

Plans d'expériences

Séance 2: Plans factoriels

Définition

Plans factoriels complets

Un plan factoriel complet est constitué de l'ensemble des objets que l'on peut constituer à l'aide des combinaisons des facteurs que l'on étudie.

⇒ On représente les plans factoriels complets par un treillis.

Notation

- On note h_i le nombre de facteurs ayant p_i modalités.
 - Un plan factoriel complet est représenté par l'expression $p_1^{h_1} p_2^{h_2} \dots$
- ⇒ Exemple : Le plan $2^3 3$ fait intervenir trois facteurs ayant deux modalités et un facteur ayant trois modalités.

Remarque : Le nombre $p_1^{h_1} p_2^{h_2} \dots$ correspond au nombre d'objets du plan factoriel complet.

Représentation des objets

- Plan 2^3 :

000	001	010	011
100	101	110	111

- Plan 2^3 :

(1)	<i>c</i>	<i>b</i>	<i>bc</i>
	<i>a</i>	<i>ac</i>	<i>ab abc</i>

- ⇒ Le signe (1) désigne l'absence de tout traitement.
- ⇒ Quand une lettre n'apparaît pas au niveau d'un objet, cela signifie que le traitement du facteur correspondant n'a pas été appliqué.

- Plan $2^2 3$:

111	112	113	121	122	123
211	212	213	221	222	223

Représentation des objets : Plans de type 2^n

Plan 2^3 (autre notation utilisée pour les plans de type 2^n dans la suite) :

$$\begin{array}{cccc} -1 & -1 & -1 & -111 \\ -1 & -1 & 1 & -11 \\ -1 & 1 & -1 & -11 \\ 1 & -1 & -1 & 111 \\ 1 & -1 & 1 & 11 \\ 1 & 1 & -1 & 11 \\ 1 & 1 & 1 & 111 \end{array}$$

- Le plan 2^3 est relativement simple.
 - Les notions et résultats énoncés se généralisent au cas général 2^n de manière relativement simple.
 - Le cas 2^n est très utilisé en pratique.
- ⇒ On considère dans la suite trois facteurs A , B et C dont les niveaux sont notés 1 et -1.

Définitions et notations

On note y la réponse (les observations) a un traitement.

On définit l'espérance de la réponse y (ou réponse théorique) par

$$E(y) = \tau(A, B, C), \quad A = \pm 1, B = \pm 1 \text{ et } C = \pm 1.$$

⇒ Par exemple $\tau(1, 1, -1)$ l'espérance de la réponse de l'unité qui a été traitée par les facteurs A et B seulement.

Définitions et notations

- La moyenne générale notée $e(1)$:

$$e(1) = \bar{\tau}(.,.,.),$$

où $\bar{\tau}(.,.,.)$ est la moyenne des 8 réponses possibles.

- On note dans la suite :

$$\bar{\tau}(A, B, .) = \frac{1}{2}(\tau(A, B, 1) + \tau(A, B, -1))$$

$$\begin{aligned}\bar{\tau}(A, ., .) &= \frac{1}{4}(\tau(A, 1, 1) + \tau(A, 1, -1) \\ &+ \tau(A, -1, 1) + \tau(A, -1, -1)).\end{aligned}$$

Définitions et notations

- L'effet du facteur A est défini comme suit

$$e(A) = \frac{1}{2}(\bar{\tau}(1, \cdot, \cdot) - \bar{\tau}(-1, \cdot, \cdot)),$$

c'est la demi-différence entre la moyenne des réponses pour $A = 1$ et $A = -1$. On remarque que :

$$e(A) = \frac{1}{8} \sum_{A,B,C} A\tau(A, \cdot, \cdot).$$

Définitions et notations

$$\Rightarrow e(A | B = 1) = \frac{1}{2}(\bar{\tau}(1, 1, \cdot) - \bar{\tau}(-1, 1, \cdot)).$$

$$\Rightarrow e(A | B = -1) = \frac{1}{2}(\bar{\tau}(1, -1, \cdot) - \bar{\tau}(-1, -1, \cdot)).$$

On définit l'interaction entre les facteurs A et B :

$$e(AB) = \frac{1}{2}(e(A | B = 1) - e(A | B = -1)).$$

- On peut également montrer que :

$$e(AB) = e(BA) = \frac{1}{8} \sum_{A,B,C} AB\tau(A, B, C).$$

Définitions et notations

- On peut également montrer que :

$$\begin{aligned}e(ABC) &= \frac{1}{2}(e(AB \mid C = 1) - e(AB \mid C = -1)) \\ &= \frac{1}{8} \sum_{A,B,C} ABC\tau(A, B, C).\end{aligned}$$

Expression des effets en fonction des réponses

⇒ En exprimant de manière similaire $e(B)$, $e(C)$, $e(AC)$ et $e(BC)$ on écrit :

$$\begin{pmatrix} e(1) \\ e(A) \\ e(B) \\ e(C) \\ e(AB) \\ e(AC) \\ e(BC) \\ e(ABC) \end{pmatrix} = \frac{1}{8} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \tau(1,1,1) \\ \tau(1,1,-1) \\ \tau(1,-1,1) \\ \tau(1,-1,-1) \\ \tau(-1,1,1) \\ \tau(-1,1,-1) \\ \tau(-1,-1,1) \\ \tau(-1,-1,-1) \end{pmatrix}$$

Note : Les erreurs n'apparaissent pas puisque leur espérance est supposée nulle.

