	B	C	D
1	3	1	1	1
2	1	4	2	1
3	3	4	1	2
4	3	2	1	2

5	3	4	2	2
6	2	2	2	2
7	4	4	1	2
8	2	4	2	1
9	4	3	1	1
10	4	2	2	2
11	2	1	1	2
12	2	3	2	2
13	1	3	1	2

```
14 2 2 2 1
15 2 4 2 2
16 4 1 1 2
17 2 3 1 2
18 3 3 2 2
19 4 1 2 1
20 2 3 2 1
21 4 4 1 1
22 2 2 1 2
23 3 4 1 1
24 3 1 2 1
25 1 1 1 2
26 3 3 2 1
27 3 1 1 2
28 1 2 1 2
29 4 3 2 1
30 2 3 1 1
31 4 3 1 2
32 1 1 2 2
33 1 1 1 1
34 4 2 1 1
35 1 1 2 1
36 1 2 1 1
37 4 3 2 2
38 2 1 2 2
39 3 1 2 2
40 1 4 1 1
41 2 1 2 1
42 2 2 1 1
43 3 2 2 2
44 3 4 2 1
45 1 4 2 2
46 1 3 2 2
47 2 4 1 2
48 1 3 1 1
49 4 4 2 2
50 1 2 2 1
51 1 2 2 2
52 4 4 2 1
53 4 1 2 2
54 1 3 2 1
55 4 1 1 1
56 3 2 1 1
57 3 3 1 1
58 1 4 1 2
59 2 1 1 1
60 2 4 1 1
61 4 2 1 2
62 3 3 1 2
63 4 2 2 1
64 3 2 2 1
class=design, type= full factorial
```

2) Le plan complet comporte trop d'essais. Réaliser à la main un plan en 16 essais (4 chacun), adapté pour l'expérience, à partir d'un plan fractionnaire. Réaliser un tel plan sur R également.

On part du plan de base $2^4 = 16$.

	I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
3	1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
4	1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
5	1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
6	1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
7	1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1
8	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1
9	1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1
10	1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1
11	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
12	1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1
13	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1
14	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
15	1	-1	-1	-1	1	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	1

Nous avons ici 4 facteurs à 2 modalités. Il s'agit d'obtenir 2 facteurs à 2 modalités et 2 autres à 4 modalités. On commence par ajouter deux facteurs E et F, que l'on confond avec les interactions d'ordre le plus élevé possible, mais de même ordre pour E que pour F ; on doit donc les confondre avec les interactions d'ordre 3. On confond E avec ABC et F avec ABD.

On va ensuite construire les facteurs 1, 2, 3, et 4 à partir de ce plan.

On construit les facteurs 1 à partir de A et B, et 2 à partir de C et D, selon le codage suivant :

$$(1 \ 1) = 1 \quad (1 \ -1) = 2 \quad (-1 \ 1) = 3 \quad (-1 \ -1) = 4$$

On obtient le plan suivant, qu'on traduit dans les termes de l'expérience :

	1 = (A B)	2 = (C D)	3 = (E)	4 = (F)	Juge	Diluant	Sirop	Glaçons
1	1	1	1	1	1 A	Eau	Grenadine	Glacons
2	1	2	1	-1	2 A	Limonade	Grenadine	Non
3	1	3	-1	1	3 A	Jus d'orange	Pêche	Glacons
4	1	4	-1	-1	4 A	Lait	Pêche	Non
5	2	1	-1	-1	5 S	Eau	Pêche	Non
6	2	2	-1	1	6 S	Limonade	Pêche	Glacons
7	2	3	1	-1	7 S	Jus d'orange	Grenadine	Non
8	2	4	1	1	8 S	Lait	Grenadine	Glacons
9	3	1	-1	-1	9 R	Eau	Pêche	Non
10	3	2	-1	1	10 R	Limonade	Pêche	Glacons
11	3	3	1	-1	11 R	Jus d'orange	Grenadine	Non
12	3	4	1	1	12 R	Lait	Grenadine	Glacons
13	4	1	1	1	13 D	Eau	Grenadine	Glacons
14	4	2	1	-1	14 D	Limonade	Grenadine	Non
15	4	3	-1	1	15 D	Jus d'orange	Pêche	Glacons
16	4	4	-1	-1	16 D	Lait	Pêche	Non

A l'œil, le plan est de bonne qualité. Les modalités de chaque variable sont bien réparties.

Sur R :

```
plan_fractionnaire <- oa.design(nlevels = c(4,4,2,2),nruns=16)
plan_fractionnaire
```

```

  A B C D
1  1 2 2 2
2  1 3 1 2
3  3 3 2 1
4  3 2 1 1
5  2 1 2 1
6  2 4 1 1
7  4 1 1 2
8  4 2 2 1
9  3 1 2 2
10 2 2 1 2
11 2 3 2 2
12 4 3 1 1
13 4 4 2 2
14 3 4 1 2
15 1 1 1 1
16 1 4 2 1
class=design, type= oa
```

3) Pourra-t-on estimer les effets d'interaction avec ce plan ?

On peut étudier les effets principaux sans aucune confusion, car ils sont parfaitement orthogonaux :

```
X <- model.matrix(~ A + B + C + D, data = plan_fractionnaire)
round(solve(t(X) %*% X), digit=4)
```

```

(Intercept) (Intercept) A.L A.Q A.C B.L B.Q B.C C1 D1
(Intercept) 0.0625 0.00 0.00 0.00 0.00 0.00 0.00 0.0000 0.0000
A.L 0.0000 0.25 0.00 0.00 0.00 0.00 0.00 0.00 0.0000 0.0000
A.Q 0.0000 0.00 0.25 0.00 0.00 0.00 0.00 0.00 0.0000 0.0000
A.C 0.0000 0.00 0.00 0.25 0.00 0.00 0.00 0.00 0.0000 0.0000
B.L 0.0000 0.00 0.00 0.00 0.25 0.00 0.00 0.00 0.0000 0.0000
B.Q 0.0000 0.00 0.00 0.00 0.00 0.25 0.00 0.00 0.0000 0.0000
B.C 0.0000 0.00 0.00 0.00 0.00 0.00 0.25 0.00 0.0000 0.0000
C1 0.0000 0.00 0.00 0.00 0.00 0.00 0.00 0.0625 0.0000 0.0000
D1 0.0000 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.0000 0.0625
```

En revanche nous ne pouvons pas étudier les effets d'interaction, car la matrice (X'X) n'est alors pas inversible.

- 4) Comme dit plus haut, les colocataires sont prêts à réaliser jusqu'à 6 essais chacun, soit 24 au total. Avec l'algorithme de Federov, tirer les 24 meilleurs essais du plan complet. Le plan obtenu est-il de suffisamment bonne qualité pour pouvoir étudier les interactions sans trop de confusions ?

```
# Extraction d'un plan en 24 essais avec l'algorithme de Federov
library(AlgDesign)
meilleur_plan <- optFederov(~ ., data = plan_complet, nTrials = 24)
as.data.frame(meilleur_plan$design)
```

```
  A B C D
4  3 2 1 2
5  3 4 2 2
12 2 3 2 2
14 2 2 2 1
15 2 4 2 2
16 4 1 1 2
18 3 3 2 2
22 2 2 1 2
23 3 4 1 1
24 3 1 2 1
25 1 1 1 2
31 4 3 1 2
34 4 2 1 1
35 1 1 2 1
48 1 3 1 1
51 1 2 2 2
52 4 4 2 1
53 4 1 2 2
54 1 3 2 1
57 3 3 1 1
58 1 4 1 2
59 2 1 1 1
60 2 4 1 1
63 4 2 2 1
```

Voyons la qualité de ce plan en calculant $(X'X)^{-1}$.

```
#Qualité de ce plan
X <- model.matrix(~ A + B + C + D + A:B + A:C + A:D + B:C + B:D + C:D, data =
as.data.frame(meilleur_plan$design))
solve(t(X) %*% X, tol = 1e-20)
```

La matrice renvoyée est très lourde, mais quoi qu'il en soit le plan n'est pas suffisant car il y a de nombreuses confusions, et la diagonale est loin d'être minimisée (toutes les valeurs de la matrice sont $> 10^{12}$ en valeur absolue).

