

Cours

Plans d'expérience

Walter TINSSON

Master MSiD 2

Année 2022 / 2023

01 Introduction à la planification expérimentale

01 Généralités

A l'époque actuelle bon nombre de procédés de fabrication ou d'expériences en laboratoire deviennent de plus en plus complexes car ils dépendent d'un grand nombre de variables difficiles à régler intuitivement. Ceci concerne, par exemple :

- le problème de la mise au point de moteurs atmosphériques dépendant d'un nombre croissant de réglages électroniques,
- le pilotage optimal de machines-outil,
- la détermination des proportions d'un mélange chimique,
- la recherche des conditions environnementales optimales pour la production agricole, etc...

Seule la réalisation d'expériences va permettre d'appréhender et de modéliser de tels phénomènes complexes. Si ces expériences sont effectuées sans une méthodologie rigoureuse il est fort probable qu'elles vont soit conduire à des impasses (modèle impossible à ajuster, résultats incohérents, etc...) soit à des résultats de qualité décevante. C'est pourquoi la **méthode des plans d'expérience** est préconisée afin d'optimiser ce type de démarche. L'objectif principal de cette méthode peut être résumé par :

« obtenir un maximum d'information en un minimum d'expériences »

Une autre vision du problème est la recherche de variations simultanées pour toutes les variables contrôlées afin, une nouvelle fois, d'extraire un maximum d'information en un minimum d'essais. Une telle problématique est primordiale dans le milieu industriel où minimiser le nombre d'expériences à réaliser est synonyme de gain de temps et de productivité. Réaliser des productions de la meilleure qualité possible au coût le plus bas est de plus une quête universelle pour tous les fabricants.

02 Notions de base

Listons ici quelques notions élémentaires utilisées dans la méthode des plans d'expérience (certaines d'entre elles étant des concepts statistiques très généraux).

1) Réponse 18.5

On appelle **réponse** la grandeur qui est observée pour chaque expérience réalisée. On supposera toujours ici que cette grandeur est numérique et qu'une seule réponse à la fois est observée.

2) Facteur températura, presión, color

On appelle **facteur** toute variable, obligatoirement contrôlable, susceptible d'influer sur la réponse observée. La différence entre la notion classique de variable (explicative) et celle de facteur tient donc dans le fait que tout facteur doit pouvoir être modifié sans difficulté. Remarquons que les facteurs peuvent être **quantitatifs** lorsqu'ils sont naturellement exprimés à l'aide de valeurs numériques (pression, température, durée, etc...) ou bien **qualitatifs** dans le cas contraire (couleur, type de matériau, sexe, etc...). Il est classique de transformer des facteurs qualitatifs en facteurs quantitatifs à l'aide d'un codage approprié (par exemple en affectant la valeur 0 pour "Homme" et la valeur 1 pour "Femme" dans le cas du sexe). Lorsque des expériences sont réalisées les différentes valeurs prises par un facteur quantitatif ou les différentes modalités prises par un facteur qualitatif sont appelés **niveaux** de ce facteur.

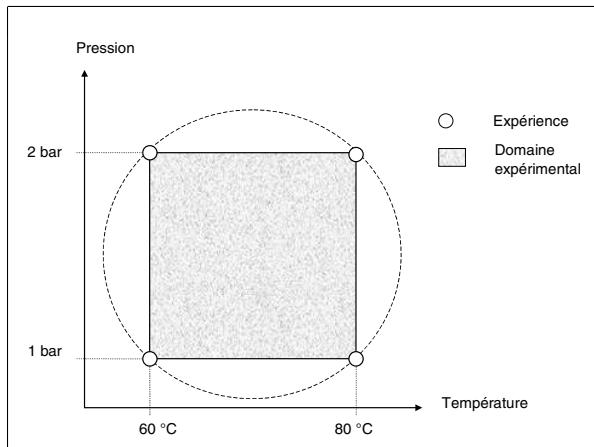
Observation
experimental
design
diseños de experimentos?

3) Domaine expérimental

Considérons un phénomène aléatoire dépendant de m facteurs (soit quantitatifs soit qualitatifs codés numériquement). Toute **expérience** est alors géométriquement identifiable au vecteur de \mathbb{R}^m contenant tous les niveaux des différents facteurs. En pratique le i -ème facteur (pour $1 \leq i \leq m$) est le plus souvent à valeurs dans un intervalle de la forme $[a_i, b_i]$ où a_i et b_i sont respectivement ses niveaux bas et haut (il s'agit de la « **plage de variations** »). On appelle alors **domaine expérimental** tout sous-ensemble de \mathbb{R}^m (noté \mathcal{E} par la suite) dans lequel il est possible de réaliser les expériences. Une méthode élémentaire afin d'obtenir un tel domaine consiste simplement à **croiser les diverses plages de variations**, on obtient donc par produit cartésien :

$$\mathcal{E} = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_m, b_m].$$

La figure ci-dessous représente le domaine expérimental cubique obtenu en croisant les deux facteurs température à pression. On considère aussi parfois le domaine expérimental sous forme sphérique, il s'agit ici du plus petit disque (limité par les pointillés) contenant le domaine cubique précédent.



03 Notion de plan d'expérience

Considérons une entreprise produisant une colle industrielle qui a tendance de se solidifier durant le processus de fabrication. Afin de s'opposer à cette tendance trois additifs sont introduits durant le procédé industriel et les débits injectés sont contrôlables à l'aide de trois vannes prenant les niveaux suivants :

Faible / Moyen / Fort.

La réponse mesurée est le taux de fluidité de la colle. Nous sommes donc ici en présence d'un phénomène dépendant, à priori, de trois facteurs qualitatifs (chacun d'eux ayant trois modalités). Supposons que l'on cherche à ajuster un modèle statistique classique comportant 7 paramètres inconnus (il s'agit du modèle dit additif, voir la suite de ce document). Supposons enfin que chaque expérience a un coût total de 500 euros (réglage des vannes, lancement d'une production et analyse de la colle produite). Vu qu'il y a 7 inconnues dans ce problème le nombre minimal de 7 expériences sera donc requis soit un coût minimal de 3500 euros ici afin d'ajuster le modèle proposé.

Comment réaliser alors efficacement les expériences afin d'ajuster les paramètres de ce modèle ? Nous présentons ci-dessous quelques stratégies classiques (dans les différents tableaux la notation DVi désigne le débit fixé pour la vanne i).

1) Réalisation de toutes les expériences

Il s'agit de la méthode en théorie la « plus sûre » afin d'étudier le phénomène aléatoire. En effet elle consiste donc en la réalisation des $n = 3^3 = 27$ expériences présentées ci-dessous et le réglage optimal sera simplement déterminé par l'expérience maximisant le taux de fluidité. Ce type de démarche est cependant souvent **impossible** à réaliser en pratique à cause du coût expérimental trop important. Dans cet exemple la réalisation de toutes les expériences correspond à un coût global de 13500 euros alors que le budget souhaité serait au voisinage de 3500 euros.

Vecindario

Exp	DV1	DV2	DV3
01	Faible	Faible	Faible
02	Moyen	Faible	Faible
03	Fort	Faible	Faible
04	Faible	Moyen	Faible
05	Moyen	Moyen	Faible
06	Fort	Moyen	Faible
07	Faible	Fort	Faible
08	Moyen	Fort	Faible
09	Fort	Fort	Faible
10	Faible	Faible	Moyen
11	Moyen	Faible	Moyen
12	Fort	Faible	Moyen
13	Faible	Moyen	Moyen
14	Moyen	Moyen	Moyen

Exp	DV1	DV2	DV3
15	Fort	Moyen	Moyen
16	Faible	Fort	Moyen
17	Moyen	Fort	Moyen
18	Fort	Fort	Moyen
19	Faible	Faible	Fort
20	Moyen	Faible	Fort
21	Fort	Faible	Fort
22	Faible	Moyen	Fort
23	Moyen	Moyen	Fort
24	Fort	Moyen	Fort
25	Faible	Fort	Fort
26	Moyen	Fort	Fort
27	Fort	Fort	Fort

2) Utilisation de la technique « un facteur à la fois »

L'expérimentateur devant faire face à une situation où la réalisation de toutes les expériences est beaucoup trop lourde se rabat souvent sur ce type de technique. Comme son nom l'indique elle consiste à faire varier chacun des facteurs, l'un après l'autre, en lui affectant toutes les modalités possibles. Puisqu'ici chaque facteur est à 3 modalités ceci conduit donc à la réalisation d'un total de 9 expériences données (par exemple) ci-dessous.

Exp	DV1	DV2	DV3
13	Faible	Moyen	Moyen
14	Moyen	Moyen	Moyen
15	Fort	Moyen	Moyen
11	Moyen	Faible	Moyen
14	Moyen	Moyen	Moyen
17	Moyen	Fort	Moyen
05	Moyen	Moyen	Faible
14	Moyen	Moyen	Moyen
23	Moyen	Moyen	Fort

Valeurs à changer
 ③ Moyen Moyen
 Moyen ③ Moyen
 Moyen Moyen ③

$$\begin{array}{r}
 \downarrow \\
 3 \\
 +3 \\
 \hline
 3 \\
 \hline
 9
 \end{array}$$

Une telle stratégie s'avère en général très pauvre du point de vue statistique (car souvent le passage d'une expérience à l'autre n'apporte que peu d'information nouvelle vu que la plupart des niveaux des facteurs restent fixés). De manière générale les inconvénients associés à cette technique sont les suivants :

- i) faire varier les facteurs un par un masque les éventuels effets d'interactions entre plusieurs facteurs,
- ii) le choix du niveau pour les facteurs ne variant pas ("moyen" ici) n'est pas évident et peut avoir un effet sur la qualité des résultats obtenus,
- iii) le plan d'expérience obtenu présente le problème d'être déséquilibré dans le sens où ici le niveau « moyen » est sur-représenté au détriment des deux autres niveaux.

3) Utilisation d'un sous-ensemble quelconque d'expériences

L'utilisateur n'ayant pas réalisé toutes les expériences pour des raisons de coût et n'étant pas satisfait des résultats donnés par la méthode "un facteur à la fois" se trouve souvent désemparé et s'oriente la

plupart du temps vers le choix d'un sous-ensemble d'expériences. Ce choix est bien souvent réalisé de manière empirique : des expériences peuvent être rajoutées à celles de la technique "un facteur à la fois" dans le but d'améliorer les résultats, un sous-ensemble d'expérience peut être déterminé de manière aléatoire, etc... Généralement, tout choix d'un sous ensemble d'expériences qui n'est pas guidé par une méthodologie rigoureuse peut entraîner les problèmes suivants :

- i) un tel choix peut conduire à sélectionner des expériences qui ne permettront pas d'estimer tous les paramètres inconnus du modèle étudié,
- ii) même si tous les paramètres inconnus du modèle étudié peuvent être estimés la qualité des résultats obtenus ne sera généralement pas optimale.

Considérons alors les $n = 9$ expériences présentées ci-dessous. Ce choix peut paraître, à priori, plus judicieux que celui de la table précédente dans le sens où la configuration présentée ici est équilibrée puisque chacun des niveaux des facteurs apparaît le même nombre de fois. Il n'en est rien en fait car les expériences sélectionnées ici ne permettent pas d'estimer tous les paramètres du modèle. Ceci est dû au fait que les facteurs 1 et 2 sont « liés » artificiellement car ils changent de niveau en même temps ce qui rend impossible l'estimation de leurs effets respectifs.

hace

Exp	DV1	DV2	DV3
01	Faible	Faible	Faible
10	Faible	Faible	Moyen
19	Faible	Faible	Fort
08	Moyen	Fort	Faible
17	Moyen	Fort	Moyen
26	Moyen	Fort	Fort
06	Fort	Moyen	Faible
15	Fort	Moyen	Moyen
24	Fort	Moyen	Fort

4) Utilisation d'un plan d'expérience

Voici maintenant une stratégie « optimale » obtenue à l'aide d'une fraction régulière du plan factoriel complet (voir la suite de ce document pour ces notions).

Exp	DV1	DV2	DV3
01	Faible	Faible	Faible
06	Fort	Moyen	Faible
08	Moyen	Fort	Faible
12	Fort	Faible	Moyen
14	Moyen	Moyen	Moyen
16	Faible	Fort	Moyen
20	Moyen	Faible	Fort
22	Faible	Moyen	Fort
27	Fort	Fort	Fort

Nous avons ici un petit nombre d'expériences ($n = 9$) pour un coût global de 4500 euros. On montre que ces expériences permettent d'estimer de manière « optimale » les paramètres inconnus des modèles classiques (avec des effets linéaires voire aussi des effets d'interactions). La qualité de ce plan d'expérience réside à la fois dans le fait qu'il est équilibré pour les niveaux des facteurs (*i.e.* chaque niveau est utilisé 3 fois pour chaque facteur) mais aussi pour les couples de niveaux (*i.e.* chaque couple de niveaux est testé une fois pour chaque couple de facteurs). Une telle propriété est qualifiée d'**orthogonalité**.