Notons que les valeurs que prend $A, B, \dots, ABC \in \{-1, 1\}$ se lisent ligne par ligne.

$$\begin{array}{l}
 . \\
 A \\
 B \\
 C \\
 AB \\
 AC \\
 BC \\
 ABC
 \end{array}
 \begin{pmatrix}
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1
 \end{pmatrix}$$

On retrouve, entre autres, la relation en utilisant le modèle précédent $e(AB) = \frac{1}{8} \sum_{A,B,C} AB\tau(A, B, C) \dots$

On écrit la relation précédente :

$$e = \frac{1}{8} Z\tau.$$

En remarquant que $Z'Z = ZZ' = 8I$ on a :

$$\tau = Z'e,$$

ou bien :

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

avec $(a, b, c) \in \{0, 1\}^3$.

Estimation et bornes de confiance-Cas général

On écrit le modèle :

$$Y = Xe + \epsilon,$$

- Avec Y le vecteur des observations et N le nombre d'observations. On suppose que les erreurs sont iid gaussiennes et telles que $Var(\epsilon) = \sigma^2 I_N$. On obtient un estimateur des MCO $\hat{e} = (X'X)^{-1}X'Y$.
- Comme les erreurs sont iid gaussiennes de variance σ^2 , on a :

$$Var(\hat{e}) = \sigma^2(X'X)^{-1},$$

et donc

$$\frac{\sqrt{N}}{\sigma}(\hat{e} - e) \sim \mathcal{N}(0, I_n)$$

Estimation et bornes de confiance

- ⇒ On obtient des intervalles de confiance pour les estimateurs
- Dans notre cas (plan factoriel complet 2^3), nous n'avons pas de degré de liberté pour estimer la variance des erreurs (**pas de ddl pour cette estimation !**). Dans ce cas on ne peut pas obtenir d'intervalles de confiance.

On peut obtenir un modèle empirique avec pour paramètres les effets et pour variable expliquée les réponses. On obtient le tableau d'ANOVA suivant :

Sources de variation	ddl	Sommes des carrés des écarts	Carrés des moyens	F	<i>p</i> -valeurs
Facteur A	1	...	SCE/ddl	CM/CM_r	...
Facteur B	1
Facteur C	1
Interaction AB	1
Interaction AC	1
Interaction BC	1
Interaction ABC	1
Variation résiduelle	...	SCE_r	CM_r	×	×

Quand une *p*-valeur d'un test *F* est importante on conclut à la significativité du facteur ou de l'interaction.

Cas où on ne peut pas estimer la variation résiduelle.

Sources de variation	ddl	Sommes des carrés des écarts	Carrés moyens	F	p-valeurs
Facteur A	1	...	SCE/ddl	CM/CM_{ABC} .	
Facteur B	1
Facteur C	1
Interaction AB	1
Interaction AC	1
Interaction ABC	1

- ⇒ On prend comme variation résiduelle la somme des carrés une (ou des) interactions non significatives.
- ⇒ On peut également faire des répétitions pour chaque objet (prendre plusieurs unités pour chaque objet).

- Dans certains cas on considère des groupes (=blocs) d'unités expérimentales tels que les unités dans chaque groupe expérimental soient le plus homogène possible.
- L'intérêt de constituer des blocs est d'assurer des conditions d'expérience équitables entre les différents facteurs à l'intérieur de chaque bloc, tout en s'assurant des conditions d'expérience différentes à travers les différents blocs.

Remarque : les blocs sont considérés comme un facteur.

On minimise la variabilité intra-blocs pour que les conditions d'expériences soient similaires pour chaque unité du bloc :

⇒ Plantes clones, terrains de même nature,..etc..

On maximise la variabilité inter-blocs pour s'assurer une expérimentation tenant compte de la variabilité des unités.

⇒ Terrain argileux/rocailleux ; plante possédant un gène ou pas...etc..

⇒ On évite la confusion des effets : Les unités similaires sont regroupées dans un même bloc. Une certaine variabilité des unités est prise en compte.

But de la randomisation

Certaines conditions d'expérience exigent une randomisation

- Méthode du double aveugle : Le malade et le médecin ignorent la nature du produit administré. (Eviter l'effet placebo).

Contrôler la confusion des effets dont la source peut être due au fait que certains facteurs ne sont pas pris en compte dans l'expérience :

- Facteurs sociaux, temporels, génétiques...etc (les caractéristiques de certaines unités peuvent biaiser l'expérience dans le cas de plans complets avec ou sans blocs).

Plans complets en randomisation totale

Parmi l'ensemble des unités disponibles, on tire au hasard et sans remise un nombre égal d'unités pour chaque traitement.

Exemple : On dispose de $n = 2k$ malades. On tire k malades qui recevront un traitement et les autres k malades recevront un placebo.