Et si l'on se restreint aux effets principaux, on introduit avec ces 24 essais des confusions :

```
X <- model.matrix(~ A + B + C + D, data = as.data.frame(meilleur_plan$design))
round(solve(t(X) %*% X), digit=4)
```

(Intercept)	A.L	A.Q	A.C	B1	C.L	C.Q	C.C	D1
(Intercept)	0.0417	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
A.L	0.0000	0.1686	0.0000	0.0039	0.0000	0.0176	0.0000	0.0059
A.Q	0.0000	0.0000	0.1765	0.0000	-0.0132	0.0000	0.0395	0.0000
A.C	0.0000	0.0039	0.0000	0.1745	0.0353	0.0000	0.0118	0.0000
B1	0.0000	0.0000	0.0000	0.0000	0.0417	0.0000	0.0000	0.0000
C.L	0.0000	0.0176	-0.0132	0.0353	0.0000	0.1765	0.0000	0.0000
C.Q	0.0000	0.0000	0.0000	0.0000	0.0000	0.1667	0.0000	0.0000
C.C	0.0000	0.0059	0.0395	0.0118	0.0000	0.0000	0.0000	0.1765
D1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0417

On préfère donc se limiter à 16 essais et réaliser l'expérience selon le premier plan.

5) Analyser les résultats :

Juge	Diluant	Sirop	Glacons	Appreciation
A	Eau	Grenadine	Oui	15
A	Limonade	Grenadine	Non	12
A	Jus_orange	Peche	Oui	13
A	Lait	Peche	Non	10
S	Eau	Peche	Non	13
S	Limonade	Peche	Oui	18
S	Jus_orange	Grenadine	Non	14

S	Lait	Grenadine	Oui	17
R	Eau	Peche	Non	10
R	Limonade	Peche	Oui	12
R	Jus_orange	Grenadine	Non	14
R	Lait	Grenadine	Oui	18
D	Eau	Grenadine	Oui	17
D	Limonade	Grenadine	Non	12
D	Jus_orange	Peche	Oui	11
D	Lait	Peche	Non	13

```
#Analyse des résultats
result<-read.table("resultats.csv", header=TRUE, sep=";", stringsAsFactors=TRUE)
mod1 <- lm(Appreciation ~ Sirop + Glacons + Juge + Diluant, data = result)
summary(mod1)
```

Call:

```
lm.default(formula = Appreciation ~ Sirop + Glacons + Juge + Diluant, data = result)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-1.9375 -0.8438 -0.1250  1.0625  2.4375
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.3125     1.4297   8.612 5.68e-05 ***
SiropPeche     -2.3750     0.9531  -2.492  0.0415 *
GlaconsOui      2.8750     0.9531   3.016  0.0195 *
JugeAldric      0.7500     1.3479   0.556  0.5953
JugeRaphael     1.0000     1.3479   0.742  0.4823
JugeSolene      3.0000     1.3479   2.226  0.0614 .
DiluantJus_orange -0.7500     1.3479  -0.556  0.5953
DiluantLait     0.7500     1.3479   0.556  0.5953
DiluantLimonade -0.2500     1.3479  -0.185  0.8581
```

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

```
Residual standard error: 1.906 on 7 degrees of freedom
Multiple R-squared:  0.7587, Adjusted R-squared:  0.483
F-statistic: 2.752 on 8 and 7 DF, p-value: 0.09998
```

Les résultats de l'expérience de la colocation ne sont pas très probants. D'après la statistique de Fisher et la p-value associée, le modèle n'est pas vraiment significatif par rapport à un modèle nul (ou alors avec un seuil de 10%). Nous pouvons dire deux choses mais sans trop en être sûr : il semblerait que les boissons soient plus appréciées avec glaçons (coefficient positif), et que le sirop de pêche ne soit pas très apprécié (coefficient négatif).

En réalité, le thème serait propice à l'étude des interactions. En effet, cela permettrait d'étudier plus en détail les données : les goûts de chacun avec l'interaction juge:diluant ou juge:sirop, quel sirop va avec le meilleur diluant et inversement avec les interactions sirop:diluant... Or, 16 expériences est un nombre beaucoup trop faible pour estimer toutes ces interactions. On aurait pu faire l'expérience en intégrant moins de facteurs (par exemple, supprimer le facteur glaçon) et moins de modalités (moins de diluants par exemple), ainsi il aurait fallu moins d'expériences pour évaluer des interactions. Dommage pour eux.

Partie 2 : le tequila sunrise

Pour ses apéros et soirées, la coloc est amateur de cocktails. L'un de leurs favoris est le tequila sunrise : mélange de grenadine, de tequila et de jus d'orange. La question se pose de savoir quelles sont les proportions idéales de grenadine, tequila et jus d'orange pour un cocktail parfait (bien que cela dépende des goûts de chacun).

Il s'agit donc dans cette partie de déterminer les proportions optimales de sirop de grenadine et de tequila, qui seront nos deux variables quantitatives. La proportion de jus d'orange elle, dépend de la

proportion des deux autres. Les colocataires décident de faire varier les quantités de tequila et de grenadine entre 8 % et 32% (20% +/- 12%). Ils ne souhaitent pas dépasser les 20 essais pour leur expérience. La variable réponse sera une note d'appréciation pour chaque colocataire, sur 20. Les expériences réalisées permettront de dessiner une surface de réponse par personne, éventuellement une surface de réponse moyenne, et de l'utiliser pour déterminer ces quantités optimales.

- 1) Les colocataires réfléchissent à quel type de plan expérimental pour surface de réponses réaliser : plan composite centré, plan de Box-Behnken, ...
 Quel est l'avantage pratique des plans de Box-Behnken ? Peut-on ici utiliser un plan de Box-Behnken, ou un autre plan similaire ?
 Réaliser les plans possibles sur R, comparer les qualités des plans et trancher.

Les plans de Box-Behnken sont avantageux en pratique car pour chaque facteur, il n'y a besoin de tester que 3 niveaux : -1, 0 et 1. Tandis que pour un plan composite centré, les plans en étoile s'ajoutent aux niveaux -1, 0 et 1, ce qui vient ajouter 2 niveaux à tester pour chaque facteur.

Mais on ne peut pas réaliser de plan de Box-Behnken ici, en effet il faut 3 facteurs minimum pour cela et nous n'en avons que 2.

Mais en observant le plan de Box-Behnken à 3 facteurs, nous pouvons nous en inspirer pour construire un plan similaire pour 2 facteurs, le plan factoriel à 3 niveaux.

```
library(rsm)
#plan box-behnken à 3 facteurs
bbd(3)

run.order std.order x1.as.is x2.as.is x3.as.is
1          1          8          1          0          1
2          2          5         -1          0         -1
3          3         10          0          1         -1
4          4          1         -1         -1          0
5          5          6          1          0         -1
6          6          7         -1          0          1
7          7         15          0          0          0
8          8          2          1         -1          0
9          9         12          0          1          1
10         10          3         -1          1          0
11         11          4          1          1          0
12         12         16          0          0          0
13         13         13          0          0          0
14         14          9          0         -1         -1
15         15         11          0         -1          1
16         16         14          0          0          0
```

Le plan de Box-Behnken est constitué de toutes les combinaisons de -1, 0 et 1 possibles, avec plusieurs essais au centre (0 0 0).

Le plan composite centré pour 2 facteurs comporte 16 essais et 8 essais au centre, comme nous pouvons le voir en le générant sur R :

```
ccd(2)

run.order std.order x1.as.is x2.as.is Bblock
1          1          2  1.000000 -1.000000  1
2          2          1 -1.000000 -1.000000  1
3          3          3 -1.000000  1.000000  1
4          4          5  0.000000  0.000000  1
5          5          7  0.000000  0.000000  1
6          6          6  0.000000  0.000000  1
7          7          8  0.000000  0.000000  1
8          8          4  1.000000  1.000000  1
9          1          5  0.000000  0.000000  2
10         2          8  0.000000  0.000000  2
11         3          2  1.414214  0.000000  2
12         4          6  0.000000  0.000000  2
13         5          1 -1.414214  0.000000  2
14         6          7  0.000000  0.000000  2
15         7          3  0.000000 -1.414214  2
16         8          4  0.000000  1.414214  2
```

Nous pouvons réaliser un plan pour 2 facteurs se rapprochant du plan de Box-Behnken, en écrivant toutes les combinaisons faisables avec -1, 0 et 1, et avec 8 points au centre comme dans le plan composite centré pour évaluer les effets quadratiques et l'erreur pure. Ce plan est le plan factoriel à 3 niveaux suivant :

	x	x2
1	-1	-1
2	-1	1
3	-1	0
4	1	-1
5	1	1
6	1	0
7	0	-1
8	0	1
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0

Pour notre expérience, il s'agit donc de trancher entre ce plan et le plan composite centré. Pour cela, nous comparons les qualités à l'aide de la matrice inverse du produit ($X'X$) (X étant la matrice des effets et X' sa transposée), pour les deux plans proposés.

