Analyse de sensibilité de Modèles: Recherche des facteurs influents

École Chercheur Mexico

Claude Bruchou

INRA — BioSp, Avignon

Giens, le 12 Mai 2009



Plan

- Introduction
- 2 ANOVA
 - Planification expérimentale
 - Plan factoriel Complet équilibré
 - Modèle
 - Analyse de la variance
 - Indices de Sensibilité
 - Bilan de l'ANOVA pour l'AS
- Méthode de Morris
 - Généralités
 - Plan d'expérience standard
 - Indices
- Plan Fractionnaire
 - Bilan
- Bibliographie
- 6 A vous d'expérimenter



Plan

- Introduction
- 2 ANOVA
 - Planification expérimentale
 - Plan factoriel Complet équilibré
 - Modèle
 - Analyse de la variance
 - Indices de Sensibilité
 - Bilan de l'ANOVA pour l'AS
- Méthode de Morris
 - Généralités
 - Plan d'expérience standard
 - Indices
- Plan Fractionnaire
 - Bilan
- Bibliographi
- 6 A vous d'expérimenter



La Fonction Code

Du modèle théorique à l'utilisation

- un modèle complexe
- formalisation mathématique
 - système d'équations
 - $\mathcal{G}: \Omega \subset \mathbb{R}^K \to \mathbb{R}$ $\mathbf{x} = (x^{(1)}, ..., x^{(K)}) \longmapsto \mathcal{G}(\mathbf{x})$
 - analyse théorique de $\mathcal{G} \Rightarrow$ table de variation, points singuliers, optimum, inversibilité,...
 - outils : mathématicien, crayon, calcul formel
 - mais voilà, analyse difficile voire impossible!



La Fonction Code

- ullet Formalisation informatique de ${\cal G}$
 - outils : analyse numérique, langage informatique
 - codage de $\mathcal{G} \Rightarrow$ Fonction Code (FC) $G \sim \mathcal{G}$, $\mathbf{x} = (x^{(1)}, ..., x^{(K)}) \longmapsto y = G(\mathbf{x})$
 - ullet arguments de G: facteurs contrôlés quantitatifs ou qualitatifs,
 - G boite noire
- FC déterministe : les mêmes entrées entraînent les mêmes sorties,
- FC stochastique : les mêmes entrées entraînent des sorties relèvant du calcul des probabilités.



Méthode non intrusive d'analyse

- ullet recherche des facteurs influents de G pour un objectif ou critère $\mathcal{C}(y)$
- échantillonnage aléatoire :
 - tirage au hasard, selon des lois de probabilité, des niveaux x du K-uplet des facteurs,
 - facteurs considérés comme des variables aléatoires X,
 - la loi de tirage des facteurs induit une loi de probabilité sur les sorties,
 - \Rightarrow analyse statistique des sorties (analyse des moments),
 - on note y une réalisation de la FC, considérée comme une variable aléatoire Y.
- échantillonnage déterministe : niveaux fixés des facteurs,
- Illustration



Plan

- Introduction
- 2 ANOVA
 - Planification expérimentale
 - Plan factoriel Complet équilibré
 - Modèle
 - Analyse de la variance
 - Indices de Sensibilité
 - Bilan de l'ANOVA pour l'AS
- Méthode de Morris
 - Généralités
 - Plan d'expérience standard
 - Indices
- Plan Fractionnaire
 - Bilan
- Bibliographi
- A vous d'expérimenter



Intérêt des plans factoriels

- étude simultanée de plusieurs facteurs en entrée,
- gain en coût expérimental et en temps,
- possibilité de détecter des interactions.

Une stratégie d'échantillonnage

- découpage des gammes de définitions des facteurs en intervalles,
- si loi d'échantillonnage non uniforme : utiliser les quantiles,
- pavage en grille de l'espace Ω ,
- R tirages uniformes indépendants dans chaque pavé
- utilisation d'un plan factoriel complet, équilibré
 - contrôle du pavage avec ou sans a priori sur la FC,
 - contrôle de la puissance,
 - valable si la FC est déterministe ou stochastique,
 - accés aux interactions d'ordre élevé.