04 Codage des facteurs

Considérons ici le cas de facteurs quantitatifs. Afin d'uniformiser les variations de chacun des facteurs l'idée est d'effectuer un changement de variables les ramenant à l'intervalle de référence $[-1, +1]$. Il vient alors :

*trayendo de
vuelta*

Définition 01

Siendo Etant donné un facteur à valeurs x^* dans l'intervalle $[a, b]$, le **facteur codé** qui lui est associé est déterminé par la transformation affine suivante :

$$x = \frac{2x^* - (a+b)}{(b-a)}$$

$$x = \frac{2x^* - (a+b)}{b-a} = \frac{x^* - \frac{(a+b)}{2}}{\frac{b-a}{2}} = \frac{x^* - c}{\frac{b-a}{2}}$$

Cette transformation est très classique et sera effectuée dans la suite de manière systématique car elle présente les avantages suivants :

- i) il est possible d'uniformiser les constructions de plans d'expérience en ramenant le domaine expérimental (cubique) au domaine de référence $[-1, +1]^m$,
- ii) la plupart des analyses mathématiques vont être simplifiées par l'utilisation des deux niveaux -1 et $+1$ ou des trois niveaux classiques $-1, 0$ et $+1$ qui s'expriment donc très simplement sous forme codée,
- iii) les effets des facteurs sont facilement comparables puisque sous forme codée ils varient tous dans le même intervalle $[-1, +1]$,
- iv) les variables codées s'expriment sans unité.

est

Remarque. Dans certains cas de figure il sera nécessaire de se ramener non pas à l'intervalle $[-1, +1]$ mais à un intervalle centré $[-\Delta, +\Delta]$. La formule devient alors :

$$x = \Delta \left[\frac{2x^* - (a+b)}{(b-a)} \right]$$

Examp

$T^*(^{\circ}\text{C})$	20	30	40
T	-1	0	1

$T = \frac{T^* - 30}{10}$ centre

Niveau bas Niveau haut

05 Exercices d'applications

Exercice 1

[Expériences optimale dans le cas de l'ajustement de la droite des moindres carrés]

On cherche ici à modéliser le rendement d'une réaction chimique en fonction de la température de la solution. Le spécialiste étudiant ce phénomène estime que la plage de variation utile pour le facteur température est donnée par :

	Min	Max
Température ($^{\circ}\text{C}$)	20	30

1] Proposer un codage pour cet unique facteur.

$$x^* \in [a, b] \Rightarrow x \in [-1, 1]$$

$$x = \frac{x^* - 25}{5}$$

On fait l'hypothèse ici que sur cette plage de variation le rendement peut être approché par une fonction affine de la température. En désignant alors par y_i le i-ème rendement mesuré (en %) et par x_i la i-ème température testée (sous forme codée) le modèle statistique considéré a la forme suivante :

$$\forall i = 1, \dots, n, y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Le budget alloué permet de réaliser uniquement $n = 4$ expériences. Les deux stratégies suivantes sont alors proposées :

Codées

Stratégie 1	Stratégie 2	St 1	St 2
Exp 1 : 22°C	Exp 1 : 20°C	-1	$\frac{22-25}{5} = -\frac{3}{5} = -0.6$
Exp 2 : 28°C	Exp 2 : 24°C	0	$\frac{28-25}{5} = \frac{3}{5} = 0.6$
Exp 3 : 29°C	Exp 3 : 26°C	0	$\frac{29-25}{5} = \frac{4}{5} = 0.8$
Exp 4 : 30°C	Exp 4 : 30°C	1	$\frac{30-25}{5} = \frac{5}{5} = 1$

2 Quelle est selon vous la meilleure des deux stratégies ? Justifier ce résultat d'au moins deux manières différentes. [LogCalc]

3 Existe-t-il une stratégie optimale en 4 expériences ? [LogCalc]

20 Intuitivement la stratégie 2 semble être la meilleure, Justification on termes d'estimation on sait que $\hat{\beta} = (X^T X)^{-1} X^T Y$

On va codée l'information $a = 20^\circ C$ et $b = 30^\circ C \Rightarrow$ centre $\frac{20+30}{2} = 25$ y $\frac{30-20}{2} = 5$

On sait que si $\hat{\beta}$ est l'estimateur des moindres carrés de β

alors : $V(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$ Justification en termes de Var

$$X_1 = \begin{pmatrix} \beta_0 & \beta_1 \\ 1 & -0.6 \\ 1 & 0.6 \\ 1 & 0.8 \\ 1 & 1 \end{pmatrix} \Rightarrow X^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -0.6 & 0.6 & 0.8 & 1 \end{pmatrix} \Rightarrow X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -0.6 & 0.6 & 0.8 & 1 \end{pmatrix} \begin{pmatrix} 1 & -0.6 \\ 1 & 0.6 \\ 1 & 0.8 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -0.6 + 0.6 + 0.8 + 1 \\ -0.6 + 0.6 + 0.8 + 1 & -0.6^2 + 0.6^2 + 0.8^2 + 1 \end{pmatrix}$$

Matriz de Covarianzas
en la diagonal se obtiene la Varianza

Stratégie 1:

$$\text{On a } X_1 = \begin{pmatrix} \beta_0 & \beta_1 \\ 1 & -0.6 \\ 1 & 0.6 \\ 1 & 0.8 \\ 1 & 1 \end{pmatrix} \Rightarrow X_1^T X_1 = \begin{pmatrix} 4 & 1.8 \\ 1.8 & 2.36 \end{pmatrix} \Rightarrow (X_1^T X_1)^{-1} = \begin{pmatrix} 0.381 & -0.290 \\ -0.290 & 0.645 \end{pmatrix}$$

Stratégie 2:

$$X_2 = \begin{pmatrix} 1 & -1 \\ 1 & -0.2 \\ 1 & 0.2 \\ 1 & 1 \end{pmatrix} \Rightarrow X_2^T X_2 = \begin{pmatrix} 4 & 0 \\ 0 & 2.08 \end{pmatrix} \Rightarrow (X_2^T X_2)^{-1} = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.481 \end{pmatrix}$$

$$= \frac{\left| \begin{array}{cc} \text{Var } \hat{\beta}_0 & \text{Var } \hat{\beta}_1 \\ 0.381 \sigma^2 & 0.645 \sigma^2 \end{array} \right| - \text{Cov } (\hat{\beta}_0, \hat{\beta}_1)}{\left| \begin{array}{cc} \text{Var } \hat{\beta}_0 & \text{Var } \hat{\beta}_1 \\ 0.25 \sigma^2 & 0.481 \sigma^2 \end{array} \right|} \Rightarrow \text{ST2 es mejor}$$

$$\text{ST1: } \sigma^2 (1 \times 2) \begin{pmatrix} 0.381 & -0.29 \\ -0.29 & 0.64 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \begin{pmatrix} -0.381 - 0.29x & -0.29 + 0.64x \\ 2 \times 1 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = (-0.381 - 0.29x - 0.29x + 0.64x^2)$$

Justification en terme de prédictions:

Rappel:

$$\text{Var } \hat{y}(x) = \sigma^2 t g(x) (t^T x x)^{-1} g(x)$$

* avec $t^T g(x)$ vecteur de régression, i.e. $t^T g(x) = (1, x)$

\Rightarrow il vient

Stratégie 1: $\text{Var } \hat{y}(x) = \sigma^2 (0,381 - 0,580x + 0,645x^2)$

Stratégie 2: $\text{Var } \hat{y}(x) = \sigma^2 (0,250 + 0,481x^2)$

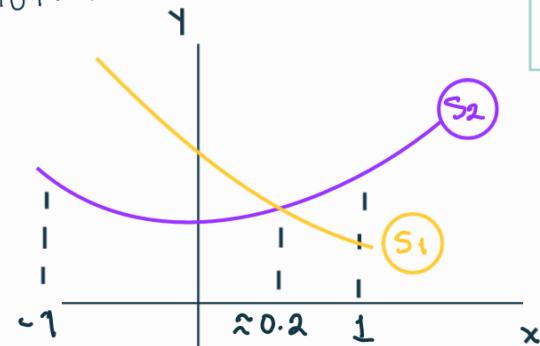
$\text{Var } \hat{y}(x)$
 $\sigma^2 = 1$

Qualité globale de la stratégie 1:

$$\frac{1}{2} \int_{-1}^1 \text{Var } \hat{y}(x) dx \approx 0,596 \sigma^2$$

stratégie 2: $0,410 \sigma^2 = \frac{\sigma^2}{2} \int_{-1}^1 0,25 + 0,481x^2 dx$

Conclusion: En moyenne la variance de prédiction de la stratégie 2 est meilleure



3^e Stratégie optimale en n expériences? $\text{Var } \hat{\beta} = \sigma^2 (X^T X)^{-1}$

Rappel) Résultats pour la droite des moindres carrés

$$\text{Var } \hat{\beta}_0 = \frac{\sigma^2}{m} \left(1 + \frac{\bar{x}^2}{\sigma_x^2} \right) ; \quad \text{Var } (\hat{\beta}_1) = \frac{\sigma^2}{m \sigma_x^2} ;$$

$$\text{Cov } (\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \frac{\bar{x}}{m \sigma_x^2} ; \quad \text{Var } \hat{y}(x) = \frac{\sigma^2}{m} \left(1 + \frac{(x - \bar{x})^2}{\sigma_x^2} \right)$$

On veut minimiser

$$\text{Var } \hat{\beta}_0 = \frac{\sigma^2}{m} \left(1 + \frac{\bar{x}^2}{\sigma_x^2} \right) \Leftrightarrow \bar{x} = 0^*$$

ceci va aussi entraîner que
 $\text{Cov } (\hat{\beta}_0, \hat{\beta}_1) = 0$

On veut minimiser $\text{Var } (\hat{\beta}_1) = \frac{\sigma^2}{m \sigma_x^2} \Rightarrow$ maximiser σ_x^2 , $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \left(\sum_{i=1}^n (x_i^2) \right)$$

comme nous avons minimisé $\hat{\beta}_0 \Rightarrow \bar{x} = 0$

Conclusion | 2 des x_i doivent valoir ± 1 les deux autres -1 ça veut dire mesurer 2 fois le minimum et deux fois

pour à température maximale

$$\Rightarrow \text{alors } X = \begin{pmatrix} 1 & -1 \\ 1 & +1 \\ 1 & -1 \\ 1 & +1 \end{pmatrix}, \quad {}^T X X = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = 4 I_2$$

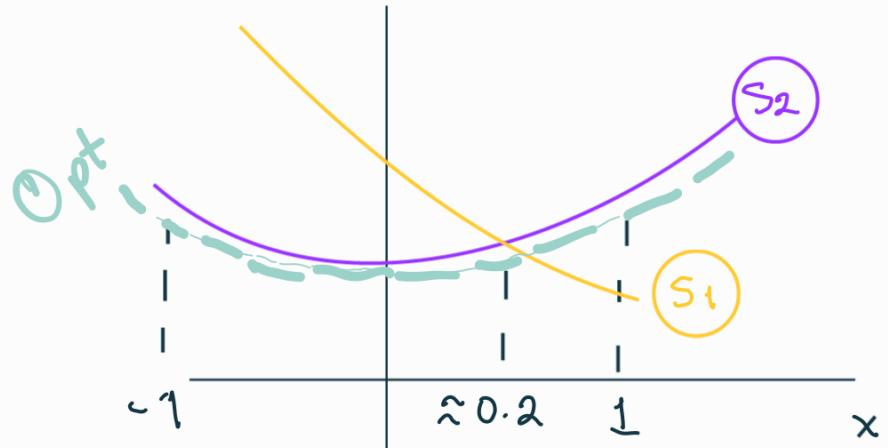
$$\text{donc } \text{Var} \hat{y}(x) = \sigma^2 {}^T g(x) ({}^T X X)^{-1} g(x) = \sigma^2 (1_x) \left(\frac{1}{4} I_2 \right) \begin{pmatrix} 1 \\ x \end{pmatrix}$$

$$\Rightarrow \text{Var} \hat{y}(x) = \frac{\sigma^2}{4} (1+x^2)$$

$$\int_{-1}^1 \text{Var} \hat{y}(x) dx = \int_{-1}^1 \frac{\sigma^2}{4} (1+x^2) dx = \frac{\sigma^2}{2} \int_{-1}^1 \frac{1+x^2}{2} dx =$$

Variance de prédiction moyenne

$$\frac{1}{2} \int_{-1}^1 \text{Var} \hat{y}(x) dx \approx 0.333 \sigma^2$$



Calcul
moindres carrés

$$S: X = \begin{pmatrix} p_0 \\ C_1 \\ C_2 \\ C_3 \end{pmatrix}$$

$$X^T = {}^T X$$

alors

$${}^T X X = \begin{pmatrix} {}^T C_1 \\ {}^T C_2 \\ {}^T C_3 \end{pmatrix} \begin{pmatrix} {}^T C_1 & {}^T C_2 & {}^T C_3 \\ {}^T C_2 & {}^T C_2 C_2 & {}^T C_2 C_3 \\ {}^T C_3 & {}^T C_3 C_2 & {}^T C_3 C_3 \end{pmatrix} = \begin{pmatrix} \|C_1\|^2 & \langle C_1, C_2 \rangle & \langle C_1, C_3 \rangle \\ \langle C_2, C_1 \rangle & \|C_2\|^2 & \langle C_2, C_3 \rangle \\ \langle C_3, C_1 \rangle & \langle C_3, C_2 \rangle & \|C_3\|^2 \end{pmatrix}$$

↑ columnas
↑ elementos

$$X^T = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 0 & 4 \\ 1 & 0 & -1 \end{pmatrix}^T C_1$$

avec $X =$

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 0 \\ 4 \\ -1 \end{pmatrix}$$

$$\beta g({}^T X X) = \beta g(X)$$

Minimos Cuadrados
demonstración

$${}^T X X = \begin{pmatrix} 5 & 6 & 2 \\ 6 & 25 & -1 \\ 2 & -1 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 2^2 + 1^2 + 0^2 & 2 \cdot 3 + 1 \cdot 0 + 0 \cdot 4 & 2 \cdot 1 + 1 \cdot 0 + (-1) \\ 3^2 + 0^2 + 4^2 & 3 \cdot 1 + 0 \cdot 0 + 4 \cdot (-1) \\ 1^2 + 0^2 + (-1)^2 & \end{pmatrix}$$

El objetivo de una investigación no es estimar un parámetro sino decidir en cuál de 2 conjuntos: θ_0 o θ_1 es más asequible que el parámetro poblacional se encuentre, donde θ_0 y θ_1 no se traslapen.

Una hipótesis estadística es una aseveración sobre el valor de uno o varios parámetros, distribución de probabilidad completa. La afirmación de que el parámetro se encuentra en el conjunto θ_0 se llama hipótesis nula y suele denotarse por H_0 , mientras que el hecho de asumir que se encuentra en H_1 se llama hipótesis alternativa. De esta forma, en una prueba de hipótesis se define una regla para decidir si se acepta la hipótesis nula ó se rechaza en favor de la hipótesis alternativa.

Dicha decisión se basa en el valor de una variable T , estadístico de prueba, obtenido a partir de una muestra, se espera que el estadístico de prueba T tenga una distribución conocida bajo el supuesto de que se cumple la hipótesis nula.

Para tomar la decisión de rechazar o no la hipótesis nula, se define un conjunto de valores C dentro del conjunto de los posibles valores de T , si el valor del estadístico de prueba está en $C \Rightarrow$ se rechazará H_0 . En caso contrario, si el valor del estadístico no está en $C \Rightarrow$ no se rechazará H_0 y se rechazara H_1 .