Plan composite centré :

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

	(Intercept)	x1	x2	I(x1^2)	I(x2^2)	I(x1 * x2)
(Intercept)	0.1250	0.0000	0.0000	-0.0625	-0.0625	0.00
x1	0.0000	0.125	0.000	0.0000	0.0000	0.00
x2	0.0000	0.000	0.125	0.0000	0.0000	0.00
I(x1^2)	-0.0625	0.000	0.000	0.1250	0.0000	0.00
I(x2^2)	-0.0625	0.000	0.000	0.0000	0.1250	0.00
I(x1 * x2)	0.0000	0.000	0.000	0.0000	0.0000	0.25

Plan fractionnaire à 3 niveaux, avec 8 points au centre :

```
X <- model.matrix(~A+B+I(A^2)+I(B^2)+I(A*B), data=plan_factoriel)
solve(t(X)%*%X)
```

	(Intercept)	A	B	I(A^2)	I(B^2)	I(A * B)
(Intercept)	0.1136	0.0000	0.0000	-0.0682	-0.0682	0.00
A	0.0000	0.1667	0.0000	0.0000	0.0000	0.00
B	0.0000	0.0000	0.1667	0.0000	0.0000	0.00
I(A^2)	-0.0682	0.0000	0.0000	0.3409	-0.1591	0.00
I(B^2)	-0.0682	0.0000	0.0000	-0.1591	0.3409	0.00
I(A * B)	0.0000	0.0000	0.0000	0.0000	0.0000	0.25

Nous voyons que les valeurs sur la diagonale sont globalement plus faibles pour le plan composite centré : la variance est davantage minimisée pour ce plan. De plus, avec le plan factoriel à 3 niveaux, nous avons non-orthogonalité entre les effets quadratiques, il y aurait donc une importante confusion pour leur estimation.

Nous nous attendons à observer des effets quadratiques : en effet l'appréciation d'un cocktail peut être moins bonne s'il n'y a pas assez de grenadine/tequila, mais aussi s'il y en a trop. Il est donc important pour cette expérience d'évaluer correctement ces effets quadratiques. Par conséquent, il vaut mieux se tourner vers le plan composite centré, bien qu'il soit moins pratique car chaque facteur doit être testé à 5 et non 3 niveaux.

Nous déterminons les proportions de grenadine et de tequila à tester avec la fonction ccd du package rsm :

```
set.seed(1234)
plan_ccd <- ccd(2, coding = list(x1~(Grenadine-20)/12, x2~(Tequila-20)/12))
plan_ccd
```

```
run.order std.order Grenadine Tequila Block
1 1 7 20.000000 20.000000 1
2 2 2 32.000000 8.000000 1
3 3 6 20.000000 20.000000 1
4 4 1 8.000000 8.000000 1
5 5 4 32.000000 32.000000 1
6 6 3 8.000000 32.000000 1
7 7 8 20.000000 20.000000 1
8 8 5 20.000000 20.000000 1
9 1 1 3.029437 20.000000 2
10 2 7 20.000000 20.000000 2
11 3 4 20.000000 36.970563 2
12 4 2 36.970563 20.000000 2
13 5 5 20.000000 20.000000 2
14 6 8 20.000000 20.000000 2
15 7 6 20.000000 20.000000 2
16 8 3 20.000000 3.029437 2
```

Data are stored in coded form using these coding formulas ...
 $x_1 \sim (\text{Grenadine} - 20)/12$
 $x_2 \sim (\text{Tequila} - 20)/12$

2) Pour réaliser correctement les expériences, à quoi devra-t-on prêter attention ? Détailler le déroulé de l'expérience.

L'expérience demande de réaliser 16 échantillons. Par soucis de précision lors du dosage, les volumes de chaque échantillon ne doivent pas être trop faibles. On se porte sur un volume de 20cL par échantillon.

De plus, les quantités à verser sont assez précises, et non dosables avec un verre doseur, et les colocataires n'ont pas d'éprouvette. En revanche ils ont à leur disposition une balance précise au g près. La mesure des quantités à verser se fera donc à la masse. Dès lors, il faut prêter attention à la masse volumique de la grenadine, de la tequila et du jus d'orange, qui sont déterminés à partir de la proportion d'alcool pour la tequila, et par la pesée de grands échantillons pour la grenadine et le jus d'orange.

Les échantillons pour les 8 points au centre devront bien entendu être faits séparément, et non pas en un gros volume séparé ensuite en plusieurs verres. De plus, la grenadine, la tequila et le jus d'orange utilisés doivent être de la même marque, pour ne pas avoir d'effet marque/qualité.

Enfin, il est mieux que les échantillons soient goûtés par chacun des 4 colocataires, dans un ordre aléatoire.

Voici le tableau des expériences réalisées :

Numéro plan	Grenadine			Tequila			Jus d'orange	
	En pourcentage	En cL	En g	En pourcentage	En cL	En g	En cL	En g
1	20	4	48	20	4	37	12	121
2	32	6,4	77	8	1,6	15	12	121
3	20	4	48	20	4	37	12	121
4	8	1,6	19	8	1,6	15	17	170
5	32	6,4	77	32	6,4	59	7	73
6	8	1,6	19	32	6,4	59	12	121
7	20	4	48	20	4	37	12	121
8	20	4	48	20	4	37	12	121
9	3	0,6	7	20	4	37	15	156
10	20	4	48	20	4	37	12	121
11	20	4	48	37	7,4	69	9	87
12	37	7,4	89	20	4	37	9	87
13	20	4	48	20	4	37	12	121
14	20	4	48	20	4	37	12	121
15	20	4	48	20	4	37	12	121
16	20	4	48	3	0,6	6	15	156

3) Les 4 colocataires ont goûté chacun des 16 échantillons et ont attribué leurs notes.

Les goûts de chacun sont tellement différents que les variances des appréciations moyennes sont très élevées et la moyenne n'a plus de sens. Nous nous attarderons sur les résultats de deux colocataires, A et S :

A : 13,11,16,7,7,8,14,14,6,12,7,15,14,13,14,11

S : 11,9,13,15,7,9,11,9,15,13,7,6,14,13,7,14

(dans l'ordre des expériences définies par le plan composite centré)

Analyser les résultats, dessiner les surfaces de réponse de ces deux colocataires, et trouver leur cocktail idéal. Commentez les erreurs pures. La tequila change-t-elle l'appréciation de la grenadine ?

Résultats de A

```
Appreciation_A <- c(13,11,16,7,7,8,14,14,6,12,7,15,14,13,14,11)
CR.rsm.A <- rsm(Appreciation_A~SO(x1,x2), data=plan_ccd)
summary(CR.rsm.A)
Call:
rsm(formula = Appreciation_A ~ SO(x1, x2), data = plan_ccd)
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.75000    0.57783  23.7957 3.902e-10 ***
x1           1.96599    0.57783   3.4023 0.0067445 **
x2          -1.08211    0.57783  -1.8727 0.0906060 .
x1:x2       -1.25000    0.81718  -1.5296 0.1570998
x1^2        -2.00000    0.57783  -3.4612 0.0061105 **
x2^2        -2.75000    0.57783  -4.7591 0.0007695 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Multiple R-squared:  0.8388, Adjusted R-squared:  0.7583
F-statistic: 10.41 on 5 and 10 DF, p-value: 0.001027
Analysis of Variance Table
Response: Appreciation_A
      Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)  2 40.289  20.144  7.5415 0.0100719
TWI(x1, x2)  1  6.250   6.250  2.3398 0.1570998
PQ(x1, x2)  2 92.500  46.250 17.3147 0.0005648
Residuals   10 26.711   2.671
Lack of fit  3 17.211   5.737  4.2274 0.0530763
Pure error   7  9.500   1.357
Stationary point of response surface:
      x1      x2
0.5952577 -0.3320325
Stationary point in original units:
Grenadine Tequila
 27.14309 16.01561
Eigenanalysis:
eigen() decomposition
$values
[1] -1.646131 -3.103869
$vectors
      [,1]      [,2]
x1 -0.8701999 0.4926988
x2  0.4926988 0.8701999
```

L'erreur d'ajustement a une p-value supérieure à 5%, donc le modèle retenu est significatif. La p-value la plus faible est obtenue pour les effets quadratiques, comme on s'y attendait, ce sont ces effets qui sont les plus présents. Nous voyons en revanche que l'effet d'interaction n'est pas significatif. Nous retenons donc le modèle sans les effets d'interaction.