Exemple:

- FC déterministe : $\Omega = [-1, 1]^2 \to \mathbb{R}$, $Y = G(X^{(1)}, X^{(2)})$
- décomposition de la gamme [-1,1] en I (resp. J) intervalles de même amplitude pour $X^{(1)}$ (resp. $X^{(2)}$),
- \Rightarrow partition de Ω en IJ pavés (unités expérimentales),
- I = J = 3 et R = 2,
- codage de $X^{(1)}$ (resp. $X^{(2)}$) \Rightarrow facteur A (resp. B)

Exemple: table des simulations

simu.	classe de A	classe de B	r	sortie Y_{ijr}
1	1	1	1	9.5
2	1	1	2	10.5
3	1	2	1	20.2
4	1	2	2	18.8
5	1	3	1	30.3
6	1	3	2	27.7
7	2	1	1	8.4
8	2	1	2	9.6
9	2	2	1	29
10	2	2	2	31
11	2	3	1	41
12	2	3	2	39
13	3	1	1	1.5
14	3	1	2	2.5
15	3	2	1	31.5
16	3	2	2	28.5
17	3	3	1	41
18	3	3	2	39

moyenne $\bar{Y}_{...}=23.28$, variance $\sigma_Y^2=180.9$



Modèle d'ANOVA à effets fixes

- $Y_{i,j,r} = \mu + A_i + B_j + AB_{i,j} + \epsilon_{i,j,r}$ i = 1, I, j = 1, J, r = 1, R
- $\mu = \text{constante}$,
- A_i : i^{eme} effet factoriel de $X^{(1)}$; $\sum_i A_i = 0$,
- B_i : j^{eme} effet factoriel de $X^{(2)}$; $\sum_i B_i = 0$,
- AB_{ij} : effet d'interaction; $\forall j, \sum_i AB_{ij} = 0$ et $\forall i, \sum_i AB_{ij} = 0$,
- ullet erreur *aléatoire* d'espérance nulle et de variance σ_{Y}^{2} .

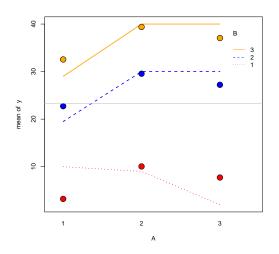


- effets factoriels
 - ullet estimateur de $A_i:\hat{A_i}=ar{Y}_{i..}-ar{Y}_{i..}$
 - ullet estimateur de $B_j:\hat{B_j}=ar{Y}_{.j.}-ar{Y}$

	A_i	B_j
1	-3.78	-16.28
2	3.06	3.22
3	0.72	13.06

- effets d'interaction
 - $\bullet AB_{ij} = (E(Y_{ij}) \mu) A_i B_j$
 - $\bullet \Rightarrow \widehat{AB_{ij}} = \overline{Y}_{ij.} \overline{Y}_{i..} \overline{Y}_{.j.} + \overline{Y}$

$i \setminus j$	1	2	3
1	6.78	-3.22	-3.56
2	-1.06	0.45	0.61
3	-5.72	2.78	2.94



 $\begin{array}{l} \text{points} = \text{mod\`ele avec effets principaux} \\ \text{courbes} = \text{mod\`ele avec effets principaux} + \text{interaction} \\ \text{(passent par les points d'ordonn\'ee $\bar{Y}_{ij.}$)} \end{array}$

Décomposition de la variabilité

$$SS_T = \sum_{i=1}^{J} \sum_{j=1}^{J} \sum_{r=1}^{R} (Y_{ijr} - \bar{Y}_{...})^2 = SS_A + SS_B + SS_{AB} + SS_{\epsilon} = \frac{3075.7}{5}$$

$$SS_A = JR \sum_{i=1}^{I} A_i^2 = 6 \sum_{i=1}^{3} (\bar{Y}_{i..} - \bar{Y})^2 = 144.8, \quad SS_B = IR \sum_{j=1}^{J} B_j^2 = 6 \sum_{j=1}^{3} (\bar{Y}_{.j.} - \bar{Y})^2 = 2674.8$$

$$SS_{AB} = R \sum_{i=1}^{J} \sum_{i=1}^{J} AB_{ij}^2 = 2 \sum_{i=1}^{3} \sum_{i=1}^{3} (\bar{Y}_{ij.} - \bar{Y})^2 = 239.6$$

$$SS_{\epsilon} = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{i=1}^{2} (Y_{ijr} - \bar{Y}_{ij.})^{2} = 253.11 = 16.6$$

$$n = IJR, \ V(Y) \simeq \frac{SS_T}{n} = \frac{\sum_{i=1}^{I} A_i^2}{I} + \frac{\sum_{j=1}^{J} B_j^2}{J} + \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} AB_{ij}^2}{IJ} + \frac{SS_{\epsilon}}{n}$$