C : región de rechazo. Una vez q se especifica el estadístico de prueba y la región de rechazo la prueba de hipótesis está completamente especificada.

El rechazo de la hipótesis nula es una conclusión fuerte de q θ_0 parece ser consistente con la muestra, mientras que no rechazarla es una conclusión débil que debe ser interpretada como evidencia q si θ_0 si fuese consistente con los datos de la muestra.

Decisión estadística	Realidad		Potencia de la prueba $\alpha = P(t \in C H_0 \text{ cierta})$
	$H_0 \text{ es } T$	$H_0 \text{ es } F$	
Rechazar H_0	α , Error tipo I Rechazar q hay un problema, no hay problema	\checkmark $1 - \beta$	$\beta = P(t \in C^c H_0 \text{ falsa})$
No rechazar	\checkmark	Error tipo II Parar si q hay un problema	

Algebra lineal

Sean V y W espacios vectoriales con $\dim V < \infty$ y sea $T: V \rightarrow W$ una transformación lineal $\Rightarrow \text{Rank}(T) + \text{Nul}(T) = \dim V$

$$\text{Rank}(T) = \dim(\text{Im}(T)) \quad \dim(\text{Im}(T)) + \dim(\text{Ker}(T)) = \dim V$$

$$\text{Nul}(T) = \dim(\text{Ker}(T))$$

Determinante de una matriz siempre es un número real y únicamente se puede calcular para matrices cuadradas. Sea $A_{n \times n}$

- $\det(A) = \det(A^T)$
- $\det(A) = 0$ si contiene filas o columnas linealmente dependientes
- $\det(AB) = \det(A)\det(B)$

$$\sum x^2 p(x) - (\sum x p(x))^2$$

$$\text{Var}(X) = E[(X - E(X))^2]$$

Varianza: La varianza es la media de los cuadrados menos el cuadrado de la media.

$$\text{Var}(X) \geq 0 \quad \text{Var}(X+b) = \text{Var}(X) \text{ bijo constante}$$

$$\text{Var}(ax) = a^2 \text{Var}(X)$$

$$\Rightarrow \text{Var}(ax+b) = a^2 \text{Var}(X)$$

Si X, Y son v.a. independientes $\Rightarrow \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

02 Outils mathématiques pour les plans d'expérience

01 Généralités

Nous présentons ici des notions algébriques, probabilistes et statistiques utiles pour mettre en oeuvre la méthode des plans d'expérience. Bon nombre de ces notions sont des rappels classiques repérés par la notation \textcircled{R} .

con referencia
referenciadas

02 Outils algébriques

Etant donné un modèle linéaire sous la forme matricielle $Y = X\beta + \varepsilon$ il est courant de manier sa **matrice d'information** tXX , et notamment de savoir si elle est inversible ou non. Les résultats suivants d'algèbre linéaire sont classiques :

- \textcircled{R} | Pour toute matrice $X \in \mathcal{M}(n, p)$ la matrice ${}^tXX \in \mathcal{M}(p, p)$ est une matrice de **produits scalaires** telle que (en notant c_i le i -ème vecteur colonne de X) :

$$X = [c_1 \mid c_2 \mid \dots \mid c_p] \implies {}^tXX \text{ a pour terme général } \langle c_i, c_j \rangle$$

Rango: # maximo de columnas
lin. independientes

Amen los sig. propiedades son equivalentes:
A tiene rango completo
No es singular
Det(X) ≠ 0

Des conditions d'inversibilité de tXX peuvent être facilement obtenues via les résultats suivants :

- \textcircled{R} | Une matrice carrée est inversible si et seulement si elle est de plein rang. $M_{n,n}$ $\text{rango}(M) = n$

- \textcircled{R} | On a toujours $rg(X) = rg({}^tXX)$ donc tXX est **inversible** si et seulement X est de plein rang.

03 Outils probabilistes

Le résultat d'une expérience réelle est toujours **aléatoire** (car possiblement entaché d'erreur humaine, d'erreur de mesure, etc ...), il peut donc être considéré comme étant la réalisation d'une **variable aléatoire réelle** (v.a.r.) Y . Nous supposerons dans la suite que Y admet toujours une espérance mathématique et une variance. Nous ferons rarement d'hypothèse de loi sur Y car des théorèmes asymptotiques de type central-limite sont inadaptés à la situation des plans d'expérience où la **taille** de l'échantillon est, par définition, **petite**.

contaminado

Lorsque n expériences sont réalisées on peut considérer cette fois le **vecteur aléatoire** Y de \mathbb{R}^n , c'est-à-dire un vecteur $Y = {}^t(Y_1, \dots, Y_n)$ tel que chacune de ses composantes est une variable aléatoire réelle. On généralise les notions d'espérance et de variance, désignées par \mathbb{E} et \mathbb{V} dans le cas vectoriel par :

- \textcircled{R} | L'espérance mathématique de Y est : $\mathbb{E}(Y) = {}^t(E(Y_1), \dots, E(Y_n))$.
- \textcircled{R} | La matrice des covariances de Y est : $\mathbb{V}(Y) = \mathbb{E}[(Y - \mathbb{E}(Y))^t(Y - \mathbb{E}(Y))]$.

En ce qui concerne la matrice des covariances remarquons que $\mathbb{V}(Y)$ est constituée par les éléments suivants ($\forall i, j = 1, \dots, n$ avec $i \neq j$) :

$$\begin{cases} Var(Y_i) = E(Y_i^2) - [E(Y_i)]^2 \text{ sur la diagonale,} \\ Cov(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i) E(Y_j) \text{ hors de la diagonale.} \end{cases}$$

$$\mathbb{E}(Y_i Y_j) - [\mathbb{E}(Y_i)][\mathbb{E}(Y_j)]$$

$$\text{Diagonal} = [\mathbb{E}(Y_i^2) - [\mathbb{E}(Y_i)]^2]^2$$

Rappelons enfin que toute transformation affine telle que le vecteur Y a pour image le vecteur AY (avec A matrice non-aléatoire de taille adéquate) conduit aux résultats suivants :

$$\textcircled{(R)} \mid \mathbb{E}(AY) = A\mathbb{E}(Y) \text{ et } \mathbb{V}(AY) = A\mathbb{V}(Y)^t A$$

04 Outils statistiques : rappels

Considérons dans la suite uniquement des modèles statistiques linéaires, pouvant donc se ramener à une écriture matricielle $Y = X\beta + \varepsilon$. Supposons que n expériences ont été réalisées, que m facteurs sont étudiés et que le modèle considéré contient p paramètres inconnus. Voici une liste de rappels de notions classiques (voir éventuellement un cours de modélisation linéaire) :

- $\textcircled{(R)}$ Les hypothèses classiques sur les résidus (centrage, indépendance et homoscédasticité) sont supposées vérifiées donc :

$$\mathbb{E}(Y) = X\beta \text{ et } \mathbb{V}(Y) = \sigma^2 I_n$$

Les estimateurs des paramètres des différents modèles proposés seront toujours recherchés selon la technique des **moindres carrés**, donc :

- $\textcircled{(R)}$ Soit le modèle linéaire $Y = X\beta + \varepsilon$ avec X matrice de plein rang. L'estimateur des **moindres carrés** de β est alors donné par :

$$\hat{\beta} = (\mathbb{X}^t \mathbb{X})^{-1} \mathbb{X}^t Y \quad \beta = (\mathbb{X}^t \mathbb{X})^{-1} \mathbb{X}^t Y$$

Concernant les propriétés de cet estimateur il vient principalement :

- $\textcircled{(R)}$ L'estimateur des moindres carrés $\hat{\beta}$ est un estimateur **sans biais** de β admettant pour dispersion

$$\mathbb{V}(\hat{\beta}) = \sigma^2 (\mathbb{X}^t \mathbb{X})^{-1} \quad \mathbb{V}(\hat{\beta}) = \sigma^2 (\mathbb{X}^t \mathbb{X})^{-1}$$

$$g(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}$$

Concernant maintenant la réponse moyenne prédite par le modèle au point $x \in \mathbb{R}^m$ nous savons qu'elle est donnée par :

$$\hat{Y}(x) = t g(x) \hat{\beta} \quad \hat{Y}(x) = g^t(x) \hat{\beta} \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

avec $g(x) \in \mathbb{R}^p$ **vecteur de régression** c'est-à-dire tel que $t g(x)$ est construit de manière identique aux lignes de la matrice X . Ceci conduit alors au résultat suivant quantifiant la qualité de toute prédition :

- $\textcircled{(R)}$ La variance de $\hat{Y}(x)$ est donnée par :

$$\text{Var}(\hat{Y}(x)) = \sigma^2 g^t(x) (\mathbb{X}^t \mathbb{X})^{-1} g(x)$$

evaluamos en ese punto

Cette formule montre que la qualité des prédictions réalisées au point x dépend donc du point x choisi, de la dispersion des résidus via σ^2 mais aussi de la **matrice du modèle X** . Ce dernier point dépend directement du choix des expériences fait par l'utilisateur. La méthode des plans d'expérience peut donc aussi être vue comme un choix optimal pour la matrice X de manière à minimiser les variances de prédition.

Effectuons maintenant quelques rappels relatifs à la technique d'analyse de la variance. Désignons dans la suite par \bar{Y} la réponse moyenne observée. La **somme totale des carrés** (centrés), la **somme des carrés due à l'erreur** et la **somme des carrés due à la régression** sont données respectivement par (la notation SS venant de l'anglais *Sum of Squares*) :

$$\begin{array}{lll} \text{Sommes Total de Carrés} & SST = \sum_{i=1}^n (Y_i - \bar{Y})^2 & , \quad SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 , \quad SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 . \\ \text{Sum of Squares Total} & & \text{Somme des carrés due à l'erreur} \quad \text{Somme des carrés due à la régression} \end{array}$$

Il en découle le résultat classique suivant (souvent qualifié de « décomposition fondamentale ») :

- $\textcircled{(R)}$ Si $\mathbb{I}_n \subset \text{Im}X$ alors les sommes de carrés SST , SSE et SSR vérifient :

$$SST = SSR + SSE$$

tomato je la mets à range

$n-1$ $n-p$ $p-1$

Les sommes de carrés SST , SSE et SSR sont associées, respectivement, à $(n - 1)$, $(n - p)$ et $(p - 1)$ degrés de liberté. Ceci permet de définir les sommes moyennes de carrés de la manière suivante :

$$MSE = \frac{SSE}{n - p} \text{ et } MSR = \frac{SSR}{p - 1}.$$

Deux applications classiques directes de l'analyse de la variance sont présentées ci-dessous :

- ④ On appelle **coefficient de corrélation linéaire multiple** le réel :

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

- ④ Lorsque X est une matrice de plein rang p , un **estimateur sans biais** de la variance des résidus σ^2 est :

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n - p} = \frac{\sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2}{n - p}$$

Notons enfin les deux tests d'hypothèse classiques suivants (attention dans ce cas il est nécessaire d'effectuer une hypothèse de normalité par rapport aux réponses observées) :

Test de **validité** du modèle :

- ④ Un test d'hypothèse de H_0 : "tous les paramètres du modèle (sauf β_0) sont nuls" contre l'hypothèse $H_1 = \overline{H_0}$ peut être réalisé à l'aide de la statistique :

$$T = \frac{MSR}{MSE} = \frac{\frac{SSR}{p-1}}{\frac{SSE}{n-p}} = \frac{(n-p)SSR}{(p-1)SSE}$$

La règle de décision est alors donnée par (avec $f_{\alpha, p-1, n-p}$ fractile de la loi de Fisher à $(p - 1)$ et $(n - p)$ ddl) :

on rejette H_0 au niveau α si $t \geq f_{1-\alpha, p-1, n-p}$

Test de **significativité** des paramètres :

- ④ Désignons par β_i le i -ème élément du vecteur des paramètres inconnus du modèle linéaire utilisé. Un test d'hypothèse de H_0 : " $\beta_i = 0$ " contre H_1 : " $\beta_i \neq 0$ " peut être réalisé à l'aide de la statistique :

$$T = \frac{\hat{\beta}_i}{\sqrt{a_{ii}MSE}}$$

avec a_{ii} i -ème élément diagonal de $(^tXX)^{-1}$. La règle de décision est alors donnée par (avec $f_{\alpha/2, n-p}$ fractile de la loi de Student à $(n - p)$ ddl) :

on rejette H_0 au niveau α si $|t| \geq f_{1-\alpha/2, n-p}$

05

Outils statistiques : compléments

$$\sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2$$

Présentons ici une méthode de décomposition plus fine de la somme des carrés due à l'erreur SSE . Cette erreur découle en effet de deux sources différentes : soit le choix d'un mauvais modèle soit une trop grande variabilité des résultats observés (erreurs humaines, *etc...*). Une technique classique afin de distinguer ces deux sources d'erreur consiste à réaliser un certain nombre de **répétitions** d'expériences. Désignons alors par n^* le **nombre total de conditions expérimentales distinctes** et supposons que l'expérience i ($1 \leq i \leq n^*$) a été répétée $c_i \in \mathbb{N}^*$ fois. Pour la i -ème expérience notons les c_i réponses observées sous la forme suivante :

condition expérimental

$$Y_i^{(1)}, Y_i^{(2)}, \dots, Y_i^{(c_i)}.$$

$$n^* = 3 \text{ # de expériences} \quad \begin{array}{r} c_1 = 3 \\ c_2 = 4 \\ c_3 = 2 \\ \hline 8 \end{array} + \quad \begin{array}{c} Y_1^{(1)} Y_1^{(2)} Y_1^{(3)} \\ Y_2^{(1)} Y_2^{(2)} Y_2^{(3)} \\ Y_3^{(1)} Y_3^{(2)} \end{array}$$

Il découle de ces hypothèses que le nombre total d'expériences réalisées est :

$$n = \sum_{i=1}^{n^*} c_i = c_1 + c_2 + c_3 = 3 + 4 + 2 = 8$$

Remarquons que dans ce contexte le vecteur Y des observations est écrit dans l'ordre suivant :

$$tY = (Y_1^{(1)}, \dots, Y_1^{(c_1)}, Y_2^{(1)}, \dots, Y_2^{(c_2)}, \dots, Y_{n^*}^{(1)}, \dots, Y_{n^*}^{(c_{n^*})}).$$