Le modèle est actualisé pour afficher la surface de réponse et trouver l'optimum :

```
CR.rsm.A2 <- rsm(Appreciation_A ~ FO(x1,x2) + PQ(x1,x2), data=plan_ccd)
summary(CR.rsm.A2)
contour(CR.rsm.A2,~x1+x2,image=TRUE)
persp(CR.rsm.A2,~x1+x2,col=rainbow(50),contours="colors")
```

```
Call:
rsm(formula = Appreciation_A ~ FO(x1, x2) + PQ(x1, x2),
data = plan_ccd)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.75000	0.61201	22.4668	1.53e-10 ***
x1	1.96599	0.61201	3.2123	0.0082712 **
x2	-1.08211	0.61201	-1.7681	0.1047308
x1^2	-2.00000	0.61201	-3.2679	0.0074931 **
x2^2	-2.75000	0.61201	-4.4934	0.0009111 ***

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

Multiple R-squared: 0.8011, Adjusted R-squared: 0.7288
F-statistic: 11.08 on 4 and 11 DF, p-value: 0.0007498

Analysis of Variance Table

Response: Appreciation_A	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	40.289	20.144	6.7226	0.0123763
PQ(x1, x2)	2	92.500	46.250	15.4347	0.0006415
Residuals	11	32.961	2.996		
Lack of fit	4	23.461	5.865	4.3218	0.0448737
Pure error	7	9.500	1.357		

Stationary point of response surface:

x1	x2
0.4914976	-0.1967467

Stationary point in original units:

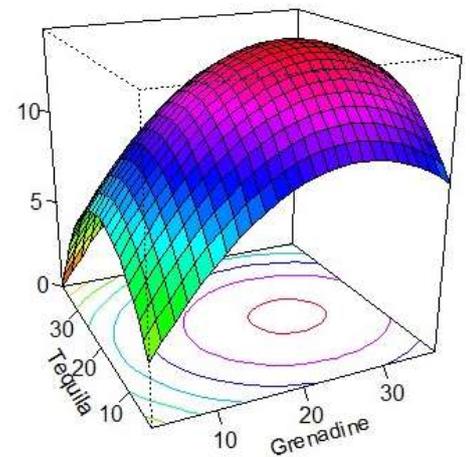
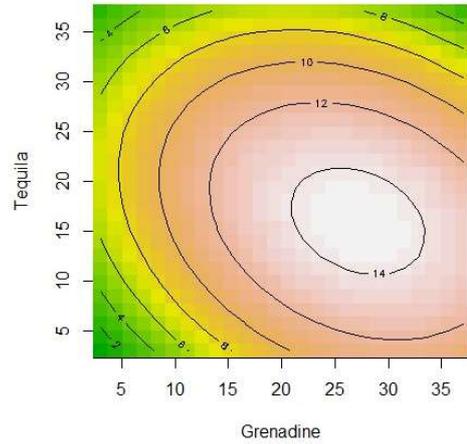
Grenadine	Tequila
25.89797	17.63904

Eigenanalysis:

```
eigen() decomposition
$values
[1] -2.00 -2.75
```

\$vectors

	[,1]	[,2]
x1	-1	0
x2	0	1



La surface de réponse présente un optimum, et cet optimum est un maximum car les deux valeurs propres sont négatives. Ce maximum théorique d'appréciation se situe pour 26% de grenadine et 18% de tequila. Pour confirmer ce résultat, une expérience complémentaire a été réalisée. Le cocktail aux proportions idéales pour A a été réalisé, et A l'a goûté parmi d'autres. Celui correspondant à l'optimum théorique a obtenu une note de 17, soit supérieure à toutes les notes déjà attribuées.

Résultats de S

```
Appreciation_S <- c(11,9,13,15,7,9,11,9,15,13,7,6,14,13,7,14)
CR.rsm.S <- rsm(Appreciation_S~SO(x1,x2), data=plan_ccd)
summary(CR.rsm.S)
```

Call:

```
rsm(formula = Appreciation_S ~ SO(x1, x2), data = plan_ccd)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	11.37500	0.73841	15.4047	2.706e-08 ***
x1	-2.59099	0.73841	-3.5089	0.005642 **
x2	-2.23744	0.73841	-3.0301	0.012676 *
x1:x2	1.00000	1.04427	0.9576	0.360841
x1^2	-0.56250	0.73841	-0.7618	0.463784
x2^2	-0.56250	0.73841	-0.7618	0.463784

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

Multiple R-squared: 0.7021, Adjusted R-squared: 0.5532
F-statistic: 4.714 on 5 and 10 DF, p-value: 0.01795

Analysis of Variance Table

Response: Appreciation_S	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	93.755	46.877	10.7467	0.003228
TWI(x1, x2)	1	4.000	4.000	0.9170	0.360841
PQ(x1, x2)	2	5.062	2.531	0.5803	0.577517
Residuals	10	43.620	4.362		
Lack of fit	3	3.745	1.248	0.2192	0.880131
Pure error	7	39.875	5.696		

```

Stationary point of response surface:
      x1      x2
-19.39690 -19.23052
Stationary point in original units:
Grenadine Tequila
-212.7628 -210.7662
Eigenanalysis:
eigen() decomposition
$values
[1] -0.0625 -1.0625
$vectors
      [,1]      [,2]
x1 -0.7071068 -0.7071068
x2 -0.7071068  0.7071068

```

La moyenne des carrés de l'erreur pure est environ 4 fois plus élevée pour S que pour A : on peut dire que sa perception du goût des cocktails est moins affûtée que celle de A, elle a donné des notes plus variées aux 8 cocktails pourtant identiques (points au centre).

La p-value de l'erreur d'ajustement est supérieure à 5%, donc le modèle retenu est significatif. Le R2 vaut 0,5532, soit le modèle explique 55% de la note d'appréciation, ce qui est moins bien que pour A. On voit que les effets d'interaction et effets quadratiques ne sont pas significatifs. Nous pouvons donc améliorer notre modèle, et essayons un modèle pour S, en intégrant uniquement les effets linéaires, et nous affichons la surface de réponse.

```

CR.rsm.S2 <- rsm(Appreciation_S~FO(x1,x2), data=plan_ccd)
summary(CR.rsm.S2)
contour(CR.rsm.S2,~x1+x2,image=TRUE)
persp(CR.rsm.S2,~x1+x2,col=rainbow(50),contours="colors")

```

```

Call:
rsm(formula = Appreciation_S ~ FO(x1, x2), data = plan_ccd)

```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.81250	0.50327	21.4844	1.536e-11	***
x1	-2.59099	0.71173	-3.6404	0.002992	**
x2	-2.23744	0.71173	-3.1436	0.007766	**

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

Multiple R-squared: 0.6402, Adjusted R-squared: 0.5849
F-statistic: 11.57 on 2 and 13 DF, p-value: 0.0013

Analysis of Variance Table

Response: Appreciation_S	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	93.755	46.877	11.5675	0.0013
Residuals	13	52.683	4.053		
Lack of fit	6	12.808	2.135	0.3747	0.8738
Pure error	7	39.875	5.696		

Direction of steepest ascent (at radius 1):

```

      x1      x2
-0.7568572 -0.6535803

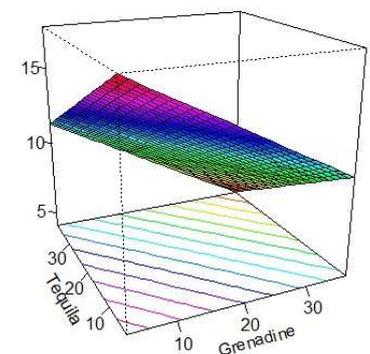
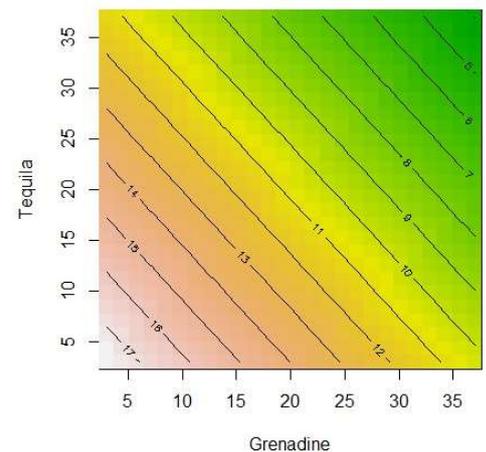
```

Corresponding increment in original units:

```

Grenadine Tequila
-9.082286 -7.842964

```



Les effets linéaires sont bien significatifs. Le R2 est plus important avec 58%, mais le modèle reste d'une qualité moyenne, sans doute à cause de l'importance de l'erreur pure.