Analyse sous R

• tirage de l'échantillon tirage.r = function(PAV, binf, bsup, Nbclass){ # TIRAGE uniforme des coordonnées d'un point dans le pavé PAV de R^K #----binf et bsup : vecteurs des bornes inf et sup des facteurs, Nbclass : nbre de classes par facteur, PAV : vecteur de K numéros de classe (1 à Nbclass) sortie = vecteur de K éléments K = length(PAV)bornes = matrix(NA,nrow=K, ncol=2) bornes[,1] = binf + (PAV-1)*(bsup-binf)/Nbclass bornes[,2] = binf + PAV*(bsup-binf)/Nbclass cc = numeric(K) for(i in 1:K) cc[i] = runif(1,min=bornes[i,1],max=bornes[i,2]) СС

}

Analyse sous R

table des simulations et anova

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
\overline{A}	2	144.78	72.39	39.294	3.573e-05
В	2	2674.78	1337.39	725.965	1.130e-10
AB	4	239.56	59.89	32.509	2.327e-05
Residuals	9	16.58	1.84		

- Quid des tests F en simulation?
- si échantillon aléatoire, le test F constitue un garde fou,
- un grand échantillon peut rendre un test significatif,
- La différence détectée est-elle importante?
- si échantillon déterministe, le test F n'a pas de sens,

- Indice = contribution de la somme des carrés d'un effet à SS_T
- si un facteur a plus de modalités que les autres, sa contribution sera plus forte
- indices utilisés en analyse de sensibilité (Saltelli) :
 - indices principaux

facteur
$$X^{(1)}$$
: $I_1 = \frac{SS_A}{SS_T} = 0.047$
facteur $X^{(2)}$: $I_2 = \frac{SS_B}{SS_T} = 0.87$

indice totaux

facteur
$$X^{(1)}$$
 : $IT_1 = \frac{SS_A + SS_{AB}}{SS_T} = 0.125$
facteur $X^{(2)}$: $IT_2 = \frac{SS_B + SS_{AB}}{SS_T} = 0.95$

ullet part d'alea de l'anova : $\frac{SS_\epsilon}{SS au}=0.005$



- grande souplesse, clarté de l'interprétation,
- donne des pistes pour un meta-modèle,
- nombreux plans dans la littérature,
- modèle linéaire limité pour certaines non-linéarités de la FC,
- Si nombreux facteurs et/ou FC coûteuse en temps, plan complet inabordable
 - exemple : 30 facteurs, découpés en 3 intervalles, sans répétitions (R=1), $\Rightarrow 2^{14}$ simulations
 - $t_{\rm exec} = 10^{-4} {\rm s}$, ferme de 1000 processeurs $\Rightarrow T_{tot} = 7.8 {\rm mois}$.
- nécessité de méthodes exploratoires moins coûteuses ⇒ Morris,
- Interactions d'ordre élevé peu importantes \Rightarrow plans fractionnaires.



Plan

- Introduction
- 2 ANOVA
 - Planification expérimentale
 - Plan factoriel Complet équilibré
 - Modèle
 - Analyse de la variance
 - Indices de Sensibilité
 - Bilan de l'ANOVA pour l'AS
- Méthode de Morris
 - Généralités
 - Plan d'expérience standard
 - Indices
- Plan Fractionnaire
 - Bilan
- Bibliographi
- A vous d'expérimenter



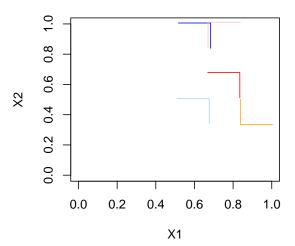
Généralités

- méthode exploratoire
- FC à temps de calcul élevé
- plan OAT (One At a Time) : on fait varier un seul facteur à la fois.