Notons enfin, pour $1 \leq i \leq n^*$, \bar{Y}_i la réponse moyenne observée pour les c_i répétitions de la i -ème expérience, donc :

$$\bar{Y}_i = \frac{1}{c_i} \sum_{u=1}^{c_i} Y_i^{(u)}.$$

Ces notations généralisent bien le cas classique, sans répétition, qui correspond à $c_1 = c_2 = \dots = c_{n^*} = 1$ (n et n^* sont identiques). Lorsque l'expérience i n'est pas répliquée ($c_i = 1$) on notera, comme précédemment, Y_i au lieu de $Y_i^{(1)}$. On définit classiquement les deux nouvelles sommes suivantes appelées respectivement **somme des carrés due au manque d'ajustement** et **somme des carrés due à l'erreur pure** (avec les notations LOF pour Lack Of Fit et PE pour Pure Error) :

$$SSLOF = \sum_{i=1}^{n^*} c_i (\hat{Y}_i - \bar{Y}_i)^2, \quad SSPE = \sum_{i=1}^{n^*} \sum_{u=1}^{c_i} (Y_i^{(u)} - \bar{Y}_i)^2.$$

On montre alors que :

Proposition 01

La somme des carrés due à l'erreur se décompose toujours en :

$$SSE = SSLOF + SSPE = \sum_{i=1}^{n^*} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n^*} c_i (\hat{Y}_i - \bar{Y}_i)^2 + \sum_{i=1}^{n^*} \sum_{u=1}^{c_i} (Y_i^{(u)} - \bar{Y}_i)^2$$

Cette décomposition permet donc de distinguer la part de SSE qui est due au choix d'un mauvais modèle (**SSLOF**) de celle qui, par contre, découle de variations non-contrôlées (**SSPE**, qui n'est autre qu'une variance intra-classe à un coefficient près). On montre de plus que $SSLOF$ et $SSPE$ sont respectivement associés à $(n^* - p)$ et $(n - n^*)$ degrés de liberté, ce qui permet de poser :

$$\text{sommes moyennes de carrés} \quad MSLOF = \frac{SSLOF}{n^* - p} \text{ et } MSPE = \frac{SSPE}{n - n^*} \quad \# \text{ de prédictoires}$$

Il en découle un nouveau test d'hypothèse si au moins une des expériences a été répliquée. En effet on peut maintenant évaluer le défaut d'ajustement du modèle, c'est-à-dire sa capacité ou non à bien décrire en moyenne le phénomène étudié. Mathématiquement, on dit que le modèle postulé est mal ajusté si :

on suppose que $\mathbb{E}(Y) = X\beta$ alors qu'en réalité $\mathbb{E}(Y) = X\beta + X^*\beta^*$.

avec β^* vecteur des paramètres négligés à tort (il s'agit souvent dans le cas polynomial de paramètres de plus haut degré qu'il n'aurait pas fallu oublier). Testons maintenant l'hypothèse :

$$H_0 : \text{"le modèle est bien ajusté en moyenne"} \text{ contre } H_1 = \bar{H}_0.$$

Remarquons que du point de vue mathématique, l'hypothèse H_0 se traduit par :

$$H_0 : "(I_n - P)X^*\beta^* = 0" \text{ avec } P = X(X^*X)^{-1}X^*.$$

En d'autres termes, si le modèle est bien ajusté alors le terme $X^*\beta^*$ ne va pas apporter d'information nouvelle par rapport à $X\beta$ et donc sa projection orthogonale sur $(Im X)^\perp$ doit être nulle. On démontre alors que si les observations sont des réalisations indépendantes d'une loi normale, la statistique :

$$T = \frac{SSLOF / (n^* - p)}{SSPE / (n - n^*)} = \frac{MSLOF}{MSPE}$$

suit, sous l'hypothèse H_0 , une loi de Fisher avec $(n^* - p)$ et $(n - n^*)$ degrés de liberté. Il en résulte que :

de prédictoires

Proposition 02

Un **test d'hypothèse** de H_0 : "le modèle est bien ajusté en moyenne" contre l'hypothèse $H_1 = H_0$ peut être réalisé à l'aide de la statistique :

$$T = \frac{MSLOF}{MSPE}.$$

La règle de décision est alors donnée par (avec $f_{\alpha, n^* - p, n - n^*}$ fractile de la loi de Fisher à $(n^* - p)$ et $(n - n^*)$ ddl) :

on rejette H_0 au niveau α si $t \geq f_{1-\alpha, n^* - p, n - n^*}$.

06 Exercices d'applications

Exercice 1

[Expériences répétées]

On s'intéresse ici au rendement (en %) d'une réaction chimique avec pour objectif de le maximiser. Il semble que ce rendement ne dépend que de la température de la solution étudiée. Un technicien a réalisé les $n = 6$ expériences suivantes :

Temp. ($^{\circ}\text{C}$)	10	10	15	20	25	30
Rend. (%)	10	20	35	40	33	10

1 Justifier, de diverses manières, que l'ajustement d'une droite n'est pas de bonne qualité. [LogStat] :

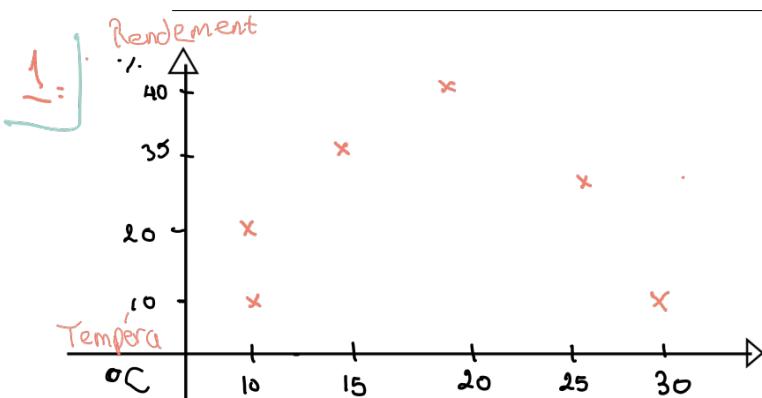
On s'oriente alors vers la mise en oeuvre d'un modèle quadratique, c'est-à-dire tel que (en désignant par x la température) :

$$\forall x \in [10, 30], Y(x) = \beta_0 + \beta_1 x + \beta_{11} x^2 + \varepsilon(x)$$

2 Réaliser une analyse « classique » (sans tenir compte des nouvelles notions de ce chapitre) à l'aide de ce modèle. Indiquer quelles sont les conclusions obtenues à priori. [LogStat]

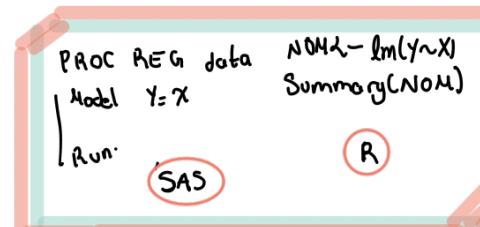
3 Réaliser maintenant une analyse plus fine (décomposer SSE , calculer la statistique de test, déterminer la p-value associée). Indiquer quelles sont les conclusions obtenues. [LogStat]

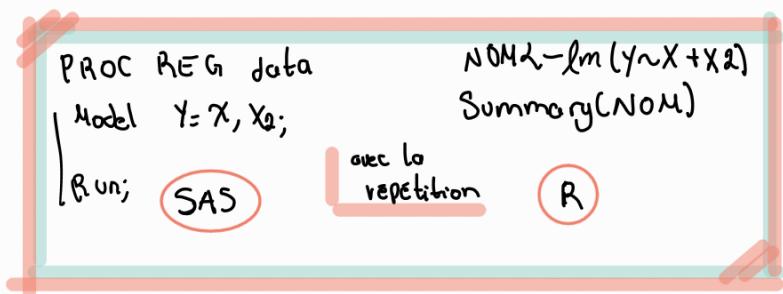
4 Quelle est alors la température optimale prédictive ? Calculer le rendement moyen prédict ainsi que sa dispersion. Commenter les résultats obtenus. [LogCalc]



- Le nuage de points a une forme pas de tout rectiligne
- Si on ajuste quand même une droite (logique) on trouve

$$R^2 \approx 0,0047$$





2 = Dans une analyse classique... $SST \approx 863,3$ avec $SSR = 812,5$
 $SSE = 50,8$

Il en découle donc que: $R^2 = 1 - \frac{SSE}{SST} \approx 0.94 < 0.95 \Rightarrow$ modèle mal ajusté?

3 = On a ici, $n^* = 5$, n'importe de températures qui sont distinctes avec $C_1 = 2, C_2 = C_3 = C_4 = C_5 = 1 \Rightarrow n = \sum_{i=1}^{n^*} C_i = 2 + 1 + 1 + 1 + 1 = 6$

avec $y_1^{(1)} = 10$ et $y_2^{(1)} = 20$ donc $\bar{y}_1 = \frac{1}{2} \sum_{v=1}^2 y_i^{(1)} = 15$

Décomposition:
On a $50,8 = SSE = SSPE + SLOF$

$$SSLOF = \sum_{i=1}^{n^*} C_i (\hat{y}_i - \bar{y}_i)^2$$

$$SSPE = \sum_{r=1}^{n^*} \sum_{v=1}^{C_r} (y_i^{(v)} - \bar{y}_r)^2$$

On a $SSPE = \sum_{i=1}^{n^*} \sum_{v=1}^{C_i} (y_i^{(v)} - \bar{y}_i)^2$ $\bar{y}_i = \frac{1}{C_i} \sum_{v=1}^{C_i} y_i^{(v)} = y_i$

$$= [y_1^{(1)} - \bar{y}_1]^2 + [y_1^{(2)} - \bar{y}_1]^2 + (y_2 - \bar{y}_2)^2 + (y_3 - \bar{y}_3)^2 + (y_4 - \bar{y}_4)^2 + (y_5 - \bar{y}_5)^2$$
 $= (10 - 15)^2 + (20 - 15)^2 = (-5)^2 + 5^2 = 25 + 25 = 50$

$\Rightarrow SLOF = 0.8 \Rightarrow$ L'erreur est principalement $SSPE$; due aux fluctuations des réponses observées (et pas au modèle!)

$$\forall x \in [10, 30] \text{ on a } y(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon(x)$$

A

$$t = \frac{MSLOF}{MSPE} = \frac{\frac{SSLOF}{n^* - p} \text{ * de prédicteurs } \beta_0, \beta_1, \beta_2}{\frac{SSPE}{n - n^*}} = \frac{\frac{0.8}{2}}{\frac{50}{1}} = 0,008$$

\Rightarrow L'hypothèse $H_0: \text{"le modèle est bien ajusté en moyenne"}$ peut être rejettée au niveau α si $t \geq F_{1-\alpha, n^*-p, n-n^*}$

Ici $0,008 = F_{1-\alpha, 2, 1}$ le test

Déterminons la p-value α^* (valeur faisant bus solution de $0,008 = F_{1-\alpha, 2, 1}$)

Notons X une VAR de la loi de Fisher à 2 et 1 ddl, alors par définition $P[X \leq F_{\alpha, 2, 1}] = \alpha$ donc on a ici

$$P[X \leq F_{1-\alpha, 2, 1}] = P[X \leq 0,008] = 1 - \alpha^*$$

$\Rightarrow \alpha^* = 1 - P[X \leq 0,008] \approx 0,992 \Rightarrow$ Rejet de H_0 impossible

Comment forcer les logiciels à décomposer SSE?

Dans PROC REG

SAS Model $y = m / lackfit;$

R Remplace LM par RSM

A) $\forall x \in [10, 30], \hat{y}(x) = -66.6 + 10.96x - 0.28x^2$ Recherche du max. théorique (déivation) $\hat{y}'(x) = 0 \Rightarrow 10.96 - 0.56x = 0 \Rightarrow x = \frac{10.96}{0.56} \approx 19.6^\circ C$

Qualité de cette prédiction:

$$\text{Var } \hat{y}(x^*) = \sigma^2 g(x^*) ({}^t x x)^{-1} g(x^*)$$

avec $\sigma^2 \approx \hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{n-p} = \frac{50.8}{6-3} \approx 16.933$

$$X_1 = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 \\ 1 & 10 & 100 \\ 1 & 10 & 100 \\ 1 & 15 & 225 \\ 1 & 20 & 400 \\ 1 & 25 & 625 \\ 1 & 30 & 900 \end{pmatrix} \quad {}^t g(x) = (1 \ x \ x^2)$$

En conclusion ① Le rendement maximal théorique est obtenu pour $x^* = 19.6^\circ C$ et vaut $\hat{y}(19.6) \approx 40.6\%$.

② La dispersion associée vaut

$$\hat{\sigma}^2 \approx 16.933$$

$$\text{Var } \hat{y}(x^*) \approx 8.16 \text{ donc } \sigma(x^*) = 2.9$$

\Rightarrow Résultats pas précis (utiliser l'approximation de la règle des 3σ) mais logiques car pour 10°C les relais sont (très) mauvais.

03 Plans factoriels complets pour modèles d'ordre un

01 Généralités

Ce chapitre concerne les plans d'expérience pour facteurs quantitatifs associés à l'un des modèles les plus simples possibles, en l'occurrence le modèle polynomial de degré un.

L'hypothèse principale utilisée ici est que le phénomène étudié peut être approché par un polynôme du premier degré à m variables. Il est clair qu'un tel modèle n'est pas d'une grande richesse mais son utilisation est cependant intéressante dans certains cas. Par exemple, utiliser un tel polynôme peut donner une bonne approximation lorsque le domaine expérimental est petit. Un autre cas d'application classique est celui où l'on dispose, *a priori*, d'un grand nombre de facteurs susceptibles d'agir sur la réponse observée. L'utilisation d'un modèle plus riche est alors généralement impossible à cause de la grande taille de celui-ci. C'est pourquoi il est courant de débuter une telle étude par un modèle facile à manipuler afin de détecter quels sont les facteurs qui semblent être les plus influents. On dit que l'on utilise des techniques de **criblage** (ou *screening* avec la terminologie anglaise).

Nous proposons ici une première approche de la planification expérimentale via l'étude des dispositifs basiques que sont les plans factoriels complets.