Il n'y a pas d'optimum dans ce modèle comportant uniquement des effets linéaires. Au vu du modèle et de la surface de réponse, nous arrivons à cette conclusion : plus il y a de tequila/grenadine, plus l'appréciation de S est faible. Et inversement, moins il y a de tequila/grenadine, et plus son appréciation est élevée. Finalement, nous pouvons conseiller à S de boire du jus d'orange.

Nous pouvons aussi faire un commentaire sur l'absence d'effet d'interaction : la quantité de grenadine n'a pas de différence sur la manière dont on apprécie la tequila, et inversement.

- 4) **Après coup, les colocataires aimeraient tester un éventuel effet du taux de sucre dans le jus d'orange. Alors que le jus utilisé pour leurs premières expériences contient 10g/L, ils ont à disposition deux autres jus, à 9g/L et 11g/L. Ils ne souhaitent pas faire plus de 10 expériences. Quel plan d'expérience réaliser ? Vont-ils pouvoir étudier cet effet en moins de 10 expériences ?**

Le facteur taux de sucre qui est rajouté dans l'expérience peut être testé à 3 niveaux, et les colocataires l'ont déjà testé au niveau 0. On peut dire que les jus à 9g/L et 11g/L correspondent respectivement aux niveaux -1 et 1. Si le facteur « taux de sucre » n'est testé qu'à 3 niveaux, les deux autres doivent aussi n'être testé qu'à 3 niveaux. Il faut donc opter pour un plan de Box-Behnken dans ce cas.

R nous renvoie le plan de Box-Behnken pour 3 facteurs :

```
library(rsm)
set.seed(1234)
plan_bbd <- bbd(3)
plan_bbd
```

	run.order	std.order	x1.as.is	x2.as.is	x3.as.is
1	1	16	0	0	0
2	2	11	0	-1	1
3	3	14	0	0	0
4	4	10	0	1	-1
5	5	4	1	1	0
6	6	3	-1	1	0
7	7	15	0	0	0
8	8	9	0	-1	-1
9	9	6	1	0	-1
10	10	2	1	-1	0
11	11	12	0	1	1
12	12	1	-1	-1	0
13	13	7	-1	0	1
14	14	8	1	0	1
15	15	13	0	0	0
16	16	5	-1	0	-1

Data are stored in coded form using these coding formulas ...

```
x1 ~ x1.as.is
x2 ~ x2.as.is
x3 ~ x3.as.is
```

Or, dans leur premier ensemble d'expériences, les colocataires ont déjà réalisé les expériences suivantes : (0 0 0) (qui apparait 4 fois) (1 1 0) (-1 1 0) (1 -1 0) (-1 -1 0)

Sur les 16, il ne reste donc que 8 expériences à réaliser. Ils vont pouvoir réaliser leur expérience en seulement 8 essais supplémentaires. Il risque cependant d'y avoir un effet « date », car leur notation pourrait hypothétiquement être différente quelques jours plus tard...

Lors de la session précédente, 8 expériences « points au centre » ont été réalisées, mais dans le plan bbd(3), il y a 4 essais au centre. La question se pose alors : est-il mieux de considérer le plan classique bbd(3) avec 4 essais au centre, et en choisir 4 au hasard parmi ceux déjà effectués, ou bien doit-on considérer un plan bbd(3) auquel on ajoute 4 essais au centre ?

Pour trancher, nous pouvons évaluer la qualité des deux plans sur R à l'aide de la matrice $(X'X)^{-1}$:

```
#plan box-behnken à 3 facteurs
plan_bbd <- bbd(3)
#Qualité de ce plan
X <- model.matrix(~SO(x1,x2,x3), data=plan_bbd)
solve(t(X)%*%X)
```

	(Intercept)	so(x1, x2, x3)x1	so(x1, x2, x3)x2	so(x1, x2, x3)x3	so(x1, x2, x3)x1:x2	so(x1, x2, x3)x1:x3
(Intercept)	0.250	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x1	0.000	0.125	0.000	0.000	0.000	0.00
so(x1, x2, x3)x2	0.000	0.000	0.125	0.000	0.000	0.00
so(x1, x2, x3)x3	0.000	0.000	0.000	0.125	0.000	0.00
so(x1, x2, x3)x1:x2	0.000	0.000	0.000	0.000	0.125	0.00
so(x1, x2, x3)x1:x3	0.000	0.000	0.000	0.000	0.000	0.125
so(x1, x2, x3)x2:x3	0.000	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x1^2	-0.125	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x2^2	-0.125	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x3^2	-0.125	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x1^2	0.00	-0.125	0.000	-0.125	-0.125	0.00
so(x1, x2, x3)x2^2	0.00	0.000	-0.125	0.000	0.000	0.00
so(x1, x2, x3)x3^2	0.00	0.000	0.000	-0.125	0.000	0.00
so(x1, x2, x3)x1:x2	0.00	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x1:x3	0.00	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x2:x3	0.25	0.000	0.000	0.000	0.000	0.00
so(x1, x2, x3)x1^2	0.00	0.00	0.250	0.000	0.000	0.00

```
so(x1, x2, x3)x2^2      0.00      0.000      0.250      0.000
so(x1, x2, x3)x3^2      0.00      0.000      0.000      0.250
```

```
#Avec 8 essais au centre au total
plan_bbd2 <- bbd(3)
center <- matrix(0, ncol = ncol(plan_bbd2), nrow = 4, dimnames = list(NULL,
colnames(plan_bbd2)))
plan_bbd3 <- rbind(center, plan_bbd2)
#Qualité de ce plan
X <- model.matrix(~SO(x1,x2,x3), data=plan_bbd3)
solve(t(X)%*%X)
```

```
(Intercept) so(x1, x2, x3)x1 so(x1, x2, x3)x2 so(x1, x2, x3)x3 so(x1, x2, x3)x1:x2 so(x1, x2, x3)x1:x3
so(x1, x2, x3)x1      0.1250      0.000      0.000      0.000      0.000      0.00      0.00
so(x1, x2, x3)x2      0.0000      0.125      0.000      0.000      0.000      0.00      0.00
so(x1, x2, x3)x3      0.0000      0.000      0.125      0.000      0.000      0.00      0.00
so(x1, x2, x3)x1:x2    0.0000      0.000      0.000      0.000      0.125      0.00      0.00
so(x1, x2, x3)x1:x3    0.0000      0.000      0.000      0.000      0.000      0.25      0.00
so(x1, x2, x3)x2:x3    0.0000      0.000      0.000      0.000      0.000      0.00      0.25
so(x1, x2, x3)x1^2     -0.0625     0.000      0.000      0.000      0.000      0.00      0.00
so(x1, x2, x3)x2^2     -0.0625     0.000      0.000      0.000      0.000      0.00      0.00
so(x1, x2, x3)x3^2     -0.0625     0.000      0.000      0.000      0.000      0.00      0.00
so(x1, x2, x3)x2:x3    0.0000      0.000      -0.0625    -0.0625    -0.0625    -0.0625
so(x1, x2, x3)x1^2     0.0000      0.0000     0.000000  0.000000  0.000000  0.000000
so(x1, x2, x3)x2^2     0.0000      0.0000     0.000000  0.000000  0.000000  0.000000
so(x1, x2, x3)x3^2     0.0000      0.0000     0.000000  0.000000  0.000000  0.000000
so(x1, x2, x3)x1:x2    0.0000      0.0000     0.000000  0.000000  0.000000  0.000000
so(x1, x2, x3)x1:x3    0.0000      0.0000     0.000000  0.000000  0.000000  0.000000
so(x1, x2, x3)x2:x3    0.2500      0.0000     0.000000  0.000000  0.000000  0.000000
so(x1, x2, x3)x1^2     0.0000      0.0000     0.21875   -0.03125   -0.03125
so(x1, x2, x3)x2^2     0.0000      0.0000     -0.03125   0.21875   -0.03125
so(x1, x2, x3)x3^2     0.0000      0.0000     -0.03125   -0.03125   0.21875
```