Plan d'expérience standard

- K facteurs définis dans [0, 1],
- Q modalités, $\{0, \frac{1}{Q-1}, \frac{2}{Q-1}, ..., 1\}$ par facteur
- variation δ d'un facteur $\delta \in \frac{1}{Q-1}, \frac{2}{Q-1}, .., \frac{Q-2}{Q-1}$,
- trajectoires aléatoires formées de K+1 points de $\mathbb{R}^K \Rightarrow K$ variations des K facteurs

Exemple : 7 niveaux, $\delta = 1/6$, N = 5 trajectoires.



• effet élémentaire :

$$\Delta_{\pmb{j}}^{(i)} Y = \frac{G(.., \pmb{\mathsf{X}}^{(i)} + \delta \mathbf{e}_{\pmb{j}}, ..) - G(.., \pmb{\mathsf{X}}^{(i)}, ..)}{\delta}$$
 avec $\mathbf{e}_{\pmb{j}} = (0, .., 0, \underbrace{1}_{i}, 0, .., 0)$

• moyenne des variations du facteur j

$$\mu_{\mathbf{j}} = \frac{1}{N} \sum_{i=1}^{N} \Delta_{\mathbf{j}}^{(i)} Y$$

- si $\mu_i \simeq 0$: le facteur j n'a pas d'effet ou variation périodique,
- la moyenne des valeurs absolues des variations est préférable

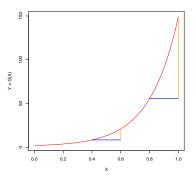
$$\mu^{\star}_{j} = \frac{1}{N} \sum_{i=1}^{N} |\Delta_{j}^{(i)} Y|$$

• écart type des variations du facteur j

$$\sigma_{\mathbf{j}} = \sqrt{\frac{1}{N} \sum_{k} (\Delta_{\mathbf{j}}^{(i)} Y - \mu_{\mathbf{j}})^2}$$

- •] $\sigma_j \simeq 0$, les effets \mathbf{e}_j ne sont pas influencés par les autres facteurs (i.e.pas d'interaction),
- $\sigma_j \gg 0 \Rightarrow$ interaction de X^j ou relation non linéaire entre $X^{(j)}$ et Y,

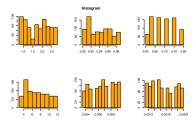
Une liaison non-linéaire entre la sortie de la FC et un facteur peut expliquer la variabilité de μ_i^{\star} ,.



Paramètres des trajectoires

- tronquer la loi si support infini,
- discrétisation à partir des quantiles,
- Pour N trajectoires, N(K+1) calculs de la FC
- nombre de niveaux Q pair $\delta = \frac{Q}{2(Q-1)} \Rightarrow$ distribution uniforme des coordonnées.

Ex. : distribution des coordonnées :



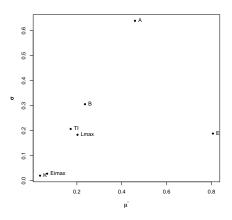




Analyse de Morris sous R :

Graphique de Morris

•
$$(\mu^{\star}_{j} \times \sigma_{j}), j = 1, K$$



Stabilité des résultats

- hypothèse d'indépendance des $|\Delta_i^{(i)}Y|, i=1, N$
- intervalles de confiance (IC) bilatéraux de probabilité $1-\alpha$:
 - $\mu_i^{\star} \pm T_{n-1}(1-\alpha/2)\sigma_{\mu^{\star}}/\sqrt{N}$ (T_{n-1} Student)
 - $\left[\sqrt{\frac{(N-1)}{v^2(N-1)\sigma/2}}\sigma_j, \sqrt{\frac{(N-1)}{v^2(N-1)\sigma/2}}\sigma_j\right]$
- IC par bootstrap : tirage avec remise des trajectoires.