02 Modèle utilisé

Considérons un phénomène aléatoire dépendant de m facteurs quantitatifs mis en œuvre au sein du domaine expérimental $\mathcal{E} \subset \mathbb{R}^m$. Le modèle statistique étudié ici s'écrit donc $Y(x) = f(x) + \varepsilon(x)$ avec la loi de réponse donnée par :

$$\forall x \in \mathcal{E}, f(x) = \beta_0 + \sum_{i=1}^m \beta_i x_i$$

Pour un tel modèle on dit que β_0 (*i.e.* la constante polynomiale) est l'**effet moyen général** et β_i ($i = 1, \dots, m$) est l'**effet linéaire** du facteur i . Le **nombre de paramètres inconnus** de ce modèle est donc :

$$p = m + 1$$

On en déduit que le nombre d'expériences à réaliser pour ajuster un tel modèle est : $n \geq p = m + 1$. Si une configuration en $n = p$ expériences est utilisée on dit alors que l'on a un plan d'expérience **saturé**.

Remarquons que matriciellement ce modèle linéaire peut être écrit sous la forme classique $Y = X\beta + \varepsilon$ avec donc ici :

$$X = [\underbrace{\mathbb{I}_n}_{\text{constante polynomiale}} \mid D]$$

La première colonne de 1 est associée à la constante polynomiale. La matrice $D \in \mathcal{M}(n, p)$ est appelée **matrice du plan** (elle contient les coordonnées des points du plan d'expérience sur ses lignes). Terminons par la définition usuelle suivante :

Définition 02

Un plan d'expérience est dit **orthogonal** si et seulement si sa matrice d'information (^tXX) est diagonale.

Nous allons avoir souvent pour objectif dans la suite de rechercher des configurations orthogonales. Il est clair que ceci va permettre de simplifier bon nombre de calculs (à commencer par la résolution des équations normales) mais nous verrons plus tard que les plans d'expérience orthogonaux sont surtout optimaux (dans un sens à préciser).

03 Plans factoriels complets : cas de 2 facteurs

Considérons un phénomène aléatoire dépendant de $m = 2$ facteurs. Une fois ces deux facteurs codés le domaine expérimental devient $\mathcal{E} = [-1, +1]^2$ et le modèle statistique étudié s'écrit donc $Y(x) = f(x) + \varepsilon(x)$ avec :

$$\forall x \in \mathcal{E}, f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

Il convient donc de réaliser ici un minimum de $p = 3$ expériences. Utiliser un plan d'expérience dit **factoriel complet** consiste alors en la mise en oeuvre des 4 expériences positionnées aux 4 sommets du carré unité. Considérons alors le protocole expérimental ci-dessous. L'ordre dans lequel les expériences sont réalisées est subjectif, elles sont ordonnées ici selon l'ordre classique dit de Yates (voir la partie 5).

Exp.	Fact. 1	Fact. 2	Réponse
1	-1	-1	Y_1
2	1	-1	Y_2
3	-1	1	Y_3
4	1	1	Y_4

La matrice du modèle associée à ce plan d'expérience est donc donnée par :

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (^tXX) = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix} = 4I_3 \Rightarrow (^tXX)^{-1} = \frac{1}{4}I_3 \quad \text{Donc ce plan est bien orthogonal}$$

→ Vérifier à titre d'exercice que ce plan d'expérience est bien orthogonal. Montrer ensuite que l'estimateur des moindres carrés des paramètres du modèle ainsi que sa dispersion sont donnés par :

$$\hat{\beta} = \frac{1}{4} (^tXY) \quad \text{et} \quad V(\hat{\beta}) = \frac{\sigma^2}{4} I_3$$

$$\text{On sait que: } \hat{\beta} = (^tXX)^{-1} X^T Y \Rightarrow \frac{1}{4} X^T Y = \hat{\beta} \quad V(\hat{\beta}) = \sigma^2 (^tXX)^{-1} = \frac{\sigma^2}{4} I_3 = V(\hat{\beta})$$

04 Plans factoriels complets : cas de 3 facteurs

Considérons un phénomène aléatoire dépendant de $m = 3$ facteurs. Une fois ces trois facteurs codés le domaine expérimental devient $\mathcal{E} = [-1, +1]^3$ et le modèle statistique étudié s'écrit donc $Y(x) = f(x) + \varepsilon(x)$ avec :

$$\forall x \in \mathcal{E}, f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

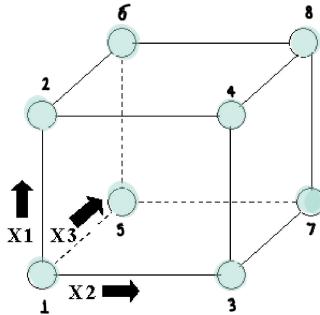
Il convient donc de réaliser ici un minimum de $p = 8$ expériences. Utiliser un plan d'expérience dit **factoriel complet** consiste alors en la mise en oeuvre des 8 expériences positionnées aux 8 sommets du cube unité. Le graphique ci-dessous représente un plan factoriel complet pour 3 facteurs avec les expériences ordonnées selon l'ordre de Yates (voir la partie 5).

Domaine expérimental

8 expériences

positionnées aux

8 sommets du cube unité



La matrice du modèle associée à ce plan d'expérience est donc donnée par :

Vérifie qu'il est orthogonal, i.e. $x^T x$ est diagonal.

Matrice de modèle selon l'ordre Yates est :

$$X = \begin{bmatrix} 1 & \beta_1 & \beta_2 & \beta_3 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \Rightarrow x^T x = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix} = 8 I_4 \Rightarrow \text{il est orthogonal}$$

$\sum_{j=1}^k \beta_j^2 = 8$

→ Vérifier à titre d'exercice que ce plan d'expérience est bien orthogonal. Montrer ensuite que l'estimateur des moindres carrés des paramètres du modèle ainsi que sa dispersion sont donnés par :

$$\text{On sait que } \hat{\beta} = (x^T x)^{-1} x^T y \quad \text{Var}(\hat{\beta}) = \sigma^2 (x^T x)^{-1} \Rightarrow (x^T x)^{-1} x^T y = \hat{\beta} = \frac{1}{8} x^T Y \quad \text{et} \quad \text{V}(\hat{\beta}) = \frac{\sigma^2}{8} I_4 = \sigma^2 (x^T x)^{-1}$$

El inverso de una matriz diagonal es el inverso multiplicativo de c/elemento de la diagonal

05 Plans factoriels complets : cas général

Considérons maintenant un phénomène aléatoire dépendant de m facteurs. Une fois ces facteurs codés le domaine expérimental devient l'hypercube $\mathcal{E} = [-1, +1]^m$ et le modèle statistique étudié s'écrit donc $Y(x) = f(x) + \varepsilon(x)$ avec :

$$\forall x \in \mathcal{E}, f(x) = \beta_0 + \sum_{i=1}^m \beta_i x_i.$$

Il convient donc de réaliser ici un minimum de $p = m + 1$ expériences.

Définition 03

Un plan d'expérience pour m facteurs est de type **factoriel complet** si et seulement si il est constitué de l'ensemble des sommets de l'hypercube $[-1, +1]^m$.

2^m

Il en résulte donc que tout plan factoriel complet est constitué par un total de $n = 2^m$ expériences.

Ordre de Yates pour les expériences

L'ordre dans lequel les expériences sont réalisées n'a aucune importance du point de vue mathématique. Cela a cependant un intérêt en pratique afin que des personnes différentes repèrent les expériences de manière identique. Yates a alors proposé d'utiliser l'algorithme suivant afin de construire la matrice D de tout plan factoriel complet :

- i) la première ligne de D n'est constituée que des valeurs -1 (i.e. la première expérience est réalisée en fixant tous les facteurs à leur niveau bas),
- ii) la première colonne de D est obtenue en changeant de signe toutes les lignes. La seconde colonne est obtenue en changeant de signe toutes les 2 lignes, ..., la k -ième colonne de D est obtenue en changeant de signe toutes les 2^{k-1} lignes.

A l'aide d'arguments géométriques basés sur le fait que ce type de configuration est invariante par de multiples symétries, on montre que :

Proposition 03

Tout plan d'expérience factoriel complet pour m facteurs est orthogonal pour le modèle d'ordre un. Sa matrice d'information est de plus donnée par :

$${}^t XX = 2^m I_p \quad p=m+1$$

$m = \# B_i > 0$
Valores que si: tendrán
máximo mínimo
los q' tienen efecto
lineal

Il en découle donc les formules simplifiées suivante pour l'estimateur des moindres carrés de β :

$$\hat{\beta} = \frac{1}{2^m} {}^t XY \quad \text{et} \quad \mathbb{V}(\hat{\beta}) = \frac{\sigma^2}{2^m} I_p$$

Comme il sera montré dans un chapitre ultérieur l'orthogonalité de ces plans d'expérience assure leur optimalité (dans un sens à préciser). Un problème se pose cependant au niveau de la taille de ces configurations puisqu'il vient pour tout plan factoriel complet lorsque m est grand :

$$n = 2^m \gg m + 1 = p$$

Il en résulte que de telles configurations ne sont absolument pas économiques lorsque le nombre de facteurs est élevé (elles ne sont utilisées en général que pour $m = 2, 3, 4$ facteurs).

Qualité de la réponse moyenne prédictive

La dispersion de la réponse moyenne prédictive est donnée de manière générale par :

$$\text{Var}(\hat{Y}(x)) = \sigma^2 {}^t g(x) ({}^t XX)^{-1} g(x)$$

avec ici $x = (x_1, \dots, x_m)$ et ${}^t g(x)$ vecteur de régression construit de manière identique aux lignes de la matrice X donc ${}^t g(x) = (1, x_1, \dots, x_m)$. Il découle donc de la proposition 03 que pour tout plan factoriel complet :

$$\text{Var}(\hat{Y}(x)) = \frac{\sigma^2}{2^m} {}^t g(x) g(x) = \frac{\sigma^2}{2^m} \left(1 + \sum_{i=1}^m x_i^2 \right) = \frac{\sigma^2}{2^m} \left(1 + \|x\|^2 \right)$$

rotability

Au vu de ce résultat on peut affirmer que tout plan factoriel complet est isovariant par rotations (pour le modèle d'ordre un). En d'autres termes la variance de prédiction n'est alors plus une fonction de m variables mais une fonction d'une seule variable réelle (le rayon $r = \|x\|$). Cette propriété est très utile en pratique car elle simplifie beaucoup l'évaluation de la qualité des prédictions réalisées.

06 Exercices d'applications

Exercice 1

[Mise en oeuvre pratique d'un plan factoriel complet]

Deux pièces plastiques sont collées entre elles dans une presse hydraulique. L'utilisateur pense que la pression exercée par la presse (en kg/m^2), la durée de cette pression (mesurée en s) ainsi que la quantité de colle déposée (en g) semblent avoir un effet sur la qualité du collage final. Cette qualité est mesurée par un coefficient quantifiant la résistance à la traction de la pièce (plus cette valeur est élevée meilleur est le collage effectué). Un plan factoriel complet est utilisé et les réponses suivantes ont été mesurées :

1°) Comme chaque B_i en $[-1, 1]$, avec B_1, B_2, B_3 on trouve les min et max \Rightarrow

avec $y = \begin{pmatrix} y_1 \\ \vdots \\ y_8 \end{pmatrix} \in \mathbb{R}^8 \quad \hat{\beta} = \frac{1}{8} {}^t X Y$

$\Rightarrow \hat{\beta} = \begin{pmatrix} \frac{1}{8} (y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8) = \bar{Y} \\ \frac{1}{8} (-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8) \\ \frac{1}{8} (-y_1 - y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8) \\ \frac{1}{8} (-y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8) \end{pmatrix} = \begin{pmatrix} 8,0 \\ 20,5 \\ 3,5 \\ 0,25 \end{pmatrix}$

$X = \begin{pmatrix} 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \Rightarrow X^T X = 8 \mathbb{I}_4$

Analyse de la variance: $SST = 3474$, $SSE = 13.5$, $SSR = 3460.5$
 Ceci donne: $R^2 = 1 - \frac{SSE}{SST} \approx 0.996$ Modèle à priori bien ajusté

Exp.	Pre.	Dur.	Qua.	rép.	Quantity
1	40	6	10	56	
2	80	6	10	98	
3	40	8	10	63	
4	80	8	10	102	
5	40	6	15	54	
6	80	6	15	98	
7	40	8	15	65	
8	80	8	15	104	+ mieux

$$\hat{y}(x_1, x_2, x_3) = 80 + 20.5 x_1 + 35 x_2 + 0.25 x_3$$

où $|B_1| > |B_2| > |B_3|$

$$\text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{2^m} I_p$$

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-p} = \frac{13.5}{8-(3+1)} = \frac{13.5}{4} = 3.375$$

$$\Rightarrow \text{Var}(\hat{\beta}) = \frac{3.375}{8} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 0.4218 I_4$$

- 1 Ajuster un modèle d'ordre un. Justifier que la qualité de l'ajustement est bonne.
 \Rightarrow modèle à priori est bien ajusté

2 Le commanditaire de l'étude pose les questions (classiques) suivantes. Le facteur les plus influent sur le collage est la
 Quels facteurs semblent être les plus influents par rapport à la résistance de la pièce? est la pression puis la durée
 Quelle est l'interprétation pratique des valeurs obtenues?

2.5) Existe-t-il un ou plusieurs facteurs à priori sans effet sur cette résistance?

Quels sont les réglages optimaux recommandés?

Lorsque la pression augmente de une unité sous forme codée
 soit $+20 \text{ kg/m}^2$ en moyenne la force de décollage de la
 pièce augmente de 20.5 units

Exercice 2

/ Rajout d'expérience centrales /

On considère ici un phénomène aléatoire dépendant de $m = 2$ facteurs et on propose de l'étudier à l'aide d'une des deux configurations suivantes :

Plan d'expérience 1 : plan factoriel complet.

Plan d'expérience 2 : plan factoriel complet avec deux expériences centrales.

- 1 Expliquer précisément en quoi la configuration 2 va donner des résultats de meilleure qualité.

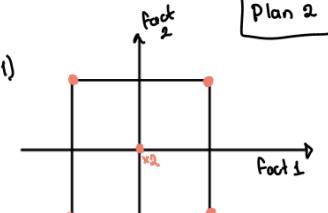
- Exercice 1) 2) Généraliser ceci au cas où $n_0 \in \mathbb{N}$ expériences centrales sont considérées. Quel choix peut-on proposer en pratique pour n_0 ?