Les valeurs sur la diagonale sont globalement plus faibles pour le plan avec 8 essais au centre. Dans le plan avec 8 essais de plus, on n'a plus d'orthogonalité entre les effets quadratiques (confusion, avec $0,03125/0,21875 = 0,14$), en revanche les confusions avec intercept sont relativement plus faibles ($0,0625/0,21875 = 0,29$ contre $0,125/0,25 = 0,5$ pour 4 points au centre). $0,29 + 0,14 = 0,43 < 0,5$, on peut donc dire qu'il y a globalement moins de confusion avec le plan avec 8 essais au centre. On se tournerait plutôt sur ce choix.

Finalement, les colocalitaires peuvent conserver 12 expériences parmi leurs 16 premières et en effectuer 8 de plus pour évaluer l'effet du taux de sucre du jus d'orange.

Exercices divers

Exercice 1. Caractérisation d'une boisson par les sensations trigéminales perçues relatives à l'éthanol (exemple fictif, par Bastien DELEUZE)

L'exercice se base sur une thématique réelle mais des données fictives. Une équipe de recherche étudie l'effet de la teneur en éthanol (7.5, 10, 12.5, 15), de la concentration en glucides sous forme d'un mélange de fructose et de glucose à 50/50 (0, 4 g/L) et de la couleur du stimuli (transparent, framboise, pourpre et tuilé) sur les sensations trigéminales perçues lors de la dégustation d'une boisson alcoolisée dans le but de savoir s'il serait possible de formuler des équivalents no-/low- à des boissons alcoolisées ayant des propriétés sensorielles semblables aux produits d'origine. La problématique est donc d'estimer l'importance de différents facteurs dans la perception des sensations trigéminales d'une boisson.

1. Quel est le nombre d'essai à réaliser si on veut utiliser le plan complet associé à cette expérience. Donner le plan complet.

Dans le cas de cette étude, on s'intéresse à l'effet de 3 facteurs qualitatifs :

- F1 : la teneur en éthanol de la solution, facteur à 4 niveaux (E₁, E₂, E₃, E₄)
- F2 : la teneur en glucides, facteur à 2 niveaux (G₁, G₂)
- F3 : la coloration de la solution, facteur à 4 niveaux (C₁, C₂, C₃, C₄)

Le nombre d'essais correspondant au plan complet est égal à $2 \times 4^2 = 32$

```
library(DoE.base)
plancomp <- fac.design(nlevels=c(4,4,2), factor.names=list(CEthanol=c("7,5", "10", "12.5", "15"),
coloration=c("transparent", "framboise", "pourpre", "tuile"), CGlucides=c("0", "4")), randomize
=F)
plancomp
```

CEthanol	coloration	CGlucides					
1	7,5	transparent	0	29	7,5	tuile	4
2	10	transparent	0	30	10	tuile	4
3	12.5	transparent	0	31	12.5	tuile	4
4	15	transparent	0	32	15	tuile	4
5	7,5	framboise	0				
6	10	framboise	0				
7	12.5	framboise	0				
8	15	framboise	0				
9	7,5	pourpre	0				
10	10	pourpre	0				
11	12.5	pourpre	0				
12	15	pourpre	0				
13	7,5	tuile	0				
14	10	tuile	0				
15	12.5	tuile	0				
16	15	tuile	0				
CEthanol	coloration	CGlucides					
17	7,5	transparent	4				
18	10	transparent	4				
19	12.5	transparent	4				
20	15	transparent	4				
21	7,5	framboise	4				
22	10	framboise	4				
23	12.5	framboise	4				
24	15	framboise	4				
25	7,5	pourpre	4				
26	10	pourpre	4				
27	12.5	pourpre	4				
28	15	pourpre	4				

```
class=design, type= full factorial
```

2. Combien d'essai doit-on réaliser au minimum pour obtenir une orthogonalité des facteurs deux à deux. Donner le plan correspondant et commenter la qualité du plan.

Pour l'orthogonalité des facteurs, il faut minimum :

- $F1 \perp F2$ et $F2 \perp F3 = 4 \times 2 = 8$ essais
- $F1 \perp F3 = 4 \times 4 = 16$ essais

$N \geq \text{PPCM}(8, 16) = 16$.

On en conclut qu'il suffirait de 16 essais pour obtenir un plan présentant une orthogonalité de facteurs deux à deux.

```
planort <- ao.design(nlevels=c(4,4,2),factor.names=list(CEthanol=c("7,5","10","12.5","15"),
coloration=c("transparent", "framboise", "pourpre", "tuile"), CGlucides=c("0","4")),randomize =F)
```

```
planort
```

```
      CEthanol  coloration CGlucides
1         7,5 transparent      0
2         15      pourpre      0
3         10       tuile      0
4        12.5  framboise      0
5        12.5       tuile      0
6         10  framboise      0
7         15 transparent      0
8         7,5      pourpre      0
      CEthanol  coloration CGlucides
9         15  framboise      4
10        7,5       tuile      4
11        12.5  pourpre      4
12         10 transparent      4
13         10      pourpre      4
14        12.5 transparent      4
15         7,5  framboise      4
16         15       tuile      4
class=design, type= oa
```

Pour étudier un modèle comprenant les interactions de degrés 2, il faudrait estimer la constante, deux facteurs à 4 modalités, un facteur à 2 modalités et les interactions de degré 2 soit un total de $(1+1+2 \times 3 + (4 \times 4 + 2 \times (2 \times 4))) = 40$ paramètres à estimer. Or, on a avec ce plan 16 essais/possibles. On ne pourra donc pas estimer les interactions de degré 2. Le modèle retenu pour étudier ces résultats sera donc :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

Avec :

α_i : effet de la concentration en éthanol

β_j : effet de la coloration

γ_k : effet de la concentration en glucides

ε_{ijk} : variance résiduelle du modèle

Pour ce modèle, la variance résiduelle a un ddl associé de $16 - (1 + 1 + 2 \times 3) = 8$.

```
x <- model.matrix(~.,data=planort)
```

```
solve(t(x)%*%x)
```

```
(Intercept) CEth10 CEth12.5 CEth15 colframboise colpourpre coltuile  CGlul
(Intercept)  0.4375 -0.25    -0.25  -0.25        -0.25    -0.25    -0.25  0.0000
CEth10      -0.2500  0.50     0.25   0.25         0.00     0.00     0.00  0.0000
CEth12.5    -0.2500  0.25     0.50   0.25         0.00     0.00     0.00  0.0000
CEth15      -0.2500  0.25     0.25   0.50         0.00     0.00     0.00  0.0000
colframboise -0.2500  0.00     0.00   0.00         0.50     0.25     0.25  0.0000
colpourpre  -0.2500  0.00     0.00   0.00         0.25     0.50     0.25  0.0000
coltuile     -0.2500  0.00     0.00   0.00         0.25     0.25     0.50  0.0000
CGlul       0.0000  0.00     0.00   0.00         0.00     0.00     0.00  0.0625
```

On peut constater que la quasi-totalité des facteurs (à l'exception du facteur concentration en glucides) sont confondus avec la constante. Par ailleurs, on a bien une orthogonalité des facteurs deux à deux, ce qui est satisfaisant, mais on peut observer une confusion entre les modalités d'un même facteur. La variabilité dans l'estimation des coefficients des facteurs est assez élevée mais reste acceptable.

Pour utiliser ces produits dans le cadre d'une séance d'analyse sensoriel, proposer un plan de service de ces produits permettant de prendre en compte l'effet rang des produits.

3. Proposer un plan de service de ces produits permettant de prendre en compte l'effet rang des produits. On considère qu'un juge ne peut tester que 7 produits.