Calcul des IC de probabilité $0.95(\alpha=0.05)$ sous R :

```
> IC.mu = apply(abs(x$ee),2, t.test)
```

- > for(i in 1:7) print(IC.mu[[i]]\$conf.int)
- > sigma = sqrt(apply(x\$ee,2,var))

μ^{\star}	Eb	Eimax	K	Lmax	Α	В	ΤI
binf	0.73	0.065	0.033	0.19	0.38	0.15	0.15
bsup	0.80	0.065 0.08	0.042	0.28	0.64	0.27	0.2
σ							
binf	0.16	0.02 0.03	0.021	0.21	0.61	0.25	0.18
bsup	0.22	0.03	0.028	0.27	0.80	0.33	0.24

Plan

- 1 Introduction
- 2 ANOVA
 - Planification expérimentale
 - Plan factoriel Complet équilibré
 - Modèle
 - Analyse de la variance
 - Indices de Sensibilité
 - Bilan de l'ANOVA pour l'AS
- Méthode de Morris
 - Généralités
 - Plan d'expérience standard
 - Indices
- Plan Fractionnaire
 - Bilan
- Bibliographi
- 6 A vous d'expérimenter



Plan complet

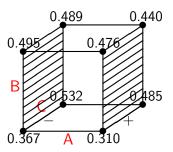
3 facteurs à 2 niveaux (-1,+1), $N=2^3$ simulations

MU	Α	В	С	AB	AC	ВС	ABC	Y
+1	-1	-1	-1	+1	+1	+1	-1	0.367
+1	-1	-1	+1	+1	-1	-1	+1	0.532
+1	-1	+1	-1	-1	+1	-1	+1	0.495
+1	-1	+1	+1	-1	-1	+1	-1	0.489
+1	+1	-1	-1	-1	-1	+1	+1	0.310
+1	+1	-1	+1	-1	+1	-1	-1	0.485
+1	+1	+1	-1	+1	-1	-1	-1	0.476
+1	+1	+1	+1	+1	+1	+1	+1	0.440

colonne AB = interaction = produit terme à terme des colonnes A et B. L'interaction d'ordre 3 est utilisée pour estimer la variance d'erreur.



Effets factoriels



Effet de A: demie-différence des Y aux niveaux + et - de A.

Plan fractionnaire

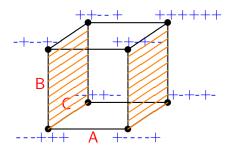
On suppose les facteurs de base A,B et C sans interactions. Construction de 3 nouveaux facteurs.

MU	Α	В	С	D = AB	E = AC	F = BC	ABC
+	_	_	_	_	+	+	+
+	_	_	+	+	+	_	_
+	_	+	_	+	_	+	_
+	_	+	+	_	_	_	+
+	+	_	_	+	_	_	+
+	+	_	+	_	_	+	_
+	+	+	_	_	+	_	_
+	+	+	+	+	+	+	+

- les effets de *D* et *AB* ne peuvent être estimés séparément ; ils sont **confondus** ou aliasés.
- 8 simulations au lieu de $2^6 = 64$
- plan 2^{6-3} = fraction 1/8 du plan complet 2^6 .



Effets factoriels



$$(-+--+-) = (A=-1, B=+1, C=-1, D=-1, E=+1, F=-1)$$

Effet : demie-différence des Y aux niveaux $+$ et $-$ d'un facteur.

Modèle d'ANOVA

• notation pour un plan à 3 facteurs à 2 niveaux : L'espérance du vecteur des sorties Y est la somme des effets $e(.) \in \mathbb{R}$

$$\mathbb{E}(Y) = \sum_{a,b,c} A^a B^b C^c e(A^a B^b C^c)$$

$$a, b, c \in \{0, 1\}, \text{ (ex. } A^1B^1C^0 = AB \text{ vecteur colonne)}$$

- Exemple :
 - plan complet $\mathbb{E}(Y_{+,+,-}) = \mu + e(A) + e(B) e(C) + e(AB) e(AC) e(BC) e(ABC)$
 - plan fractionnaire
 - $\mathbb{E}(Y_{+,+,-,+,-,-}) = \mu + e(A) + e(B) e(C) + e(D) e(E) e(F) e(ABC)$