Plan en blocs complets

On effectue une randomisation au sein de chaque bloc. Cette randomisation se fait de façon indépendante d'un bloc à l'autre.

Exemple : On dispose de $n = 3k$ malades, k individus maigres, k moyens et k en surpoids. On désire tester k traitements sur ces individus. Dans le tableau ci-dessous on prend $k = 3$.

individu	1	2	3
<i>maigre</i>	T_3	T_2	T_1
<i>moyen</i>	T_2	T_1	T_3
<i>surpoids</i>	T_1	T_3	T_2

Ecriture du modèle d'analyse de l'exemple précédent.

$$\begin{pmatrix} R_{1,1} \\ R_{2,1} \\ R_{3,1} \\ R_{1,2} \\ R_{2,2} \\ R_{3,2} \\ R_{1,3} \\ R_{2,3} \\ R_{3,3} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \textit{maigre} \\ \textit{moyen} \\ T1 \\ T2 \end{pmatrix} + \begin{pmatrix} \epsilon_{1,1} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \epsilon_{3,3} \end{pmatrix}$$

On a 2 ddl pour la variable poids, 2 ddl pour la variable traitement et 4 ddl pour la estimer la variance résiduelle.

Carrés latins

Définition

Un carré latin est un plan comprenant n^2 unités pour trois facteurs ayant le même nombre de niveaux, et tel que le nombre de répétitions d'une paire de niveaux pour deux facteurs est toujours de 1.

Carrés latins : Exemple général pour $n = 4$.

modalités	1	2	3	4	•	1	2	3	4	•	1	2	3	4
1	1	2	3	4	1	2	3	4	1	1	4	3	1	2
2	2	3	4	1	2	3	4	1	2	2	1	4	2	3
3	3	4	1	2	3	4	1	2	3	3	2	1	3	4
4	4	1	2	3	4	1	2	3	4	4	3	2	4	1

Trois facteurs : **bleu**, **rouge**, **vert** ayant chacun 4 modalités (ou niveaux). Les facteurs bleu et rouge sont de type bloc. Le facteur vert est le facteur d'intérêt.

A gauche : Construction du carré.

Au centre : mélange des lignes (Randomisation).

A droite : mélange des colonnes (Randomisation).

Remarque : Les étapes de mélange colonnes et ligne peuvent être interverties.

Exemple illustratif avec $n = 3$.

On applique trois traitements insecticides en prenant en compte l'immeuble (premier facteur bloc) l'étage (deuxième facteur bloc) et le facteur d'intérêt : Traitement.

immeuble	A	B	C
étage 1	T_2	T_1	T_3
étage 2	T_1	T_3	T_2
étage 3	T_3	T_2	T_1

Ecriture du modèle d'analyse de l'exemple précédent.

$$\begin{pmatrix} R_{1,1} \\ R_{2,1} \\ R_{3,1} \\ R_{1,2} \\ R_{2,2} \\ R_{3,2} \\ R_{1,3} \\ R_{2,3} \\ R_{3,3} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ e1 \\ e2 \\ A \\ B \\ T1 \\ T2 \end{pmatrix} + \begin{pmatrix} \epsilon_{1,1} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \epsilon_{3,3} \end{pmatrix}$$

On a 2 ddl pour la variable étage, 2 ddl pour la variable traitement, 2 ddl pour la variable immeuble et 2 ddl pour estimer la variance résiduelle.

Cross-over.

Dans certains cas on est amené à considérer des dispositifs analogues au carré latin mais qui ne sont pas carrés. On applique un traitement sur des périodes successives à deux groupes de patients (A et B).

temps	1	2	3	4
A	T2	T1	T1	T2
B	T1	T2	T2	T1

Remarque : Un traitement peut apparaître plusieurs fois sur une même ligne.

Analyse de la variance.

- Dans la suite on suppose que les termes d'erreurs sont de même variance, de distribution normale et indépendants.
- Les paramètres doivent être interprétés comme des écarts d'effets entre les facteurs apparaissant dans la régression et le facteur qui n'y apparaît pas.

Par exemple la nullité du paramètre "T1" dans le modèle s'interprète comme l'absence d'une différence entre l'effet de T1 et T3.

- ⇒ En présence de blocs, on fait apparaître le facteur "blocs" dans les sources de variation.

- On construit un tableau d'analyse de la variance où figurent les p -valeurs des tests de Fisher.

Sources de variation	ddl	Sommes des carrés des écarts	Carrés moyens	F	p -valeurs
Facteur A	SCE/ddl	CM/CM_r	...
Blocs
Variation résiduelle	...	SCE_r	CM_r	×	×

Quand une p -valeur d'un test F est importante on conclut à la significativité du facteur ou de l'interaction.

Remarques

- Quand il ne reste pas de ddl pour estimer la variance, les tests F ne peuvent pas être menés.
- On peut tester la pertinence de former des blocs (ou du choix de ces blocs) avec les tests F .
- Dans le cas "Cross-over" ou "Carré" latin, on remplace la source de variation "blocs" par les sources "lignes" et "colonnes".