On utilise ici un carré latin (plan symétrique) pour proposer un plan de service des échantillons à un jury avec comme paramètres du plan :

- 3 facteurs :
 - F1 : Le Juge (16 modalités)
 - F2 : L'ordre de présentation tel que F2 = 1 correspond au premier service (16 modalités)
 - F4 : Le produit testé (16 modalités). Chaque produit correspond à un stimuli caractérisé par la combinaison des trois facteurs tel que décrit dans la question 2

		Rang															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Juge	1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1
	3	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2
	4	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3
	5	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3	4
	6	6	7	8	9	10	11	12	13	14	15	16	1	1	3	4	5
	7	7	8	9	10	11	12	13	14	15	16	1	2	2	4	5	6
	8	8	9	10	11	12	13	14	15	16	1	2	3	3	5	6	7
	9	9	10	11	12	13	14	15	16	1	2	3	4	4	6	7	8
	10	10	11	12	13	14	15	16	1	2	3	4	5	5	7	8	9
	11	11	12	13	14	15	16	1	2	3	4	5	6	6	8	9	10
	12	12	13	14	15	16	1	2	3	4	5	6	7	7	9	10	11
	13	13	14	15	16	1	2	3	4	5	6	7	8	8	10	11	12
	14	14	15	16	1	2	3	4	5	6	7	8	9	9	11	12	13
	15	15	16	1	2	3	4	5	6	7	8	9	10	10	12	13	14
	16	16	1	2	3	4	5	6	7	8	9	10	11	11	13	14	15

4. Discuter des limites d'un tel plan

Compte tenu des propriétés du plan, on ne peut étudier que les effets principaux, ce qui est dommage pour ce type d'étude où une interaction entre les facteurs aurait été très intéressante à étudier (c'est d'ailleurs un des objectifs principaux de cette étude).

En analysant ce plan de distribution des stimuli, on peut trouver certaines limites à son utilisation.

Tout d'abord, l'utilisation d'un carré latin ne permet pas d'étudier l'effet succession des produits, chaque produit étant toujours servi soit en premier soit après toujours le même produit. De plus, pour l'étude, il sera préférable, pour que tous les stimuli soit étudié un même nombre de fois, que le nombre de participant soit un multiple de 16, ce qui, dans le cas d'une analyse sensorielle, n'est pas idéal (le nombre de poste étant souvent de 12).

Par ailleurs, ce plan est équilibré car chaque produit est vu un même nombre de fois et à chaque rang de présentation.

Exercice 2. Cauchemar en cuisine (thématique réelle - DAVID Solène)

Partie I

Après avoir bu beaucoup de « tequila sunrise », Aldric, Raphaël, Alban et Solène se fixent l'objectif de manger plus sain. Pour cela Solène propose de remplacer les chips traditionnelles par des chips de chou. Cependant les premiers essais en cuisine sont infructueux ; la recette doit être optimisée ! 4 variables influençant l'appréciation des chips ont été relevées : le type de chou utilisé (vert ou blanc), l'assaisonnement (paprika ou mélange sel poivre), la température de cuisson (170 ou 200 °C) et le juge qui déguste les chips (4 possibilités). Disposant de peu de temps, Solène ne souhaite pas faire plus de 8 expériences.

1) Combien d'expériences faudrait-il pour réaliser le plan complet ? Quel plan fractionnaire doit on réaliser pour se contenter de 8 essais ?

Le plan complet est le plan $L_8 2^3 4$: il y a 2 variables à deux modalités et une variable à 4 modalités, il faudrait donc $2^{3+2} = 32$ expériences. Comme on souhaite se contenter de 8 expériences, on va réaliser un plan 2^{5-2} .

2) Donner la matrice des essais.

On peut considérer le facteur juge à 4 modalités comme 2 facteurs à 2 modalités, A et B qui seront confondues avec les interactions « type-assaisonnement » et « type-température » puis on construira un quatrième facteur avec l'interaction de A et B.

On commence par écrire le plan de base 2^3 :

I	1	2	3	12	13	23	123
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

On confond ensuite le facteur A avec l'interaction 12 et le facteur B avec l'interaction 13. Puis on utilise l'interaction AB pour représenter le facteur « juge », on obtient alors la matrice des essais suivante.

T°	Assais.	Type	Juge
170	sel-poivre	blanc	4
170	sel-poivre	vert	3
170	paprika	blanc	2
170	paprika	vert	1
200	sel-poivre	blanc	1
200	sel-poivre	vert	2
200	paprika	blanc	3

200	paprika	vert	4
-----	---------	------	---

Raphaël propose d'ajouter de l'huile ou du beurre fondu, car il ne se sent pas encore prêt à manger un légume non accompagné de matière grasse. Cela rajoute donc une variable.

3) En conservant un plan en 8 expériences ainsi que toutes les anciennes variables en plus de celle-ci, quelle serait la résolution du plan ? Combien y a-t-il de générateurs d'alias ?

D'après le cours, la résolution d'un plan à 6 facteurs et 8 expériences est de 3, cela signifie que des effets principaux sont confondus avec des interactions d'ordre 2 ou plus.

Le nombre de générateurs d'alias est égal à 2nombre de facteurs ajoutés -1, ici nous avons 6 facteurs en tout et un plan de base 2^3 , le nombre de facteurs ajoutés est donc de 3. Le nombre de générateurs d'alias est $2^3-1=7$

4) Quel modèle servira à l'analyse des résultats après réalisation des expériences ?

C'est le modèle d'analyse de la variance à 5 facteurs :

$$\forall(i,j,k,l,m) Y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \kappa_m + \epsilon_{ijklm},$$

avec α_i l'effet de la température, β_j l'effet de l'assaisonnement, γ_k l'effet du type de chou utilisé, δ_l l'effet de la matière grasse utilisée, κ_m l'effet juge et ϵ_{ijklm} les résidus.

Partie II

L'analyse des résultats de la partie I montre que la combinaison optimale est paprika-huile-chou vert-200°. Cependant les notes données par les juges restent basses, révélant un manque d'enthousiasme pour les bonnes chips de chou. Solène, ne désespérant pas de faire manger des légumes à ses colocataires décide de pousser l'expérience encore plus loin en se concentrant sur deux nouveaux facteurs : la taille des chips et la durée de cuisson. Mais attention il s'agit cette fois de déterminer précisément la taille et la température idéale ; on travaillera donc avec des variables quantitatives : taille entre 20 et 40 cm² et durée cuisson entre 5 et 15min.

1) Quel plan choisir : box-benkhen ou bien composite centré ?

L'avantage principal du plan de type box-benkhen est qu'il permet d'ajouter facilement des facteurs supplémentaires ce qui est pratique si l'on est pas sûr de ce que l'on veut vraiment tester. Mais ce plan est aussi moins précis car il n'utilise pas de points en étoile, testant chaque variable à 3 niveaux et non 5. Dans notre cas le plan de box-benkhen n'est donc pas pertinent car on n'envisage pas de rajouter des facteurs en cours d'expérience.

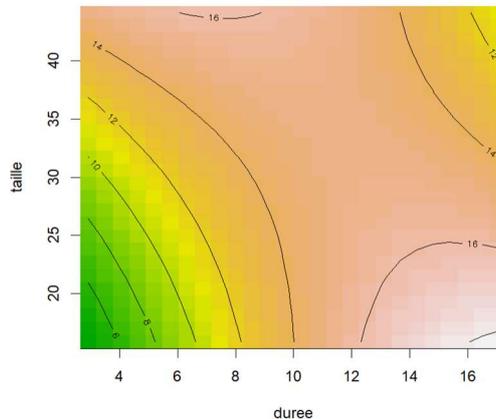
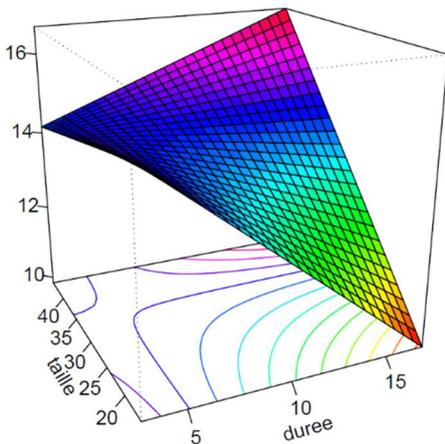
2) Le nombre maximal d'expériences est toujours de 16. Donnez les valeurs à tester pour chaque expérience à l'aide du logiciel R.

```
library(rsm)
set.seed(1234)
plan<-ccd(2,coding=list(x1~(duree-10)/5,x2~(taille-30)/10)) #données à étudier,
#donne l'instruction de faire le plan
plan #donne le plan des expériences à suivre
run.order std.order      duree  taille Block
1          1            2  15.000000 20.00000 1
2          2            4  15.000000 40.00000 1
3          3            8  10.000000 30.00000 1
4          4            1   5.000000 20.00000 1
5          5            7  10.000000 30.00000 1
6          6            6  10.000000 30.00000 1
7          7            3   5.000000 40.00000 1
8          8            5  10.000000 30.00000 1
9          1            2  17.071068 30.00000 2
```