Modèle d'ANOVA

- effet $e(A^aB^bC^c) = 1/N < A^aB^bC^c$, Y > (p.s. de colonnes) avec $a, b, c \in \{0, 1\}$ Ex. : N = 8, $e(A^1B^1C^0) = e(AB) = 1/8 < AB$, Y >.
- Effet principal de A: $e(A) = \frac{1}{2}(\bar{Y}_{+..} \bar{Y}_{-..}) = \frac{1}{8}[(Y_{+--} + Y_{+-+} + Y_{++-} + Y_{+++}) (Y_{---} + Y_{--+} + Y_{-+-} + Y_{-++})] = \frac{1}{9} < A, Y >$
- Effet d'interaction (plan complet) $e(AB) = \frac{1}{2} [e(A|B = +1) e(A|B = -1)] \\ e(AB) = \frac{1}{2} {\frac{1}{4} [Y_{++-} + Y_{+++} Y_{-+-} Y_{-++}] \frac{1}{4} [Y_{+--} + Y_{+-+} Y_{---} Y_{--+}]} = \frac{1}{8} < AB, Y > 0$



Résolution d'un plan fractionnaire

- ullet un plan dît de résolution ${\mathcal R}$ permet d'analyser :
 - toutes les interactions d'ordre (R-1)/2 si R est impair,
 - toutes les interactions d'ordre (R-2)/2 si R est pair.
- Ex. : le plan 2^{6-3} , de résolution $\mathcal{R}=3$, est noté 2^{6-3}_{III} .Il ne peut estimer que les effets principaux



Analyse sous R

Analyse sous R

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
А	1	0.0037	0.0037	18.49	0.15
В	1	0.0053	0.0053	26.52	0.12
C	1	0.0111	0.0111	55.5	0.085
D	1	0.00016	0.00016	8.0	0.53
E	1	0.00005	0.00005	0.25	0.70
F	1	0.018	0.018	91.203	0.066
Residuals	1	0.0002	0.0002		



Majoration du nombre K de facteurs :

N de simulations,

• si $\mathcal{R} = III : N \geq 1 + K$,

• si $\mathcal{R} = IV : N \geq 2K$,

• si $\mathcal{R} = V : N \ge 1 + K + K(K - 1)/2$.

Nombre $maximal \ \kappa$ de facteurs pour un plan :

• de $\mathcal{R} = V$,

• de taille $N = 2^s$,

• donnant des estimateurs de variance minimale,

<i>S</i>	4	5	6	7	8	9
N	16	32	64	128	256	512
κ	5	6	8	11	17	≥ 23



Bilan

PROBLEME	METHODE		
Très nombreux facteurs, sélec-	Plans de screening (résolution		
tionner les plus influents	III ou <u>IV</u>)		
Etudier l'influence simultanée	Plans factoriels à 2 niveaux		
de nombreux facteurs avec	complets ou fractionnaires, ré-		
peu d'observations, détecter les	solution $\geq V$		
principales interactions			
Etude plus détaillée de facteurs	Plans factoriels 3^n ou 4^n , cf sur-		
quantitatifs	faces de réponses		

Plan

- Introduction
- 2 ANOVA
 - Planification expérimentale
 - Plan factoriel Complet équilibré
 - Modèle
 - Analyse de la variance
 - Indices de Sensibilité
 - Bilan de l'ANOVA pour l'AS
- Méthode de Morris
 - Généralités
 - Plan d'expérience standard
 - Indices
- 4 Plan Fractionnaire
 - Bilan
- Bibliographie
- A vous d'expérimenter



Bibliographie

- Droesbeke JJ, Fine J., Saporta G. ed. (1997), Plans d'expérience, applications à l'entreprise, Editions Tecnip
- Saltelli A. et al. Sensitivity Analysis, (2000), Wiley
- Saltelli A. et al. (2008), **Global Sensitivity analysis The primer**, Wiley
- F. Fabre, C. Bruchou, A. Palloix, B. Moury, Key determinants of resistance durability to plant viruses: insights from a model within-and between-host dynamics, Virus Research 141 (2009) 140-149.

Plan

- Introduction
- 2 ANOVA
 - Planification expérimentale
 - Plan factoriel Complet équilibré
 - Modèle
 - Analyse de la variance
 - Indices de Sensibilité
 - Bilan de l'ANOVA pour l'AS
- Méthode de Morris
 - Généralités
 - Plan d'expérience standard
 - Indices
- Plan Fractionnaire
 - Bilan
- Bibliographie
- 6 A vous d'expérimenter



TP

• Mettez en oeuvre les 3 méthodes sur le modèle wwdm.