2.b) Effect significatif de la colle?
 On peut tester l'hypothèse $H_0: \hat{\beta}_3 = 0$ contre $H_1: \hat{\beta}_3 \neq 0$
 Statistique de test $t = \frac{\hat{\beta}_3}{\text{Var}(\hat{\beta}_3)} = \frac{0,25}{0,65} \approx 0,385$. Au niveau $\alpha = 5\%$ le rejet de H_0 est

possible si: $t > t_{1-\alpha/2, n-p} = t_{0,975, 4} \approx 2,776$ como no pasa que $0,385 > 2,776 \Rightarrow$ no se rechaza H_0

\Rightarrow Pas d'effet significatif de la colle détecté à 5% (p-value de 0,72)

Exercice 2)



Matrice du modèle

$D = 2$ facteurs = m

$\hat{\beta}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_2$

$1 \quad -1 \quad -1$

$1 \quad 1 \quad -1$

$1 \quad -1 \quad 1$

$1 \quad 1 \quad 1$

$0 \quad 0 \quad 0$

Plan Var $\beta_0 > \text{Plan Var } \beta_0$
 l'estimation de la constante β_0

Approche: Qualité des prédictions: $\text{Var} \hat{y}(x) = \sigma^2 g(x)(t_{xx})^{-1}g(x)$, avec plan 1

$$\Rightarrow \text{Var} \hat{y}(x) = \frac{\sigma^2}{4} t g(x) g(x)^T = \frac{\sigma^2}{4} (1, x_1, x_2) \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}^T = \frac{\sigma^2}{4} (1 + x_1^2 + x_2^2) = \frac{\sigma^2}{4} (1 + \|x\|^2)$$

04 Fractions régulières pour modèles d'ordre un

01 Généralités

expériences = $n = 2^m$

Nous venons de voir que les plans factoriels complets sont intéressants dans la mesure où ces configurations sont toujours orthogonales (ce qui induit des simplifications calculatoires importantes et surtout des optimalités qui seront présentées au chapitre 05). Cependant la taille très importante de ces plans (en $n = 2^m$ expériences) rend leur utilisation impossible lorsque le nombre de facteurs m est grand (puisque le modèle n'a que $p = m+1$ paramètres inconnus). Il est donc naturel de s'interroger sur l'utilité d'utiliser tous les sommets de l'hypercube unité. L'objectif de ce chapitre est alors de construire des sous-ensembles des plans factoriels complets, appelés fractions régulières de plans factoriels complets, aptes à conserver la propriété d'orthogonalité.

Le contexte de ce chapitre est le même que celui du chapitre 03 que ce soit pour les notations ou pour le modèle polynomial d'ordre un utilisé.

02 Présentation d'un exemple

Reprendons l'exemple du chapitre précédent concernant un phénomène dépendant de $m = 3$ facteurs. Une fois ces trois facteurs codés le domaine expérimental devient $\mathcal{E} = [-1, +1]^3$ et le modèle statistique étudié s'écrit donc $Y(x) = f(x) + \varepsilon(x)$ avec :

$$\forall x \in \mathcal{E}, f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

On en déduit qu'un minimum de $n = 4$ expériences est nécessaire afin d'ajuster ce modèle. La configuration proposée dans le chapitre précédent était un plan factoriel complet composé des $n = 8$ sommets du cube unité. Si chacune des expériences a un coût très élevé cette stratégie va être jugée alors globalement trop coûteuse dans le sens où l'on réalise le double du nombre minimal d'expériences nécessaires. Designons par D la matrice du plan factoriel complet et ayons pour objectif de diviser par deux le nombre d'expériences (donc de rechercher une configuration saturée).

Critère de choix:

$$x_1 x_2 x_3 = 1 \times 0$$

que les produits $\text{cat}_D =$

-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
1	-1	1
-1	1	1
1	1	1

Critère de choix:

$$1 \oplus 2 \oplus 3 = \overline{t}_{\overline{t}_4}$$

$$= 123 = \overline{t}_{\overline{t}_4}$$

Une fraction régulière de ce plan complet est, par exemple, donnée par la matrice D_f ci-dessous :

$$D_f = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Seules les lignes de D dont le produit des trois coordonnées est égal à +1 ont été conservées. On a donc divisé par deux ici le nombre d'expériences tout en conservant une configuration orthogonale. Nous allons détailler et expliquer les fondements de cette méthode de sélection des expériences dans la suite de ce chapitre.

- ☛ Vérifier à titre d'exercice que ce plan d'expérience est bien orthogonal. i.e $X^T X$ est une matrice diagonale
- Exemple d'application pratique (noter que les trois facteurs sont ici qualitatifs).

Fractional Factorial Designs

L6O

Selecting the Half Fraction

Standard Order	Button Size	Offering Graphic	Discount Field
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

- Each factor makes two comparisons for each of the 3 factors (balanced)
- Covers the most experimental space using four trials
- Collapses into a full factorial if one of the factors is found not significant

© 2012 Sigma Consulting Resources, LLC 72

03 Fractions régulières : cas d'un seul générateur

Présentons ici la théorie des fractions régulières en commençant par le cas particulier où un seul générateur intervient.

1) Produit d'Hadamard

On appelle **produit d'Hadamard** (noté \odot) l'opérateur qui à deux vecteurs u et v de \mathbb{R}^n associe le vecteur $u \odot v$ de \mathbb{R}^n tel que (si $u = (u_i)_{i=1,\dots,n}$ et $v = (v_i)_{i=1,\dots,n}$) :

$$u \odot v = (u_i v_i)_{i=1,\dots,n}.$$

Le produit d'Hadamard définit donc une loi de composition interne sur \mathbb{R}^n et les propriétés suivantes sont immédiates ($\forall u, v, w \in \mathbb{R}^n$) :

- le produit d'Hadamard est **associatif** $((u \odot v) \odot w = u \odot (v \odot w))$,
- le produit d'Hadamard est **commutatif** $(u \odot v = v \odot u)$,
- \mathbb{I}_n est **élément neutre** pour le produit d'Hadamard $(u \odot \mathbb{I}_n = \mathbb{I}_n \odot u = u)$.

Remarquons enfin que pour tout vecteur $u \in \{-1, 1\}^n$ il vient :

$$u \odot u = u^{\odot 2} = \mathbb{I}_n. \quad \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \odot \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} \approx \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix}$$

Appliquons maintenant tout ceci aux vecteurs colonne composant la matrice D du plan factoriel complet et utilisons par la suite les deux conventions simplificatrices suivantes :

Convention 1

On désigne par $\mathbf{1}, \dots, \mathbf{m}$ les m vecteurs colonnes de la matrice D du plan factoriel complet.

Convention 2

L'opérateur \odot ne sera désormais plus noté ($u \odot v$ sera remplacé par uv).

2) Notion de générateur

Reprendons l'exemple du paragraphe 02 et définissons clairement (avec les notations présentées ici) la fraction régulière construite. Le critère de choix était basé sur le produit des trois coordonnées des points

du plan c'est-à-dire d'un point de vue global sur le vecteur **123**. Seules les expériences dont le produit des coordonnées est égal à 1 ont été conservées. Toutes autres termes seules les expériences (c'est-à-dire les lignes de la matrice D) vérifiant l'équation suivante ont été gardées :

$$\mathbb{I} = \mathbf{123}$$

On dit alors que la fraction régulière construite a pour **générateur 123**.

3) Notion de résolution

Etant donnée une fraction régulière définie à l'aide d'un générateur comment savoir maintenant si l'analyse statistique va être réalisable ou non ? Et si elle est réalisable le plan d'expérience obtenu est-il orthogonal ? Afin de répondre à ces questions il est nécessaire d'introduire la notion de résolution de la fraction régulière :

Définition 04

Considérons une fraction régulière définie à l'aide d'un seul générateur. On appelle résolution de cette fraction régulière (notée classiquement en chiffres romains) la longueur du générateur utilisé.

→ Vérifier à titre d'exercice que l'utilisation de fractions régulières de résolution I ou II est impossible dans le cadre de l'ajustement d'un modèle d'ordre un.

Le résultat général est présenté ci-dessous (sa démonstration repose sur la théorie algébrique de la représentation linéaire des groupes finis) :

Proposition 04

Considérons une fraction régulière d'un plan factoriel complet pour m facteurs, obtenue à l'aide d'un générateur. Alors :

- 1) cette fraction régulière a un total de $n = 2^{m-1}$ expériences,
- 2) si cette fraction régulière est de résolution III (ou plus) elle constitue un plan d'expérience orthogonal pour le modèle d'ordre un avec :

$${}^t X X = 2^{m-1} I_p$$

Il en résulte donc que l'estimateur des moindres carrés ainsi que sa dispersion sont donnés par :

$$\hat{\beta} = \frac{1}{2^{m-1}} {}^t X Y \quad \text{et} \quad \mathbb{V}(\hat{\beta}) = \frac{\sigma^2}{2^{m-1}} I_p$$

04

Fractions régulières : cas général

Les résultats du paragraphe 03 ne sont pas satisfaisants en pratique puisque pour un grand nombre de facteurs m réaliser 2^{m-1} expériences va toujours être trop coûteux. L'idée est alors de généraliser ce qui a été vu avant en ne se limitant plus à un unique générateur. *mais columns que entre si complon con i Obj = II*

1) Exemple introductif

Considérons ici un phénomène aléatoire dépendant de $m = 5$ facteurs et supposons que le coût de chaque expérience est très élevé. Le nombre minimal d'expériences nécessaires est donc $p = 6$. Dans une telle situation utiliser un plan factoriel complet en $2^5 = 32$ expériences n'est pas envisageable mais utiliser une fraction régulière obtenue à l'aide d'un générateur va nous conduire à $2^4 = 16$ expériences ce qui risque aussi de se révéler trop coûteux. Utilisons alors la fraction régulière définie à l'aide des deux générateurs suivants :

$$\mathbb{I} = \mathbf{123} = \mathbf{345}$$

En ne conservant que les expériences compatibles avec ces deux relations on obtient la matrice du plan suivante :

$$D = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -1 & -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 \\ -1 & 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Attention au fait que le lien supplémentaire suivant existe entre les colonnes de D :

$$\mathbb{I} = 1245$$

DCJLto

Ceci montre alors que 1245 peut aussi jouer le rôle de générateur, il s'agit donc d'un générateur caché. Une des difficultés lorsque deux générateurs (ou plus) sont utilisés est la non-unicité des générateurs utilisés. Remarquons enfin que ce générateur caché pouvait être obtenu sans même construire la matrice D puisqu'il vient algébriquement :

$$\begin{cases} \mathbb{I} = 123 \\ \mathbb{I} = 345 \end{cases} \Rightarrow (123)(345) = 123^2 45 = 1245 = \mathbb{I}$$

2) Notion de générateurs

Plaçons nous ici dans le cas général avec m facteurs et une fractions régulière construire à l'aide de $q \in \mathbb{N}$ générateurs. Il convient dans un premier temps de déterminer l'ensemble contenant tous les candidats générateurs possibles pour cette fraction régulière. Comme vu dans l'exemple introductif un générateur est un produit d'Hadamard d'un certain nombre de colonnes étant égal à l'élément neutre. Il en découle que tout produit de générateurs va encore donner un générateur, d'où la définition suivante (où le groupe est muni du produit d'Hadamard comme opérateur) :

équipado

Définition 05

On appelle **groupe des contrastes de définition** d'une fraction régulière, noté \mathcal{G} , le groupe engendré par ses q générateurs.

Exemple. Concernant l'exemple introductif il vient alors :

$$q=2$$

$$\mathcal{G} = \{\mathbb{I}, 123, 345, 1245\}$$

On a 4 éléments, 3 générateurs plus l'identité

Cette définition étant posée la question se pose maintenant de savoir combien de candidats générateurs peuvent exister. On montre alors que (raisonnement identique à celui pour, par exemple, déterminer le cardinal de l'ensemble des parties d'un ensemble) :

On a premièrement l'élément de l'identité, après on pris le couples, etc ... et les singuliers

Proposition 05 Longueur du générateurs

Le groupe \mathcal{G} engendré par les q générateurs d'une fraction régulière est un **groupe fini** constitué de 2^q éléments.

3) Notion de résolution

Une nouvelle fois la question de l'utilisation et de l'orthogonalité du plan d'expérience ainsi construit se pose. Ceci va dépendre de la résolution dans un sens qu'il faut généraliser puisque la définition 04 devient inopérante maintenant. On généralise alors cette notion de la manière naturelle suivante :

La fraction est définie de manière équivalente par :

$$\left. \begin{array}{l} \mathbb{I} = 123 = 345 \\ \mathbb{I} = 1245 = 345 \\ \mathbb{I} = 123 = 1245 \end{array} \right\}$$

Définition 06

Considérons une fraction régulière définie à l'aide de q générateurs. On appelle **RÉSOLUTION** de cette fraction régulière (notée classiquement en chiffres romains) la **longueur** du plus petit des éléments du groupe \mathcal{G} des contrastes de définition de la fraction régulière (\mathbb{I} exclu).

On généralise enfin la proposition 04 pour obtenir le résultat primordial suivant :

Proposition 06

Considérons une **fraction régulière** d'un plan factoriel complet pour m facteurs, obtenue à l'aide de q générateurs. Alors :

- 1) cette fraction régulière a un total de $n = 2^{m-q}$ expériences,

2) si cette fraction régulière est de résolution III (ou plus) elle constitue un plan d'expérience orthogonal pour le modèle d'ordre un avec :

$${}^tXX = 2^{m-q} I_p$$

Il en résulte donc que l'estimateur des moindres carrés ainsi que sa dispersion sont donnés par :

$$\hat{\beta} = \frac{1}{2^{m-q}} {}^tXY \quad \text{et} \quad V(\beta) = \frac{\sigma^2}{2^{m-q}} I_p$$

Remarque. Il est classique de désigner une fraction régulière de résolution (par exemple) III, obtenue à l'aide de q générateurs, par la notation générique : 2_{III}^{m-q} . *dans l'exemple on a* $2^{\frac{5-2}{3}}$

Cette théorie des fractions régulières est très efficace en pratique car elle permet de réduire les tailles des plans d'expérience factoriels complets de manière très conséquente tout en conservant l'orthogonalité. A titre d'exemple pour $m = 10$ facteurs la fraction régulière de résolution III suivante peut être considérée :

$$\text{II} = 125 = 136 = 147 = 238 = 249 = 34\overline{10} \quad \frac{1}{12} \text{ le produit de column 1 et 2}$$

Le modèle contient alors $p = 11$ paramètres inconnus, le plan factoriel complet est en $2^{10} = 1024$ expériences et cette fraction est construite à l'aide de $q = 6$ générateurs, elle est donc de type 2_{III}^{10-6} c'est-à-dire composée de seulement $2^4 = 16$ expériences.