10	2	8	10.000000	30.00000	2
11	3	4	10.000000	44.14214	2
12	4	5	10.000000	30.00000	2
13	5	7	10.000000	30.00000	2
14	6	3	10.000000	15.85786	2
15	7	1	2.928932	30.00000	2
16	8	6	10.000000	30.00000	2

3) Vous trouverez les résultats des expériences dans le vecteur en annexe. Représentez la surface de réponse. Commentez.

```
library(rsm)
set.seed(1234)
plan<-ccd(2,coding=list(x1~(duree-10)/5,x2~(taille-30)/10)) #données à étudier,
#donne l'instruction de faire le plan
plan #donne le plan des expériences à suivre
Y <-c (12,16,16,10,14,17,14,14,7,14,14,15,17,15,15,14)#vecteur contenant les
#résultats du test
CR.rsm <- rsm(Y~SO(x1,x2),data=plan) #construit le modèle du second ordre, avec
#interactions etc
summary(CR.rsm) #donne le résumé de la surface de réponse, valeurs optimales,
#valeurs propres
contour(CR.rsm,~x1+x2,image=TRUE)
persp(CR.rsm,~x1+x2,col=rainbow(50), contours="colors") #trace la surface de
#réponse
```



D'après la représentation, ainsi que les valeurs propres (une positive et une négative), la surface de réponse présente un point selle, ce qui signifie qu'il n'existe pas de couple de valeurs idéales. Cependant on aurait sûrement pu obtenir un résultat différent en utilisant une plus grande plage de valeurs pour la durée de cuisson, l'appréciation finissant nécessairement par diminuer si la chips est complètement crue ou au contraire cramée, on aurait ainsi pu obtenir des effets quadratiques.

4) Commentez l'ajustement du modèle. Un tel résultat était-il prévisible ?

```
Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.62500    0.64966  22.5118 6.731e-10 ***
x1           1.78921    0.64966   2.7541  0.02034 *
x2           0.62500    0.64966   0.9620  0.35871
x1:x2       -2.25000    0.91876  -2.4490  0.03432 *
x1^2        -1.37500    0.64966  -2.1165  0.06038 .
x2^2         0.12500    0.64966   0.1924  0.85127
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	28.735	14.3676	3.5175	0.06277
TWI(x1, x2)	1	20.250	20.2500	4.9577	0.04589
Residuals	12	49.015	4.0846		
Lack of fit	5	33.140	6.6279	2.9226	0.09731
Pure error	7	15.875	2.2679		

D'après la sortie R, seuls les effets de x1 c'est-à-dire la durée de cuisson, ainsi que l'interaction x1*x2 sont significatifs (p-values de 0.02 et 0.034). Cela signifie que la taille seule n'a pas d'effet, en revanche la durée de cuisson doit être adaptée à la taille de la chips, ce qui semble logique étant donné que des chips de tailles différentes ne cuiront pas à la même vitesse.

5) Quels changements apporter au modèle pour que celui-ci soit mieux ajusté ?

D'après les résultats de la question précédente, les effets quadratiques ne sont pas significatifs, on va donc les supprimer afin d'obtenir un modèle plus ajusté.

Le modèle se réécrit de la manière suivante :

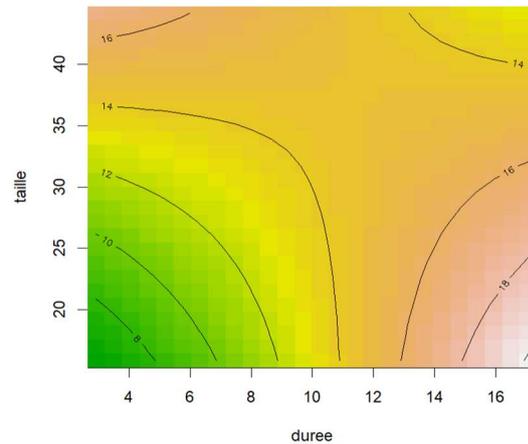
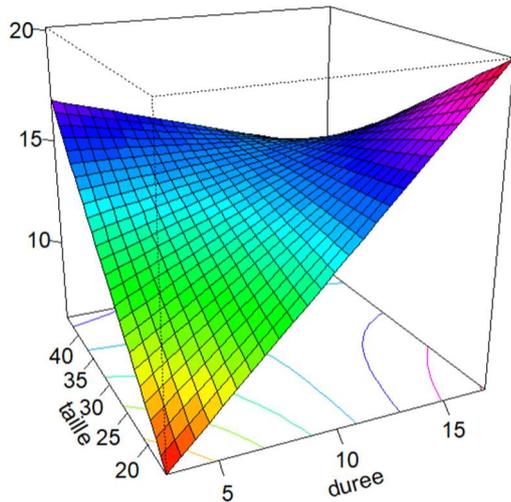
```
library(rsm)
set.seed(1234)
plan<-ccd(2,coding=list(x1~(duree-10)/5,x2~(taille-30)/10))
plan
Y<-c(12,16,16,10,14,17,14,14,7,14,14,15,17,15,15,14)
CR.rsm<-rsm(Y~FO(x1,x2)+TWI(x1,x2),data=plan)
summary(CR.rsm)
contour(CR.rsm,~x1+x2,image=TRUE)
persp(CR.rsm,~x1+x2,col=rainbow(50), contours="colors")
Call:
rsm(formula = Y ~ FO(x1, x2) + TWI(x1, x2), data = plan)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.00000	0.50526	27.7087	3.016e-12 ***
x1	1.78921	0.71454	2.5040	0.02771 *
x2	0.62500	0.71454	0.8747	0.39891
x1:x2	-2.25000	1.01051	-2.2266	0.04589 *

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2)	2	28.735	14.3676	3.5175	0.06277
TWI(x1, x2)	1	20.250	20.2500	4.9577	0.04589
Residuals	12	49.015	4.0846		
Lack of fit	5	33.140	6.6279	2.9226	0.09731
Pure error	7	15.875	2.2679		



La sortie R nous indique que la p-value du lack of fit est passée de 0,13187 pour le premier modèle à 0,09737 avec ce modèle. Le modèle sans effets quadratiques est donc mieux ajusté. De plus cette nouvelle valeur est suffisamment faible pour considérer que le modèle est assez bien ajusté.

Etant donné qu'il n'y a pas d'interactions quadratiques, la surface de réponse ne présente pas de point optimal, mais un point selle (durée 11.38889, taille 37.95206), d'où les deux valeurs propres de signes différents.

6) Vérifiez la qualité de ce modèle à l'aide du calcul matriciel

```
plan<-ccd(2)
X<-model.matrix(~x1+x2+I(x1*x2),data=plan)
t(X)%*%X #calcule x'x
solve(t(X)%*%X) #solve = inverse => calcule (X'X)^-1
```

```
(Intercept) x1 x2 I(x1 * x2)
(Intercept)      16  0  0          0
x1                0  8  0          0
x2                0  0  8          0
I(x1 * x2)       0  0  0          4
> solve(t(X)%*%X)
      (Intercept)      x1      x2 I(x1 * x2)
(Intercept)  0.0625  0.000  0.000      0.00
x1           0.0000  0.125  0.000      0.00
x2           0.0000  0.000  0.125      0.00
I(x1 * x2)   0.0000  0.000  0.000      0.25
```

Ces deux matrices montrent une orthogonalité parfaite entre les facteurs. Ainsi, aucune confusion entre les variables et leur interaction ne peut avoir lieu. Cela est notamment dû au fait que l'on a supprimé les effets quadratiques, ce qui limite le nombre d'essais à tester. Ce plan est donc de très bonne qualité.