05 Exercices d'applications

Exercice 1

[Construction de fraction régulière]

On considère ici $m = 4$ facteurs et la fraction régulière définie par :

II = 123 = 234

Analyser cette fraction régulière.

Premièrement on a que
puisque il y a 2 générateurs

$$m=4 \quad \begin{matrix} 123 = 234 \\ \text{on a } (123)(234) = 12^2 3^2 4 = 14 \end{matrix} \quad \text{donc } G = \{1, 123, 234, 14\}$$

Résolution II. analyse impossible parce que
les lignes sont: $2^4 = 2^2 = 4$ avec le générateur donné,

$$1 = 14 \Rightarrow t = 4$$

Exercice 2

[Construction de fraction régulière] $a=3 \Rightarrow 2^3=8$

On considère ici $m = 6$ facteurs et la fraction régulière définie par :

$$\mathbb{I} = 124 = 135 = 236$$

Analyser cette fraction régulière.

$$\begin{array}{r} 124 \cdot 135 = 2345 \\ \hline Q = \quad 124 \cdot 236 = 1346 \\ \hline 135 \cdot 236 = 1256 \end{array}$$

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \\ -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & \end{matrix}$$

Exercice 3

[Construction de fraction régulière]

On considère ici $m = 5$ facteurs et la fraction régulière définie par :

$$\mathbb{I} = 123 = 345 = 1245$$

Analyser cette fraction régulière.

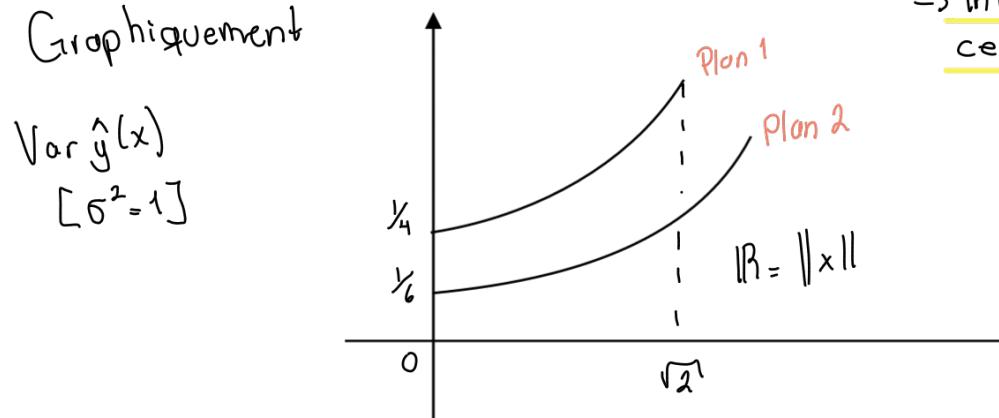
Exercice 4

[Construction de fraction régulière]

Construire la matrice du plan associée au dernier exemple vu en cours. [LogCalc]

$$\begin{aligned} \text{Exercice 2: b) Plan 2: } \mathbf{x}' \mathbf{x} &= (6, H, 4) \Rightarrow (\mathbf{x}' \mathbf{x})^{-1} = \begin{pmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \\ \Rightarrow \text{Var} \hat{y}(x) &= \sigma^2 g(x) \begin{pmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} \end{pmatrix} g(x) = \sigma^2 \left(\frac{1}{6}, \frac{1}{4} x_1, \frac{1}{4} x_2 \right) \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = \\ &= \sigma^2 \left(\frac{1}{6} + \frac{1}{4} x_1^2 + \frac{1}{4} x_2^2 \right) = \sigma^2 \left(\frac{1}{6} + \frac{1}{4} \| \mathbf{x} \|^2 \right) \\ \Rightarrow \text{Plan 2 Var} \hat{y}(x) &< \text{Plan 1 Var} \hat{y}(x) \end{aligned}$$

Graphiquement



\Rightarrow Intérêt d'avoir 2 expériences centrales:

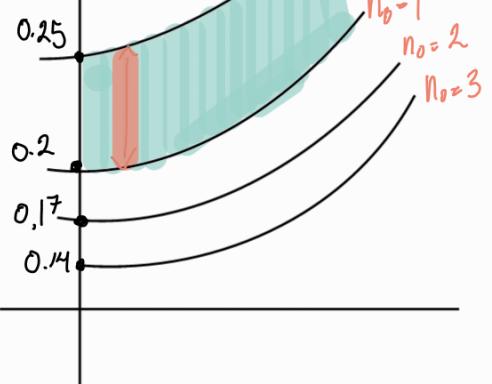
- ① meilleur estimation de β_0
- ② meilleures prédictions
- ③ analyse plus fine de l'erreur en décomposant SSE

c) Généraliser : Combien d'expériences centrales faut-il réaliser ?

Cos général avec $n_0 \in \mathbb{N}$ expérience(s) centrale(s)

$$n_0 = 0$$

Variance de prédiction
 $\Rightarrow \text{Var } \hat{y}(x) = \sigma^2 \left(\frac{1}{4+n_0} + \frac{\|x\|^2}{4} \right)$



Si possible réaliser $n_0 = 2, 3, 4$ expériences centrales

Mais note que il y a une grande différence entre $n_0=0$ et $n_0=1$, donc il est meilleure de faire au moins une expérience central

Chapitre 4

Def 4. Exercices

Fraction régulière de résolution I

Pour $m=3$ considérons, par exemple, la fraction régulière définie pour

$$\mathbb{I} = 2$$

$$X = \begin{pmatrix} & 1 & 2 & 3 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Tomamos estos renglones para cumplir con $\mathbb{I}=2$

$$X_F = \begin{pmatrix} & 1 & 2 & 3 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

La 1^{er} column et 2 ont la même information

En plus, le facteur 2 ne varie jamais

X_F n'est pas de plein rang $\Rightarrow {}^t X_F X_F$ n'est pas inversible
parce qu'il y a 2 vecteurs dépendantes.

de résolution II $\mathbb{I} = 12$

$$X = \begin{pmatrix} & 1 & 2 & 3 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Tomamos estos renglones

$$\Rightarrow \begin{pmatrix} & 1 & 2 & 3 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

iguales

Pb

- Les facteurs 1 et 2 sont liés artificiellement
- d'un point de vue mathématique 2 colonnes de X_F sont identiques $\Rightarrow {}^t X_F X_F$ n'est pas inversible

$$\mathbb{I} = 12 \Leftrightarrow 1(\mathbb{I}) = 1(12) \\ \Leftrightarrow 1 = 1^2 2 \\ \Leftrightarrow 1 = 2$$

Il y a une confusion d'effets entre les effets linéaires des facteurs 1 et 2

Exemple

$$m=5 \quad \mathbb{I} = 123 = 345$$

$$\begin{pmatrix} & 1 & 2 & 3 & 4 & 5 \\ -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 \end{pmatrix}$$

On a aussi la relation

$\mathbb{I} = 1245$ générateur caché

$$(123)(345) = 123^2 45 = 1245 \\ (123)(345) = \mathbb{I} \cdot \mathbb{I} = \mathbb{I}$$

$$D = \begin{pmatrix} 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 \end{pmatrix}$$

Travaux Dirigés

Ex 1

$$m=4 \quad \text{et } \text{I} = 123 = 234, q=2$$

$$\text{On a } (123)(234) = 12^2 3^2 4 = 14 \quad \text{donc } G = \{1, 123, 234, 14\}$$

Résolution II: analyse impossible parce que:

lignes sont: $2^2 = 2^2 = 4$ avec le générateur donné,

$$D = \begin{pmatrix} 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{Pb: On constate que } I = 4, X \text{ n'est pas de plein rang}$$

\hat{x}_X n'est pas inversible, \hat{B} n'existe pas

$$\text{et } \text{I} = 14 \Rightarrow I = 4$$

Ex 2

$$m=6 \quad \text{et } \text{I} = 124 = 236 = 135, q=3 \quad \text{Primero encontramos a } G$$

$$\text{On a } (124)(236) = 12^2 436 = 1436 = 1346$$

$$(124)(135) = 2435$$

$$(135)(236) = 123^2 56 = 1256$$

$$(124)(135)(236) = (2345)(236) = 456$$

$$G = \{1, 124, 135, 236, 2345, 1346, 1256, 456\}$$



Résolution III analyse possible

On essaie de remplir avec les plus petits avec 124, après on remplir avec l'autre option 135, finalement on fait la concordance d'une forme que 236 marche aussi $\text{I} = 124 = 135 = 236$

Résolution

$$D = \begin{pmatrix} 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

hevisor

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 \\ \hline 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 \end{pmatrix}$$

Ex 3

$m=5$ et $\Pi = 123 \cdot 345 = 1245$, $q=3 \Rightarrow q=2$ porque no es de longitud mínima

On cherche des générateurs caché:

$$(123)(345) = 123^245 = 1245 \text{ / déjà dans } G \Rightarrow \begin{array}{l} \text{On a ici seulement } q=2 \\ \text{générateurs indépendants} \end{array}$$

$$(345)(1245) = 3124^25^2 = 312 = 123 \text{ / déjà dans } G$$

$$(123)(1245) = 12^2345 = 345 \text{ déjà dans } G$$

\Rightarrow pas possible de construire

Ex 4

On va partir de construire le générateur complet

1: Générer pas ou pas les expériences du plan factoriel complet selon l'ordre de Yates

2: Si les conditions des q générateurs sont vérifiées alors sortir la ligne dans la matrice D.

packages FrE2, rsm, mixexp

Conc = sous forme codée

Plan Complet \Leftarrow FrE2 (nruns = 24)

nfactor = 2
randomize FA (SI),
ncenter = 3

$$E = A\beta \Rightarrow 5 = 12 \lambda \Rightarrow \lambda = 125$$

$$\Pi_{min} \rightarrow 1^{\text{ordre}} \quad 2 = 4^q \quad \beta J = D - AH$$

$$\Pi_{16} = 24^q \Rightarrow$$

de graphes des effets linéaires:
plus incliné ça veut dire il est plus significative

05 Optimalité des plans d'expérience

01 Généralités

Face à un problème donné il n'existe pas en général de solution unique en terme de configuration expérimentale. Si, par exemple, deux plans d'expérience sont proposés et sont équivalents du point de vue du coût expérimental (c'est-à-dire avec le même nombre d'expériences) lequel choisir ? Il est clair que le « meilleur » des deux va être privilégié mais quel sens peut-on donner à cette notion ? Nous proposons ici quelques pistes afin de répondre à ce problème complexe qui est celui de la quantification de l'**efficacité** d'un plan d'expérience.

Une fois cette notion posée il est naturel de rechercher un éventuel plan d'expérience plus efficace que tous les autres (au sein d'une classe donnée) qui sera alors qualifié de **plan d'expérience optimal**.

Remarquons que ce problème est très simple à résoudre en dimension un. Supposons en effet que deux plans d'expérience de même taille \mathcal{D}_1 et \mathcal{D}_2 soient mis en oeuvre afin d'ajuster un modèle n'ayant qu'un seul paramètre β inconnu. Notons $\hat{\beta}_1$ et $\hat{\beta}_2$ les estimateurs sans biais de β obtenus par les deux plans. Il est évident que l'on va alors dire que le plan d'expérience \mathcal{D}_1 est plus efficace que le plan d'expérience \mathcal{D}_2 si et seulement si :

$$\text{Var}(\hat{\beta}_1) < \text{Var}(\hat{\beta}_2).$$

La généralisation de ce raisonnement est complexe car un plan d'expérience est en général utilisé pour estimer non pas une valeur mais un vecteur β de paramètres inconnus. Deux plans d'expérience vont alors conduire à deux **matrices des covariances** $\mathbb{V}(\hat{\beta}_1)$ et $\mathbb{V}(\hat{\beta}_2)$ qu'il n'est pas simple de comparer puisqu'il n'existe pas d'ordre naturel sur les matrices.

02 Optimalités alphabétiques

Introduisons ici deux critères d'optimalité très classiques que sont la D et la A optimalité des plans d'expérience. D'un point de vue mathématique ces deux critères ne sont rien d'autre qu'une façon d'introduire un ordre sur les matrices de covariance des estimateurs obtenus.

Attention, nous supposerons dans toute la suite de ce chapitre que les plans d'expérience utilisés sont réguliers (c'est-à-dire que la matrice d'information tXX est toujours inversible).

1) Notion de A-optimalité

Une méthode afin de quantifier la qualité d'un plan d'expérience donné est présentée ci-dessous (la lettre « A » venant de « average ») :

Définition 07

Considérons un plan d'expérience \mathcal{D} utilisé afin d'estimer les p paramètres inconnus d'un modèle linéaire. Sa **A-efficacité** est définie par :

$$\Phi_1(\mathcal{D}) = \frac{1}{p} \text{Trace}\left[({}^tXX)^{-1}\right]$$

Au sein d'une classe donnée de plans d'expérience s'il existe un plan \mathcal{D}^* minimisant la fonction Φ_1 il sera alors qualifié de plan d'expérience **A-optimal**.

Trazo de una matriz = suma de los elementos en la diagonal

$$\begin{array}{c} \text{Dispersion } (X^T X)^{-1} \\ \text{Information } X^T X \end{array}$$

Remarque. Cette stratégie consiste donc à ordonner les matrices des covariances (au coefficient σ^2 près puisqu'on travaille en fait avec les **matrices de dispersion - inverses des matrices d'information**) à l'aide de la **moyenne** des variances de tous les estimateurs obtenus. Cette approche est la plus simple possible mais elle présente cependant l'inconvénient de ne porter aucune attention aux termes extra-diagonaux de la matrice de dispersion (donc ignore complètement les **corrélations** éventuelles entre estimateurs).

2) Notion de D-optimalité

Une méthode afin de quantifier la qualité d'un plan d'expérience donné est présentée ci-dessous (la lettre « D » venant de « déterminant ») :

Définition 08

Considérons un plan d'expérience \mathcal{D} utilisé afin d'estimer les p paramètres inconnus d'un modèle linéaire. Sa **D-efficacité** est définie par :

$$\Phi_0(\mathcal{D}) = \left(\text{Det}[(^T X X)^{-1}] \right)^{1/p}$$

Au sein d'une classe donnée de plans d'expérience s'il existe un plan \mathcal{D}^* minimisant la fonction Φ_0 il sera alors qualifié de plan d'expérience **D-optimal**.

Remarques. Ce critère est en pratique le critère d'optimalité alphabétique le plus couramment utilisé, ceci est dû principalement aux raisons suivantes :

i) le critère de D-efficacité est **facilement calculable** puisque par propriété des déterminants il vient :

$$\Phi_0(\mathcal{D}) = \left(\text{Det}[(^T X X)^{-1}] \right)^{1/p} = \left(\frac{1}{\text{Det}(^T X X)} \right)^{1/p}$$

En d'autres termes ce critère peut être directement évalué à l'aide de la seule **matrice d'information** $^T X X$ (aucune inversion n'est nécessaire). Minimiser le critère Φ_0 équivaut donc à maximiser le déterminant de la matrice d'information.

ii) ce critère est **invariant par reparamétrisation affine**. Ceci est particulièrement intéressant par rapport à l'opération affine de codage des facteurs (voir les exercices pour une démonstration).

iii) d'un point de vue pratique minimiser le critère Φ_0 équivaut à **minimiser le volume** de l'ellipsoïde de confiance du vecteur des paramètres inconnus β (voir un cours de modélisation linéaire, la région de confiance est alors une forme quadratique associée à la matrice $^T X X$, il s'agit alors d'un ellipsoïde dont le volume est à une constante multiplicative près égal à la racine carrée de $\text{Det}(^T X X)^{-1}$).

03

Généralisation

Les deux critères présentés précédemment ont en commun le fait d'être directement liés aux valeurs propres de la matrice de dispersion $(^T X X)^{-1}$ (donc aux valeurs propres de la matrice d'information $^T X X$ puisqu'on passe de l'une à l'autre par simple inversion). On propose alors de généraliser ces deux critères via la classe de critères suivante :

Définition 09

Considérons un plan d'expérience \mathcal{D} utilisé afin d'estimer les p paramètres inconnus d'un modèle linéaire. Notons $\lambda_{\mathcal{D}}^{[1]} \geq \dots \geq \lambda_{\mathcal{D}}^{[p]}$ la suite pleine décroissante des **valeurs propres** de la **matrice d'information** de ce plan d'expérience. Pour tout réel q (avec $0 < q < +\infty$) sa **Φ_q -efficacité** est alors définie par :

$$\Phi_q(\mathcal{D}) = \left(\frac{1}{p} \sum_{i=1}^p (\lambda_{\mathcal{D}}^{[i]})^{-q} \right)^{1/q}$$

Au sein d'une classe donnée de plans d'expérience s'il existe un plan \mathcal{D}^* minimisant la fonction Φ_q il sera alors qualifié de plan d'expérience **Φ_q -optimal**.

Remarque. Si $\lambda_{\mathcal{D}}^{[i]}$ est une des valeurs propres de la matrice d'information tXX alors il est bien connu que $\mu_{\mathcal{D}}^{[i]} = 1/\lambda_{\mathcal{D}}^{[i]}$ est aussi une valeur propre de la matrice de dispersion $({}^tXX)^{-1}$ (associée au même vecteur propre). Il en résulte que si $\lambda_{\mathcal{D}}^{[1]} \geq \dots \geq \lambda_{\mathcal{D}}^{[p]}$ la suite pleine décroissante des valeurs propres de la matrice d'information alors $\mu_{\mathcal{D}}^{[1]} \leq \dots \leq \mu_{\mathcal{D}}^{[p]}$ est la suite pleine croissante des valeurs propres de la matrice de dispersion.

Il est immédiat que pour $q = 1$ on retombe sur le critère de A-efficacité présenté précédemment. La fonction Φ_q peut de plus être prolongée par continuité en 0 et en $+\infty$ (voir les exercices pour plus de détails). Le prolongement par continuité en 0 permet de retrouver le critère de D-efficacité présenté précédemment :

$$\Phi_0(\mathcal{D}) = \left(\text{Det} \left[({}^tXX)^{-1} \right] \right)^{1/p}$$

Le prolongement par continuité en $+\infty$ conduit lui à un autre critère classique d'efficacité qualifié de **E-efficacité** (la lettre « E » désignant cette fois « extremal ») :

$$\Phi_{\infty}(\mathcal{D}) = 1/\lambda_{\mathcal{D}}^{[p]} = \mu_{\mathcal{D}}^{[p]}$$

L'objectif de ce critère est alors de minimiser la plus grande des valeurs propres de la matrice de dispersion $({}^tXX)^{-1}$ (ou encore, de manière équivalente maximiser la plus petite des valeurs propres de la matrice d'information tXX). D'un point de vue pratique on cherche alors ici à minimiser la longueur du plus grand axe de l'ellipsoïde de confiance du vecteur des paramètres inconnus β (voir un cours de modélisation linéaire).

04 Optimalité universelle

Les critères de A, D et E-optimalité ayant été présentés, puis généralisés via le critère de Φ_q optimalité, il est naturel de rechercher maintenant une condition d'optimalité « plus forte » entraînant toutes les optimalités vues précédemment. Cette recherche n'admet pas une unique réponse, une façon usuelle d'atteindre cet objectif consiste cependant à s'intéresser à l'optimalité qualifiée d'**universelle**. Nous n'entrons pas ici dans les détails concernant la définition de cette optimalité. Lorsque deux plans sont considérés l'un des deux va être plus efficace que l'autre selon le critère considéré ici si et seulement si le vecteur des valeurs propres de sa matrice de dispersion est le plus petit des deux selon « **l'ordre faible de Schur** » (voir des ouvrages spécialisés pour plus de détails). Nous allons nous limiter ici au résultat applicatif suivant :

Proposition 07

Soit une classe Θ de plans d'expérience et $C_{\mathcal{D}}$ la matrice d'information de \mathcal{D} . Soit $\mathcal{D}^* \in \Theta$ un plan d'expérience tel que :

- 1) $C_{\mathcal{D}^*}$ est multiple de l'identité,
- 2) $\text{Trace}(C_{\mathcal{D}^*}) = \max_{\mathcal{D} \in \Theta} (\text{Trace}(C_{\mathcal{D}}))$.

Le plan d'expérience \mathcal{D}^* est alors **universellement optimal** dans la classe Θ .

Il en découle alors qu'un tel plan d'expérience est aussi optimal selon le critère de Φ_q optimalité et ceci quelle que soit la valeur de q . Voilà donc pourquoi un tel intérêt est porté en pratique à la recherche de plans d'expérience **orthogonaux**.

Application aux plans factoriels

Considérons un plan factoriel complet \mathcal{D}^* pour m facteurs (la démonstration est en tous points identique pour des fractions régulières) et montrons que la proposition 07 est bien vérifiée. Concernant le premier point nous savons que pour ce type de plans (orthogonaux) il vient :

On a fait dans les chapitres antérieurs
(3)

$$C_{D^*} = {}^t X^* X^* = 2^m I_p$$

Pour le second point soit \mathcal{D} un plan d'expérience quelconque en $n = 2^m$ expériences. Désignons par $(x_u)_{u=1,\dots,n}$ les points de ce plan avec $x_u = (x_{u1}, x_{u2}, \dots, x_{um})$. Il faut alors prouver que l'on a toujours :

$$\text{Trace}(C_{D^*}) \geq \text{Trace}(C_D) \iff n + nm \geq n + \sum_{u=1}^n \sum_{i=1}^m x_{ui}^2 \quad \text{plus petit que } m$$

Ce résultat est immédiat puisque, d'après le codage des facteurs réalisé, on a toujours $x_{ui}^2 \leq 1$. Ceci montre bien alors que tout plan **factoriel complet** est universellement optimal au sein de la classe des plans en $n = 2^m$ expériences.

05 Exercices d'applications

$$x = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

Exercice 1

[Invariance de la D-optimalité par reparamétrisation affine]

Soit un modèle linéaire dont les paramètres inconnus sont regroupés dans le vecteur β . Supposons qu'un changement de variables est effectué (par exemple via le codage des facteurs) et que le nouveau vecteur des paramètres inconnus est $\theta = H\beta + h$ avec H matrice fixée et h vecteur fixé. Justifier que tout plan d'expérience D-optimal relativement au vecteur β est encore D-optimal relativement au vecteur θ .

Exercice 2

[Prolongement par continuité de la fonction Φ_q]

Prolonger par continuité, en 0 et en $+\infty$ le critère d'efficacité Φ_q . Retrouver les critères de D et E efficacité présentés en cours.

Exercice 3

[Optimalité d'un plan factoriel complet]

Vérifier à l'aide de simulations que le plan factoriel complet pour $m = 2$ facteurs est bien **D-optimal** (ou A-optimal) pour le modèle linéaire d'ordre un au sein de la classe des plans d'expérience en $n = 4$ expériences. [LogCalc]

Exercice 1:

Notons $y = x\beta + \epsilon$ le modèle linéaire initial. Relativement au vecteur β un plan d'expérience D-optimal minimise la quantité $\text{Det} \left[\underbrace{({}^t x x)}^{-1} \right] V(\hat{\beta}) = \sigma^2 ({}^t x x)^{-1}$

$$\frac{1}{\sigma^2} V(\hat{\beta}) \Rightarrow \frac{1}{\sigma^2} V(\hat{\beta}) = {}^t (\hat{\beta}) ({}^t \hat{\beta})$$

Considérons la reparamétrisation de β et $\theta = H\beta + h$.

Que devient la D-efficacité de ce plan par rapport au vecteur θ ?

On a $V(\hat{\theta}) = V(H\hat{\beta} + h) = H V(\hat{\beta}) {}^t H = \sigma^2 H ({}^t x x)^{-1} H$

Conclusion La D-efficacité du plan relative au vecteur θ est donnée par

$$\text{Det} \left[H ({}^t x x)^{-1} H \right]$$

Las propiedades del determinante incluyen a la commutatividad del producto

Si las dimensiones de los matrices lo permiten

$$\Rightarrow \text{Det} [H(t_{xx})^{-1} t H] = \text{Det} [t^T H H (t_{xx})^{-1}] = \underbrace{\text{Det}(t^T H H)}_{\text{constante}} \text{Det} [(t_{xx})^{-1}]$$

Exercice 2

$$\Phi_q(D) = \left(\frac{1}{P} \sum_{i=1}^P (\lambda_0^{[i,j]})^{-q} \right)^{\frac{1}{q}}$$

Rmq

$$\lim_{x \rightarrow \infty} (1+x)^{\frac{1}{x}} = e$$

$$\Rightarrow \ln(\Phi_q(D)) = \ln \left(\frac{1}{P} \sum_{i=1}^P (\lambda_0^{[i,j]})^{-q} \right)^{\frac{1}{q}} \quad q(D) = \frac{1}{P} \sum_{i=1}^P [\lambda_0^{[i,j]}]^{-q}$$

$$\Rightarrow \ln[\Phi_q(D)] = \frac{1}{q} \ln[q(D)] = \frac{1}{q} \ln \left[1 + (q(D)-1) \right]$$

$$\text{Comme } \lim_{q \rightarrow 0^+} q(D) = 1 \text{ on a donc } q(D)-1 = \frac{1}{P} \sum_{i=1}^P [(\lambda_0^{[i,j]})^{-q}-1]$$

qui tend bien vers 0 lorsque h tends vers 0⁺.

Sachant que $\ln(1+t) \approx t$ on a donc: $\ln[\Phi_q(D)] \approx \frac{1}{P} \sum_{i=1}^P f_i(q)$
avec $f_i(q) = \frac{(\lambda_0^{[i,j]})^{-q}-1}{q}$

Remarquons que $f_i(q) = \frac{f_i(q) - f_i(0)}{q}$ avec $f_i(q) = (\lambda_0^{[i,j]})^{-q}$

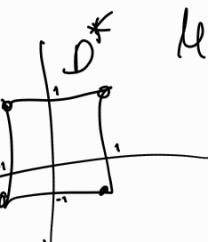
On en déduit que: $\lim_{q \rightarrow 0^+} f_i(q) = f'_i(0)$ avec $f'_i(0) = [-\ln \lambda_0^{[i,j]}] [\lambda_0^{[i,j]}]^{-1}$

Conclusion: $\lim_{h \rightarrow 0^+} [\ln(\Phi_q(D))] = -\frac{1}{P} \sum_{i=1}^P \ln \lambda_0^{[i,j]} = \ln \left[\frac{P}{i=1} \lambda_0^{[i,j]} \right]^{\frac{1}{P}}$

donc, $\lim_{h \rightarrow 0^+} \Phi_q(D) = [\text{Det}(t_{xx})]^{\frac{1}{P}}$

Exercice 3 Machine

1°



Montre que cette
modèle est
le meilleur

Tirer au hasard N points expérimentaux
dans $[-1, 1]^2$. Calculer la D-éfficacité du
plan est le comparer à la D-éficacité de D^*
*(Remplir un matrice $H \times H$ au hasard)

$$\overline{\Phi}_0(D^*) = [\text{Det}(t_{xx})]^{\frac{1}{3}} = (4^3)^{\frac{1}{3}} = 0.25$$

+ Los lanzamientos al azar en una gráfica

truccos los otros vienen
Cuando es ortogonal todos son igual

10/10/2022

Verificar que para $m=2$ factores el plan factorial completo es bien D-optimal (ou A-optimal) à l'aide de simulaciones

$$\underline{\Phi}_0(D) = [\text{Det}((X^T X))^{1/p}]^{1/p} = [\text{Det}(X^T X)]^{1/p}$$

à minimiser

D-eficacité du plan factorial complet D^* :

$$\underline{\Phi}_0(D^*) = [\text{Det } 4I_3]^{1/3} = \frac{1}{4} = 0.25$$

Idee - simular un plan en 4 experiencias
- Calcular $\underline{\Phi}_0$ - representación gráfica

- Para cada simulación graficar $\underline{\Phi}_0$, veremos que ese valor siempre es superior a 0.25
- no olvidar adjuntar la columna de 1's a X para calcular $[\text{Det}(X^T X)]^{-1/p}$

